THE ELEMENTS OF ECONOMICS

TARSHIS

The Elements of Economics

An Introduction to the Theory of Price and Employment

LORIE TARSHIS

Stanford University

HOUGHTON MIFFLIN COMPANY

BOSTON · NEW YORK · CHICAGO · DALLAS · ATLANTA · SAN FRANCISCO

The Riverside Press Cambridge

1947

Contents

Editor's Introduction

THIS BOOK reflects Professor Tarshis's conviction that the major purpose in teaching economics is to enable the student to understand the urgent problems of our national economy and to participate in their solution. The materials provided for use in the classroom are so selected and organized as to serve this ultimate objective, as well as the more immediate aim of acquainting the beginning student with the vocabulary and fundamental concepts of economic theory.

There are, however, different ways of dealing with economic problems in the classroom. One method, too frequently used, is to tackle them directly and in isolation from the normal functioning of the economic system. This method results in a symptomatic treatment of the ills and maladjustments of the national economy, but leaves the student with very little enduring knowledge of the essentials of the economy itself. Professor Tarshis avoids this method. He believes that before teacher and student come to grips with problems there must first be a patient, dispassionate investigation of the entire economy.

This book contains the best that the expert economist has to offer regarding the economic problems of our times, and in particular the over-all problem of full employment and the optimum functioning of the system, but these topics are postponed until the necessary introductory investigation has been completed. The student is first given a conspectus of the economy as a whole (in Part One); he is then taken (in Part Two) through a realistic analysis of the institutions of the business world where the actual decisions are made and the dynamic power is developed which move the entire economy toward a better or inferior adjustment to the requirements of general welfare; only on this basis of thorough understanding are the fundamental

problems defined and their solutions explored in the later sections. In the preliminary chapters, the student encounters the technical terms of economic theory, and he comes upon them in such close context with the operations of the business world that these terms, so often bewildering to the beginner, acquire real meaning for him.

The book thus combines the best features of the older systematic treatment of economics with the more recent approach based both on problem or case studies and on the national income as a whole. A mastery of the book will give the student as thorough a grasp of economic theory and the fundamentals of economic institutions as could be desired by advocates of the traditional method. At the same time it will provide him with the specialized knowledge and the techniques of analysis necessary to equip him to take an active part as a citizen in the determination of policy in economic matters.

EDGAR S. FURNISS

Preface

TWO CONVICTIONS guided me in the writing of this book. First, I believe that we cannot hope to enjoy lasting peace and prosperity until an understanding of the nature of our economy has become widespread. Second, I believe that this understanding can be gained only by applying the techniques of investigation developed by the scientist — not those of the lawyer, the philosopher, or the theologian — to economic problems. These convictions have dictated my choice of what to write about and how to write about it.

I realize that these guiding principles are shared by most economists. Every economist would agree that economics has a real and a valuable social function. The hope that we can learn how to do away with economic misery is the most important reason for carrying on economic research. And the knowledge that our findings must be communicated to the ordinary citizen, who after all determines the economic policy of the country, makes us keenly aware of the importance of teaching our findings to others. As economists we are justifiably impressed by the value of the contribution we can make to social welfare, though perhaps most of us are also sobered by the realization that up to now our contribution has been relatively small.

This point of view has been expressed many times before. A particularly sympathetic statement of it may be found in a paper read on March 9, 1939, to a faculty group at Tufts College by the late F. St. L. Daly, who died in 1944 while serving with the Canadian Army. Those who were privileged to work with him and to learn from him will appreciate how sincerely he meant and how successfully he followed this statement of his belief in the role of economics.

> Economic problems will be faced and, we hope, faced successfully by our graduates and their contemporaries. In one sense, this is truer of

> economics than of most fields. The cure for cancer will come from doctors and the responsibility will rest mostly with medical science, but the cure for idle factories will come from the ordinary citizen. Decisions regarding economic policies will be taken by the individual, partly in his role in industry, partly in his role in government as voter. The economist can give advice, but the businessman, the union leader, the Congressman, the voter will make the final decision. This puts much of the responsibility upon our colleges.[1]

To say that one believes that the scientific method must be used in economic research is now rather like saying that one believes in the good and the beautiful. It would be very hard to find an economist who would defend the use of non-scientific method. But the term is interpreted by each economist to cover precisely those practices which he himself finds most congenial and helpful. In keeping with my own preferences, I have tried to avoid introducing concepts which are immeasurable — for example, *utility*. I have also endeavored to introduce statistical data where by doing so I could conveniently illustrate how, in my view, economic problems should be treated.

A word about the contents of the book. In addition to the conventional material on the firm, monetary institutions, and so on, a good deal of attention has been given to analyzing the determination of the national income and employment. This has been done because I feel that the problems which this analysis is designed to clarify are the most important ones facing our economy today. This book devotes less attention to distribution theory than most books do, chiefly because I am not satisfied that the problems which conventional distribution theory attempts to illuminate can possibly be handled apart from an analysis of the determination of the national income. How, for instance, can an acceptable wage theory be devised which is separated from a theory of employment? It is my sincere belief, however, that what is here said on these matters adds as much to an understanding of distribution as the more ambitious analysis contained in most texts.

The scheme of the book was devised in order to make the student aware of the essential unity of economics. Part One is descriptive, and simply provides a background for the analysis which follows. Part Two analyzes the way in which the individual firm determines

[1] This paper was later published in *The Canadian Journal of Economics and Political Science*, Volume II, No. 2. I am indebted to the editors for permission to quote from it.

its price and output. Part Three shows how the financial sector of the economy functions, and pays special attention to the determination of the interest rate. Part Four, which is based on the results of the two preceding sections, analyzes the determination of the national income and employment. The adaptation of these results to an economy which has relations with other economies is considered in Part Five. Finally, Part Six treats some of the problems that arise in connection with the distribution of the economy's output and of the national income.

To the friends, colleagues, teachers, and students who have assisted in the preparation of this book I am deeply indebted. My former colleagues at Tufts College, especially Professors Lewis F. Manly and George Halm, gave me material aid both in organizing the material and in writing it, and provided an opportunity for me to try a first draft on their students. Dr. Carl Shoup of Columbia University read the manuscript with great care and pointed out many errors and inaccuracies in it. My colleagues at Stanford University were patient with me during the hectic period in which I was completing the work, and for their sympathy and forbearance I owe them my heartfelt thanks. I should like to acknowledge the assistance of Miss Ellen Ronning, Miss Patricia Gorman, and Mrs. Rhoda MacKenzie Jones in typing the manuscript. I wish to thank especially Miss Nancy Smith for her very great help both editorial and clerical, and Miss Patricia Brown, who found and corrected many examples of incorrect and unclear writing. Finally, to those in the editorial offices of Houghton Mifflin Company who assisted in the final preparation of the manuscript, I wish to acknowledge my debt.

I also wish to acknowledge the courtesy of Harcourt, Brace and Company, and that of the editors of *The Canadian Journal of Economics and Political Science*, for their permission to reprint copyrighted material. Most of the statistical data I have used were obtained from government publications. Every economist must feel indebted to his colleagues in Washington and in government offices elsewhere for their efforts to give us a good picture of the functioning of the economy, and I am particularly grateful for the material which has been made available through these agencies.

The writer of an introductory text cannot claim that what he says is original. Indeed, I hope that what I have written is not thought to be so; for if it is, the chances are that it is wrong. It would

be difficult indeed to trace the parentage of most of the ideas which this book expresses. Certainly my teachers at the University of Toronto and at Cambridge University made me acquainted with many of them, and numerous friends who were at Harvard University between 1936 and 1941 have helped me to follow recent developments. From all these, and from the works of others which I have read and studied, the pattern of my thought has been formed. To all of them I give my sincere thanks.

LORIE TARSHIS

STANFORD UNIVERSITY
December 30, 1946

Introduction

IT IS USUAL at the beginning of a study of any science to present a definition of the subject — a clear-cut, gemlike statement that in two or three sentences gives a precise account of its meaning and content. Unfortunately, the formal definition usually means very little to the student until he has gained such an insight into the subject that the definition is no longer necessary. To say at this stage that "economics is the scientific study of the operations of an economy" may make sense to an economist, but it will convey very little meaning to one who is just being introduced to the field. Until we have gained some understanding of economics, such a definition raises as many questions as it answers. "What is the economy?" "Which operations of the economy are to be investigated?" and so on. Rather than start with a definition which will, in any case, be evolved through the following pages, let us see whether we cannot find out something about economics by approaching the problem indirectly. Let us go around the subject, and reconnoiter it in order to learn something about its outlines. A general description of economics and of what the economist is trying to do is perhaps the most useful way to begin. We can derive the clearest preliminary understanding of the nature of the subject by examining the kinds of questions and problems with which the economist is ordinarily concerned. What problems is he normally interested in? What does he try to do?

In the first place, the economist does not pick the problems he studies at random. Generally speaking, he gives his chief attention to those which, within the special field of his competence, are the

most urgent in the society in which he lives. In this tendency to seek remedies for "real" difficulties, the economist behaves no differently from scientists in other fields. Even the purest of sciences are conditioned by the desire of their practitioners to find solutions to actual problems. So the needs of navigators affected astronomy; the early development of physics was stimulated and guided, in some degree at least, by the engineering difficulties of mining; and obviously medical science grew out of the desire to combat disease. Sciences are not born in a test tube. They grow out of man's desire to overcome difficulties in his environment. These difficulties may be of many kinds; physical, medical, psychological, or social. It is not surprising, then, that this imperfect world should provide the incentive for a study of problems which are social in nature; problems which have to do with such maladies as poverty, unemployment, and inflation. For these difficulties have plagued us consistently and seriously, and their existence provides the incentive for the work of the economist. Out of his attempts to secure an understanding of the social or economic institutions that give rise to such social ailments, he has developed a body of analysis which is known as economics.

The problems with which the economist has been concerned in recent years have been numerous and often very serious. Unemployment, inflation, strikes, economic nationalism, shortages of housing and of many other commodities, severe depression in agriculture, and the plight of small business — simply to list a few — gives an idea of their variety and their importance to our well-being.

The economist analyzes unemployment, inflation, strikes, and so on, because he wants to guide action designed to cure these things. The mere fact that they are recognized as problems about which something should be done carries an implication which it is desirable to develop. It implies that we set a certain high standard of performance for our economy. We expect it to deliver the goods in ways to be considered shortly. And when it fails to do so, we attempt to discover the causes of these failures and to correct them. In an economy which functioned perfectly without social controls of any sort, the economist would be out of a job. Indeed if people were completely satisfied with their economy, no matter how it really functioned, there would be no such thing as economics. A healthy man seldom calls a doctor; nor does a sick man who is completely resigned to his sickness.

How then do we want our economy to operate? What standards or criteria do we set up by reference to which we can judge its performance? In general we want it to satisfy the following conditions:

1. The total output of goods and services should be as high as possible with the physical resources and technical knowledge available to the society and consistent with the attitude toward work and leisure of the men and women in the society.

2. The division of the output among members of the society must be compatible with the society's standards of justice.

3. The composition of that output — that is, the number of suits and coats, tons of steel, bushels of wheat, and so on — must be in some sense appropriate to the desires and tastes of the individuals who comprise the society.

4. There must be a satisfying rate of growth in output per head and therefore in physical well-being.

5. And these objectives must all be secured without forcing the society into such a course of action that war becomes inevitable.

In an economy which satisfies all the above conditions, there would be nothing more for an economist to do. But in one which fails to satisfy any of them, there is still work for him. To say that such an economy is imperfect does not, of course, mean that it is a failure and has to be discarded. It means only that a strong effort should be made to enable it to perform efficiently. But let us examine more carefully the conditions listed above so that we may see more clearly why we want the economy to satisfy them. Why is it important that these conditions be met? Because they can be regarded as defining the conditions for our economic well-being. Let us see why.

The more goods and services there are, the more we have available to divide among us. And the greater the output of goods and services, the greater is the amount available for our enjoyment. Hence we want as large an output as possible. But it is not enough that the total output be at the maximum permitted by our resources and technical knowledge. That output must be distributed in a way that satisfies us. We do not want all of the output to go to ten people, leaving nothing for the rest. We want everybody to have a fair share of what is produced. Moreover, the content of that output must satisfy us. An economy which used all of its resources and all its technical skill to produce, let us say, steel and steel alone would obviously not perform properly. We should soon starve, even though

our output of steel were enormously higher than that of any other country. We want bread and butter, clothing and houses, radios and automobiles, and so on, in the appropriate proportions. In other words, we want a balanced output. Moreover, in a properly functioning economy, provision must be made to set aside some of the current output in order that the output in the future should be somewhat larger. For example, some of labor's effort must go to producing and expanding factories and equipment so that in the future our labor force can produce more than it does today. Otherwise, we should grow poorer with the years, and obviously we do not want that. Instead, we want more and better factories, more and better houses, and so on. That is to say, we insist upon a certain amount of progress and advance in our economic well-being. Finally, while it may seem unnecessary to urge the importance of avoiding war as a result of our economic activity, the point must be emphasized nevertheless. It is now clear that economic factors have a good deal to do with creating the conditions out of which war develops. And we do not want our economy to function in such a way that wars are the natural outgrowth of its operations. In short, we want our economy to provide maximum prosperity and peace. If we can make it do these things, our job as economists is finished. Other problems may still plague mankind, but the economic problems will have been licked.

The economist does not devote his time to drawing up blueprints for a new economic order. He resembles the mechanic who repairs the old motor rather than the designer who drafts plans for a new one. We have inherited our economic institutions from the past. The economist conceives his job as that of making these institutions work properly. He would be interested in changing them only if he should decide that they could not be made to work efficiently.

It is, of course, not surprising that our economy sometimes fails to operate at peak efficiency. Most of the economic institutions we have inherited were not designed by economists; certainly the basic ones were not. In fact, they were not designed at all. Students of economic history can trace their gradual evolution under the pull and tug of various interest groups: of the landlord and the businessman, the merchant and the Church, the wage earner, the investor, and the bureaucrat. Anyone who has studied the development of these institutions will not be surprised to find that they do not always per-

form efficiently. After all, it is rarely enough that what we plan turns out as we planned it; it would be remarkable indeed if something which grew without planning should perform in just the way we want it to. If doctors had designed the human body, there would probably be no diseases. Likewise, if economists had designed the economy, the chances are that there would be no economic problems to worry about. As it is, the economist believes his job is to understand the existing economy in order that he can properly guide efforts to make it work efficiently.

The understanding of the economy is therefore the economist's first task. It is, however, a very difficult one. The complexity and variety of economic institutions places a severe demand upon our intellectual ability, for, as we shall see, each part of the economy is intricately related to every other part. Of course, if this were all, the job of the economist would be no more demanding than the job of the physicist, the research worker in medicine, or the psychologist. But the economist has other difficulties which he must face. The physicist can investigate the nature of the atom, at least in the modern age, without having to meet the opposition of anyone who is interested in perpetuating certain strongly held ideas about its structure. The research worker in medicine and the psychologist can investigate the operation of the human body and mind without having to be concerned about the prejudices of the rest of the community. There is no pressure in these sciences to make the results conform to someone else's prejudices and preconceptions. But the economist is not so fortunate. His investigations touch the pocketbook, always a sensitive spot. Consequently, he is exposed to all kinds of pressures: the pressure of those who want him to prove that free trade is the only way to prosperity, and of those who want him to show that without protective tariffs, the American worker will be deprived of his job. To satisfy one vocal group he must prove that high wages are the cause of unemployment; to satisfy another, he must show that unemployment is bound to grow unless wages are raised. He must prove that the gold standard should be maintained; likewise he must show that it ought to be abolished. The difficulties of being objective in such an atmosphere are, of course, enormous. The economist must seek the truth, when from all sides he is under pressure to defend causes.

The economist who wants to be scientific has still other difficulties

to face. Most people do not pretend to be physicists. Few of us doctor our own illnesses. When we have to cope with the problems of physics or chemistry, we call in the experts. But we all feel that we are economists. It is worth while to examine the furniture of our own minds on this subject. We will find, if we are normal, that we have rather strong opinions on economic matters. We may believe that strikes should be outlawed, or that the budget should be balanced; that a loan to Britain would help us economically, or that inflation can only be prevented if prices are allowed to rise. Or we may hold the opposite view on any of these matters. The point is that only rarely does the man in the street admit to ignorance on matters of economic policy. Amusing illustrations of this are found in the Public Opinion Surveys conducted by several prominent magazines and other agencies. The following question was asked in 1944 of a representative sample of the population: "Regardless of whether you approve of some of these things, which one do you think would do the most to help prevent unemployment after the war?" The alternatives were: prevent married women whose husbands have jobs from working; shorten the working week; encourage the development of new products; build up the market for all products; lower the age for social security pensions to sixty; and, don't know. Approximately 8 per cent of the people interviewed admitted that they did not know. The other 92 per cent were sure they knew the answer. Likewise, when asked whether the next administration should or should not balance the budget, only 14.8 per cent answered that they did not know. On matters of economic policy, most of us feel competent to give advice. Ironically enough, on questions about economic facts, we are less ready to call ourselves experts. When asked approximately how much income tax a man who earned $25,000 a year and who had two children would have to pay under the existing rates, 30.4 per cent said they did not know. On being asked which of a series of figures came closest to the number of International Labor Union members in the United States, 23.5 per cent said they did not know. We tend to be modest about our competence in matters of economic fact, but we are much less willing to admit that we do not know how to solve the bigger problems, like unemployment, or inflation, or whether the budget should be balanced, or whether labor unions should be controlled. And yet these are most complex questions which, one would think, can be settled only after knowing all the facts.

Unfortunately for the economist, the man in the street is a walking encyclopedia of economics. This is not to say that his knowledge of economics is necessarily sound or good. It may be no more satisfactory in treating our economic difficulties than is the mumbo-jumbo of the witch doctor in treating sickness. But because all of us have ideas and opinions about economics, the task of the economist is made the more difficult. For one thing it is hard for the economist or anyone else to divest his own mind of prejudices. It is difficult to believe that our opinions may simply be the unconsciously assimilated opinions of others who knew no more about the subject than we know ourselves. But once we realize this, we can understand the importance of studying the operations of the economy with humility and of realizing that we probably do not know the truth about them by instinct.

The economist has still another problem to face. In physics, in medicine, and in most other sciences it is possible to perform experiments in a laboratory where they can be controlled. The whole purpose of a properly designed experiment is to let us isolate and study the influence of some one factor in which we are interested. But the economist's laboratory is the actual economy, and it is anything but controlled. Obviously it is much harder to get valid results in an uncontrolled experiment than in a well-designed laboratory. The objection to the economist's laboratory is that too many things are happening at once. The economist may, for instance, be interested in determining the effects of an increase in wages. Now wage rates frequently vary in our economy, but unfortunately for the economist, numerous other changes are also taking place at the same time, so that the effect of the change in the wage rate is very often obscured by the effects of changes in these other factors. But even though the economist cannot make controlled experiments, it does not follow that he should not observe what happens. Rather, it means that his observations have to be very careful and very numerous. Moreover, interpreting his observations is much more difficult for the economist than it is for most other scientists. When fifty different variables are all subject to change, it is hard to determine the precise influence of any one of them. Economics in this respect is rather like astronomy or geology. Indeed, the astronomer is no better able to experiment than is the economist. But that, of course, does not prevent the astronomer from using his telescope. Likewise, the economist must

use his. Only by patient and careful observation can the economist overcome the difficulties of having to work without controlled experiment.

The science of economics, therefore, is difficult on four counts. The economy itself is exceedingly complex; the economist is subject to the pressures of various interest groups; like everyone else, the economist may early acquire tenacious prejudices about economics which are very hard to get rid of; and finally, the economist is not free, as are the practitioners of many other sciences, to make controlled experiments. But while it is very difficult for all these reasons to secure an objective understanding of how the economy works, it is nonetheless important that we do so. We have seen enough of the effects of unemployment, of inflation, of war, of depression, and of poverty to persuade us that the work of the economist must be done. And it must be done properly and accurately, if it is going to help us in protecting the economy from these disasters. For only a clear and objective understanding of how our economy works can guide us properly in devising methods for improving its workings.

In summary, then, the economist concerns himself with such problems as unemployment, inflation, and the maldistribution of income. He does so because he recognizes that only if these conditions are remedied can ordinary people have material well-being. His methods are the methods of other scientists. He frames hypotheses, reaches conclusions, and checks them by observation. Unfortunately his subject matter, and the environment in which he works, do not favor analysis. Hence economics is not simple. Much more than the application of what the newspaper editorials call "common sense" is needed to solve the complex problems of the economy. Indeed, if this common sense were so very common we should all be expert physicists, expert biologists, and expert psychologists. The science of economics places as great a demand upon the economist as any other science does upon its practitioner. Uninstructed common sense will seldom guide us to a cure for a disease. Nor is it likely to assist us in curing an economic ailment. Study of the economy requires the will to think without prejudice, a strong desire to observe conscientiously and carefully, and a great deal of work. But the problems which this study is directed to solving affect us all deeply enough to make this effort worth our while.

PART ONE

The Structure of the Economy

Introduction

BEFORE we can begin our analysis of how the economy functions, we must make a preliminary survey of its structure or its physical features. In Part One, we shall present a simplified and therefore to some extent an impressionistic sketch of the economy, noting its more important institutions, the resources at its disposal, and the variety and amount of its product. With this survey as a background, we shall in subsequent parts of the book devote our attention to a more detailed analysis and discussion of various sectors of the economy.

1

An Over-all View of the Economy

THE AIM OF THE ECONOMIST is to analyze the operations of the economy. But what is the economy? In the next four chapters we shall try to answer that question, at least in general terms, by presenting a series of pictures of the economic landscape, first spotting only the most important landmarks, then gradually filling in the details. In this chapter we shall examine the economy as though from a long way off, as seen perhaps in a single reconnaissance photograph. In the second chapter we shall use not a still- but a motion-picture camera in order to discover how the economy operates when viewed over a period of time. In the third chapter our observations will be more detailed, and we shall concentrate on one institution, perhaps the most critical one in the economy — the business firm. Finally, in the fourth chapter we shall examine some of the complicated interrelationships that exist in the economy, specifically the part that government plays; for government embodies our social control over the economy, and no description of the economy would be complete without a record of how we exercise this control.

The Factors of Production

If our observations were made during a working day, here are some of the things we should see. We should see large numbers of factories clustered here and there over the country, as well as farms, retail and wholesale stores, barbershops, mines, office buildings, banks, and so on. Many of the things we can see would be in the picture because the government had put them there — things like highways, post-offices, schoolhouses, and power dams. There would be all kinds of

equipment in these factories, mines, farms, and schools: tractors and barber chairs, blast furnaces and display windows, derricks and typewriters, generators and electric motors, and other sources of power. In all the factories there would be stocks of raw materials, of goods on the assembly line, and of finished goods waiting to be shipped. Warehouses would be packed with canned peas, soap, radio tubes, tires, tin cans, and baby bottles. The shelves of retail stores would be filled with shoes and shirts and food and toothpaste. These items of equipment, these plants and highways and stores and buildings, and these stocks of goods in the possession of business firms, are known collectively as "capital goods." They exist in our economy in great profusion and in fantastic variety.

The mere extent of these capital goods is astonishing. Because of their variety, we can only measure their volume in dollars, for after all, you cannot add six locomotives, eighteen thousand pairs of shoes, and an electric furnace except in money terms. The National Resources Committee, an agency of the United States government, prepared the following estimate, which as they emphasized gives only a crude approximation. In 1935 plants engaged in manufacturing, together with their equipment and goods in process, were valued at $31 billion. Since there were about 30 million families in the country in that year, the value of manufacturing plant, equipment, and inventory came to about $1000 a family. In mining, the value of the capital goods was set at about $6 billion; in agriculture, at about $39 billion; in public utilities — railroads, electric power plants, telephone companies, and so on — the value of capital goods was estimated at $51 billion. Retail and wholesale stores and their contents were valued at about $13 billion. And other kinds of capital goods, including banks, public schools, barbershops, office buildings, and so on, were valued at about $92 billion. Thus the total value of our capital goods was estimated at approximately $232 billion in 1935, or almost $8000 for each family.

Our reconnaissance view would show us men and women working in the plants and stores and on the farms, tending the machinery, selling, typing, moving raw materials, inspecting, sowing and reaping. These people would constitute our labor force. If our photograph had been made in May 1946, we would have found about 55 million people at work. About 39 million of them were men, and the other 16 million were women. More than 8.5 million of them were engaged in agriculture. The self-employed and other businessmen

numbered about 9 million. The other 37 million were employees in manufacturing, mining, construction, trade, and so on. Some idea of the types of industry in which they were engaged can be obtained from the following table.

TABLE 1

Employees in Non-Agricultural Establishments: May 1946

(in millions)

Industry	Number of Employees
Manufacturing	12.6
Mining	.8
Construction	1.8
Transportation	3.9
Trade	7.7
Financial, service, and miscellaneous	5.1
Government	5.5
All employees:	37.4

The labor, as we have implied, would be of all kinds, from the most skilled to the least. A list of the jobs performed — that is, a catalogue of the kinds of work being done — would fill many pages.

In addition to men and machines, something else is needed to make the combination productive. Work must be planned and organized. Labor and capital goods must be brought together and their work must be mapped out. Decisions must be made as to the methods of production to be followed. It is easy to find the capital goods and the labor force on the reconnaissance photograph. But we must look more closely if we are to spot those who take the final authority in planning and organizing. For the non-government part of our economy, we call these people the businessmen. Sometimes we shall find them as the general managers of large corporations; sometimes as the owners of retail stores; sometimes as farmers. There are roughly 17 million of them in this country. Our labor supplies, our stock of capital goods, and the services of those who organize the work are our productive factors. In combination, they are responsible for our production.

Production

Now let us see what is produced in our economy. In a good year like 1945 we produce goods and services valued at about $200 bil-

lion.[1] This is the equivalent of about $5000 worth of goods and services for each family in this country, a figure enormously high compared to anything we have done in the past or to the output of any other country. This output is in many forms: food and clothing, automobiles, washing machines, sewing machines, radios and other durable consumers' goods, books, newspapers, magazines, haircuts, permanent waves, radio programs, new houses and house furnishings, factories, new machines and the repairs to old ones, the lectures of a professor, and baseball games. We cannot attempt to list the variety of goods and services that make up our annual output, though some measure of the relative importance of the various categories which constituted the total in 1943 may be found in the following table.

TABLE 2

National Income Produced by Industrial Source: 1943

(in billions of dollars)

Industry	Value of Production
Agriculture	14.0
Mining	2.5
Manufacturing	48.1
Contract construction	4.3
Transportation	9.5
Power and gas	1.6
Communications	1.2
Trade	17.4
Finance	9.2
Government	25.1
Service	10.3
Miscellaneous	4.6
Total	149.4

This table shows, for instance, that in 1943 firms engaged in manufacturing were responsible for about $48 billion worth of the total output of the economy (exclusive of the effort which went into repairing and maintaining capital goods). For recent figures on the production of any particular commodity, the reader should refer to the *Survey of Current Business* of the United States Department of Commerce. In that monthly magazine he will find production data on such various items as sodium silicate, glycerine, explosives, animal fats, coconut oil, linseed oil, soy beans, dried skim milk, distilled spirits, electric

[1] Part of this output (about $8.2 billion) had to be used for repairs and maintenance of plant and equipment; hence the figure $200 billion slightly exaggerates the value of goods and services available.

power, wheat, apples, steel ingots, steel for castings, paper, pulpwood, electric overhead cranes, synthetic rubber, tires and tubes, coke, Portland cement, and hundreds of other commodities. It is hard to say much more about our output except to repeat that it is very large and almost infinitely varied.

The Consumer

All we produce is either consumed by ourselves, sold to foreigners, or purchased by business firms or government bodies. In broad categories, our goods were purchased in the following ways in 1945:

TABLE 3

The Gross National Product and Its Components: 1945

Consumers bought	$104,900,000,000
Business firms * bought	9,100,000,000
Foreigners bought †	300,000,000
Government bought	83,000,000,000
Total purchases	$197,300,000,000

* Including purchasers of new houses, but excluding raw materials used up in production.
† Subtracting what we bought from them.

Now let us see how these goods were distributed among the families in this country in 1942, the most recent year for which figures are at present available. The following table indicates how the goods were divided by showing the number of families in various income classes:

TABLE 4

Distribution of Families and Single Consumers by Money Income Level: 1942

Income Level	Number of Families
Under $500	3,488,000
$500– 1,000	6,652,000
1,000– 1,500	6,601,000
1,500– 2,000	6,008,000
2,000– 2,500	4,618,000
2,500– 3,000	3,272,000
3,000– 4,000	4,620,000
4,000– 5,000	2,633,000
5,000– 7,500	1,901,000
7,500–10,000	628,000
$10,000 and over	789,000
All incomes	41,210,000

There were 41,210,000 consumer units — families and individuals — in the country in 1942. Of these, 33,360,000 were family groups; the other 7,850,000 were single individuals living alone. Of the 41 million consumer units in the country, 3,488,000, or 8.5 per cent of the total, received under $500 of income in 1942; 16.1 per cent received between $500 and $1000; and 789,000 units or 1.9 per cent received $10,000 or over. Using this table and certain additional data, it is possible to secure the following results. In 1942, the less than 2 per cent of all consumer units who received incomes of $10,000 or above, received about 15.8 per cent of the total income of the country. Thus, the wealthiest 800,000 families in the country could claim more than one-seventh of our total output of goods and services. Only 8 per cent of all consumer units received incomes of over $5000, and this group claimed about 34 per cent of the total output. At the other end of the scale, the 8.5 per cent of consumer units who received incomes of under $500 were able to claim only 1.1 per cent of the total income of the country. From these figures one fact is strikingly evident: the total output of our economy is distributed most unevenly. Some families get a very large share; others get, comparatively speaking, very little.

It is also instructive to see how income is distributed in relation to the way it is earned. Men and women who work in factories or offices receive wages and salaries. Those who lend money or rent land receive interest and rent. Individuals in business for themselves receive an income which is a combination of wages, salaries, rent, interest, and profit. Incidentally, for some purposes it is important to distinguish the income of those who work on farms from the income of those business and professional men who are engaged in retail and wholesale trade, medicine, or elsewhere. Finally, individuals who buy securities in corporations — and who, as we shall see, are the owners of the corporations — receive as income the profit made by these firms. Table 5 shows how the income was divided among employees, businessmen, farmers, and so on. Employees, who may range from day laborers to the general managers of our largest corporations, get almost three-fourths of the total income earned, but it must be remembered that there are almost 40 million of them in the economy. The rest of the income is divided among people whom we designate as farmers, businessmen, professional men, lenders, and investors.

TABLE 5

National Income by Distributive Shares: 1945

Item	Per cent of Total
Salaries and wages	71.1
Income of agricultural proprietors	7.8
Income of other private businessmen	8.5
Interest and rents	7.3
Net profits of corporations	5.6
Total	100.0 per cent, approximately

The Business Firm

We shall see on inspection that the three productive factors — labor, capital goods, and management or ownership — are organized for the most part into business firms. We say "for the most part" advisedly, because a significant proportion of all our factors of production is actually organized by government bodies. But about six-sevenths of the total is engaged in what we call business. On December 31, 1945, there were about 3,235,000 business firms in existence, exclusive of agricultural concerns. These were distributed among the following industries.

TABLE 6

Number of Operating Business Firms: December 1945

Industry	Number of Firms
Mining and quarrying	26,300
Contract construction	189,100
Manufacturing	262,500
Transportation, communication, and public utilities	205,500
Wholesale trade	141,800
Retail trade	1,504,200
Finance, insurance, and real estate	285,600
Service industries	619,800
Total	3,234,800

In addition to these 3 million business firms, there were more than 6 million farms, which should be regarded as separate business firms engaged in agriculture. Thus there were about 10 million firms in the economy. As we have already pointed out, about six-sevenths of all employees were connected with business firms; the other one-seventh worked for various government bodies.

Now 10 million business firms, including farms, is a very large number of firms among which to divide even $900 billion worth of business.[1] But we should not infer from this that they are all relatively small in size. There are giants among them like the American Telephone and Telegraph Company, the United States Steel Corporation, and the General Motors Corporation. And on the other side of the scale, there are such small firms as the corner grocery or drugstore, the barbershop, and the newsboy. In some industries, almost the entire production is provided by a few very large firms. For instance, in 1944, the four largest copper mines supplied something like 82 per cent of the total output of copper in the country. That year the Aluminum Company of America furnished about 95 per cent of all the aluminum we produced. On January 1, 1945, the three largest companies in the iron and steel industry controlled about 60.9 per cent of the total ingot capacity. In 1945 the four largest firms in the industry produced 81.6 per cent of all the tin cans. The five largest firms accounted for about 91 per cent of the total production of ball bearings, and the largest of these, the New Departure Division of General Motors, was responsible for more than 50 per cent of the total. It is clear that there are many industries in the country which are practically dominated by a few very large firms.

There are other industries in which the dominance of the big firm is not at all marked — for example in retail trade, construction, and the service industries — barbershops, garages, and so on. In 1935, as much as 30 per cent of all sales in retail trade were made by independent stores with annual sales of under $30,000. In the same year, about a third of all contract construction work was carried on by firms with less than $50,000 worth of business annually. The pattern is thus very complex — ranging from the telephone and aluminum industries, where one firm is practically synonymous with the industry, to agriculture and retail trade, where no one firm exercises an appreciable influence. We have millions of firms in the country, and most of them are very small. But in some industries, and indeed in most, there are a few which overshadow all the rest.

Summary

More than anything else, perhaps, this first quick reconnaissance view of the economy has shown us its immense diversity. Approxi-

[1] The approximate value of checks transferred in 1944.

mately 50 million men and women are at work with about $250 billion worth of goods and equipment organized into something like 10 million firms. In a good year they produce about $200 billion worth of goods and services of the widest possible variety and type. The work they do is varied in the extreme, and the amounts they earn are almost equally varied. The $200 billion worth of output is shared quite unequally by the approximately 40 million consumer units in the country.[1] While most of the firms are relatively small, most of the men and women, and most of the capital goods, are parts of a relatively few very large firms. More detailed inspection of the economy will of course bring to light further interesting and important information. Much of it will be presented in the following chapters. So far we have made only a first quick reconnaissance. Now let us look a little further.

[1] After allowing for depreciation.

2

The Changing Economy

AN INSTANTANEOUS PHOTOGRAPH can convey a good deal of information about an object, but not necessarily all we should like to have. For some purposes it is desirable to have a motion-picture record as well. This hapter is intended to provide such a record — an account of the development of our economy over time. For our economy is anything but static and unchanging. There are changes in the output of one commodity or another, in the distribution of that output, in the kinds of work we do, and in the importance of the various industries that make up the economy. There are even changes in the institutions of which our economy is composed. We cannot hope to understand how the economy works unless we know something about the ways in which it has altered.

Factors of Production

To bring some order out of the welter of information before us, we shall once more begin with our productive resources, though this time our main interest is in how they have grown. We have seen that in 1935 the value of our capital goods came to about $230 billion. While accurate figures are not available for other years, we do know that we generally add rapidly to the stock of capital goods in existence. In 1925 we added about $5.5 billion to the value of the plant and equipment held by business firms, and in 1940 about $7 billion. About $2 billion worth of new plant was constructed in 1940, and $11.4 billion worth of new plant and equipment was added to the total in 1941. The value of our capital equipment has increased

vastly through the years. As we shall see later, however, the increase has not been regular. It has been very small in years of deep depression such as 1931, 1932, and 1933, and very great in years of prosperity such as 1929, 1940, and 1941.

Dr. Simon Kuznets has estimated the rate of growth of capital goods in this country for various decades. According to his figures, in the sixty years between 1879 and 1939 we increased our physical stock of capital goods by almost eight times. In the first four decades of that period, the rate of growth was very even. Thus between 1879 and 1889 the stock of capital goods increased by 51 per cent; between 1889 and 1899, by an additional 61 per cent; between 1899 and 1909 by another 50 per cent; and between 1909 and 1919 by 47 per cent. With the next decade the rate of growth slackened. In the period between 1919 and 1929, the stock of capital goods increased by only 38 per cent. And between 1929 and 1939, when depression was severe, our stock of capital goods grew by only 6 per cent. But between 1940 and 1946, immense additions were made to our stock of capital goods.

Our labor force has also grown quickly. In 1870 there were only 12 million men in the labor force. By 1900 this number had increased to about 27 million men, and by 1945 to about 60 million. Thus the size of the labor force has expanded fivefold during the period of seventy-five years. The number of people who were actually at work has of course varied too, and unfortunately, not always in the same direction as the number of people who were seeking jobs. For instance, in 1929 about 42 million people were at work, a figure very little below the number of people who wanted to work. But in 1932, while about 48 million people wanted jobs, only 37 million actually had them. Generally, however, we have had a rapid, though not a steady, rise in employment. In short, the stock of capital goods and the size of the labor force have both increased rapidly over the years.

There have also been profound changes in our ability to make productive use of our capital goods and our labor. There have been striking improvements in the techniques of production. We have been able to put into effect new methods that depended on the use of ball and roller bearings, the steam turbine, the gasoline engine, the Diesel motor, high speed tool steel, the electric motor, and a number of other important inventions. Production, now highly mecha-

nized, is guided by the assembly line in many of our factories. Agriculture has its tractors, gang plows, combines, reapers, and binders. Almost every business office now has its typewriters, dictating and computing machines, and mechanical aids to filing. In industry after industry, methods of production have markedly improved over the last fifty years.

The effect of these improvements can be seen in many ways. In manufacturing industries, the output from one hour of work increased by about 75 per cent between 1919 and 1934.[1] Our productivity in the boot and shoe industry expanded by about 42 per cent between 1923 and 1934. In the cane sugar industry, the increase in productivity is even more striking — between 1919 and 1936 it amounted to 120 per cent. The output of cement per man-hour increased by about 65 per cent between 1923 and 1936. Numerous other examples of increased productivity could be cited. In short, there is overwhelming evidence that we have improved our methods of production substantially, and that in consequence a given supply of labor and capital goods can produce much more now than it could even twenty years ago. It is estimated that productivity has increased at a rate of about 2.5 per cent per annum. Such a rate of increase means that our productivity doubles in about twenty-eight years. The amount of our productive resources and their efficiency has grown enormously.

Changes in Output

An eightfold increase in our stock of capital goods, a fivefold expansion in our labor force,[2] and enormous improvements in our methods of production between 1875 and 1945 can only mean enormously increased ability to produce. Putting all these things together, we can scarcely be surprised to find that the actual output of goods and services has expanded with great rapidity. In 1945 we produced goods and services valued at about $200 billion, whereas, in 1875 the goods and services produced would be valued at only $12 billion in the same prices. In these seventy years, therefore, our production of goods and services increased sixteen times over. This rate of growth is equivalent to a doubling of production every twenty years. If it

[1] Since we have selected years in which total output reached about the same level, the increase in output per man-hour measures chiefly the improvement in methods of production.

[2] Offset in part by a reduction in the length of the work-week.

continues, we should be able to produce about $400 billion worth of goods and services in 1970. Whether we actually do so or not will of course depend on whether we use all the labor and capital goods available to us at that time.

But the use of the productive factors has varied enormously, as the fluctuations in employment clearly show. If we had also examined the changes in hours of work, the variations would have been even larger. Hence we should not be surprised to find that the value of our total output has also fluctuated widely. In 1920 we produced $86.6 billion worth of output; in 1921, only $70.7 billion worth. In 1929 the total rose to $99.4 billion, but in 1932 it fell to $55.4 billion. Then by 1939 it had climbed to $88.6 billion, by 1941 to $120.5 billion, and by 1944 it had increased to the unprecedented figure of $198.7 billion.[1] Since it measures how much we produce, this series provides one of the best possible indications of our economic well-being. And it has obviously been far from stable over the years.

The kinds of goods and services that make up this output have of course also changed enormously. One of the most striking shifts has been the decline in the relative importance of agricultural production. In 1870, more than half the people gainfully employed in the United States were engaged in agriculture. Hence we may suppose that about half our total output was agricultural. By 1900, the proportion so employed had fallen to about 35 per cent, and by 1930 to about 20 per cent. It is fair to conclude that agricultural production has declined in importance, when set against our total output, from something like 50 per cent in 1870 to below 20 per cent in 1930.[2] But while agriculture has become less important in our economy, manufacturing, trade, transportation, and the service industries have grown in importance. Manufacturing accounted for about 20 per cent of the total number gainfully employed in 1870, and for about 30 per cent in 1930. Trade and transportation accounted for about 9 per cent of the total in 1870 and something over 20 per cent in 1930. Our country, which in 1870 was predominantly agricultural, had by 1930 become specialized in manufacturing, trade, the service industries, and transportation.

[1] All these figures are measured in current prices, that is in prices of the year in question. This means that the variation in physical output is less since high prices are found when the value of output is high.

[2] Because productivity is relatively low in agriculture, agricultural products make up an even smaller percentage of the total output.

The most spectacular changes in the composition of our total output have of course arisen from the development of new products. Our output in 1900 did not include a single radio, a single vacuum cleaner, a single aeroplane, or a single pair of nylon stockings. In 1900 we produced no more than 4000 automobiles, whereas in 1920 we produced about 2 million, and by 1929 over 5 million. New industries, such as the electrical equipment industries, new textile industries, the automobile and aircraft industries, and many others, have made greater and greater contributions to the stream of our total output. Not only has our output grown immensely in the last seventy years; it is now composed of many things not even known earlier and of many other things whose relative importance in the total has changed enormously.

Changes in the Distribution of Income

There are adequate data on changes in the amount of labor and capital goods, and in the size and kind of production, but variations in the distribution of income are rather more difficult to measure, for unfortunately the figures on the distribution of income by families for one year are not strictly comparable with those of other years. There is some reason to believe, however, that the pattern of distribution did not alter greatly between 1935 and 1945. The poorest families, though better off in the latter year than in the former, received about the same proportion of the total income. Likewise the wealthiest families received about the same proportion of the total in the two years. But because the total income was much higher in 1945 than in 1935, the actual amount of income received was greater at both ends of the scale.

We can, though, measure more precisely the changes in the distribution of income by function. In 1944, about 72 per cent of the national income went to wages and salaries, as compared with 64 per cent in 1929. In 1944 corporation profits accounted for 6 per cent of all income, while in 1929 corporation profits had accounted for 9 per cent. The farmer received about 7 per cent of the total in 1944, and only 6 per cent in 1929. Other private businessmen received 8 per cent of the total in 1944, but they received 10 per cent in 1929. And the recipients of rent and interest received 7 per cent of the total in 1944 as against 11 per cent in 1929. It appears that labor, which in this connection includes all salaried employees as well as wage

earners, earned in 1944 more of the total income than it did fifteen years earlier. Shareholders, other businessmen, and the recipients of rent and interest got a smaller share of the total. It is doubtful, however, whether this change represents a real trend; it may be simply the result of the fact that 1944 was an abnormal year. At any rate, in the distribution of income as in other ways, we should expect to find the economy constantly changing.

Changes in the Number of Firms

Our resources and productivity have increased. The composition of our output has altered. And as we shall now see, there have been sharp changes in the number and identity of business firms in the country. The number of firms listed by Dun and Bradstreet stood at 1,335,000 in the decade between 1900 and 1909. By 1929 there were 2,213,000 firms on their list. During the depression the number fell to just under 2 million, and by July 1938 had risen again to 2,102,000. The number remained very steady between 1939 and 1945. But while the number of firms was about the same, the make-up of this total had evidently altered considerably. We can get some measure of this by noting the figures for new and discontinued businesses in these years. Between 1939 and 1945, almost 2 million firms went out of existence, and about as many new ones were organized. Probably a good number of the firms that were discontinued between 1941 and 1943 had been organized between 1940 and 1942, for the mortality rate for new businesses was a great deal higher than for established ones. Nevertheless there is no doubt that a considerable change occurs from year to year in the personnel and identity of the firms in the economy. And as we have seen, the total number of firms in the country also varies considerably.

Changes in the Structure of the Economy

It would be most surprising, in view of the changes which have taken place within our economy, if its over-all structure had remained unchanged through all these years. In actual fact, the changes in its form have been striking. In the late eighteenth century the economy was chiefly agricultural, and most of its products were destined for consumption on the farm or for trading within short distances. There was some trade in the modern sense, of course, particularly with foreign countries. But manufacturing and the service industries, which

by 1945 bulked very large in the total, were of little or no importance. Ours was then a subsistence economy; goods were produced chiefly for the use of the producer. Today, on the other hand, goods are produced to be sold on the market, and subsistence production has practically disappeared. This change by itself is a significant and profound revolution. That it should have occurred within one hundred and fifty years, and the greater part of it within the last seventy-five years, is a fact which is particularly worth noting.

One hundred and fifty years ago most of our production was on small farms. Today the greater part of it is in large business firms. As the subsistence economy has given way to the modern economy in which production is mainly for the market, and as there has been an increase in the importance of the large capital-using industries — the iron and steel and automobile industries, the railroads, and many others — so the place filled by the large business firm in our economy has become greater and greater. The change in the form of our economy is not more important than the change in the size of the atoms that make up the economy. While one hundred and fifty years ago most of our production was spread over a relatively large number of very small firms, today the bulk of it is concentrated in a relatively small number of very large firms.

Moreover, as the size of firms has grown, so has their structure altered. The small subsistence farm was essentially a one-man enterprise, and hired labor was relatively unimportant. The farmer did his own work, supplied the very small amount of capital he needed, and carried on his own business by himself. As a corollary of this, the wage-earning class was relatively unimportant in this economy. Today, however, the usual business firm presents quite a different picture. The large corporation hires thousands or even hundreds of thousands of wage earners. Funds are made available to it by hundreds or even hundreds of thousands of individual investors. The modern corporation is a form of business organization which did not really develop until after 1870, though by now it is by far the most important type of business concern in this country. The large corporation has displaced the small one-man firm.

Summary

Our economy should not be regarded as static and unchanging. Not only have its resources grown enormously during the last hundred

years, but perhaps even more notable, its very structure has changed drastically. With the decline in the relative importance of agriculture and the increase in that of manufacture, trade and commerce, and the service industries; with the rapid growth in the stock of capital goods available to us; and with the rapid improvements in methods of production, the simple economy of a hundred years ago has passed away. The modern one which has taken its place is dominated by business firms most of which are corporations. For while the majority of the business firms are relatively small, the bulk of the business is carried on in almost all industries by very large corporations. The very fact that the form of the economy has changed enormously in the last hundred years gives us no reason to suppose that it has now become immutable and will forever remain as it is today. It is hard to believe that after a century of rapid change, the economy should suddenly freeze into its present form in the middle of the twentieth century. We may expect changes in the future as we have had them in the past.

3

The Business Firm

IF WE WERE TO COMPARE a reconnaissance photograph of the American economy with one of the Russian economy, we would be impressed by one striking difference. In our economy, there are millions of independent business firms. In the Russian economy, the private business firm does not exist. The presence of the private, independent business firm is the most important feature of a capitalist economy. Indeed, the firms are the building blocks of which a capitalist economy is made. Since this is primarily an inquiry into the structure of our own economy, it is therefore important to consider certain characteristics of these building blocks. How many business firms are there in the economy? How big are they? What is their structure? And what do they do? Questions like these are the subject of the present chapter.

The Business Firm and Capitalism

We have said that the business firm is the basic unit in a capitalist economy. Its importance derives from the fact that decisions about output, price, the level of employment, the methods of production, and a great number of other things that influence our economic well-being are made within the individual firm. When the output of a commodity is increased, this happens because the firms that produce it have decided to produce more. When prices rise, they do so because the managements of business firms have decided to raise them. In a capitalist economy there is no over-all government planning board which determines the total output of an industry, the prices

it will charge, or the level of employment. These decisions are made within the millions of independent business firms throughout the country, and what happens in the economy is no more than the sum of what these millions of business firms decide to do. To understand the capitalist economy we must therefore understand the business firm.

This point cannot be overstressed. Perhaps the easiest way to bring it home is to contrast the situation in our economy with that in a socialist economy, such as the Russian or Czechoslovakian. In the Russian economy the decision to produce, let us say, 20 million tons of pig iron, is made by the Central Planning Board, which presumably takes into account the needs and resources of the Russian economy before it comes to a decision. The same board determines how many automobiles to produce, how many pairs of socks to manufacture, and how many acres to put into wheat. In our economy no such institution exists. No one group or person determines how much steel to produce, how many tractors to make, or how much land to plant in cotton. The United States Steel Corporation determines how much steel it will produce, the Bethlehem Steel Corporation determines how much steel it will produce, and the other steel companies make their decisions independently. The total output of the steel industry is simply the sum of what the several steel companies decide to manufacture. The output of a commodity like wheat is even more complex in its determination. Instead of five or ten producers who effectively control the output of the industry, there are millions of independent wheat growers, each of whom determines, on his own, how much land he is going to put into wheat. In a socialist economy, important questions of output, price, employment, and so on are planned collectively. In a capitalist economy, these decisions are made separately by individual firms. That is why it is of critical importance in a capitalist economy that we understand the operations of the private business firm.

How does the business firm determine the amount it will produce? The answer to this question is to be found in the fact that the business firm in this country is privately owned. No matter what its form, whether a partnership, a proprietorship, or a corporation, the firm is owned by private individuals. The determination of how much to produce, or of the price to be charged for the product, is made with one interest in mind — that of the owner. The owner's interest is to secure as large a profit as possible out of the business firm, for this

profit is the source of his income. Thus in the business firm decisions are made with a view to maximizing profits.

An extract from testimony presented before the Senate Committee on Patents in 1942 will illustrate this fact. Mr. C. Fath was counsel for the committee; Mr. Z. Jeffries appeared for the Carboloy Company, a subsidiary of the General Electric Corporation. This is a part of the examination:

> Mr. Fath (*interposing*): You try to make as much profit as you can at the same time, of course?
>
> Mr. Jeffries: Naturally. . . . I think one must say that individuals by and large try to receive for their services as much as they can. . . . And so it is with corporations. They are in business primarily for the purpose of making money so far as their charters are concerned.

The business firm exists to make money for its owners — to make as large a volume of profit as possible. But it must not be supposed that to seek profits is an act of villainy. As we have said, the owners of the business firm derive their income from the profits of the business they own. Naturally everyone wants to make as much income as he can. A man will drop one job to take another with higher pay. The investor seeks to put his money where it will earn the highest returns. These actions are not censured. And the businessman who attempts to maximize the profits of his firm is doing essentially the same thing; he is attempting to maximize his own income. But his income is of special importance in a capitalist economy. Because he determines how much his firm will produce and the price it will charge, his efforts to maximize profits have a unique significance for all of us. It is because of this motive that a capitalist economy is often described as a profit economy. The first thing to realize about the business firm, then, is that its decisions are made with a view to maximizing profits.

The Number and Size of Firms

Privately owned, profit-seeking concerns — the structural units of a capitalist economy — are very numerous, as we have seen. Business firms numbered between 10 and 12 millions in 1937 if farmers and private professional men are included in the total. Omitting these two categories, the number of business firms in December of 1945 was estimated, as noted earlier, at about 3,235,000. And these firms are widely different in size, for most of them are very small though the relatively few large ones are immensely important. Including

farms and professional units, the make-up of the economy by size of business firms was about as follows in 1937:

TABLE 7

Distribution of Producing Units and Their Employment by Number Employed: 1937 *

Number Employed	Number of Producing Units	Per cent of Total Employed
1 to 5	9.4–10.9 million	30–35
6 to 299	696,564	28–35
300 to 999	11,762	9–12
1000 to 9999	3,549	12–16
10,000 and over	246	11–14

* Source: *Structure of the American Economy*, Part 1, National Resources Committee, 1939, p. 99.

Thus in 1937 there were roughly 10 million firms that employed between one and five people. But though these accounted in terms of numbers for perhaps 93 per cent of all firms, they employed only 30 to 35 per cent of the active labor force. At the other end of the scale, 246 firms, each with 10,000 or more employees, gave work to between 11 and 14 per cent of all those employed in 1937. Only one-tenth of one per cent of the total number of firms employed as many as 300 workers in 1937, though they gave work to as many people as the 93 per cent of firms employing one to five persons. The obvious importance of the very large firm is clearly brought out in this table.

If we disregard the farmer, the non-profit public services, and the very large number of enterprises which had no employees at all, the importance of the large firm in our economy becomes even clearer. According to evidence presented by the Department of Commerce to the Temporary National Economic Committee of the United States Congress, 50 per cent of the 1,730,000 firms surveyed had three employees or less, but these firms employed only 4 per cent of all the workers in the firms covered by the survey. And 75 per cent of the firms surveyed employed only 11 per cent of the workers. In contrast, one one-hundredth of one per cent of the firms employed 12 per cent of the workers. Thus, the 195 firms represented by this tiny percentage employed a slightly larger number of workers than 75 per cent — or 1,260,000 — of the smaller firms. The importance of these

titans in providing employment and in producing goods and services can hardly be overemphasized.

In some industries the big firm is predominant; in others it plays a very small role. Its place in specific industries is clearly shown by the following table, which gives the percentage of the total output that a small number of large firms controlled in the years designated.

TABLE 8

Industries Where Large Firm Is Dominant

Product	Number of Firms	Percentage of Output
Tin cans	4	88.8 (1939)
Magnesium	1	100. (1938)
Sewing machines	4	79. (1935)
Automobiles	4	88. (1935)
Nitric acid	4	87. (1945)
Synthetic nitrogen	2	49.6 (1945)
Adding machines	10	100. (1944)
Cigarettes	4	90. (1935)
Rayon	7	86. (1945)

Profitability of Small and Large Firms

Small firms tend to be more profitable than large ones when the total income of the economy is high. But when the total income is low large firms do better than small ones.[1] Thus, in 1932, almost all firms suffered losses, but generally speaking, in that year the smaller firms lost by a greater percentage than the larger ones.[2] For firms with assets under $500,000 the average profit was minus 6 per cent in that year. For firms with assets between $500,000 and $1 million, the rate of profit was minus 3.3 per cent. And for firms with assets in excess of $50 million, the rate of profit was plus .7 per cent. Thus in a bad year like 1932, the larger the firm the greater is its profitability. In a good year like 1940, the situation is reversed. Firms with assets below $50,000 earned profits (including compensation of officers) of 36.2 per cent. Firms with assets between $500,000 and $1 million earned profits of only 11.9 per cent, while firms with assets exceeding $50 million earned profits of only 6.7 per cent. Generally, if the compensation of the firm's officers is included in the profit figure, profitability is higher for small firms than for large ones in a good

[1] *Survey of Current Business*, United States Department of Commerce, January, 1946.

[2] We measure profitability as the ratio of net profit before payment of taxes and officers' compensation to the total equity of the firm. It is desirable to include with total profit the salaries of corporation officers, since particularly for small firms some of these salaries can be regarded as profits.

year like 1940. If the officers' salaries are not counted as part of profit, profitability does not greatly vary with respect to the size of firms except for the smallest, which even in good years make relatively low profits.

The Organization of the Firm

Firms differ widely in type of organization. There are three main classifications and in addition several subtypes. The main types are the individual proprietorship, the partnership, and the corporation. The relative frequency of these types varies greatly from industry to industry. As would be expected, the proprietorship is the most common type in agriculture and the professions; almost all farms, and doctors' and dentists' offices are organized in this way. Omitting agriculture and the professions, there were about 1,500,000 firms organized as proprietorships in 1937, while there were about 230,000 partnerships and about 540,000 corporations. Thus, even in other sectors of the economy, the proprietorship is the most common form of business. But even though proprietorships are more numerous than either corporations or partnerships, the corporation carries on the bulk of the business in most industries. The following table illustrates clearly the importance of the corporation in different parts of the economy.

TABLE 9

Importance of Corporate Activity by Branches of Industry: 1937 *

Industry	Per cent of Business Done by Corporations in Each Industry
Agriculture	7
Mining	96
Electric light and power and manufactured gas	100
Manufacturing	92
Contract construction	36
Transportation	89
Communication	100
Trade	58
Finance	84
Service	30
Miscellaneous	33

*Source: U.S. Department of Commerce; *Hearings before Temporary National Economic Committee*, Part I, 1939.

The corporation is seen to be predominant in every major industry except agriculture, contract construction, and the service industries. This, of course, does not mean that incorporated firms are more numerous in these industries than are other types of organization; it means only that most of the business is carried on by incorporated firms.

The Individual Proprietorship

The individual proprietorship, we have seen, is the most common type of business firm. Most farms are operated under this form of ownership, and so are many firms in retail trade, the service industries, and the professions. The structure of such a firm is very simple. Essentially, it is a one-man business. It has an individual owner, who generally supplies most of the funds required, who directs the firm, and who receives all the profits since he owns and controls the business. Technically, the firm is not distinguished from the man who owns it. It lives just as long as he wishes it to live. If he retires from business, it goes out of existence. Its assets are his assets, and its debts are his liabilities. Thus it is simply the business aspect of a man in business for himself.

Because the firm and its owner are not separated, its creditors need not distinguish between the assets of the firm and the assets of the owner. They can claim the owner's assets if the firm owes more than it can pay out of its own. Thus the creditors of the firm can take the owner's private automobile or his golf clubs to satisfy a debt of the firm. For this reason men of wealth are unlikely to put any of it into an individual proprietorship. And this is understandable, since the whole of their wealth could be taken to satisfy outstanding claims.

Naturally, this type of liability limits the size of individual proprietorships. Since such a firm is owned by only one man its assets cannot be greater than the assets he can command. Hence these firms are generally small to begin with, and the fact that they ordinarily have difficulty in borrowing large sums tends to limit their growth. It is for this reason that individual proprietorships are most common in industries such as agriculture, barbering, and others where a relatively small firm compares favorably in efficiency with larger ones. But in most industries, the firm must be moderately large in order to take advantage of modern production techniques, and therefore the individual proprietorship is not found. Consequently the great bulk

of the country's business is done by other kinds of firms. The fact that control is undivided and that the profit of the firm belongs only to the owner makes the proprietorship attractive to many. But the fact that all the owner's personal as well as business assets are available to satisfy the claims of creditors is an unattractive feature. The one-man business is common, but its role is severely limited. It cannot be expected to play an important part in most industries.

The Partnership

The partnership is modeled closely on the individual proprietorship, though it has two, three, four, or many more owners who share in its control and in its profits. Since a partnership calls upon the wealth of more than one man, it can obviously be a good deal larger than an individual proprietorship. Hence it can often overcome one of the prime difficulties of the one-man firm in the modern economy — that of raising enough money to finance a large plant. But the partnership faces certain other difficulties that in general limit its role in the economy. All the partners in the firm are liable for its debts. In the first instance they are liable only to the extent set forth in their partnership agreement. Thus, one partner may be responsible for one-third of the debts of the partnership, a second for one-ninth, and a third for one-sixth, while the fourth may be responsible for the rest. But the validity of the terms of such an agreement for any one of them depends upon the ability of the others to pay. If the assets of the partner responsible for one-third of the firm's debts should be insufficient to meet his share of the claims of the creditors, then the other owners can be compelled to make good. In the final analysis, therefore, each partner is responsible to the full extent of his wealth. Moreover, the business acts of any one partner can bind the firm, so that each partner is dependent upon the good sense and integrity of all the others. For that reason, a wealthy individual is generally unwilling to enter a partnership unless he is quite certain of the character and business sense of the other partners. This means that a partnership encounters difficulties in getting adequate funds which are not very different from those encountered by an individual proprietorship.

The partnership has still other disadvantages. Like an individual proprietorship, it is mortal. If a partner dies or withdraws from the business, the firm itself comes to an end. This, of course, does not mean that the firm's equipment disappears into dust. But it does

mean that a new agreement has to be worked out, and that the claim of the ex-partner upon the net worth of the firm has to be satisfied. Hence when a partner dies, it may be necessary to sell certain assets in order to get the funds to pay his estate for his share in the business. It may, however, be unprofitable to sell the assets at such a time. Furthermore, it may prove difficult to attract other funds into the firm. Hence there is a danger that the business life of the firm may be harmfully interrupted.

But a partnership also has certain advantages over a proprietorship. Since more than one man is engaged in the operations of a partnership, control can be specialized. One partner may devote his attention to production, another to financing, a third to selling, and a fourth to labor. With such specialization, production may be more efficient than in an individual proprietorship. In this respect, then, the partnership may have a definite advantage. Its greater ability to secure funds has already been indicated. In fact, these features may sometimes offset the disadvantage of the unlimited liability of each partner for all the debts of the firm.

The Corporation: Introduction

The corporate form of business organization was devised to make possible the raising of the immense sums of money needed to run a large modern firm. Though this type of business organization, as we have seen, is not the most common in this country, it carries on by far the greatest volume of business. It is therefore worth while to give a more detailed description of the corporation than of the other two types of business organization.

The characteristic feature of the corporation is that it is separate from those who own it. It is an institution created by the state, it owns its own assets, and it must meet its own liabilities. The owners of the corporation — the stockholders — are not the corporation. They are not responsible for its debts, nor can they claim its assets while it is in operation. The American Telephone and Telegraph Company had over 650,000 stockholder-owners in 1935. But they were quite separate from the company. The company itself had an independent existence. Because the corporation is not identical with its owners, their liability is limited. This is a fact of the greatest importance. Unless the stockholders owe money to the corporation on account of shares for which they have not yet paid in full, their

property cannot be touched to meet the claims of its creditors. If the assets of the corporation are not enough to meet its liabilities, the owners are not liable for the debt, as are the owners in both the individual proprietorship and the partnership. The owners of the corporation are protected by this limited liability feature, which makes it possible for them to put money into the corporation without having to fear the consequences of its failure. They will, of course, lose their investment if the firm fails. But the rest of their wealth will not be liable to seizure by the creditors. Hence a wealthy individual who may be reluctant to invest in a partnership might be willing to put his money into the same firm if it had been organized under the corporate form. Since such a firm can attract the wealth of very large numbers of people, it can easily command far more wealth than the other forms of business organization.

For this reason the corporation is common in all industries where a large plant and heavy investment are necessary. In many industries, it is uneconomical and inefficient to operate a very small plant. A steel mill could never be as small as a corner drugstore. Likewise in the automobile industry, the aluminum industry, copper refining, shipbuilding, the aircraft industry, the public utilities industry, and in many others, the corporate form of business is practically essential. For only the corporation with its limited liability feature can bring together the immense wealth needed by firms in many types of industries.

But for all its advantages, the corporate form also has certain drawbacks. For one thing, the federal government and most state governments tax the income of corporations, frequently at a high rate. Moreover, the owners of the corporation have to pay an additional tax on their personal income received as dividends, whereas if they had invested the same amount of money in a proprietorship or a partnership, they would have had to pay only one income tax, that on their income as owners. Perhaps the tax on the income of the corporation can be looked upon as the price its owners pay for the limited liability protection granted by the state. A second possible disadvantage is that the corporation may have to make a public report of its financial status, whereas neither of the other types of business has to do so. Moreover, certain expenses have to be met to incorporate a firm. For these reasons, many businesses do not choose the corporate form. The added costs may more than offset the advantages of limited liability.

The Control of the Corporation

Now, what about the control of the corporation? Who runs it? The owners, in principle. For they elect the directors, who appoint the managers, who make the day-to-day decisions on which the corporation operates. And thus, in theory, the owners of the corporation run it just as the owners of a partnership run their business. But the facts of the matter are very different. The American Telephone and Telegraph Company, as we saw, had 650,000 owners in 1935. Most of these owners knew nothing about the affairs of the company and were satisfied as long as they received their dividends. They were invited to the stockholders' meetings, and if they attended they were able to vote. But their interest in these meetings was less than lukewarm. Although the stockholders have the right to vote at meetings of the corporation and to elect the directors, they very rarely exercise that right directly. It is quite uncommon for the ordinary stockholder to be present at the annual meeting of his corporation. Incidentally stockholders do not share equally in the right to vote, for the individual vote actually adheres to the share of stock rather than to the shareholder. Thus if a person owns one share in the corporation, he has one vote; if he owns ten thousand shares in the corporation, he has ten thousand votes. Hence, especially for the small shareholder in this "dollar democracy," who is not likely in any case to be informed about the operations of the corporation, this right to vote is practically meaningless.

How then is the business of electing directors carried on? With the announcement of the annual meeting, each stockholder receives a slip of paper known as a proxy form, by the use of which he is invited to assign his vote to one of the officials of the corporation. If he returns the proxy form, the designated official may vote his shares If he fails to return the form, his shares are not voted, unless he himself appears at the meeting. Thus the proxy form is rather like the ballot used by the Nazis in Germany with room on it only for a "Ja." Yet is is hard to see how in the corporation any other device could be more democratic than this.[1] Because so large a number of the shareholders in a corporation own so few shares individually, and because their knowledge of the operations of the corporation is necessarily

[1] The Securities and Exchange Commission now requires a full disclosure of information about any proposal to be raised at the annual meeting, so that the shareholder may vote properly. Moreover, the proxy form must give him an opportunity to vote "no."

restricted, it is not clear that their inability to vote personally is a great loss, either to themselves or to the economy.

As a consequence of the way in which the voting machinery works, real control of the corporation is very often in the hands of management — the officers and directors. The management may own a relatively small amount of the total stock. Yet, because it controls the voting machinery, it may be in a position to dominate the affairs of the corporation. So long as the management is able to earn profits for the corporation, it is unlikely to meet with much opposition from the other stockholders. And even if the stockholders were not satisfied, they might be unable to vote effectively against the management. They would have to pay to send literature to the other stockholders to ask for their support and permission to vote their stock at the annual meeting, out of their own pockets. The cost of doing this is so high that usually stockholders who do not like the way the company is being run, simply sell their stock. Unless they have many shares of stock and consequently a large number of votes, they have very little chance of defeating the management.

In large corporations the separation of ownership and control is particularly marked, for even though the stockholders' right to vote is not questioned, the control of the corporation is effectively in other hands. And yet management's share in the total ownership of the corporation is frequently very small. In 61 out of 155 very large corporations, the management owned less than 1 per cent of the total amount of stock outstanding.[1] In another 30, the management owned between 1 and 3 per cent of the total. Thus in 91 out of 155 of the largest corporations in the country, the management owned less than 3 per cent of the total stock outstanding. And in only two of these corporations did management own more than 50 per cent of the total stock. But generally, in spite of its small ownership stake, management was in a position to control these firms. The National Resources Committee summarizes the situation in this way:

> It is clear, therefore, that for most of the largest corporations ownership and control have become largely separated. This condition appears to be particularly characteristic of the corporations which have travelled furthest along the road of corporate development, such as the railroads and others of the older corporations. The lack of significant

[1] Compare *Structure of the American Economy*, National Resources Committee, Part I, p. 157, 1939

> stockholder control over corporate policy may be regarded as the typical condition toward which the large corporate unit has been tending. . . . Since the owners of the larger corporations do not in most cases exercise a significant degree of control over corporate policy, attention must be shifted to the management which is at the centre of the sources influencing the policy formation. The officers and directors of the corporation are responsible for the development of policies and their execution. Together, the officers and directors are usually in a position to exercise a large measure of control over corporate affairs.

We may therefore conclude that although the control of the corporation is legally in the hands of the stockholders, in many large corporations it is actually in the hands of the management. And when control is not in the hands of management, it is usually exercised by stockholders who are in a position to dominate the firm's affairs because they own more stock than anyone else, though perhaps much less than 50 per cent of the total.

Types of Stock

The owners of corporations may be divided into a number of types. A broad classification is that between the owners of common stock and the owners of preferred stock, though in many corporations there are also subtypes. The essential difference between common and preferred stock is that the owners of preferred stock have a first claim upon any dividends which are to be distributed. It is as if priorities were set up, one class of stockholder having first claim upon dividends, another class having a second priority, and so on. The common stockholder who has the lowest priority therefore takes the greatest risk of not receiving dividends. But as we shall see, he also has the greatest chance of receiving very high dividends. Preferred stock usually has a par value, for example $100, and it bears a set rate of return, such as 6 per cent per annum. Common stock, on the other hand, may or may not be given a par value.

Let us consider an actual case to see how all this works. The Minneapolis-Moline Power Implement Company, which manufactures agricultural equipment, had outstanding 100,000 shares of cumulative preferred stock with no par value and with a dividend rate of $6.50.[1] In addition, it had 700,000 shares of common stock with a

[1] If the preferred stock had had a par value of $100 a share and had carried a dividend rate of 6.5 per cent, the situation, in respect to dividends, would have been the same.

par value of $1 outstanding. In 1938 it made profits of $727,000. Because the owners of preferred stock had a first claim on dividends, the corporation would have had to pay them $650,000, or $6.50 a share, before it could pay anything to the owners of common stock. Actually it declared dividends of only $642,000, and the owners of the preferred stock, since they had first claim, had a right to the whole of this sum, so that no dividends could be paid to the owners of common stock. If the firm had found it possible to declare $1 million in dividends, it would have been obliged to pay $650,000 to the owners of preferred stock before it could pay its other shareholders the remaining $350,000. And since there were 700,000 shares of common stock outstanding, the dividend rate would have been 50 cents a share. The owner of common stock thus receives dividends only if the total amount paid is large enough to satisfy the full claims of the preferred stock, and to leave something over.

But actually the situation in this company was not quite so simple. It will be noticed that the preferred stock of the company is called cumulative. This means that if dividends are not paid in any one year the claim from that year must be satisfied before the owners of common stock can secure any dividends even in later years. To illustrate: In 1934 and 1935, the company declared no dividends, and as a result the owners of preferred stock piled up a claim of $1.3 million, that is $650,000 a year, for two years. Then in 1936 the dividends amounted to only $296,000, so that the owners of the preferred stock built up their claims for the three years to $1,654,000, or about $16.54 a share, which would have to be paid before the owners of common stock could be paid anything at all. Thus, cumulative preferred stock not only has priority rights in each single year; the priority rights carry over into subsequent years. And the owners of common stock in a corporation which has issued cumulative preferred stock have but a very small chance of receiving dividends if for some years previously none have been paid on the preferred stock.

We have seen that if the Minneapolis-Moline Company declared dividends of $650,000, the owners of common stock would receive nothing, and if it declared dividends of $1 million, the owners of common stock would receive 50 cents a share, assuming that the preferred stock had accumulated no unpaid dividends. If total dividends of $3,450,000 are paid, the owners of common stock will receive $4 a share, since only $6.50 is paid on every share of preferred stock.

And obviously, the higher the total, the more the common stock will earn. In short, the owners of preferred stock are much more certain to get something; but their dividends cannot exceed a fixed amount. The owners of common stock, on the other hand, are quite uncertain about receiving any dividends at all; but when dividends are paid, they may be very high.

The pattern of the capital structure or of the security issues outstanding for some companies may be very complex. The Associated Gas and Electric Company at one time had three kinds of common stock, six kinds of preferred stock, and four kinds of preference shares outstanding. In addition there were no less than twenty-five different types of bonds and debentures on which it owed money. Armour and Company (Illinois) had outstanding in 1938 the following securities:

TABLE 10

Securities of Armour and Company (Illinois): December 31, 1938

Type of Security	Dividend Ratio per Share	Number of Shares
Preferred Stock — cumulative, convertible, prior, no par value	$6	532,996
Preferred Stock — cumulative, $100 par value	7 per cent	33,715
Preferred Stock — cumulative, guaranteed, $100 par value	7 per cent	557,825
Common Stock — $5 par value	–	4,965,992

Each of these types of preferred stock occupied a certain position in the hierarchy of priorities. But it is clear that it would take a rather careful reading to determine exactly what rights and privileges each issue carries.

We have already mentioned that the control which the owners of different kinds of shares exercise in the corporation may vary. Frequently the owners of preferred stock have no vote whatsoever. Sometimes there are various classes of common stock, some of which carry a vote while others do not. In 1938 the Hercules Powder Company had outstanding 1,341,000 shares of common stock which carried voting rights, and about 14,000 shares of common stock which did not. Sometimes the owner of preferred stock is allowed to vote

on certain broad questions, such as changing the privileges of the preferred stock. Thus the privileges as well as the types of stock are widely varied.

Corporation Bonds

Companies may do a great deal of their financing by selling bonds rather than stock. The purchaser of a bond, however, is not an owner of the corporation; rather, he is a creditor. Interest paid on bonds is not regarded as a distribution of profit as are dividends on stocks. Instead, it is an expense or cost of the corporation. It follows that failure to pay interest on bonds is the same as failure to pay any other creditor. The owners of the bonds can foreclose on the company and force it into bankruptcy. But since they are not owners of the corporation, they generally have no voice in its control. Sometimes, however, if interest on the bonds falls due and is not paid, the bondholders are given the right to elect some members to the Board of Directors of the corporation. But the owners of the bonds have no more voice in controlling the corporation than any other creditors so long as the interest due them is paid.

Bonds are of many different kinds. For instance, the Bethlehem Steel Corporation (Delaware) had outstanding in 1938 the following types in the amounts given below:

TABLE 11

Bonds of Bethlehem Steel Corporation (Delaware): Dec. 31, 1938

Type of Bond *	Face Amount
Serial gold bonds due to 1941, 4½%	$ 3,517,000
Consolidated mortgage, series "D," due 1960, 4¼%	$52,590,000
Fifteen-year sinking fund convertible debentures, due 1952, 3½%	$48,000,000
Consolidated mortgage, series "E," due 1966, 3¾%	$53,022,000
Purchase money mortgage gold bonds, due 1998, 6%	$ 7,500,000

* It is not important at this stage to understand the distinction between these various kinds of bonds.

Conflicts of Interest Within Corporations

Although the corporation is regarded in law as a legal person, it sometimes looks more like a battle in a civil war. Because ownership and control in the modern corporation are separated, and because there are so many kinds of ownership, conflicts of interest may develop between one kind of owner and another or between owners and management. These conflicts seldom lead to serious and open disagreement except when the corporation is being reorganized or when it is earning very low profits. Nevertheless, they may arise at any time. The basis for these conflicts is not hard to understand. Stockholders naturally want their corporation to earn maximum profits. Management, on the other hand, often holds very little stock, and derives most of its income from salaries or from other business interests. Now, higher salaries for management do not obviously lead to higher profits for stockholders. So when management's salaries are raised from $707,000 in 1936 to $848,000 in 1937, as they were for a well-known motion-picture producing firm, stockholders' profits are lowered by about $141,000, or so it would seem. This is not to say that management is not justified in getting salary increases, or, for that matter, that the stockholders necessarily oppose the payment of higher salaries. The stockholders may be persuaded that by paying higher salaries they can get better management and so earn even higher profits in the long run. But generally stockholders are not asked for their opinion on such matters. Certainly they are not asked in a great number of very large corporations, in which after all no single stockholder controls a substantial part of the total stock and which therefore are practically dominated by management. In short, the interests of management and of the stockholders do not always coincide. Indeed, such conflicts have sometimes become so bitter that stockholders have appeared at corporation meetings in order to object to management salaries. Usually, however, since they cannot muster enough votes to overrule the management, they have to be satisfied with the opportunity to voice their objections.

There are other ways in which the interests of management may be at variance with the interests of the stockholders. The corporation officers may derive the major part of their income not as salary but from other sources. For instance, it is not uncommon for large corporations to have on their boards of directors, and possibly in positions of executive control, representatives of large banks. When the cor-

poration requires financial services, as in floating new securities, there is a natural tendency for it to give its business to the bank represented on its board of directors, even though the interests of the stockholders might be better served if the business were given to some other bank. But since the advice of stockholders is rarely sought in such matters, the conflict of interest may here also be resolved in favor of management. Again, management may provide legal services, engineering services, or possibly raw materials. This also can be a source of conflict, for management will seek to get as large a volume of business from the corporation as possible, whereas the interests of the stockholders may best be served by bargaining for the services or materials required. Management, if it serves two masters, can hardly be expected to have the same objectivity or interests as ownership.

There is also likely to be a conflict of interest between the different kinds of owners of the corporation, most obviously between the owners of preferred and common stock. The owners of common stock may receive extraordinarily high dividends when total dividends are high, and they may get nothing when total dividends are low. If a corporation declares $1 million in dividends over a two-year period, owners of common stock will receive much more in total if the whole sum is paid in one year. Owners of preferred stock, however, can receive only so much in any one year, and the more the company pays out in that year, the less it is likely to have for the next.[1] Naturally, therefore, they will favor minimum dividends every year, so long as their own are paid. If $1 million is to be paid out in dividends over a two-year period, the owners of preferred stock would like to have the firm declare dividends of $500,000 in the first year, (or less, if it would meet their claim), and then wait. The owners of common stock, however, do best if the whole sum is paid out in a single year. When there are not just two kinds of owners but five or six, as there may well be, the conflict can become even more intense. Naturally those with the highest priority for dividends would like to keep the total dividends of the corporation low. If they could express their interest coherently, they would urge the corporation not to "squander" all its profits at once, but to accumulate for a rainy day. But the lower on the priority ladder the owner is perched, the more insistent would be his claim for a full distribution of profits in the year in which the profits are earned.

[1] When preferred stock has participating privileges, it shares to a limited extent in very high dividends.

Corporations that are not able to meet the claims of their creditors have to go into bankruptcy and to be reorganized. When this happens, creditors and owners have to revise their original agreements. The conflicts of interest at such times are bound to be intense. The bondholders want to lose as little as possible; the owners of preferred stock want to keep their priority rights for dividends, and they want the creditors of the corporation to scale down their claims as far as possible. The owners of common stock are also interested in maintaining their maximum possible claims to dividends. The greater the revision in creditors' claims, and the sharper the reduction in priorities on preferred stock, the greater will be the value of any ownership rights that may be left to the common stockholders. Furthermore, each group, including of course the original management, has a strong reason for wishing to assume control of the corporation. What happens at such times is more like a free-for-all than an orderly process of bargaining. The individual interests may sponsor their own advertising campaigns; "protective committees" may be set up — sometimes by "outsiders" who hope to use them to further their own interests — to protect the rights of the various claimants; the ordinary bondholder or stockholder is subjected to strong pressure to throw his weight on one side or another. Normally, of course, the conflicts of interest between different kinds of owners and between owners and management do not become sufficiently intense to be evident to the public. Like the well-bred family, the corporation does most of its quarreling in private.

Summary: Picture of the Typical Business Firm

The typical business firm in our economy is most complex. The business firm is anything from the little boy who shines shoes or delivers newspapers to the American Telephone and Telegraph Company. By far the largest number of business firms are very small, but most of the business of the country is carried on by very large firms. In only a few industries, such as contract construction and some service industries, does the small firm play the chief role. In almost all others, the large firm predominates. Business firms vary both in size and in nature. Numerically most of them are simply one-man businesses, run by the owner, who very often does all the labor and supplies all the funds which the firm may require. Such businesses are called proprietorships, and are found principally in agriculture,

the professions, service industries, and retail trade, but since they have certain disadvantages in the modern economy they are unimportant in most other industries. Their principal difficulty is in securing funds, for they are limited to the funds of the individual owner and the money he can borrow, and so they can seldom grow to any great size. Moreover, they have one further disadvantage. A person of considerable wealth is not likely to invest his funds in such a firm for the simple reason that all his wealth can be seized by creditors if the firm becomes bankrupt.

The second type of business organization, the partnership, is in most respects similar to the individual proprietorship, but there is this difference, that two or more owners generally share the control of the firm, and its profits, on some prearranged basis. The partnership also is limited in size, for it has access only to the wealth of the partners, and a man of considerable wealth will probably hesitate to enter a partnership since he may be held liable for all of its debts. Moreover, the life of the firm is limited to the lives of its owners. When one dies or retires, the firm itself comes to an end. New arrangements have to be made between the remaining partners, and often some of the assets of the firm have to be sacrificed to meet the claims of the retiring member.

The incorporated firm is designed to meet these objections and difficulties. The distinguishing characteristic of the corporation is that its owners are not responsible for its debts, since it has a legal existence of its own and it may hold assets and owe money. If its assets are not adequate to meet all claims against it, the creditors can not lay claim to the wealth of the owners of the corporation. The protection given by legally limited liability means that those with money to invest in a business may put it in a corporation with much less risk than in any other kind of firm. For this reason the corporation has grown immensely popular. All the large firms in our economy are incorporated; indeed they have to be in order to raise the vast sums of money they need. The owners of a corporation are divided into various classes depending upon their priority rights in claiming dividends and upon their voting rights. Owners of common stock usually have voting rights, but they have the last priority in claiming dividends and may only be paid after the owners of preferred stock have been paid in full. When the corporation pays large dividends, the owners of common stock do very well indeed, but when

dividends are small, they may receive nothing at all. Thus they take the greatest risk of getting nothing, but they also have the greatest chance of getting a large amount. The owners of preferred stock, on the other hand, have greater security, for they must be paid in full before the owners of common stock get anything. Sometimes there are various classes of preferred stock owners, whose claims are arranged in a hierarchy of priorities.

Normally, the owners of preferred stock do not have a vote in the corporation, though often this rule is waived when the corporation fails to pay dividends on its preferred stock for a number of years. But generally voting rights in a modern corporation are relatively unimportant for the ordinary stockholder. Because he owns a very small fraction of the total stock, and because ownership is dispersed over a very large number of stockholders, management is usually able to dominate the affairs of the corporation. From its strategic position it can maintain itself in power unless it does something so flagrant as to arouse the opposition of thousands of individual stockholders. Since this rarely happens, ownership and control in the modern corporation are generally separate. But this is not always true — witness the General Motors Corporation, the Standard Oil Company of New York, and the Sun Oil Company, in each of which a substantial amount of the common or voting stock is owned by a single person or family. Finally, because of the diversity of interest between management, common stockholders, preferred stockholders of various kinds, and bondholders, who are really creditors, conflicts are not uncommon.

The business firm is the critical unit in our economy. But though its structure is immensely varied, we must not overlook the underlying similarity between firms of all kinds, from the corner drugstore to the largest corporations in the country: the fact that all exist to earn profit, and that they make all their decisions as to output, price, employment, and so on, with this end in view. It is because of this that we may make useful generalizations about their operations, and accordingly about the way in which the economy functions.

4

Government and the Economy

WE HAVE ALL SEEN newspaper cartoons in which a large, villainous-looking ruffian marked Government is shown tyrannizing over a little fellow labeled Taxpayer, or Consumer, or Labor, or Business. While the dominance of government is generally exaggerated in these cartoons, there is no doubt whatever that government plays a prominent part in the modern economy. Our picture of the economy is therefore incomplete until we have sketched in the role played by government. For, as we shall see, government performs an important function in even the most extreme capitalist economy. And in our own, the part it has played has become more and more important with the passage of the years.

There are various ways of measuring the importance of government in the economy. One way is to see how many jobs it provides. In March, 1946, a total of 36.3 million men were employed in non-agricultural establishments. About 5.5 million of them were on government civilian payrolls — almost one man in every six. Only twice as many as that were employed by manufacturing concerns, though most of us think of the economy as being mainly engaged in manufacturing. The public utilities, railroads, bus lines, and trucking concerns employed only about three-quarters as many as the government. And at that date another 4.4 million men were in the armed forces, and they also were in a sense in the employ of the government. Another way of showing government's role in the economy is to look at the total income payments made to individuals. In March, 1946, income payments to individuals were running at the

rate of about $158 billion a year. Wages and salaries were being paid by the government alone at a rate of $19.2 billion a year.[1] They thus accounted for about 8 per cent of the total income received. Government is clearly an important employer and an important source of our total income.

How does the government use the 5.5 million civilians it employs? In the first place, we must clarify a term. "The government" means not simply the federal government, but in addition all the state and local government bodies in the country, of which there are about twenty thousand. How are these 5.5 million men used by government bodies? A good number of them are employed by the Post Office Department. Many more are employed as teachers in schools, which are organized and financed by state and local governments. Still others are employed by police departments, fire departments, and so on. The federal government runs the largest publishing institution in the country — the Government Printing Office. The federal government also produces a certain amount of electric power and light, and many local governments do too. Government employees look after our national parks, maintain our highways, dispose of garbage, supply water to our cities, clean streets, and in many other ways provide services for which we as citizens have voted. Most government employees are of course not the bureaucrats that one sees pictured in the unfriendly cartoons. Most of them are just the policemen, or the postmen who deliver your mail, or the teachers you had in school.

In addition to being an extremely large employer, government has also been an extremely important customer of many of our industries, especially in recent years. In the nineteen-twenties, government bought between 10 and 15 per cent of all the goods and services produced in our economy,[2] and in the nineteen-thirties its purchases reached an even higher figure. In 1938 they amounted to 18 per cent, and by 1941 to 22 per cent of our total output. Finally, in 1944 the government purchased almost exactly 50 per cent of our total output of goods and services. The importance of the government as a customer for our business firms cannot be overestimated.

The regulations of government are also of great importance to our economy. The government, which is after all no more than an em-

[1] In addition, the government paid a large sum in interest on outstanding government securities held by private individuals.

[2] This figure, since it covers the services provided by government (education, protection, and so on) includes the amounts paid to government employees.

bodiment of our social conscience, has extended its regulations and controls over a wide range of our economic activities. State governments control the prices charged by public utilities; whenever a privately owned electric power plant or water supply system sends you a bill, you can be sure that the rates it charges have been scrutinized by a state regulating commission. State governments also set the rules by which corporations are governed. Corporations are organized in accordance with state law; their charters are granted by the state, and the activities in which they are permitted to engage are defined in the charter. The federal government controls the sale of a corporation's stocks and bonds on the securities market. When a new issue of securities is offered to the public, the terms of the offer must first be approved by the Securities and Exchange Commission. Our federal government grants monopoly rights in the exploitation of patents and restricts certain monopoly practices in other ways. It also controls the advertising of foods and drugs. It takes a part in determining the relations between labor and employer; it provides, for example, that employees shall be free to form unions of their own choosing and to bargain collectively about wages and conditions of work with their employers. The federal government also provides that the wages of those employed in firms engaged in interstate commerce shall not fall below a certain level. Government regulation has for years been important in agriculture. Through the Federal Reserve Banks the government exercises a considerable degree of control over commercial banks and the money market. During the war the government set maximum prices on large numbers of commodities, administered ration schemes, and in other ways extended its control over a large number of the activities of business firms. The federal government regulates our economic relations with foreign powers. Tariffs to keep out foreign goods have been an important factor from the earliest days. In fact, there is hardly a field of economic activity in which the government does not take some part. It is an important employer, an important customer — at times extremely important; and by reason of its regulations, it is a critical factor in the operation of business firms.

In view of all this, it may be asked whether it is correct to call ours a capitalist economy any longer. For is not the ideal form of capitalism a *laissez-faire* economy — one in which the government's role is at a minimum? It is impossible to answer this question, because the

answer must depend upon the way in which we choose to define a capitalist economy. However, we can learn from our own history, and from the experience of other countries, some facts about the role of government in a capitalist economy.

It is clear that our economy occupies an extreme position in regard to the part played by government. We have today the nearest thing to a *laissez-faire* economy in the world. The easiest way to illustrate this is to examine an economy at the other extreme, such as the Russian or the Czechoslovakian. In those economies, instead of hiring perhaps 10 per cent of all the employees in the country, the government is the chief employer. In the U.S.S.R., everybody works for it, and the worker in a steel mill is as much a civil servant as the postman or the bureaucrat in our own. Since the government hires all workers, it disperses all income. It sets wages and prices. There are no private firms in those economies, and so the government is the only determining factor. The world's other economies fall somewhere between the extreme positions occupied by ourselves and Russia. In Great Britain, for example, the government runs some of the most important industries, such as the coal industry, air transportation, and the public utilities; and the number is likely to grow. The pattern of government control is similar to this in most of the economies of western Europe. Thus it is clear that while the government is an important employer in this country, it is by no means as important as in other economies. To worry about whether we are a capitalist economy is of course only to worry about words; the fact of the matter is that we are much more nearly a capitalist economy than any other.

Some of us like to think that the part government now plays in the economy is something newfangled and abnormal. Nothing could be further from the truth. Government has never been a negligible factor in our economic life. The student of our economic history will recognize that government determination of tariffs and of other aspects of our economic relations with foreign powers began very early indeed The settlement of the West was greatly influenced by the Homestead Act and the land grants to the railroads. The extension of transportation over the country first by canals, then by railroads, and more recently by highways and airlines, has been greatly stimulated by government action. Government concern with the banking system does not date from 1913 when the Federal Reserve System was founded; our first central bank, the United States Bank, was founded

in 1791. In short, our economy has never been free from government control or uninfluenced by government activities; we have never known true *laissez faire.* It is clear, however, that the importance of government in the economy has obviously grown in recent years. Such a development is not unique to this country; governments have extended their participation in all other economies as well. A quick review of the development of our economy will suggest some of the reasons for this extension of government control. It will be remembered that a hundred years ago most of our economy was agricultural, and the rest of it was dominated by small firms. There was no large iron and steel industry, no big insurance companies, no automobile industry. The wage earner at that time was a rare specimen, and he had every prospect of not remaining a wage earner for very long. The economy was loosely knit, and every man was more or less independent and self-sufficient. Great depressions, like those of recent years, were simply not known. The range of problems that occupy the attention of modern government did not exist. But times have changed. About 75 per cent of all persons now engaged in economic life are employees. Giant industries have grown up, and in most sectors of the economy small firms are no longer very important. As the relative importance of agriculture has declined, that of manufacturing, transportation, and trade and commerce has grown. Our economy is now complex in the extreme. Partly because so many people are employees, and partly because the economy is so very wealthy, we are now liable to very severe depressions. With the growth of the large business firm and the increasing importance of hired labor, the relations between employer and employee have become more difficult and more complex. In such an economy, it is not surprising that the role of government has grown in importance.

Nevertheless, we have given up our psychological and philosophical predilection for *laissez faire* reluctantly. Most of us have not welcomed government intervention in economic life. We have called for it only when it seemed that the abuses and difficulties were too great to be borne. For example, we had a severe depression in 1920, some unemployment during the nineteen-twenties, and very heavy unemployment between 1930 and 1934. Yet we did not even inaugurate a system of unemployment insurance until after that year. We had banking crises, bank failures, and monetary disturbances for years before we created the Federal Reserve System. Regulation of the

security market was imposed only after a long record of abuse and corruption. We have adopted government controls reluctantly rather than eagerly. But it must be emphasized that it is we ourselves who have adopted the controls and sponsored the extension of the government's activities. The government does only what its legislative bodies — federal, state, and local — determine that it shall do. If they determine that there shall be regulation of the securities market, regulation of prices, or regulation of the monetary system, they do so because such regulation meets with the approval of the majority of our citizens. The majority may be right or wrong; that for the moment is not the issue. It is important to remember that in a democracy the scope of the government's activities is not determined by some dark power outside of us, but by our own majority will.

The government has played an increasingly important part in our economy. It is now, as we have seen, the employer of 5.5 millions of our workers. It is also one of our most important customers, and the source of many of our incomes. In short it is the controlling influence in many of our economic activities. It has come to this position not because the American people were predisposed toward an extension of government powers. Quite the contrary. As the nature of our economy has changed and as the problems that it has been compelled to face have altered and grown in gravity, we have been compelled to call upon the government — in other words, Ourselves Incorporated — to meet these new situations.

SUGGESTIONS FOR FURTHER READING

Listed below are references to a few of the most interesting and readable works on subjects discussed in the Introduction and Part One of this book. On the whole, this list includes only important books and articles which the reader, surrounded and indeed almost suffocated by textbooks, may overlook. This does not mean that reference to other texts is undesirable; obviously, many of these texts are very well worth reading. But on the whole it is believed that the student should become familiar with the original contributions of outstanding economists as soon as possible. On that account, this bibliography and subsequent ones at the ends of other parts of this book list, not textbook references, but rather readable and important contributions by recognized economists.

Arnold, Thurman W. *The Folklore of Capitalism.* New Haven: Yale University Press, 1937.

This book, though written by a lawyer, is a fascinating study in economics. Chapters 8–12 are particularly interesting. They should be read in connection with Chapter 3 of this text.

Berle, A. A., and G. C. Means. *The Modern Corporation and Private Property.* New York: The Macmillan Company, 1933.

The six chapters of Book I present the factual basis for recognition of the divorce between ownership and control.

Chase, Stuart. *Goals for America.* New York: The Twentieth Century Fund, 1942.

This is easy reading, interesting and useful. A clear account, because it is a clear example, of what economics is about.

Daly, F. St. L. "The Scope and Method of Economics," *The Canadian Journal of Economics and Political Science* (Toronto: May, 1945).

An economist writes about economics.

Pigou, A. C. *Economics in Practice.* London: Macmillan & Company, Limited, 1935.

The first lecture, "An Economist's Apologia," and the fifth, "State Action and *Laissez-Faire*," are especially recommended.

Robertson, D. H. *The Control of Industry.* New York: Harcourt, Brace and Company, 1923.

Especially chapters 1, 3, 5, 6, and 7. While the author writes chiefly about the British economy, much of what he says is also applicable to our own.

United States Government: National Resources Committee. *The Structure of the American Economy*, Part 1. Washington: Government Printing Office, 1939.

This report gives a relatively detailed account of the make-up of the economy. The student is advised to examine chapters 2, 3, 5, 6, 7, and 9.

United States Government: Temporary National Economic Committee. *Investigation of Concentration of Economic Power.* Hearings before the Temporary National Economic Committee, Congress of the United States. Washington: Government Printing Office, 1939–41.

The hearings before the TNEC (37 volumes and over 18,000 pages) contain more information about the economy than any other source, and for this reason they are invaluable. The student should look through Part One of the Hearings, and especially the testimony of Isador Lubin, Willard Thorpe, and Leon Henderson.

United States Government: Temporary National Economic Committee. *Investigation of Concentration of Economic Power.* Monographs for the Temporary National Economic Committee, Congress of the United States. Washington: Government Printing Office, 1939–41.

All together, 43 monographs were prepared for this committee. The most useful ones for the student are Number 11, "Bureaucracy and Trusteeship in Large Corporations" (chapters 1–3), and Number 27, "The Structure of Industry" (for reference).

Veblen, Thorstein. *The Theory of Business Enterprise.* New York: Charles Scribner's Sons, 1936.

This is neither up-to-date (in a superficial sense, at any rate) nor easy reading. However, it is very important, as are Veblen's other writings. Read especially chapters 2 and 3. Veblen's *The Engineers and the Price System* may interest engineering students.

PART TWO

The Operations of the Business Firm: Price and Output

Introduction

WE HAVE SEEN that the actions of business firms are of decisive importance to the functioning of the economy. When firms decide to hire more men, employment grows; and when they decide to raise prices, the general price level increases. Generally what happens in the economy, or at any rate in the private sector of the economy, then, is no more than the sum of what all the individual constituent firms decide to do. Hence it is of first importance that we analyze how a firm determines the price it will charge, the size of its output, its methods of production, and so on. Problems of this kind are discussed at some length in Part Two.

Since the analysis in these chapters is somewhat detailed, it may prove desirable in some brief courses not to study it *in extenso*, but to treat it briefly or instead to concentrate on Chapter 22, which in summary form sets forth the major conclusions derived in this part of the book.

5

The Study of the Business Firm: Introduction

A PHOTOGRAPH of a Diesel engine would not convey much information about how and why the engine operates. Similarly, the examination of what we might call the physical features of our economy does not tell much about the operations of the economy, although it might describe some of the results of these operations. To understand how the Diesel engine operates, it is necessary to understand certain principles of physics — for instance, how oil is vaporized, what happens when vaporized oil is heated under very great pressure, and so on. Likewise, to understand how the economy operates, it is first necessary to understand something of the ways in which economic actions are determined.

In our economy, though not in all economies, the most important economic activities take place in business firms. Indeed, the firm in our economy is rather like the hydro-carbon molecule in the fuel of the Diesel engine. Therefore, in order to gain some understanding of how our economy functions, we must begin our analysis with an inquiry into the operations of the business firm. More concretely, we have to investigate how the business firm adjusts itself, or reacts, to a variety of changes in economic conditions. And since the adjustments that are of special significance to society are those that have to do with the output or volume of production undertaken by the firm, we shall focus our attention on this aspect of the problem.

Let us clarify some of the issues involved. In order to understand the operation of the firm, we shall first have to see what factors it

takes into account in deciding upon the quantity to produce. Next we shall have to determine what changes it will make in its scale of operations if, for example, the demand for its product should increase, or if the government should impose a sales tax, or the union should succeed in getting an increased wage rate. And since, as we shall see, firms are not all alike, either in their relation to other firms in the industry, in the kinds of commodities they produce, or in the level of their output in relation to capacity, we shall have to analyze the effects of changes like those listed above on firms in a variety of circumstances.

There are two ways of undertaking such an inquiry. One procedure would be to examine statistically the actual behavior of business firms confronted by the changes which we are anxious to investigate. We could, for instance, determine how one hundred similarly placed business firms altered their output when the demand for their products increased. But this would be a task of enormous complexity, and there is some doubt whether it would produce meaningful results. The difficulty is that, in the real world, firms rarely experience a simple change in demand, or in wage rates, or in anything else taken by itself. The economy is so complicated, and its parts are so interwoven that, for example, an increase in the demand for the product of a firm will almost certainly be accompanied by changes in other economic variables. Hence, unless great care is taken, the observed results are likely to be compounded not only of changes in demand, but also of changes in wage rates, in expectations, in the relations between the firms being investigated and their competitors, and in many other factors. What we want to see is not the result of so complex a series of changes, but the effect upon output of an increase in demand when *all other factors remain unchanged*. This is not to say that actual observations of the output adjustments to a change in demand or in wage rates are unnecessary or even undesirable. Quite the contrary; they are the foundation upon which all hypotheses are based. Furthermore, it is only by appeal to the observed facts that we can determine whether to accept or reject an hypothesis. But it does suggest that, as with physics, or chemistry, or astronomy, or meteorology, actual observations in raw form are unwieldy, and are even inconclusive until marshaled in the light of theory.

The alternative way to determine the effect upon output of a change in any variable is to begin not with a mass of facts but with an

hypothesis. Those who oppose this method dismiss it as "theorizing," though actually this is a type of approach which we all use in handling everyday problems. We are "theorizing," for instance, when we calculate what will happen if we draw fifty dollars out of our bank account. We are "theorizing" when we attempt to assess before trying it the effects upon our digestion of eating a very heavy dinner. And again we are "theorizing" when we decide that it would be cheaper in the end to turn the old car in and buy a new one. For in each of these instances we deliberately set up in our mind a simplified model of the real situation, and then attempt to work out the probable results of a change in one of the significant factors. The physical sciences depend upon "theorizing" in the same way, for they too sometimes set up a hypothetical situation in order to throw light on a real one, and they sometimes erect a new hypothesis on a minimum of data in order to provide a better explanation of known phenomena. For instance, even though a perfect vacuum cannot be obtained in the real world, we are nonetheless taught in elementary physics that in a perfect vacuum a body falls at a rate of speed that depends only upon the time of fall and the value of the gravitational constant. And the Copernican theory that the earth revolves around the sun was first worked out almost entirely by mathematical calculations, though the evidence of direct visual observation had led men to the very opposite conclusion.

In economics we must also rely upon "theorizing" — or to call it by a more dignified name, deduction. As an example of what this involves in economics, we can set up an idealized picture or model of a business firm. This model must be realistic in the sense that it embodies all that is important for our purpose, but it can nevertheless be considerably simplified. Provided that the *essential* features are kept, it need not duplicate faithfully every detail of an actual business firm. Thus, while it would involve a very serious departure from reality to suppose that, in the model firm, price and output decisions were made by reading the stars or analyzing a cup of tea leaves, it would not matter for most purposes whether the firm had its annual meeting in New York, Chicago, or San Francisco. Using such a simplified model of the firm, and of the economy, we can treat the effects of changes that otherwise would be too complicated to handle.

Naturally, such a process involves some loss of reality. We cannot

expect that the results we get from our model will exactly duplicate those in the real world. But provided that our model is suitably conceived, this will not greatly matter. Using it, we can be 90 per cent right with ease when to be 100 per cent right would require enormous effort.

This does not mean that in economics, any more than in physics or any other science, we can be content with the deductive approach alone. The results obtained from studying the model must be constantly checked against reality. Thus, for instance, if our analysis of the model leads us to conclude that an increase in the demand for a firm's product will generally lead it to expand its output, this is only part of a proof. We cannot accept this statement as demonstrated until we have checked our theory against the facts and satisfied ourselves that in the real world this actually happens.

Theorizing, then, does involve a deliberate departure from reality, a deliberate simplifying of the problem. But this does not mean that theorizing is useless or impractical. Theorizing is useless and impractical when it is bad — that is to say, when we make errors in thinking, or when it is based upon a model which, in essentials, is unrealistic. Otherwise, it is a useful and indeed an indispensable tool. It is necessary to make this point before we begin our study of the business firm, because we shall soon be in the middle of what will perhaps appear to be theorizing of a very abstract kind. The model of the business firm which we shall use may seem but a poor representation of the United States Steel Corporation or the drugstore on the corner. But it will not do to call it unrealistic unless the model behaves quite differently from existing firms in essential ways.

The chief characteristic of our model firm is that the decisions about price and output are made with a view to maximizing profits. We assume, in other words, that whoever determines the scale of output, the price to charge, or the methods of production to be used, chooses whatever course of action he thinks will be most profitable to the firm. Now, while this is not a perfectly realistic assumption to use in describing the business world, it is substantially so. Business firms may deliberately sacrifice profits for, let us say, patriotic motives, but generally no such actions are expected. Thus, to assume that a firm chooses the most profitable of several alternative policies is, essentially, a realistic assumption.

But what the economist means by the term *profit* must be clearly

understood. For in the science of economics, as in the business world, it has a definite and precise meaning quite different from that in common usage. In its strict sense it means the difference between the receipts of a firm and its costs. There are problems, of course, in connection with the concepts of receipts and costs, many of which will be discussed in later chapters. But there should be no confusion about the technical meaning of the term *profit.* Misunderstanding arises because we sometimes, in non-technical conversation, use the term in the much broader sense of "benefit" or "advantage," as in such a statement as "You wouldn't get any profit out of that course." In economics, however, the term is always used in its technical sense.

While we assume that a business firm in our economy will normally do whatever promises the highest possible level of profit, this does not imply that the firm always knows what policy will be most profitable. In actual fact, a firm rarely has *definite* knowledge about either future receipts or costs. But it can still make *estimates* — and indeed, striving for maximum profits implies that it does so. These estimates may be accurate or not; but so long as the firm bases its actions upon them, they are of critical importance even when they are inaccurate.

Thus our basic assumption about the model we are to employ amounts to this: the businessman uses his best judgment to determine the most profitable course of action and follows this course to the extent that he is free to do so. To work on this assumption is of course to "theorize," but our model seems realistic enough to promise useful results.

The profits earned in a period of time we have already defined as the difference between the receipts of that period and the costs. The businessman will normally adopt that policy toward price and output which promises the greatest profits. Hence we must examine carefully the factors that affect profits — or, to rephrase it, the factors that affect the difference between receipts and costs.

A numerical illustration will show how complicated it is to determine the most profitable output level. For while it is difficult to estimate receipts and costs at any one level of output, it is laborious and burdensome to make such assessments for a large number of alternatives — and the number of alternatives for which these estimates must be made is enormous. Let us suppose that the manager of a firm has made the following series of estimates.

TABLE 12

Determining the Most Profitable Output Level: Simple Estimate of Receipts, and Costs

If in a Period the Firm Produces	Its Total Receipts Will Be	Its Total Costs Will Be
10 units	$100	$80
15	147	117
20	190	150
25	232	190
30	270	231
35	311	273
40	350	316
45	387	360
50	420	410

Inspection of this table will show at which of the listed outputs the profits would be at the maximum: At a production level of 25 units a period, profits would be $42. But further refinement is obviously possible. On the basis of the table above it is reasonable to suppose that the most profitable level of output lies between 20 units a period and 30. But is it 29 units or 28 or 23? Thus to determine precisely what scale of output is most profitable would be relatively complicated — even if the critical estimates could be as simply set out as they have been above.

It will be seen that profits are at their maximum not where receipts are greatest nor where costs are least. Ten units of output could be produced at a much lower cost than 25 units, but profits would be only $20 instead of a possible $42. Receipts would be higher at an output of 50 units than at an output of 25, but profits would be only $10 as against $42. In determining the output which will yield the maximum profit, it is necessary to consider both receipts and costs. Neither alone will give the answer. Thus before we can consider such problems as how the imposition of a certain type of tax will affect the firm's output, we must first analyze carefully the factors upon which the firm's receipts and its costs depend. Once these issues are grasped, we can bring the results together and treat the problem of output determination.

6

The Costs of a Business Firm

IN THE SUCCEEDING CHAPTERS we must deal with the problem of how the individual business firm adjusts its output and its price to prevailing economic conditions. How much does it produce, for example, when the demand for its product is high, when it must pay $1 an hour for labor, when it is able to produce one thousand units a day with one hundred employees, and so on? This question can be answered because each firm chooses the most profitable course of action. This implies that it compares receipts and costs at various levels of output, for its profit is the difference between its receipts and its costs. Our analysis in the next few chapters will be basically similar to that which the firm itself uses. That is, we shall first analyze the firm's costs and then its receipts at various levels of output. That done, we shall be able to determine for any combination of economic conditions the level of output that provides the highest profit. In this chapter we shall look into the nature of the costs of a firm.

The Need for Goods and Services in Production

If it is going to produce anything at all, a firm must take certain steps that will necessarily involve it in costs.[1] It must acquire the services of factors of production, as we have called them. First of all, it must have an administrative organization, a framework represented by the top officials. Second, it must have the necessary raw material. For some firms, such as a hydro-electric power plant, the raw material requirements are very small. For others, such as a leather tannery, the requirements are very large. A certain amount of "fixed" equipment, plant and buildings and machinery, is generally necessary. Labor must be employed in greater or less quantity depending upon

[1] Even when its current production is zero, it would normally have to meet certain costs, as noted below.

the processes. Some of the needed labor will be engaged to do manual work, requiring different degrees of skill from the lowest to the highest; some will be engaged in clerical tasks; and some, like straw bosses and foremen, will be supervisory. There will have to be power to operate machinery, whether supplied by the firm's own plant or by a public utility. Heat and light must be provided if production is to be efficient. Finally, as a condition for carrying on business, the firm may be required to pay taxes to federal, state, and local governments. Provision for these and other items will oblige the firm to make money outlays which must be treated as costs.

Current Money Outlays and Other Costs

Each of these requirements can be met only at a cost to the firm, often requiring "current" money outlays. Wages and salaries represent the cost of securing the necessary labor. Money spent to purchase the raw materials used in a given period also represents a necessary cost of production. Rent on buildings, interest payments to creditors, and payments for taxes, light, heat, and power are also to be counted as costs of production and are measured by money out-payments.

But current expenses do not always reflect total costs, for in addition to the costs represented by money disbursals,[1] there are others which do not involve current payments. The most important is the allowance which the firm must make to cover the cost of using durable capital equipment. For instance, a manufacturer of woolen suitings may be using a good deal of machinery which was purchased five years before and is expected to have a useful life of, say, twelve years. It would be quite misleading for the firm to add the whole outlay for the new machinery to its costs for the year in which the purchase was made. Because the equipment will be productive for many years, the entire capital expenditure made to secure it should be spread over the whole life of the equipment and added to the costs in each of these years. The firm should thus make an allowance each year for the amount of the equipment "used up" during that time. This is known as an allowance for depreciation. If the equipment cost $96,000, and is expected to serve for twelve years, it would be plausible to charge only one-twelfth of the purchase price; that is, $8000, against the output of each year. This charge is just as much a cost

[1] Purchases of raw materials or other supplies on credit rather than for cash should also be regarded as purchases involving money disbursals.

of production as the president's salary or a check to pay for a thousand pounds of wool, even though it does not force the firm to *spend* any money until replacement is necessary. Other allowances which are similar in that they do not require a disbursal of funds by the firm must also be included as a part of costs: for example, the allowance for bad debts.

Determining Costs in a Small Firm

Particularly in small firms, an additional factor complicates the attempt to determine costs. The owner himself may do all the work and provide all the money and property. If so, he will not be likely to break down the amounts he takes from the cash register for his own use into payments for the various services he renders and for the money he provides. Besides, he may take more or less than the value of these services. Yet in figuring the costs of the firm, appropriate amounts should be included to pay him for his labor and for the use of that part of his wealth which is devoted to his business. Just how he should compute his own wages and his other returns is a problem we shall leave for later discussion. In general, however, it seems clear that the sums should approximate what he could earn as an employee and by investing his money elsewhere.

Costs — An Example

In a very large firm, on the other hand, costs are analyzed in great detail, and they may be published in more or less summarized form. The United States Steel Corporation, like most large companies, issues a statement of its costs for each year. In 1937, the total costs incurred by this firm amounted to $614,533,572. The breakdown of the total shows the importance of the various items of cost:

TABLE 13

Breakdown of Total Costs for One Year, United States Steel Corporation

Payroll	$275,364,898
Goods and services purchased from others	228,718,329
Social Security taxes	11,309,216
Other taxes (except those on income)	34,602,015
Interest	8,262,327
Pensions	7,743,046
Depreciation and depletion	48,532,841
Total	$614,533,572

Costs That Increase as Output Increases

On inspecting these various cost elements, we can see that some of them will alter when output is altered and that others will not change at all. Rarely can the amount of production be significantly expanded without a corresponding increase in payrolls. When the steel industry increased its monthly production by about 20 per cent between August, 1940 and August, 1941, man-hours of work went up by about 33 per cent. In other words, to expand output by roughly one-fifth, four men were needed where three had done before. Occasionally an increase in output is possible without a corresponding increase in the labor force or the number of hours worked, though only in certain industries; but it will mean that the men will have to work harder. A retail store may not need more sales clerks in order to meet an exceptionally large demand because the normal work load of its employees is usually well below capacity. In general, however, a firm can raise its production level only by hiring more employees.

Certain other costs also generally rise as production is expanded. Obviously, more raw materials will be required to produce a larger output than are needed for a small one. Probably the cost of power will also be heavier. Finally, some taxes may be increased, particularly those like the cigarette tax, which is levied on the units of the commodity produced, or the Social Security tax, which increases as employment rises.

Costs That Do Not Vary with Output

Certain other charges, however, generally do not alter as output changes, though whether they do or not depends on the practice of the individual firm. With certain exceptions, the cost for rent, the interest on long-term debt, the salaries of the principal officials, a good part of the allowance for depreciation, and the wages of the maintenance staff, are not dependent upon the current volume of production. The bonds may have been floated ten years ago, the machinery acquired over the last twenty years, the property leased much earlier, and the executive officers may have permanent tenure; and these costs do not vary with changes in output. Whether the firm produces at only ten per cent of its capacity or produces as much as it can, these costs are likely to continue at a constant level.

But even costs of this sort may vary with output if we consider them within a framework not of one year, but of, let us say, ten years.

For if there is enough time for all adjustments, the firm will not only adapt its labor force and its use of raw materials to the level of output; it will also modify the size of its plant, the amount of its equipment, and the size of its supervisory and executive staff. And such alterations will involve changes in the amounts paid as salary, rent, interest, depreciation, and perhaps taxes. Thus, in the framework of the long period all these costs would have to be treated as varying with output. The following classification of variable and fixed costs is therefore appropriate chiefly in considering adjustments to take place over relatively short periods of time.

Variable and Fixed Costs

Costs which are independent of the scale of output are called fixed costs; those which vary with output are called variable costs. Whether a particular cost is to be treated as fixed or variable is, however, a matter of business policy and of the length of time under review. If firms showed as little hesitation in hiring and firing vice-presidents when they wish to alter their scale of output as they do in varying their employment of wage earners, the salaries of vice-presidents would be treated as a variable cost. Conversely, if wage earners were hired for three-year periods and kept on the payroll whether there was work for them to do or not, their wages would constitute a fixed cost. The usual practice is to classify as variable costs such items as wages, raw material costs, some taxes, part of the cost of power and light, and possibly the interest on short-term loans. The fixed costs, on the other hand, are usually composed of rents, interest on long-term debt, property taxes, and the greater part of the allowance for depreciation. The designation of a cost as *fixed* does not, of course, mean that it may not change at all. The vice-president of a corporation may have his salary increased; the property tax may be lowered; the rent for a piece of land or a building may be raised. But though these costs have changed, they are still to be treated as fixed so long as their change is not *the result of* a change in the output of the firm.

Average Total Cost

The *cost per unit of output*— that is, the *average total cost* of any output — is obtained by dividing the total cost of production at that level by the number of units produced. Thus if it costs a firm $150 to produce 150 units, the average total cost (or cost per unit) is $1. If it costs

$180 to produce 200 units, the average total cost for that level of production is $.90. In algebraic terms, then, we may say that if a firm's costs are y when it produces n units, the average total cost corresponding to n is y/n. It follows from the definition that we can, when given the average total cost at any level of output, compute the total cost of producing that output. If, for instance, the average total cost of 50 units is $2, the total cost of 50 units is 50 × $2, or $100. The total cost can always be computed if the average total cost is known.

Marginal Costs

As we have seen, the costs of a firm depend in part upon its output. Thus, when a firm is not producing at all, its only costs are its fixed costs. When production is low, it has in addition to fixed costs a relatively small payroll, the cost of raw materials, and certain other variable costs. Finally, when output is high, variable costs are also high, and fixed costs, being independent of output, remain at the same level as when output is zero or very low. Hence it follows that the higher the output, the higher the total cost.

A measure of which we will make considerable use in our analysis of the firm can well be introduced at this stage. This is the concept of *marginal cost.* The marginal cost is defined as the change in costs that results from a one-unit increase in output. Thus if the total cost of producing 100 units of output is $50, and the total cost of producing 101 units is $52, the marginal cost for the 101st unit of output is $2. It should be noted that the marginal cost refers to the change in *total cost* resulting from a *one-unit change* in output, and not to the change in *average total cost.* Incidentally, since the total cost is bound to increase with increasing output, marginal cost is always positive. The computation of marginal cost is illustrated by the following data:

TABLE 14

The Determination of Marginal Cost

Units of Output	Total Cost	Marginal Cost
10	100	
11	108	8
12	115	7
13	123	8
14	132	9
15	142	10

The marginal cost of the 11th unit is the difference between the total cost of 11 units, and the total cost of 10 units; the marginal cost of the 15th unit is the difference between the total cost of 15 units (142) and the total cost of 14 units (132), and so on.

Factors upon Which Costs Depend

We have already seen that costs depend upon output. The higher the output, the higher is the cost of production — other things being equal. The precise nature of the relation between cost and output will be examined in the next chapter. Meanwhile, certain other factors which, together with the level of output, determine a firm's costs, will be discussed here.

Costs arise because a firm, in order to produce at all, must pay for the labor, raw material, and capital equipment it uses. Hence any change, either up or down, in the price paid for these factors of production will affect costs in the same direction. If wage rates are raised, the cost of the labor used to produce a given output will increase. If raw materials drop in price, the total cost and average total cost of a given output will be reduced. Thus the price paid for the factors of production — for labor, raw materials, and so on — is one determinant of cost.

An improvement in the technique of production would also alter the cost of producing a given output. Since the improvement would be adopted voluntarily only if it added to the profits of the firm, we can be certain that costs would be reduced, at least for a certain range of output, as a result of the improved method of producing.

Finally, a change in the amount of plant and equipment used, or more generally, in the proportions in which the factors of production are employed, will affect costs. Normally, for instance, a firm will build a larger plant or add extensively to its equipment only if it expects to be able to produce a high level of output at a reduced cost because of these improvements. While its fixed costs are likely to be higher the larger its plant, its variable costs are almost certain to be lower. For otherwise the firm would have no inducement to expand its plant. Thus total costs depend not only upon the amount of production, but also upon the price to be paid for the factors of production, the techniques of production, and the proportion in which the various factors are employed.

In the next chapter we shall be particularly interested in how costs

vary with output. This is of special interest because a firm can vary its scale of output at will, but it is not able to change with any speed the three other factors upon which its costs depend, and it may not be able to change them at all. A firm can reduce its output from 1000 units a day to 500 units in a very short period of time. To do so it need only dismiss some men and cut down its order for raw materials. But it cannot change the size of its plant so quickly. It may take a year or more to complete an extension to its existing factory, or six months to install some machinery. Likewise it cannot change its methods of production overnight; many months of patient work may be needed before it can develop and put into effect an improved process. And certainly the firm cannot at will change wage rates or the price it has to pay for raw materials. It has to bargain with the union or with the supplying firms, and it may not succeed even then in reducing the price.

Four factors determine a firm's costs, as we have seen: its output, the price it pays for the factors of production, its method of production, and the size of its plant and equipment. Over the first-named factor the firm has complete control; over the others it has practically none, at least in relatively short periods of time. These other factors are, in a sense, background factors. They determine the nature of the relation between costs and output. For convenience we shall refer to them as the *cost determinants*. As we have just seen, they are not liable to great changes over short periods. In the next chapter we shall examine how costs vary with output, assuming that these cost determinants are given.

Summary

A firm must use certain factors of production in order to produce anything, and it must pay to use them. The payments made are its costs, though for certain kinds of payments, such as the purchase of long-lived equipment, we should want to spread the total sum over the life of the equipment. Some costs vary with output, being high when output is high and vice versa — payrolls and costs for raw materials, for example. Others do not change because of changes in output, though they may of course change for other reasons. Rents and management's salaries are examples of these fixed costs.

Two concepts have been introduced in this chapter which we shall use later — the *average total cost* and the *marginal cost*. The former,

often known as the total cost per unit, equals the total cost of any output divided by the output itself. The latter, which is sometimes known in business firms as the incremental cost, measures the addition to total cost brought about because of a one-unit expansion in output.

The costs of a firm depend upon four factors: its output, the price it must pay for the factors of production, the method of production, and the size of its plant and equipment. The way in which cost and output are related is of special interest, since we are concerned with the problem of how a firm determines its output. It is convenient to group the other three factors together as *cost determinants*, which are not likely to change sharply but which affect the firm's operation by altering the relation between its price and output when they do vary.

The Use of Graphs — An Appendix

Before concluding this chapter it may be well to say a few words about the construction and use of graphs, since much of the material which the economist uses can be presented quickly in the form of graphs or charts. A short explanation of the procedures employed is given below. For the reader who has even a fragmentary knowledge of analytical geometry or graphical analysis, this appendix will prove of no value. Other readers should study it with care, since it describes a technique which will be used extensively in the following chapters.

The chief advantage of graphical presentation of data is its con-

TABLE 15

A Sample Table of Costs

When Output in the Period Is	The Average Total Cost Is	Marginal Cost Is Therefore
1	10.0	–
2	9.5	9
3	9.0	8
4	8.5	7
5	8.2	7
6	8.0	7
7	8.0	8
8	8.5	12
9	9.0	13
10	9.5	14

venience. Instead of long, involved tables of figures, this method employs what is essentially a kind of pictorial shorthand. The use of this shorthand saves time in the presentation of data, and as will become clear later, in its interpretation. Since the concept of the costs of a firm is now familiar, we may illustrate the preparation and interpretation of charts by reference to this subject. Suppose we are given the information about a firm's costs shown in Table 15, page 73. The marginal costs were computed from the data on average total costs.

To plot such material graphically, we prepare a figure similar to the following:

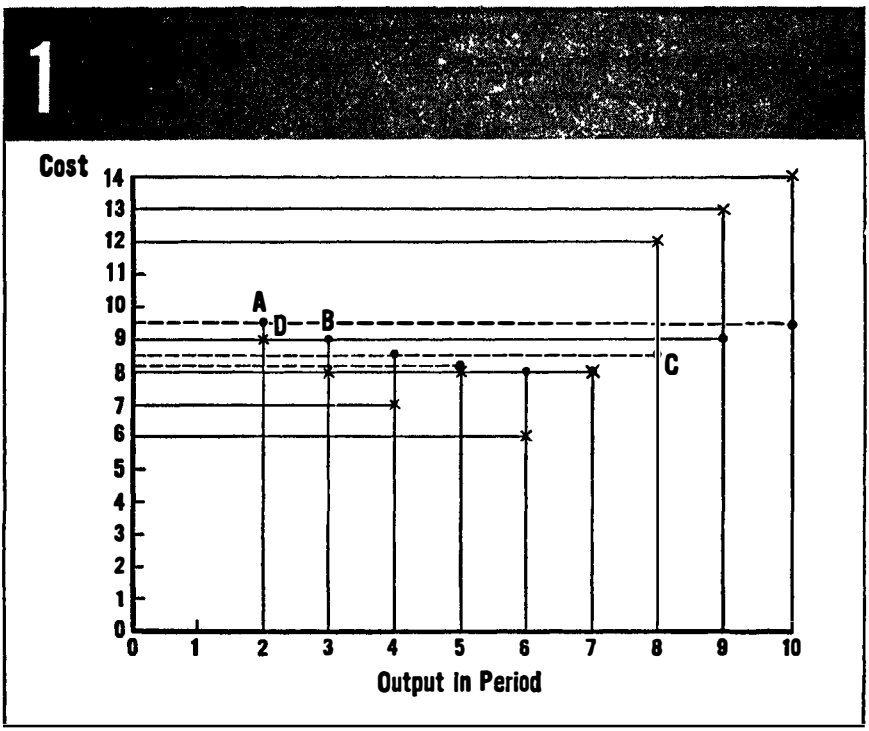

Figure 1. *Relation Between Cost and Output*

On this figure is recorded the data provided in the table. Each part of the relation is represented by a single point. For instance, the *average total cost* corresponding to two units of output is 9.5. To plot this, count along the horizontal axis from the origin (0) a distance of

two units. Then measure vertically a distance of 9.5 units. The point so found, designated by A on the figure above, indicates that the cost of each unit, when two units are produced in the period, is 9.5. Another point, designated by B on the figure, 3 units horizontally from the origin and 9 units above the origin, shows that the average cost for 3 units is 9. Similar points can be plotted for 4, 5, 6, 7 units, and so on. In each we can measure horizontally a distance that indicates the level of output and vertically a distance that shows the average total cost for that output. Thus the point C, which is situated 8 units to the right of the origin and 8.5 units above it, shows that the average total cost for 8 units of output is 8.5.

The information relating to marginal cost can be presented on the same diagram. Thus, against 2 units of output we must measure vertically a distance of 9, designated by point D. Opposite 3 units of output, we must set a point 8 spaces above the horizontal axis, and so on.

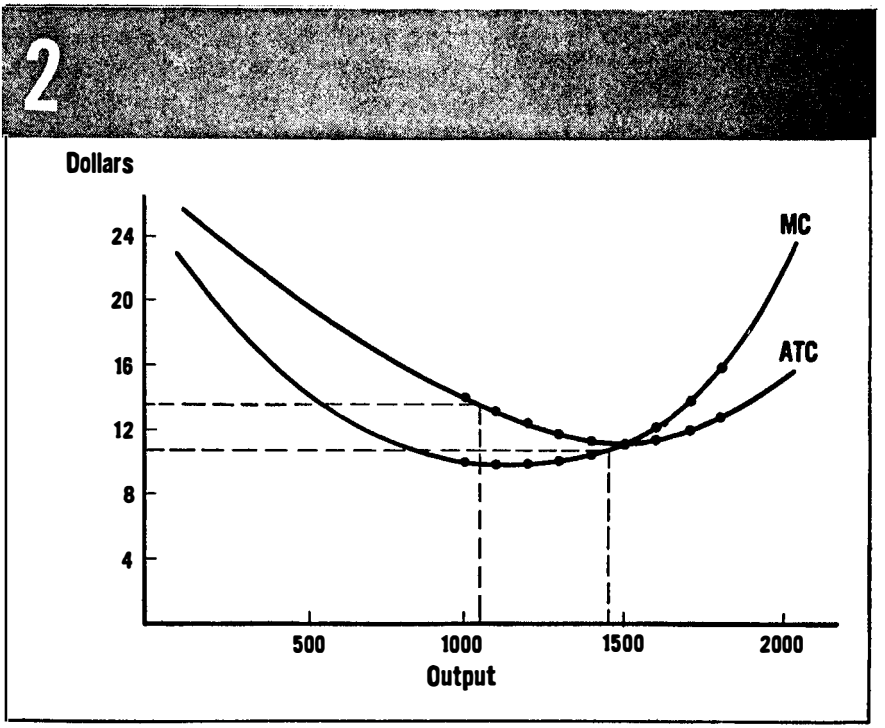

Figure 2. *Marginal and Average Cost Curve*

Normally, of course, we shall be dealing not with outputs of 2, 3, 4, 5 units, and so on, but with much larger outputs, perhaps of 2000, 10,000, or 1,000,000 units. We would not in that case plot every point — 1000, 1001, 1002, 1003, 1004; but we might plot the average total and marginal cost for 1000, 1100, 1200, 1300, 1400, and so on. This procedure is illustrated in Figure 2. These points are normally connected by smooth curves. The curve which joins the points that represent the relation between output and average total cost, is called the Average Total Cost (ATC) Curve. The other, which joins the entries representing marginal cost, is known as the Marginal Cost (MC) Curve.

If the curves are drawn by plotting the average and marginal costs corresponding to outputs 1000, 1100, 1200, 1300, and so on, it is generally safe to read from the curves themselves the average cost or marginal cost for any output between these values. Thus we can estimate from Figure 2 that the average total cost of 1050 units is $13.40, that the marginal cost of 1440 units is $10.80, and so on. It should be apparent that information of the kind presented in the table above can be given much more conveniently in the form of a graph. Because of their convenience, we shall use graphs frequently in this book. It should of course be clear that the graph does not say anything *more* or *different* from the table of figures on which it is based. But it does say as much, and it says it much more simply. Graphical presentation should therefore be looked upon as a convenience, not as a mystery.

7

Costs and Output

IN THIS CHAPTER we shall examine the nature of the relation between costs and output. We have already seen that *total* costs increase as the level of output rises; we must now explore the ways in which the *average* total cost and *marginal* cost are affected by changes in the amount produced. Is the average total cost for a large output typically greater or less than for a small one? Is the marginal or incremental cost of the 10,000th unit characteristically higher or lower than that of the 2000th unit? The answers to these questions are important if we are to determine the effects of a change in economic conditions upon price and output.

Engineering and Statistical Studies Necessary

While the economist must set the problem and define precisely what he wants to know, he is not competent as an economist to answer it. For the exact relation between costs and output is something that depends upon the technical processes of production and comes within a field of knowledge quite different from economics. Here the assistance of the engineer or the production manager is particularly valuable. However, a statistician can determine the relation between output and costs by analyzing the cost data of a firm. We shall examine some of the statistical data later in this chapter. Before doing so, however, we shall discuss certain elementary considerations which will prepare us for the results that engineering and statistical studies have secured.

Costs When Output Is Very High or Very Low

How should we expect costs to vary with output? Will the average cost be higher as output increases, or will it be lower? We may begin by comparing the cost situation for a firm operating at a time of full prosperity with the situation it faces during a period of depression.

In prosperity, not only is our firm producing very close to capacity, but most other firms in the economy are doing the same. All the equipment available to the firm is thus pressed into service — the best, most efficient machinery it has, of course; but also a certain amount of inefficient machinery which is being used only as a last resource. The labor employed is likely to be uneven in its composition. In addition to the more or less permanent body of employees, there are many who are on the job only because the company wants to produce as much as possible. Of this latter group, a number are likely to be notably less efficient than the men who have been with the firm for a long time. They may be inexperienced at this particular kind of work, and so in need of training — an expensive provision. Finally, because of the enormous pressure of work, and the reduced incentive when profits are high, there may be breakdowns at the administrative level and inefficiencies in purchasing, delivering finished goods, personnel relations, or supervision, any of which would add further to costs. Thus when output is large, costs are likely to be abnormally high, reflecting not only the ordinary increase which comes with increased output, but also the inescapable added expense resulting from the use of poorer machinery, less efficient labor, and less successful management.

But these phenomena can easily be exaggerated. There has been a tendency to suppose that whenever output is increased, there is a decline in labor efficiency because of the employment of less capable workers. But while this consideration may be valid when the operation is largely a handwork job, it is much less applicable when labor works in close conjunction with the machine or the conveyor belt. Operations may be easily standardized and quickly learned, and the speed of the machine or the belt then determines how much a man can produce in a given length of time. Individual differences in efficiency lose much of their importance in many branches of modern industry.

This does not imply a general denial of the above conclusion that, *beyond a certain point*, a high level of output should be expected to in-

volve abnormally high costs as compared to those prevailing at a lower level of output. It only suggests that the critical level above which costs are unduly high may in fact be a very high level indeed.

When the firm is producing at a low level, the situation is of course reversed. Only the most efficient equipment is likely to be in use. Since the firm needs less labor, it can weed out the poorer, less productive workers and retain only the most able. Management at such a time has an opportunity to iron out the kinks, and to avoid the difficulties that make for high costs when output is near capacity. For each dollar of labor, equipment, and administrative personnel used, output should be somewhat higher than when the firm is producing nearly at the peak.

At very low levels of output we can therefore expect to observe a reaction of variable costs to output opposite to what we should expect at very high levels. That is, in the lower range, variable costs may increase at a slower rate as production expands than production itself. Thus, if the weekly production rate is raised from ten units to fifteen, variable costs may increase from, let us say, twenty to no more than twenty-five. Perhaps the most obvious explanation for such a slow increase in variable costs as output is raised from a very low level is that the labor force is not being used to anything like full capacity. It is impossible to hire a fraction of one man, though when production is low, only a fraction of a man's time may be needed — or only a fraction of the full time of a group of men, all of whom are nonetheless required as a single operating unit. Hence it may not be necessary to expand employment by a large amount in order to increase output, but merely to use more efficiently the men already on the payroll. Thus total variable costs, in that case, would increase only slightly.

The picture so far, then, is this: some costs are fixed and go on at exactly the same rate whether output is high or low. Others, however, are variable; at normal outputs they may vary proportionately with output, but at very high outputs they will rise to an exceptionally high level compared to their level when output is low. Variable costs for some distance below the critical point may thus vary proportionately with output. That is, a 10 per cent increase in output will be accompanied by a 10 per cent rise in variable costs. But beyond that critical point, the relation between output and variable costs changes, and in this higher range variable costs change in the same direction as output but at a more rapid rate. For example, beyond a

certain very high level, variable costs may rise by 30 per cent when output rises by 10 per cent. In contrast, at very low outputs, variable costs may rise less quickly than output, perhaps going up by only 5 per cent as output rises by 10 per cent.

Average Total Cost and Output

If these conclusions are accurate, what can we say about the relation between average total cost and output? Since the total cost equals the sum of the fixed and variable cost, it follows that the average total cost equals the sum of the average fixed and average variable costs. If, for example, fixed costs are $100 and the variable costs of producing 400 units are $300, the average total cost is $400 (the total cost) divided by 400, or $1 a unit. The average fixed cost at this point is equal to $100/400 or 25 cents per unit, and the average variable cost is equal to $300/400, or 75 cents per unit. The sum of the average fixed cost and the average variable cost is of course equal to $1, which as we have seen is the average total cost. We must then examine the relation between average fixed costs and output, and average variable costs and output.

The average fixed cost is lower, the higher the level of output. In the above illustration, the average fixed cost of 400 units is equal to $100/400 or 25 cents. If, however, 500 units were produced, the average fixed cost would be only $100/500, or 20 cents. Since fixed costs (as distinguished from average fixed costs) are constant no matter what the output, the lower will be the share of fixed cost charged against each unit, the larger the number of units produced. If on a graph we measure output along the horizontal axis, and average fixed cost along the vertical axis, we should have an average fixed cost curve which would fall gradually to the right, as the accompanying diagram (Figure 3) shows.

Unfortunately, it is not possible to reach so definite a conclusion concerning the behavior of average variable costs. It has been stated that once output passes a certain rather high level, variable costs increase rather rapidly. If that critical level is called 100 and the corresponding variable costs equal $200, the average variable costs equal $2. A production figure of 120, then, will necessarily produce a variable cost figure in excess of $240 — say $264 — since production is now well past the critical point at which such costs begin to rise disproportionately. The average variable costs corresponding to that higher

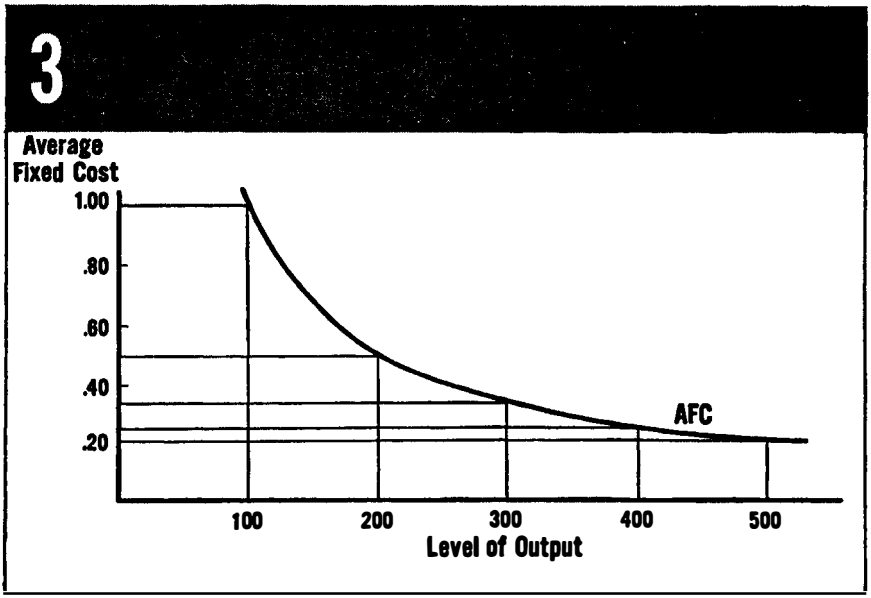

Figure 3. *Variation of Average Fixed Cost with Output*

level of output would therefore be somewhat higher too, $2.20 as against $2. Beyond what we have called the critical level, then, the higher the output, the higher is the average variable cost.

In the middle range of output, where variable costs change proportionately with output, rising by 15 per cent when output increases by that amount, or dropping by 25 per cent as output falls by the same percentage, average variable costs will remain constant. Whether this range of output is wide or narrow depends on the nature of the productive process, the facilities available to the firm, and so on.

Finally, at a very low output level, where variable costs increase at a slower rate than production, the average variable cost declines as production increases. When output is 10, for example, the total variable costs are $25, and hence the average variable cost is $2.50. When output is 15, total variable costs are only $35 (rising by less than 50 per cent), and accordingly the average variable cost is $2.33, or lower than at an output of 10.

Expressing these results diagrammatically, and measuring output along the horizontal axis and average variable costs along the vertical axis as we did in earlier graphs, we should have a flat U-shaped curve as illustrated in Figure 4.

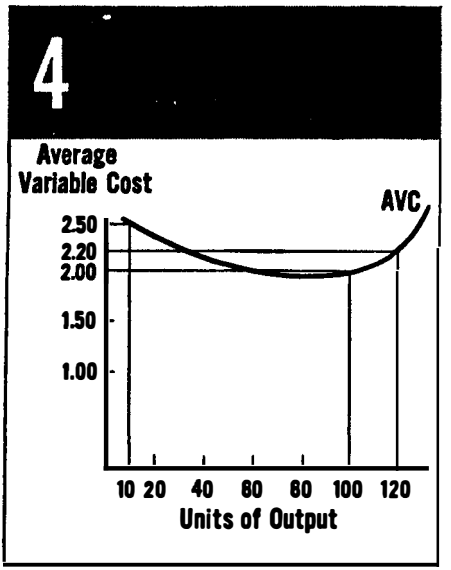

Figure 4. *Variation of Average Variable Cost with Output*

Summary: Average Total Costs and Output

Up to that level of output at which average variable costs begin to rise, it is clear that *average total costs* will decline as production expands, for both average fixed costs and average variable costs are falling; or at any rate, while the former is falling, the latter is not increasing. But once this critical output figure has been exceeded, the change in average total cost depends on the size of the increase in the average variable cost compared to the fall in the average fixed cost. So long as the rise in the average variable cost is not great enough to offset the fall in the average fixed cost, the average total cost will continue to decline. But eventually, as output comes closer and closer to the firm's capacity, the rise in the average variable cost will more than offset the decline in the average fixed cost, and average total cost will rise. We should therefore expect to find that average total cost declines as output rises from a very low level; that with further increases in output the level of average total cost is approximately constant; and that finally as output nears the peak, the average total cost begins to rise.

These considerations seem reasonable and obvious. But the economist, like any other scientist, should not be content with conclusions that seem reasonable. Where observation is possible, one accurate observation is worth a hundred armchair musings about what we should expect to observe — if only we would take the trouble. The dangerous habit of relying on intuition rather than eyesight is very tempting; but it is just as likely to be misleading in economics today as it was in medicine or physics five centuries ago. In other words, we should treat the considerations just advanced about average total costs as *hinting* at a conclusion, but certainly not as *proving* one. We

cannot be sure how average total costs are related to output until we have observed the relation in practice. And if we do not find what we expected to see, then once we are satisfied about the accuracy of our observations, we should not deny the evidence of our eyes but instead revise our expectations.

What, then, has been *observed* about the response of average total costs to changes in output? In general, careful statistical observations confirm our expectations, although we may be surprised at the very wide range of output over which average variable costs remain at about the same level. It will be well worth our while at this point to examine in some detail the published cost data for one of the largest firms in our economy in order to see whether these data confirm the conclusions reached above.

Cost Data for the United States Steel Corporation

Information about the costs of the United States Steel Corporation was presented to a government commission in 1940. The basic data were the actual cost figures of the firm for the period 1927–1938. These data were first adjusted to 1938 conditions, translating the results of other years, for example 1937, into figures based upon 1938 wage rates, taxes, and so on.[1]

TABLE 16

Composition of Total Costs of Operation in Relation to Volume of Business, United States Steel Corporation and Subsidiaries

Item	Costs That Must Be Met Regardless of Operating Rate	Additional Costs for Each Additional Weighted Ton of Product Shipped
Interest	$8,300,000	$0.00
Pensions	7,700,000	0.00
Depreciation and depletion	29,500,000	2.37
Taxes other than Social Security and federal income	24,200,000	1.43
Payrolls	62,100,000	29.10
Social Security taxes	2,500,000	1.16
Goods and services purchased, etc.	47,800,000	21.67
Total costs	$182,100,000	$55.73

[1] This procedure is tantamount to correcting for changes in the cost determinants. The results obtained are therefore meant to show how costs vary with output with given (1938) cost determinants.

The "Costs That Must Be Met Regardless of Operating Rate" we should call the fixed costs. They amount, under 1938 conditions, to $182,100,000 annually. The variable costs, or as they are identified in the table, "Additional Costs for Each Additional Weighted Ton of Product Shipped," are $55.73 a ton under 1938 conditions. That is to say, average variable costs were constant within the whole range of output of the United States Steel Corporation in this period. Output varied between $4\frac{1}{2}$ million tons in 1932 and 15 million tons in 1929, or from about 17 per cent to 90 per cent of capacity; yet over this whole range, average variable costs were $55.73.

The way in which average total costs vary with output can now be computed. The average total cost of 4 million tons is $101.25 a ton, or

$$\frac{\$182{,}100{,}000 \text{ (fixed cost) plus } \$55.73 \times 4{,}000{,}000 \text{ (variable cost)}}{4{,}000{,}000}$$

The average total cost for other levels of output is as follows:

TABLE 17

Variation of Average Total Cost with Output, United States Steel Corporation and Subsidiaries

When Weighted Tons of Product Shipped Are (in millions)	Average Total Costs Are
4	$101.25
5	92.15
6	86.08
7	81.74
8	78.49
10	73.94
12	70.91
14	68.74
16	67.11

It will be seen that in the range of output of the United States Steel Corporation between 1927 and 1938, average total costs are lower, the higher is the production. Unfortunately these figures do not show what happens to average total costs when output is increased beyond 90 per cent of capacity. Perhaps if output in this period had been raised to a point even closer to capacity, as it has been since that time, we should have found evidence of increasing average variable and average total costs for increases in output. Other analyses of costs in

different kinds of industries suggest that this would indeed be the case.

We may therefore conclude that over a wide range, the higher the level of output, the lower is the average total cost, but that beyond a certain point which apparently is rather close to the capacity level of the firm, average total costs will rise for every further expansion in output. If we present this situation diagrammatically, we have a U-shaped curve similar to that shown in the diagram below. This result, it will be noted, agrees with our expectations.

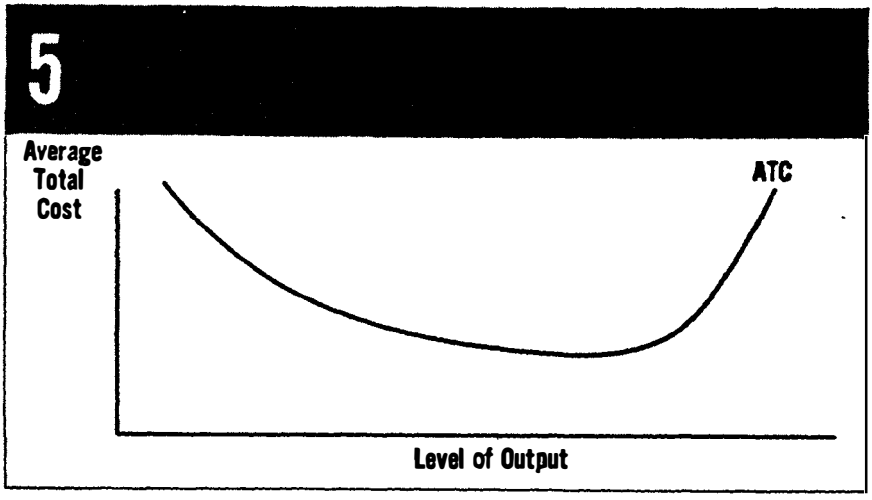

Figure 5. *Average Total Cost and Output*

Marginal Costs and Output

From these observations, much can also be learned about the behavior of marginal costs. In the case of the United States Steel Corporation, it is clear that the marginal cost per ton of output is $55.73, since costs increase by that amount with every additional ton of output. The marginal cost curve in that event is a horizontal line for output between 20 and 90 per cent of capacity. Since data are lacking we can only guess what marginal costs would be for very low and very high levels of output. If, as capacity were approached, the additional costs of expanding output were in excess of $55.73, the marginal cost would rise and the curve would incline upward in this range. If at very low levels of output the firm could not fully utilize its labor force, its power, or its raw materials, marginal costs would probably be high.

but would decrease with each rise in the level of production. In other words, in the lower range, marginal costs may be expected to decrease as output increases; beyond a certain point, there is a wide middle range within which marginal costs remain approximately unchanged; and finally, as capacity output is approached, marginal costs would be expected to rise. Therefore the marginal cost curve, like the average total cost curve, will be U-shaped, as in the following illustration.[1]

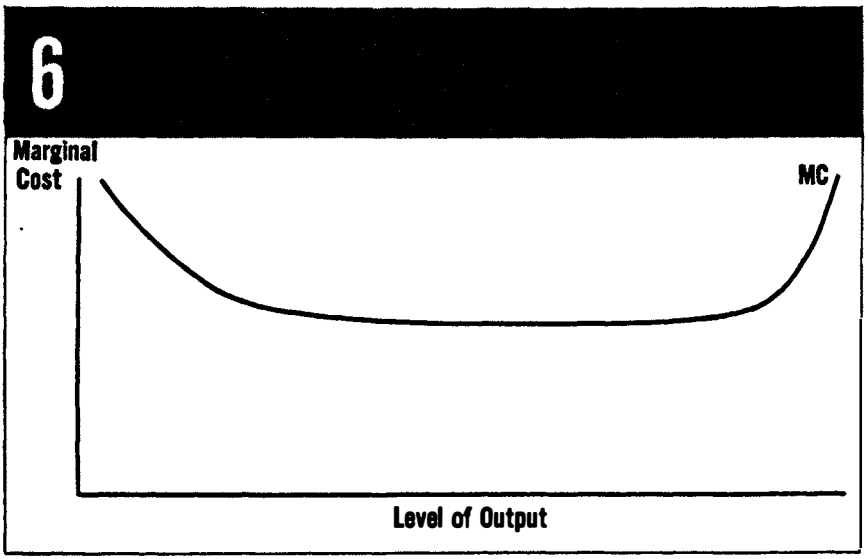

Figure 6. *Marginal Cost and Output*

Other Statistical Studies

Studies of costs in other industries, though there are not many, confirm the experience of the United States Steel Corporation. An analysis has been made of costs in a hosiery mill, and another of costs in a furniture factory; and these studies also show that within a very wide range of output average total costs decline when output expands and marginal costs remain approximately constant. Not much information has been obtained relating to very low and very high outputs, though it is at such production levels that we should expect to find exceptions to the rule noted in the previous paragraphs.

[1] Strictly, if marginal cost were abnormally high at low outputs, average variable cost would not be constant in the range in which marginal costs were level.

Summary

Cost statistics from a number of firms show that the average total cost is related to output in the following way: up to a point near the capacity of the firm, the higher the output, the lower is the average total cost. Beyond that point, the higher the output, the higher is the average total cost. The decline in average *total* cost results mostly from the fact that the higher the output, the lower is the average *fixed* cost. The average variable cost evidently does not vary with output except when the firm is operating near capacity. The cost data also show that the marginal cost does not vary with output except possibly at very low and very high levels. These statistical results agree with the conclusion we reached on the basis of general observations about the nature of production for various levels of output.

8

Some Relations Between Average, Marginal, and Total Costs

SINCE both the average total cost function and the marginal cost function depend in some way upon total costs, we should expect to find a connection between any two of these concepts. And because a knowledge of some of these relations will be useful for an understanding of later chapters, it is well worth our while to discuss them at some length here. The exact connections between these variables can be most easily demonstrated by the use of the calculus, but we can by simple arithmetic illustrate, even if we do not prove, the most significant of their relations. They will all follow logically from the definitions of average, marginal, and total costs.

Marginal and Average Costs

We may begin by examining the relation between the average [1] and the marginal costs. First let us see how the average total cost varies with output when the marginal cost is below the average total cost.

If the average cost of 10 units is 6, the total cost of 10 units is 60. The marginal cost is assumed to be less than the average cost; therefore let the marginal cost of the eleventh unit be 5. The total cost of 11 units is then 65, and the average cost for 11 units is 65/11, or 5.91. Notice that the average cost of the larger output is less than the average cost of the smaller. From this we may conclude that whenever the marginal cost of a certain output is lower than the average cost, the average cost of that output is less than it would be for a smaller

[1] Where there is no possibility of confusion, we shall use the term average cost for average *total* cost.

output. The average cost will continue to fall as output increases, so long as the marginal cost is below the average cost for corresponding outputs.

On the other hand, when the marginal cost is greater than the average cost of any output, the average cost is greater, the larger the output. Thus, to use the example given in the paragraph above, if the average cost of 10 units is 6 and the marginal cost of the 11th unit is 7, then the average cost of 11 units is 67/11, or 6.09 which is greater than 6. Generally, therefore, if the marginal cost exceeds the average cost, the average cost is higher, the greater the output. The average cost rises when output expands so long as the marginal cost is above the average cost.

Graphs of These Relations

This relation between, on the one hand, the positive or negative sign of the difference between marginal and average cost and, on the other hand, the reaction of average cost to changes in output, is particularly useful in presenting cost data diagrammatically. For it means that in those output ranges where the average cost curve slopes downward to the right, the marginal cost curve must be below it; and in those ranges where the average cost curve slopes upward to the right, the marginal cost curve must be above it. It is therefore easy to see that the two curves will intersect at the lowest point of the average curve, for at that point the slope of that curve is zero. The diagram (Figure 7) illustrates the nature of this relation.

Relation Between Marginal and Total Costs

Perhaps even more important than the relation between the marginal and average costs which has just been discussed is that between the marginal and total costs, for as we shall see, this relation will be extremely useful in our subsequent analysis. Briefly, the connection between them is that the sum of the marginal cost for the first unit, the second, the third, and so on, to and including the nth unit, is equal to the total variable cost of producing n units. Since the total variable cost equals the total cost minus the fixed cost, we have, to take a specific example, the following equality: the marginal cost of the first unit plus that of the second and third is equal to the total cost of producing three units minus the fixed cost.

Suppose the costs of production in the firm are as represented in

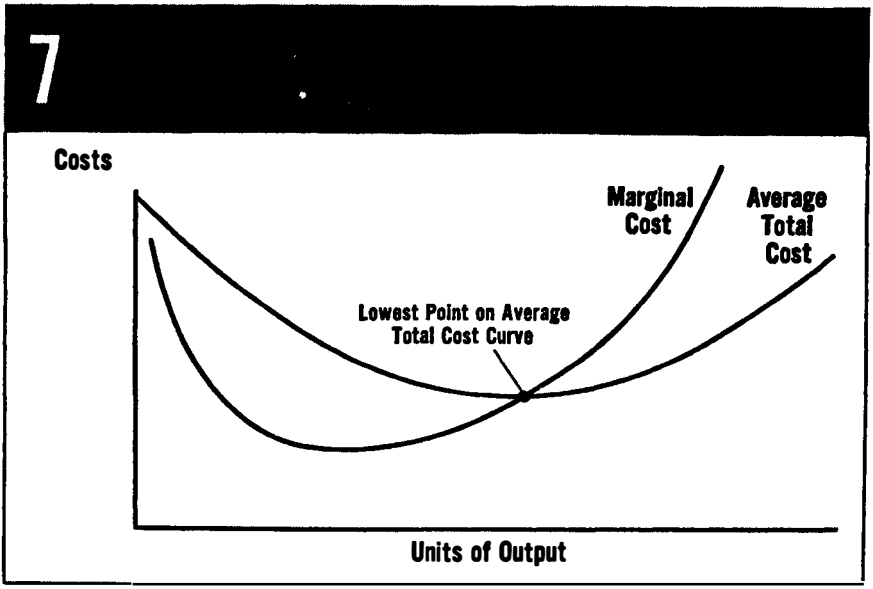

Figure 7. *Relation Between Average Total and Marginal Cost Curves*[1]

the following table,[1] where the fixed cost, since it does not vary with output, remains at F, and the marginal cost is derived from the data on fixed and variable costs:

TABLE 18

Relation Between Marginal Costs and Total Cost

Output	Fixed Costs	Total Variable Costs	Marginal Costs
0	F	0	–
1	F	V_1	$V_1 + F - F$ or V_1
2	F	V_2	$V_2 + F - (V_1 + F)$ or $V_2 - V_1$
3	F	V_3	$V_3 + F - (V_2 + F)$ or $V_3 - V_2$
4	F	V_4	$V_4 + F - (V_3 + F)$ or $V_4 - V_3$
5	F	V_5	$V_5 + F - (V_4 + F)$ or $V_5 - V_4$
6	F	V_6	$V_6 + F - (V_5 + F)$ or $V_6 - V_5$
7	F	V_7	$V_7 + F - (V_6 + F)$ or $V_7 - V_6$
8	F	V_8	$V_8 + F - (V_7 + F)$ or $V_8 - V_7$
$n-2$	F	V_{n-2}	$V_{n-2} + F - (V_{n-3} + F)$ or $V_{n-2} - V_{n-3}$
$n-1$	F	V_{n-1}	$V_{n-1} + F - (V_{n-2} + F)$ or $V_{n-1} - V_{n-2}$
n	F	V	$V_n + F - (V_{n-1} + F)$ or $V_n - V_{n-1}$

[1] In this and subsequent figures, the Average Revenue and Average Total Cost Curves should be assumed to begin with the first unit.

[2] Some readers may find it easier to substitute numbers for these symbols, for example 30 for F, 10 for V_1, 17 for V_2, and so on.

It is easy to demonstrate by a simple calculation that the sum of the marginal cost of the first, second, third, and so on . . . to the nth unit equals the total variable cost of producing n units. Let us consider an output of 8 units. The sum of the marginal costs of the first, second, third, fourth, . . . seventh, and eighth units is $V_1 + (V_2 - V_1) + (V_3 - V_2) + (V_4 - V_3) + \ldots (V_8 - V_7)$. This equals V_8, which is the total variable cost of producing 8 units. If we had taken not 8 but n units, the same equality would hold. The sum of the marginal costs of the first, second, and so on up to and including that of the nth unit, would be V_n which, as may be seen, is the total variable cost of producing n units. And since the total cost of n units is $V_n + F$, the sum of the marginal costs of the first, second, third . . . nth units equals the total cost of producing n units minus the fixed cost.

Graph of Marginal and Total Cost Relation

It is important to see how this equality is to be represented graphically. Suppose that the marginal costs of the first, second, third, fourth, and so on, units are represented in order by the columns A, B, C, D, E, etc., in the accompanying diagram.

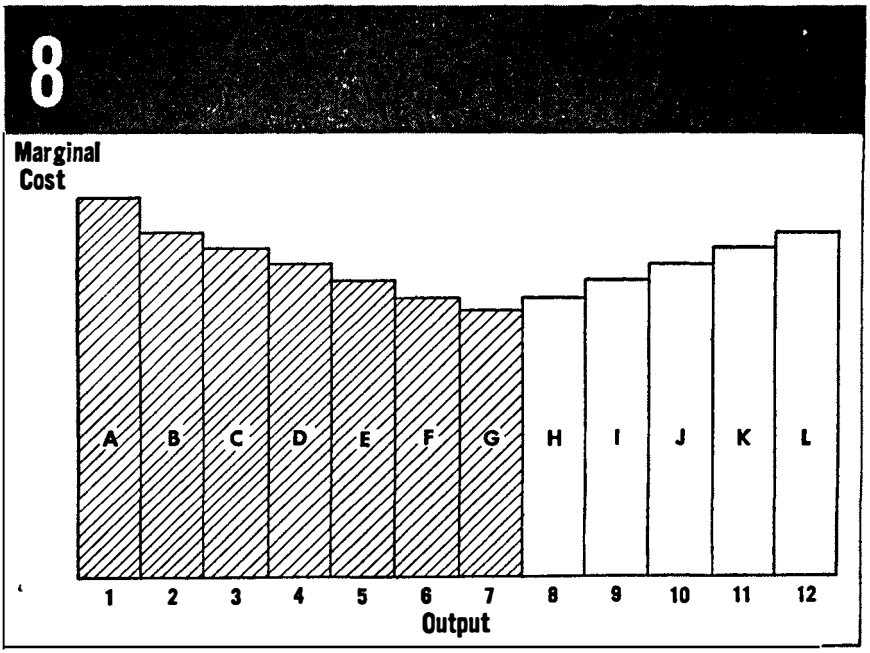

Figure 8. *The Marginal Cost Curve and Total Variable Costs*

The sum of the marginal costs of, let us say, the first 7 units is then equal to the sum of the areas of the first 7 columns; that is to say, A + B + C + D + E + F + G equals the shaded area. But this shaded area is therefore equal to the total variable cost of producing 7 units. Hence we may denote the total variable cost of any output by summing the marginal cost columns A, B, C, D, and so on.

Ordinarily, we shall be concerned with outputs very much larger than 4 or 7 or 10 units — perhaps with outputs of 10,000, 100,000, or 1,000,000 units. The columns in that case have to be drawn on an extremely narrow base, if we are to have, say, 100,000 of them represented in a single diagram. The area that represents the sum of the marginal costs of the first, second, third, fourth, . . . 10,918th and 10,919th units will be simply the area (cross-hatched) under the

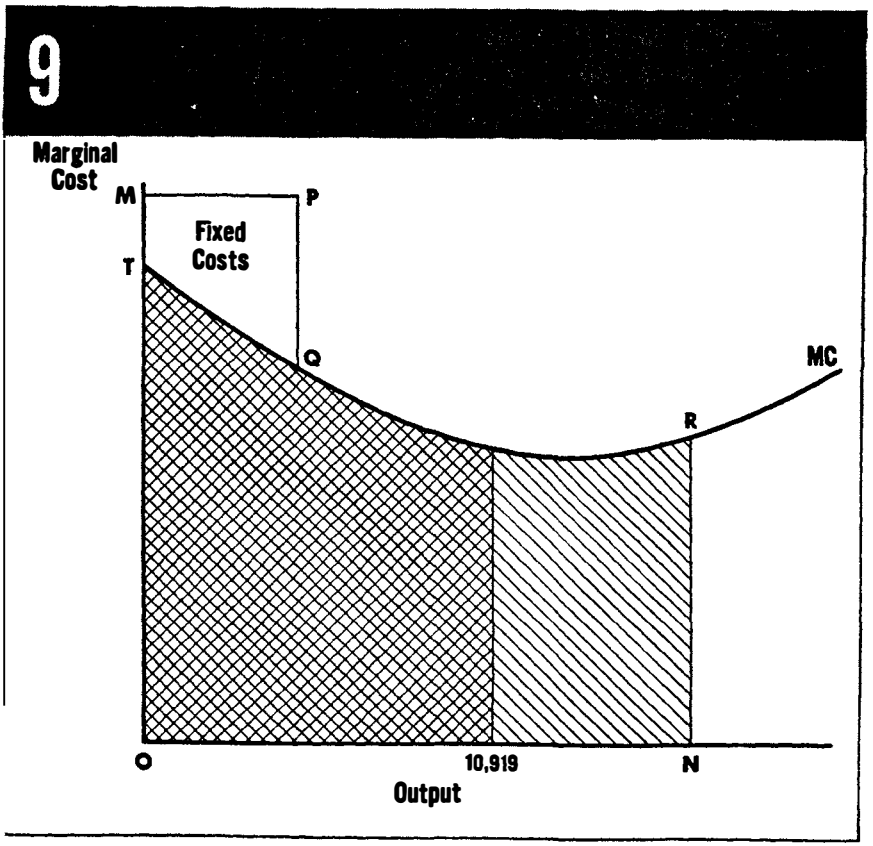

Figure 9. *Variable and Fixed Costs for Any Output*

smooth marginal cost curve up to the output 10,919, as illustrated in Figure 9. Thus the total variable cost of producing N units is the shaded area under the MC curve up to N units of output, or in the diagram, the area of $ONRT$. We can show the total cost of producing N units by adding an appropriate amount for the fixed cost — represented in the diagram (Figure 9) by the figure $TQPM$. Since the fixed cost does not vary with output, the area of this figure will be correct no matter how much is being produced.

Total Cost and the Average Total Cost Curve

Total costs may also be measured by reference to the average total cost curve. The total cost of producing any output is equal to the average total cost times the number of units produced. If 1000 units are being produced at an average total cost of 6, the total cost is 6000. This way of showing total cost is illustrated in Figure 10. The total costs are represented by the area of the rectangle $OLMN$.

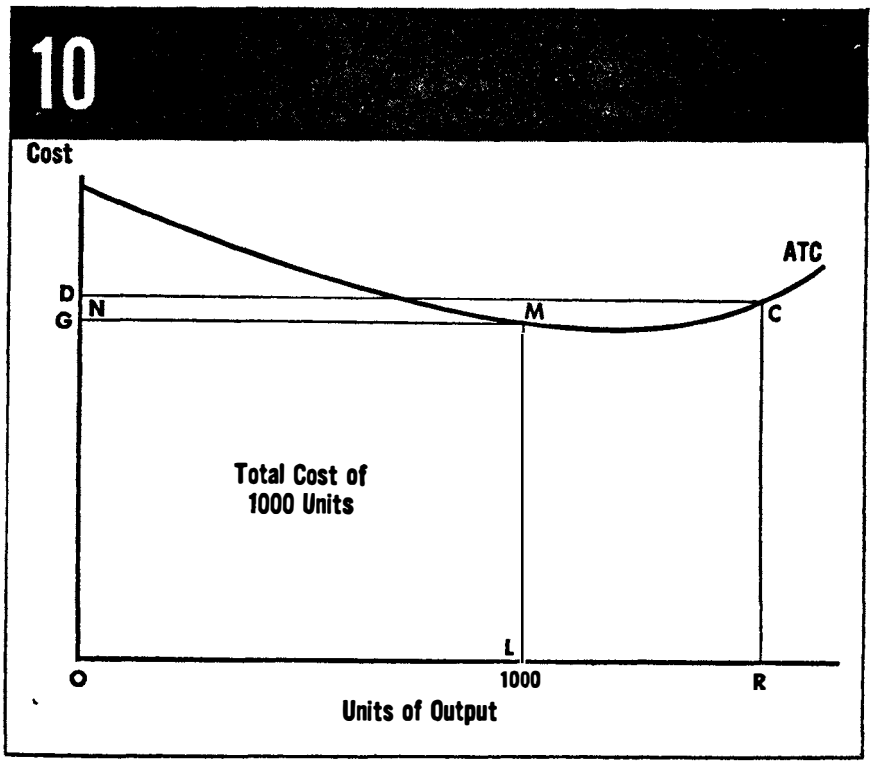

Figure 10. *Average Total Cost Curve and Total Costs*

The total cost of the output R is simply the area of the rectangle $ORCD$, that is, the product of R and RC, the average total cost corresponding to R units. Thus, using the average total cost curve, the total cost is represented by a rectangle subtended by this curve; using the marginal cost curve, the total cost is represented by the area under that curve plus an allowance for the fixed cost. We show the two ways of representing the total cost of an output in Figure 11. The area $BCDA$ is then equal to the area under the marginal cost curve $BCFG$, plus $GNML$, which represents the fixed cost. These alternative methods of measuring total costs will give identical results — but for certain purposes, the marginal cost method will be preferable; while for others the average cost method will be the better.

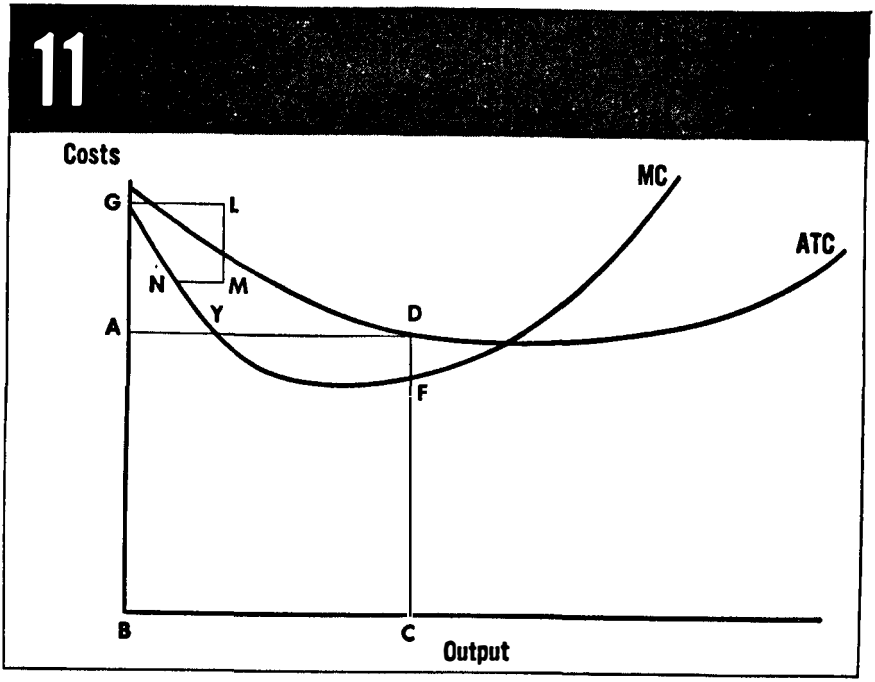

Figure 11. *Total Costs Shown by Average and Marginal Cost Curves*

Summary

When the marginal cost of any output is below the average cost, then the average cost of that output is greater than the average cost of a larger one. For example, if the average cost of N units of output

is 30, and the marginal cost of the $(N+1)^{th}$ unit is 25, it follows that the average cost of $N + 1$ units is less than 30 — perhaps 29.2. When the marginal cost of any output is greater than the average cost of that output, then the average cost of that output is less than the average cost of a larger one. If, for example, the average cost of N units is 30, and the marginal cost is 35, then the average cost of $N + 1$ units is above 30 — perhaps 30.6. From this it follows that the marginal cost curve will be below the average curve when the latter is falling, and above it when the latter is rising; therefore the two curves must intersect at the lowest point on the average curve. (Compare Figure 7.)

It follows from the definition of marginal cost, that the sum of the marginal costs of the first, second, third, . . . to the Nth unit, equals the total cost of N units minus the fixed cost. This relation gives us a useful method for showing graphically the total cost of any output (compare Figures 9 and 11). Alternatively, we can represent the total cost of an output by using the average total cost curve, in which case the area of the rectangle that it subtends at the output in question is the proper measure.

9

Changes in the Cost Determinants

IN ORDER TO KEEP our analysis of the problem as simple as possible, we have so far analyzed the relation of costs to output as if the cost determinants — that is, such items as salary and wage rates, the price of raw materials, tax rates, the size of the firm, and methods of production — remained constant. But obviously these cost determinants are liable to change in the real world, and we shall be concerned with the consequences of such changes for the economy. In order to determine these consequences, we have to understand how they affect the cost functions. What, for example, happens to costs when there is a change in the price of raw materials or in the tax rate? How will costs be altered when the firm introduces an improved technique of production? And what will happen to costs if the firm expands its plant? These questions describe the range of problems to be discussed in this final chapter on costs.

The Effect of Change in Price of Productive Factors

First let us consider the consequences of a change in the price which the firm has to pay for any of the factors of production, such as labor or raw materials. So long as the technique of production is not altered in response to such a change, the answer is perfectly clear. With a given technique of production, a certain amount of labor and raw materials are required to secure any particular level of output. If wages are raised, labor costs will be higher for each output figure, and therefore total cost and average total cost will also increase. An increase in raw material prices will produce a similar result.

With a rise in the price of either of these productive factors, there would also be an increase in the marginal cost at each level of output. The marginal cost represents the cost of the additional factors that must be used in order to expand production by one unit. These additional factors may consist, for example, in one man-hour of work plus six pounds of raw material. If wage rates are raised from 60 to 70 cents an hour, or if the cost of six pounds of raw material rises from $1 to $1.10, the marginal cost that corresponds to that output would now be 10 cents higher — $1.70 instead of $1.60. Hence, an increase in the wage rate or in the price charged for raw materials will mean that for each level of output, the average total cost and the marginal cost will be somewhat higher than formerly.

Moreover, the original average total cost and marginal cost curves will no longer represent the situation, and entirely new curves will have to be drawn. Since the average total cost corresponding to each level of output will be higher than formerly, a new point will have to be plotted at each level of output which is appropriately higher than the original one. Joining these points will give a new average total cost curve that lies above the original one. We will

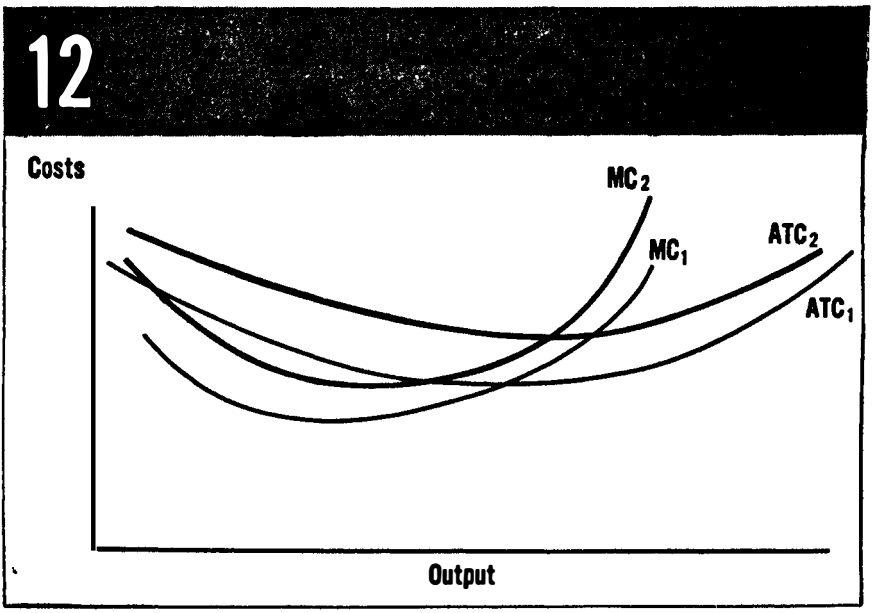

Figure 12. *Change in Average and Marginal Costs*

have to follow the same procedure in replotting the marginal cost curve. Increases in the average total cost and marginal cost curves are illustrated in Figure 12.

If the firm, because of a higher wage rate or raw material price, modifies its technique of production, the situation will be different only in degree from that found above. If, for instance, wage rates increase but raw material prices do not, the firm may now find it desirable to produce a given output by the use of slightly less labor and a somewhat larger raw material quota. The firm will be able to reduce its costs slightly, or rather it will be able to avoid a part of the increase in costs resulting from the rise in wage rates, if it is able to substitute raw material for labor. Such a substitution could, in a case like this, be accomplished by reducing inspection during the process, by reducing efforts to salvage recoverable raw materials, or in many other ways. But a moment's reflection will show that even so, costs will rise somewhat, though less than if the process had not been changed. For if this were not so, we should be compelled to explain why this substitution of raw materials for labor had not been made *before* wage rates were increased.[1] Hence we may conclude that even with an induced change in the technique of production, the total cost will increase, as it did when the price of productive factors rose but no such changes were made; and that in consequence both average total cost and marginal cost will rise also. Costs will increase, whether or not there is a substitution of one factor for another, and again they will have to be represented by new cost curves, as in Figure 12.

A change in the price of any of the factors comprising fixed costs will give somewhat different results than a change in variable costs, for although the average total cost will change, the marginal cost will not. Suppose, for instance, that the salary of the vice-president is increased or that the interest charges are raised on bonds issued by the firm. Obviously the total cost of any output in this case also will be higher, just as it was when variable costs were raised, and if total costs are increased, the average total cost will also be increased.

But what about marginal cost? Remember that the marginal cost of any output is the addition to total cost resulting from the produc-

[1] It is of course possible that management had not previously given attention to the possibility of reducing cost by adopting a different technique and that the increase in wages forced it to consider this problem. In that case, a rise in wages might have the ultimate effect of reducing costs below their original level.

tion of one additional unit. What then can bring about a change in the addition to total cost? Will the salary of the vice-president be greater *because* output is 1001 units instead of 1000 units? Obviously not — even if the vice-president receives $15,000 a year instead of the $12,000 he formerly received. For his raise would normally not depend upon the amount produced. It would ordinarily be a part of fixed costs. The marginal cost of the 1001st unit has to cover the addition to the wage and raw materials bill that results when output is increased from 1000 to 1001 units. But since the change in executive salaries has not been due to the expansion in output, we can hardly burden the 1001st unit with any part of this increase in salary. The increased output will entail the *same addition* to costs whether the vice-president receives $2000 a year or $50,000. Likewise, the increased output will involve the firm in exactly the *same increase* in costs whether the bondholders receive $60,000 or $70,000 in interest. Because the change in the salary of the vice-president is not the result of the change in output, such a salary change will not affect the marginal cost for any particular unit. This can be illustrated most clearly by working through a numerical example, as in the following table.

TABLE 19

Relation Between Output and Marginal Cost

When Output Is	Variable Costs Are	Fixed Costs Are	Marginal Cost Is Therefore
100	1000	400	–
101	1009	400	9
102	1018	400	9
103	1026	400	8
104	1034	400	8

If, because of an increase in executive salaries, the property tax, rents, or interest payment on outstanding bonds, the *fixed* costs were raised to $500, the *marginal* cost would not be in any way affected. For example, the marginal cost of the 101st unit would be 1509 minus 1500, or 9, where originally it was measured as 1409 minus 1400. Thus only a change in any of those items which enter into variable cost would affect marginal cost. Hence when salary rates or other fixed cost elements change, we have a change in the average total cost but not in the marginal cost. In fact, we should expect a

change in the latter only if there has been a change in the price of one of the productive factors which enter into variable cost. Changes in fixed cost have no bearing on the level of marginal cost. This fact is important in the development of economic policy, as we shall see later.

Costs and Improved Techniques of Production

A new process of production will almost certainly bring about a change in costs. And since it is nearly always adopted voluntarily, we can be reasonably sure that it will reduce costs; for otherwise the firm would have no reason for altering its production methods. If the new technique reduces the amount of labor that will be needed, the firm's payroll will of course be lowered. If so, there will probably be an increase in certain other charges, for the new technique may require the use of more expensive machinery or more raw materials. We may call such an improvement in technique a "labor-saving invention." There have been some spectacular examples of such technological developments in the economic history of this country. But not all new techniques allow the firm to save labor. On occasion "capital-saving inventions" have been introduced, and in this case labor charges may be somewhat increased while other costs are considerably reduced. In both cases, however, since it is in the interest of a firm to adopt only those new processes that will lower the level of total costs, we can be sure that total costs and average costs will be lower than they would have been without the new technique.

The Importance of Technological Change in the Economy of the United States

Technological improvements are important in an economy like ours. Certain major innovations such as the use of the assembly line have profoundly changed methods of production in most industries. But these changes are relatively rare. Others, perhaps unimportant when taken singly, occur so frequently and are applied so generally that their cumulative effect, in a period of only a few years, makes for drastic reductions in cost in a great many industries. The use of semi-automatic control equipment is an example of such an improvement. These improvements in technique may not be noticed by the outside observer, but their total effect is nonetheless very great.

In many industries in the decade before the Second World War, increases in the productivity of labor reduced real labor costs by as

much as 25 or 30 per cent; and this improved efficiency was the result of a large number of "small" improvements. In many industries there is a long-term and persistent tendency for costs to fall as a result of developments and improvements in methods of production. It is not always clear that in these cases marginal costs also decline. To determine whether they do or not, each situation must be examined in detail. In some cases it may be found that even though average costs are somewhat reduced, marginal costs are raised. But generally it may be expected that marginal costs are reduced too.

Effect on Cost of Change in Size of Plant

It will be remembered that in treating the variation in cost with respect to output, we supposed that some of the firm's productive factors were fixed in amount while others, such as labor and raw materials, were variable. It was assumed that the firm has a plant of a given size under the management of a certain group of executives. In such a firm, output is altered by employing more or fewer wage earners, or by purchasing more or fewer raw materials, rather than by adding to plant, or scrapping part of it, or by hiring or firing several vice-presidents. That is, certain factors of production were regarded as variable, and others, like the plant, were assumed to be relatively fixed. As a general rule, and especially when we are concerned with adjustments over short periods of time, such an assumption is realistic.

Suppose, however, that there is a change in one of these "fixed" factors, such as the size of the plant. How do the "cost functions" of the new larger plant compare with those of the smaller plant which the firm formerly maintained? In other words, how will this change affect the average total and marginal cost for each given output?

To begin with, we can safely assume that the plant will not be enlarged unless it is clear that the expansion will reduce costs for a *certain range of output*, though not necessarily for every level. If, for instance, the firm expects to produce at least 100,000 units a year after the new addition to its plant, it would certainly not be deterred from expansion by the fact that the average cost of producing 20,000 units in the larger plant will be higher than it would have been in the original plant. In the expanded plant, fixed costs will almost certainly be higher at every level of output, but variable costs should be lower for certain levels, and particularly for those which are rather

close to the capacity of the original plant. Comparing the average cost curve in the expanded plant with that in the original, we should therefore expect that after the expansion, costs would be somewhat higher at very low levels of output, but somewhat lower at higher levels. This situation is represented graphically in Figure 13.

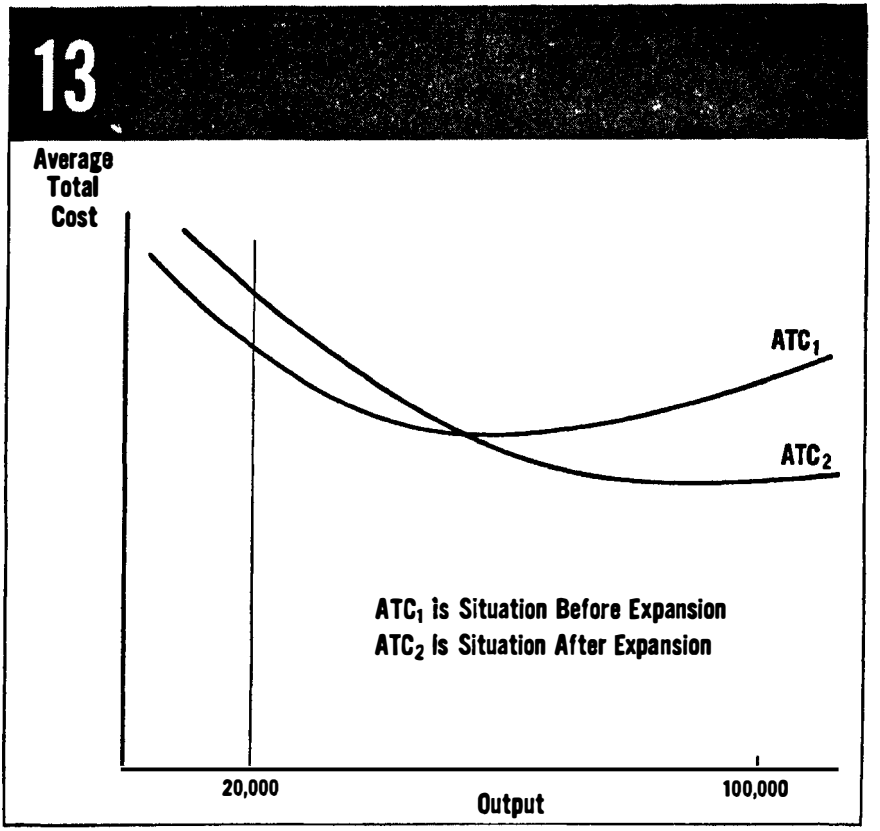

Figure 13. *Effect of Plant Expansion on Average Costs*

The effect of an increase in the size of the plant on the marginal cost function is more difficult to determine. At output levels beyond the former capacity of the plant, marginal cost would now be lower than before the expansion, simply because the marginal cost of any expansion beyond capacity is infinitely high. But in the middle and lower ranges of output, the effect on marginal cost of an expansion in the size of the firm cannot be definitely forecast.

Summary: The Cost Determinants

We have seen in analyzing costs that there are three determinants which, if they are changed individually or together, will alter average cost and probably marginal cost as well. Any change in the price which the firm has to pay for the productive factors it employs will usually alter the average cost of producing any given output and will probably alter marginal cost also. The exception arises when the price of a fixed cost factor is changed, for marginal costs are then not affected. In the same way, any change in the method of production — that is, the exploitation of a new process of production — will alter both average and marginal costs. Finally, a change in the size of a firm — that is, a variation in the use of the "fixed" factors — will mean a different level of average and marginal costs for wide ranges of output. The three cost determinants are: (1) the price of the factors of production, (2) the technique of production, and (3) the size of the firm, and a change in any of them will bring about a change in average and probably also in marginal cost. Furthermore, average and marginal cost functions will not change except as the result of a change in at least one of the cost determinants.

The analysis of costs in the individual firm which has been presented in this and the preceding chapters is a necessary prerequisite to an understanding of the way in which a firm determines its output. We must now analyze along similar lines the factors that affect a firm's receipts. Then we shall be able to see how a firm determines output and how it changes output in response to changes in the economic situation.

10

The Demand for the Product of a Firm: Introduction

NATURALLY, since the aim of every business is profit, the business firm endeavors to produce that amount of goods, and to sell them at the price, that will yield the maximum profit. One of the elements in this problem, as we have seen, is cost. But in determining profits the firm must also concern itself with sales receipts, the other major element in the profit formula. In the chapters which follow we shall therefore give the same kind of attention to the question of determining sales receipts as we have given to the consideration of costs.

Receipts, Price, and Number of Units Sold

On what factors do the total sales receipts of a firm depend? The answer to this question depends on how far back we wish to carry the analysis. Perhaps the most obvious statement we can make about this problem is that sales receipts are determined by the number of units of a commodity that can be sold at any given price. Thus, if at a price of 10 cents a unit, the firm is able to sell 10,000 units in a week, its sales receipts would be $1000. Or, if at a price of p dollars a unit, it is able in a given period to sell n units, its sales receipts would be $\$pn$. Therefore in the simplest and most obvious terms the sales receipts are the product of the price and the volume of sales. Let us then first consider the effect of changes in price on the volume of sales and consequently on sales receipts.

Relation Between Price and Receipts

We have seen that if at a price of 10 cents a unit, the firm should sell 10,000 units a week, it would have sales receipts of $1000. But a

little reflection will surely convince us that if, in that week, the price had been not 10 cents a unit but 15 cents, the firm normally could not have sold as many units. Perhaps at 15 cents a unit its sales would have amounted to only 6000 units instead of 10,000. In that event, its total receipts would not be $1000 but only $900. Again, if the firm had charged a still higher price, let us say 20 cents a unit, the number of units sold in that week would have been still less, perhaps only 4000 units, and consequently its total receipts would have been only $800. One factor, then, on which total sales receipts depend is the price charged.

Prices Charged by Competing Firms and Receipts

Certain other factors must also be taken into account in our analysis of the firm's sales receipts. One of these is the price charged by competitors, for a firm usually must compete with other firms for customers. Let us suppose that our firm is in competition with others which sell a similar commodity. Now if the other firms raise the price for their products, the sales of our firm, and its sales receipts, will be somewhat higher at each price, since some of its competitors' customers would switch to it. In the preceding paragraph we supposed that our firm could sell 10,000 units at a price of 10 cents a unit and 6000 units at a price of 15 cents. But if our competitors should all raise their prices, our firm might find itself able, if it still charged 10 cents, to sell not 10,000 units but perhaps 12,000; and if it charged 15 cents, it might sell not 6000 units but perhaps 7500. The increase in sales at each price comes about because, as our competitors charge more, some of their customers transfer their trade to our firm. Hence the sales receipts of our firm depend not only on its prices, but also on those of its competitors. If the price charged by our firm is raised, its sales will decline, and its sales receipts may fall too; while if the prices charged by its competitors are raised, the volume of its sales and its total sales receipts are both bound to rise.

Prices Charged for Complementary Products and Sales Receipts

Sometimes, of course, other firms are related to ours not as competitors but as suppliers of complementary products. For example, the producers of plumbing supplies and the manufacturers of electrical fixtures are so related, since both kinds of products are needed jointly when new houses are built. If an increase in the price of plumbing

supplies retarded the construction of new houses, it would also cause a decline in the number of electrical fixtures that would be sold at a given price, for sales of this kind of product are also closely geared to the number of houses under construction. Hence, if the relationship between firms is not competitive but complementary — in the way that ham and eggs are complementary — a rise in the price charged by related firms will reduce the sales and sales receipts of our firm; whereas, as we saw in the last paragraph, if the relationship is competitive, our sales receipts would rise following a price increase by our competitors. In conclusion, then, we see that our sales and sales receipts are affected not only by the price our firm charges, but also by the prices charged by related firms, whether complementary or competitive.

Sales Receipts and Level of Income

The volume of sales and the sales receipts of our firm also depend on the income of our customers and potential customers in the period under consideration. Normally, the higher the income the more our firm could sell at each price. To illustrate: if in moderately prosperous conditions our firm is able, as we have assumed, to sell 10,000 units a week at a price of 10 cents a unit, it might be able to increase its sales to 12,000 units a week with the same price in a period of great prosperity. On the other hand, during a deep depression it might find that at a price of 10 cents it could sell only 6500 units a week. Thus sales receipts are likely to vary directly with consumers' income.

For certain commodities the above relation may be reversed. This would apply particularly to commodities such as very cheap clothing, which is bought by families with very low incomes. During depressions, firms that produce low-cost clothing find that their sales and hence sales receipts are high, because in such periods a very large number of families have very little money to spend — certainly none for expensive clothing. But when prosperity is restored, the number of families in the very low income group is reduced; and thus, even though there should be no change in the price charged for such clothing, there will be a decline in sales. For most products, when incomes rise, sales receipts also go up, but for some sales receipts fall. Hence whatever the direction of the change, there will be some relation between the level of income in the consuming market and the sales that

can be made at each price — and therefore between the income level and the firm's sales receipts.

Sales Receipts and Desire for the Product

Finally, the sales of our firm will change whenever there is a change in the customers' desire for its product. If for any reason buyers and potential buyers want the product more than formerly, the volume of sales will be higher at each price than it was before. The desire for the commodity may of course change for reasons over which our firm has no control at all, such as the effect of impending cold weather on the desire for coal, anti-freeze, galoshes, or mittens. Likewise shifts in style — from short skirts to long, for instance — or changes in requirements because of uncontrollable events like flood, drought, or war might also bring about a change in the buyers' desire for the product. Or, the firm may deliberately set out to create or increase the desire for its product. It may, for instance, adopt a positive policy to increase sales by advertising more vigorously or by packaging its product more attractively. So many examples of both these techniques will instantly come to mind that none need be cited here. Provided that these sales efforts are successful, the volume of sales will increase at each price. Where formerly it sold 10,000 units at 10 cents a unit, it may now be able to sell 20,000 at that price. Hence the firm will enjoy higher sales receipts. Public desire for the commodity may, of course, decline just as rapidly and just as fortuitously as it may rise. A change in fashion, a sudden turn in the weather, a transportation strike which prevents people from getting to their usual vacation spots, or the development of a new product which lures customers away from the one sold by our firm, may cause a sharp and sudden drop in the desire for it. Or, as with some durable consumers' goods, such as electric refrigerators, washing machines, automobiles, or radios, we naturally expect a decline in the desire for new ones as soon as consumers have acquired adequate stocks. For example, most families have no desire for an additional washing machine when they have recently purchased one. And in varying degrees this is also true of most household and electric appliances, automobiles, lawnmowers, and durable goods of many kinds. The result of such a change in desire is of course that the volume of sales and sales receipts falls off at each price. The firm sells less at 10 cents than before, less at 15 cents than before, and so on.

Summary: The Demand Determinants

It may thus be seen that the number of units of a good that a firm can sell in a certain period, or the amount demanded, at any given price, depends upon a large number of factors of the sort described above. If the income of the buyers alters, if there is any change in the price of related goods, or if the desire for an article changes, the volume of sales *at any one price will vary*. We shall call these factors — buyers' income, buyers' preferences, and the prices of related goods — the *demand determinants*.

Demand Determinants, Price, and Volume of Sales

When the demand determinants are fixed, price alone determines the volume of sales and hence the sales receipts. Let us suppose that the demand determinants are given and constant — that is, that customers' incomes and tastes are fixed for a certain period and also that the prices of all related goods are maintained at a given level. In this situation the only factor that could cause a change in the sales of our firm would be a change in the price it charged. For when we assume that all the demand determinants are fixed, we have only one possible source of a change in the volume of sales; namely, a change in the price of the product itself.

Since price and the three demand determinants all determine a firm's sales receipts, it may be asked why we treat price differently from the other variables; why, that is, we treat the level of buyers' income as a demand determinant, but treat price as one of the variables of the demand function? From a strictly mathematical point of view there is no reason for this distinction. Since all four variables affect the sales receipts of the firm, we could treat price also as a demand determinant, and could then consider separately the relation, let us say, between buyers' incomes and amount demanded. But such a procedure would not be sensible as economics. We are interested in analyzing the factors which affect the actions of the business firm. The firm has practically no influence in determining the national income; it has no direct influence in setting the prices charged by competing firms; and it has little influence in formulating buyers' preferences. But it does set its own prices. Hence we select the relation between price and the volume of sales (or sales receipts) for special consideration, and we group together the other variables over which the firm has much less control as demand determinants, factors

which operate upon the volume of sales and sales receipts through their effect on the relation between price and the amount demanded.

Relation Between Price and Amount Demanded

The relation between the amount demanded and the price is generally inverse. This means that with the demand determinants constant, the higher the price the less is the amount sold, and the lower the price the greater is the amount sold.[1] But this generalization must be interpreted and applied with very great care. In particular, we must not expect it to hold *if the demand determinants are not constant.* For instance, in August, 1941, clothing cost about 7 per cent more than the average figure for the period from 1935 to 1939. But sales of clothing on an annual basis, instead of being lower in 1941, as a careless reading of the above generalization might suggest, were in fact 65 per cent higher then than from 1935 to 1939. Does this phenomenon contradict the rule that price and quantity sold vary inversely? Most certainly not. For the demand determinants in August, 1941, were very different from those prevailing in the years from 1935 to 1939. The most obvious change was in income. Incomes were received in the United States in August, 1941, at a level of about $87 billion a year. But between 1935 and 1939, the yearly income payments amounted to only $65 billion. Undoubtedly the other demand determinants also changed considerably between these two periods.

In fact, it is almost certain that the demand determinants will vary between any two periods in a changing, dynamic economy like ours. Consequently our generalization must be phrased so that the crucial condition of unchanging demand determinants is an integral part of it. But the only way in which we can insure constancy in our demand determinants is by confining ourselves to a particular period of time. Therefore, we should treat the prices that are subject to change as *alternative* and not as *successive* prices. In other words, the relation between price and volume of sales should properly be expressed as follows: "If in a given period a certain price is charged for the product of a firm, the sales, or amount demanded, will stand at a certain level. But if *in that same period* a higher price had been set, the sales would have been lower." In other words, if the price charged by a firm in a certain period is high, its sales, or the amount demanded, will be lower than if in that same period it had charged a lower price.

[1] The terms *quantity* or *amount sold* and *amount demanded* are used interchangeably.

The Demand or Average Revenue Curve

The relation between the price a firm charges and the volume of its sales can be shown in a table like the following, which we shall call a demand table:

TABLE 20

Relation Between Price and Volume of Sales

If the Price Is	The Volume of Sales Will Be
$1.50	100
1.45	110
1.40	119
1.35	129
1.30	140
1.25	152
1.20	165
1.05	200
.75	280

To describe the situation fully, this table would have to cover every price from 1 cent per unit up to the very highest price at which one unit could be sold, say $5; and it would have to do this at intervals of one cent. But there is a shorthand method of recording all this information which, as we saw in our discussion of costs, economists generally employ. It says no more and no less than the cumbersome, never-ending table, but it says it much more simply. We shall generally use this more convenient method, which consists in plotting the information on a graph. On one axis of a graph similar to the one on which we earlier measured costs, we shall now measure price, and on the other axis we shall record the amount demanded. The price axis could be marked off at intervals of five cents, and the volume-of-sales axis at intervals of single units or, as we shall do here, of 50 units. Our graph will then look as shown on the next page.

In this graph we could record all the information that is presented in the table. For instance, we read from the table that at a price of $1.20 per unit, the firm is able to sell in the given period 165 units. On the graph, a single point shows this, a point located at $1.20, or 24 intervals along the vertical axis, and 165 units, or 16.5 intervals along the horizontal axis. This is point *A* in our diagram. The information that the firm could sell 152 units if it had charged not $1.20

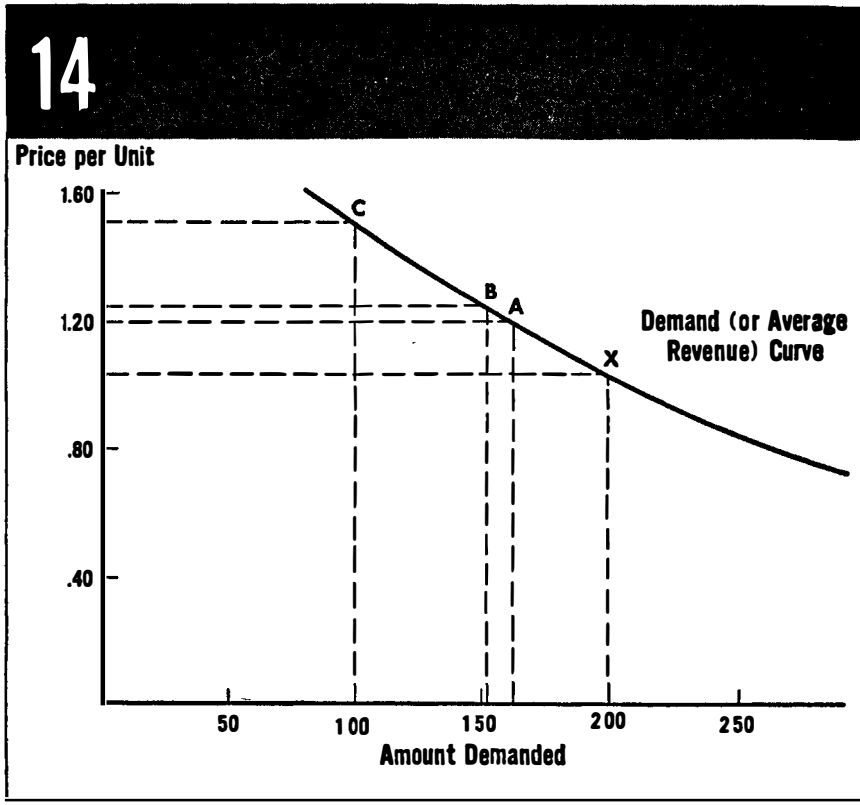

Figure 14. *The Average Revenue Curve*

but $1.25 would be provided by another point, *B*, somewhat higher and to the left of point *A*. Similarly, the information that at a price of $1.50 per unit the firm could sell 100 units would be shown by point *C*, which is somewhat higher than point *B* and to the left of it. Unless we wanted a quite unnecessary accuracy in the completed picture, we would have to record only a few more of these relationships. When enough of them had been recorded, perhaps six to ten, depending on how accurate we wished our results to be, a smooth line drawn through these points would give closely enough for most uses the information which would be contained in a very detailed table.

The value of such a curve is that it will show at a glance the approximate volume of sales at any given price. Thus point *X*, for in-

stance, is 105 units above the horizontal axis and 200 units to the right of the vertical axis. It thus tells us that at price $1.05, the volume of sales will be 200 units in each period.

A Change in Demand or Average Revenue

We must remember that all the relations shown in the above table and diagram exist only so long as the demand determinants are constant. That is, the table and the diagram were constructed on the assumptions that (*a*) the prices of all related commodities are given, (*b*) the incomes of the customers are fixed, and (*c*) the desirability of the good in the eyes of the buyers is given and constant. Suppose, however, that one of these determinants changes — that, for example, customers' income increases. We saw earlier that as a result, the volume of sales at each price would normally be somewhat higher than formerly. As the table was first drawn up, it was assumed that the firm was able to sell 100 units in the period at a price of $1.50. At the higher level of income it will be able to sell, let us say, 115 units at that price. And this increase in the volume of sales will occur not only at $1.50 but also at all other prices. Therefore, the right-hand column of the table would have to be completely rewritten. Since our graph is based on the table, we will also have to redraft the diagram. The new curve will have to be drawn in such a way as to show that at each price the volume of sales now exceeds the former level. Such a curve will be to the right of, and above the original curve, as in Figure 15.

These curves are referred to as *average revenue* or *demand* curves, and the relation they illustrate may be called the average revenue function or the demand that faces the firm. We may define the average revenue function as the relation that exists between the price charged and the volume of sales, given the demand determinants. If the demand determinants change, the relation between price and sales of course alters. A change in this relation is referred to as a change in demand or in *average revenue*.[1] This is the kind of change illustrated in the previous paragraph.

Distinction Between Change in Demand and Change in Amount Demanded

The term *demand* has an exact meaning for the economist, though

[1] We shall generally use the term *average revenue* instead of the longer term *the average revenue function.*

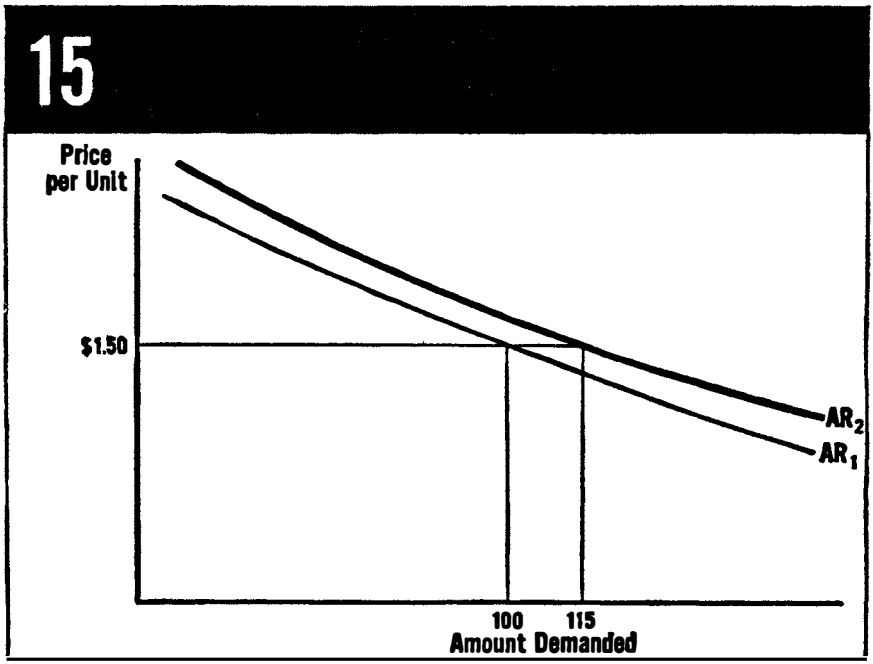

Figure 15. *Change in Demand*

in ordinary conversation it is frequently used much less precisely. Hence there is a real danger of confusion unless we are careful to use the term in its technical sense. The greatest difficulty arises from the practice of using the term *demand* to stand for the *amount demanded*. Let us try to make the distinction clear.

The demand for the product of a firm refers to the relation between price and the amount demanded. The following table illustrates the demand for a commodity:

TABLE 21

Relation Between Price and Amount Demanded

When the Price Is	The Amount Demanded Is
10	200
15	185
20	160
25	140
and so on	and so on

The whole table is needed to describe the demand for this item. If the table has to be revised, then it signifies the demand has altered. For instance if, instead of the above figures we had the following, the demand would be higher.

TABLE 22

Variant: Relation Between Price and Amount Demanded, When Demand Determinant has Altered

When the Price Is	The Amount Demanded Is
10	220
15	200
20	180
25	155
and so on	and so on

Such a change in demand can only occur because of a change in one or more of the demand determinants.

But does a change in price change the demand? According to certain writers for the newspapers, the answer is yes, for we may frequently read such a statement as "Of course it is well known that when the price rises, the demand falls." But let us look at the table immediately above more carefully. When the price is 10, the amount demanded is 220. When the price is 15, the amount demanded is only 200. There was a change in the amount demanded, but the change in price does not force us to draw up a new table; we have already allowed for the effects of a change in price in the table as set out. A change in demand occurs only if at a given price, say 10, there is a change in the amount demanded. And when the price

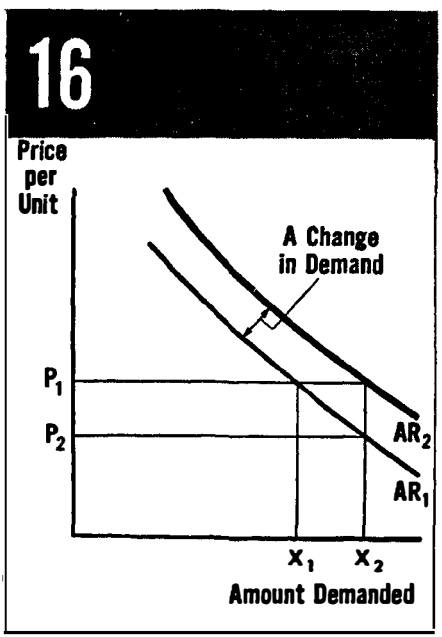

Figure 16. *Change in Amount Demanded and in Demand*

changes, it causes not a change in demand but simply a change in the amount demanded.

The diagram (Figure 16) illustrates graphically the distinction between a change in demand and a change in the amount demanded. A change in demand is portrayed by a shift from the AR_1 curve to the AR_2 curve, or the reverse. A change in the amount demanded is shown by the change from X_1 to X_2 on the horizontal axis. This change may be due simply to a change in price from P_1 to P_2, demand remaining the same, or it may be due, as is also shown in the diagram, to a change in the demand itself, in this case with no change in price.

Summary

The sales receipts of the firm depend, therefore, on the demand for its product or the average revenue function that faces the firm, as well as on the price it charges. The average revenue curve or table provides the data required for computing sales receipts: price for each unit and the number of units that can be sold at this or any other price. A change in the average revenue function or in the demand occurs as the result of a change in any of the demand determinants; that is, as a result of a change either in the prices of related goods or in the incomes or desires of the buyers.

11

Elasticity of Demand and the Marginal Revenue

BEFORE WE CONSIDER THE FACTORS that affect the demand for the product of a firm, we must examine certain characteristics of the demand or the average revenue function. There are obviously a number of things that can be said about a demand function — it can be said, as we have already seen, that it is high or low, or that it is falling or rising. But there are other important things too; that part of the description of the demand function which is conveyed by the measure of its elasticity, a concept to be introduced in this chapter, is an especially useful one.

Change in Price, Volume of Sales, and Total Sales Receipts

We have seen that the demand for the product of a firm expresses the relation between the price it charges and the volume of sales, or, in other words, between price and the amount demanded. A change in the price will normally cause a change in the opposite direction in the volume of sales. But what happens to the total sales receipts if the price is changed depends on the size of the change in the volume of sales. For instance, if a rise in price from $1.20 to $1.25 leads to a decline in the amount demanded from 162 to 156 tons, sales receipts will increase from $194.40 to $195. But if the decline in sales had been either larger or smaller than this, total sales receipts would have decreased, or they would have increased by a greater amount than they did, as a simple arithmetical calculation will show.

The Responsiveness of Volume of Sales to Price Changes

When the volume of sales is very responsive to a price change, total sales receipts move in the *opposite* direction from the change in price. In such a case, if the price rises, sales will drop sharply and total sales receipts will likewise decline. And if the price is reduced, the amount demanded, or the volume of sales, will increase so greatly that sales receipts will also rise. On the other hand, when the volume of sales responds only slightly to price changes, total receipts will move in the same direction as price, for in that case there will be but a small offset through the change in sales to the effect of the price change itself on total receipts. Thus, the responsiveness of the amount demanded to a change in price determines the direction and amount of the change in total receipts. The degree to which the volume of sales responds to price changes is measured by the *elasticity of demand* or of *average revenue*.[1] When the response is great, the average revenue or demand is very elastic. When the response is slight, the demand has a low elasticity or is said to be inelastic.

Elasticity of Demand

The actual measurement of elasticity is simple. Since we want the measurement of elasticity to be high when the volume of sales responds *very freely* to price changes, it would seem reasonable at first glance simply to measure it by the ratio of the change in the volume of sales to the change in price. But this ratio would have little meaning, for we would, so to speak, be dividing six tons by five cents — and if the six tons were re-expressed as 12,000 pounds, our answer would be quite different and equally meaningless. That is to say, such a measure of elasticity would depend on the kind of units chosen. If, however, we compared not the *absolute* changes in sales and price, but the *relative*, or *percentage*, changes in these two variables, we would have a meaningful ratio, which does not depend on the kind of unit in which we measure. We shall therefore measure the elasticity of average revenue or demand by the ratio of the relative change in volume of sales (or the amount demanded) to the relative change in price.

By way of illustration, let us for a moment return to the situation in which the amount demanded drops from 162 to 156 tons when the price rises from $1.20 to $1.25. In this case the relative change

[1] As noted above, we use the terms *demand* and *average revenue* interchangeably.

in the volume of sales is 6/162, and the relative change in the price is 5/120. The ratio of these two values, or 6/162 ÷ 5/120, is 8/9, and the elasticity of the average revenue is therefore 8/9, a figure which is less than one. Note that in this situation an increase in price brings about a rise in sales receipts, while a decrease in price reduces receipts.

Now let us suppose, to treat a different situation, that when the price is raised from $1.20 to $1.25, sales fall by very much more — let us say from 162 units to 150 units. In this case the elasticity of demand would be 12/162 (the relative change in the volume of sales) divided by 5/120 (the relative change in price) or $1\frac{7}{9}$. This time, with the elasticity of demand greater than one, total sales receipts would decline with the increase in price, whereas if the price had been reduced, sales receipts would have increased. Sales receipts vary in the opposite direction from price when the *elasticity is greater than one.*

If the demand is such that a given percentage change in price occasions an equivalent percentage change in the volume of sales, we have *unit elasticity.* For in this case the ratio of the relative change in the amount demanded to the relative change in price equals 1. This value, 1, marks the boundary between an elastic and an inelastic demand. When the ratio is greater than 1, the demand is said to be elastic, and when it is less than 1, the demand is said to be inelastic. Now since, when the elasticity of demand is equal to 1, the relative change in the volume of sales is equal to and opposite in direction from the relative change in price, there will be no change in total receipts as price varies. For example, if as price is raised by 1 per cent, the volume of sales falls by 1 per cent, total receipts do not change. For the effect on receipts of the change in price is exactly offset by the change in the amount demanded.[1] With unit elasticity, therefore, total receipts are not changed when there is a change in price or the amount sold.

The final limiting case remains to be discussed. As demand becomes more and more elastic, a smaller and yet smaller decline in price will bring about a given increase in the volume of sales. If the elasticity of demand is 1, the percentage change in price will be as great as the percentage change in the volume of sales. If the elasticity

[1] This will not be exactly true unless the percentage changes in each case are very small, but this is because of difficulties involved in measuring *relative* changes by reference to the original values.

of demand is 10, we need a price change only one-tenth as great to bring about the desired increase in sales. As the extreme of elasticity is approached, an increase in sales can be secured by an infinitesimally small reduction in price. The elasticity of demand is then said to be infinite, or alternatively, the average revenue function is said to be perfectly elastic.

Graphic Representation of Elasticities of Different Values

Let us see how demand curves of these various elasticities are drawn. When the demand is very elastic, a small change in price has a great effect upon the volume of sales, and consequently the demand curve will incline toward the horizontal. Such a demand curve is shown in Figure 17. It will be seen that in order to increase the amount demanded by a large figure, say from 100 to 140, only a small reduction in price need be made.

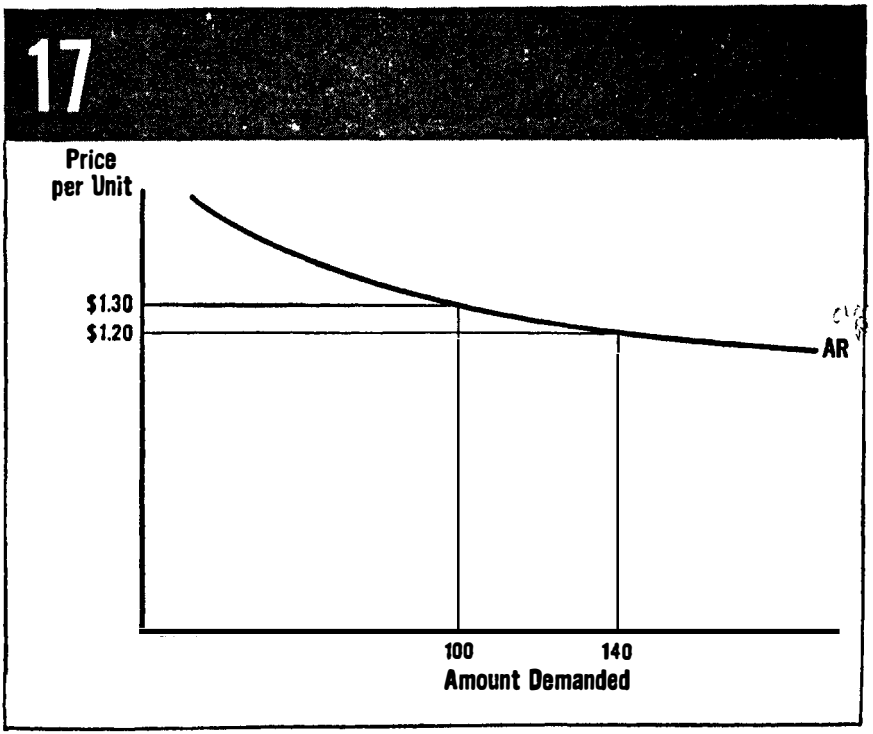

Figure 17. *Elastic Demand*

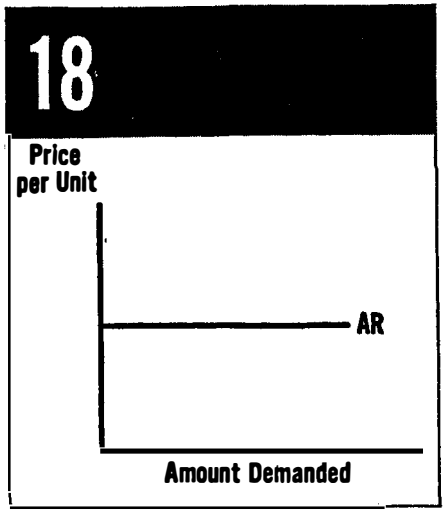

Figure 18. *Perfectly Elastic Demand*

When the demand is perfectly elastic, no reduction in price is necessary to bring about an increase in sales, and therefore the demand curve has a zero slope. In other words, it is horizontal. Such a demand curve is illustrated in Figure 18.

A relatively large price change is needed to bring about a small change in the amount demanded when the demand is inelastic. Such a situation is shown by a demand curve which has a very great slope, as in Figure 19.

We must note, however, that the slope of the demand curve is not sufficient by itself to show the elasticity, since the slope measures the ratio of the *absolute*, and not the *relative*, changes in amount demanded and in price. Hence, in addition to the slope, it is also necessary to know the actual co-ordinates of the point on the demand curve at which the elasticity is to be measured. In general, however, the greater the slope, other things being equal, the lower is the elasticity.

When the elasticity of demand is equal to one, a given percentage change in price will produce an equal change in the amount demanded. On a graph this function when plotted takes the form of a rectangular hyperbola. Such a demand curve is illustrated in Figure 20.

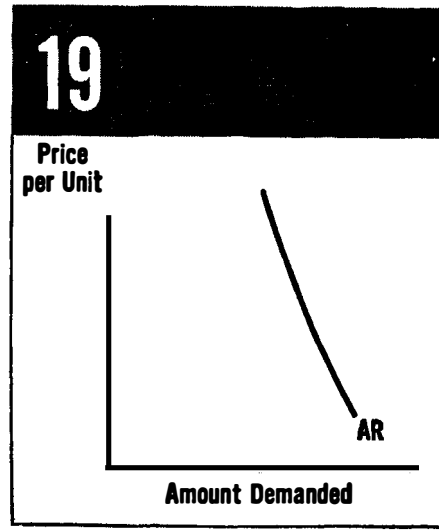

Figure 19. *Inelastic Demand*

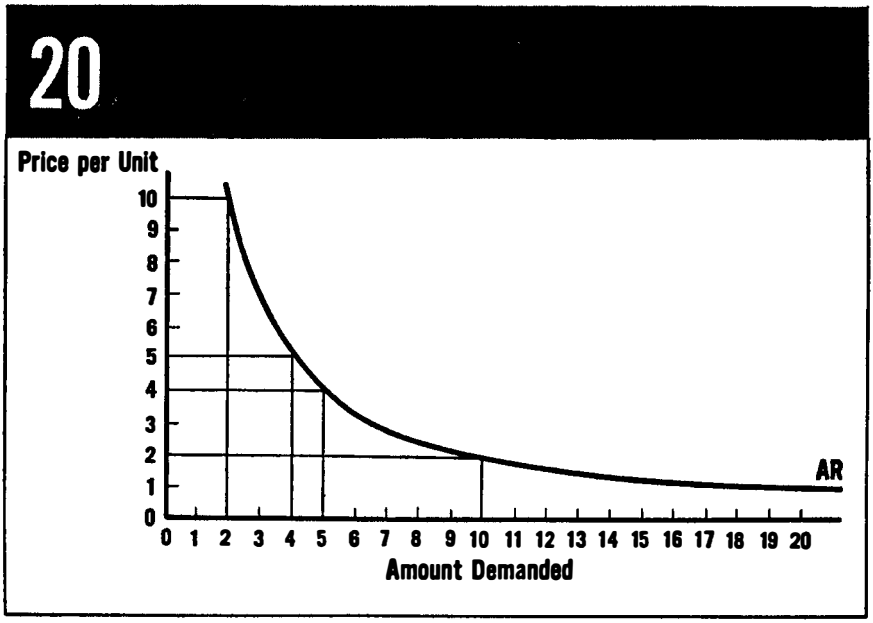

Figure 20. *Demand of Elasticity Equal to One*

Changes in Sales Receipts, Different Elasticities of Demand

We have seen that when the price alters and the demand is constant, sales receipts are likely to change. Since, with a constant demand, a change in price produces an opposite change in the volume of sales, we may equally well say that when the volume of sales alters, sales receipts are likely to change. When the volume of sales increases, total sales receipts will *increase* if the elasticity of the average revenue function is in excess of one: they will *decrease* if the elasticity of the average revenue function is less than one. When the elasticity of demand is equal to one, there will be, as we have seen, no change in sales receipts whatever the number of units sold.

The Marginal Revenue

When a firm cuts its price enough to increase its sales by one unit, its total receipts will generally change. This change in its total receipts which results from a one-unit increase in the volume of sales is known as the *marginal revenue*. It measures exactly the same feature of a firm's sales receipts as the marginal cost does for its costs. Each

of these concepts refers to the difference in a total as the result of a one-unit increase in sales. The unit in which we measure sales should, of course, be determined realistically and in accordance with business practice. We are not concerned with the change in receipts that occurs when the sales of the American Tobacco Company are raised by one carton of cigarettes, for it is clear that to produce one more carton is not the kind of decision that the management of such a company bothers about. Probably they do not think even in terms of single cases. Instead, the meaningful unit for such a firm might be twenty-five or fifty cases or even more. The unit in which we measure sales and quote prices should correspond to that used by the firm in making its decisions as to output.

Let us see how the marginal revenue is determined. Suppose the demand for the product of the firm is such that at a price of 5 it will sell 10 units. At that price its total sales receipts are therefore equal to 50. If it can sell 11 units in the period at a price of 4.8, then its total receipts will equal 52.8, and the difference of 2.8 is the marginal revenue corresponding to the eleventh unit. And if it can sell 12 units at a price of 4.5, its total receipts will be 54 and the marginal revenue for the twelfth unit would be 54 minus 52.8, or 1.2. In each case the marginal revenue measures the change in total receipts that results from expanding the volume of sales by one unit, and it is computed by subtracting from the total receipts of a certain output the total receipts of an alternative output one unit smaller. The following table illustrates the computations required.

TABLE 23

Determination of Marginal Revenue

Amount Demanded	Price per Unit	Total Receipts	Marginal Revenue
10	5	50	—
11	4.8	52.8	2.8
12	4.5	54.0	1.2
.			
.			
17	R	R × 17	
18	S	S × 18	(S × 18) − (R × 17)

The Elasticity of Demand and Marginal Revenue

We have already seen that if the elasticity of demand is greater than 1, a decrease in price brings about an increase in the volume of sales and an increase in total receipts. Hence the marginal revenue is *positive* when demand is elastic. But if the elasticity of demand is less than 1, an increase in sales brings a decline in sales receipts, and the marginal revenue is accordingly *negative*. If the elasticity of demand is equal to 1, the marginal revenue is zero, since total receipts are constant no matter how much is sold. Finally, if the elasticity of demand is infinite, the marginal revenue is equal to the average revenue. Thus, when the price at which n units are sold is p per unit, total receipts are equal to np. Since it is not necessary to reduce the price in order to sell $(n + 1)$ units, total receipts from this expanded volume of sales will be $p(n + 1)$, or $np + p$. The difference between the total receipts from the sale of n units and from the sale of $(n + 1)$ units (the marginal revenue corresponding to the $(n + 1)^{st}$ unit) is equal to $(pn + p - pn)$ or p — which is also the unit price. In this limiting case, then, the marginal revenue and price are equal. In any other case the marginal revenue for each unit is less than the price that can be charged for that volume of sales.

The proof of this last statement is simple. As before, let p refer to the unit price at which n units can be sold. If $(n + 1)$ units are to be sold, the price per unit will be lower, or $(p - r)$, *where r is a positive number*. Total receipts, when sales are $(n + 1)$ units, are $(p - r)(n + 1)$, or $pn + p - r(n + 1)$. When sales are n units, total receipts are pn. The difference in total receipts, that is, the marginal revenue corresponding to the $(n + 1)^{st}$ unit, is $(p - r)(n + 1) - pn$ or $p - r - rn$. But the price for each unit when sales are $(n + 1)$ is $p - r$. Hence the marginal revenue $(p - r - rn)$ will be less than the price $(p - r)$ by the amount rn. Or we may put it in numbers. Let the price at which a firm can sell 100 units be 1000, and 101 units be 995. The total receipts when 100 units are sold are 100,000; when 101 units are sold, 100,495. The marginal revenue is then 495, which is less than 995 — the price. Hence, when the demand is less than perfectly elastic, the marginal revenue is less than the average revenue.

We can plot the marginal revenue on a graph just as we plot the average revenue or demand. If the demand is inelastic, the marginal revenue will be negative, as in Figure 21. If the demand is relatively elastic, the marginal revenue curve will be positive but it will be be-

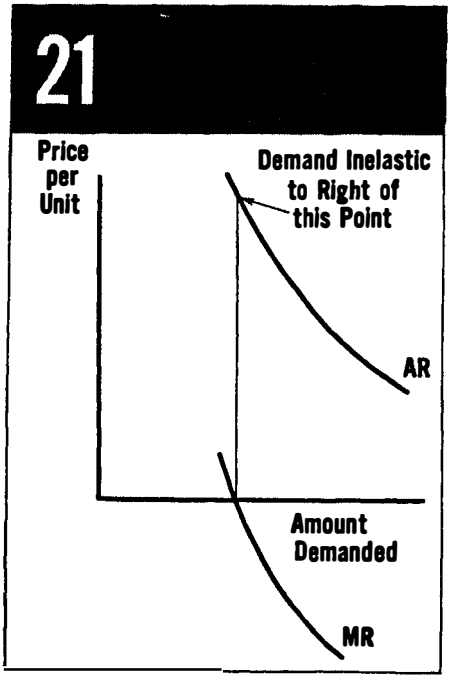

Figure 21. *Average and Marginal Revenue: Demand Inelastic*

Let us suppose, for instance, that we have drawn up an average revenue table — or, as we called it above, a demand table — for the firm. Such a table is shown below (Table 24). The first two columns of the table are similar in content to those used in the previous chapter. In column III we record the total receipts and in column IV the marginal revenue computed from column III.

Obviously the marginal revenue for the first unit of output, 100, is equal to the total

low the average revenue curve shown in Figure 22. Finally, if the demand is perfectly elastic, the average revenue curve and the marginal revenue curve will coincide, and both will be represented by the horizontal straight lines in Figure 23.

Total Receipts and Marginal Revenue

Total receipts from the sale of a given number of units are, of course, equal to the product of that number and the unit price. This is one way of representing total receipts, but it is not always the most useful way. The sales receipts can also be expressed in terms of marginal revenue, and we shall find that this form of expression is sometimes very convenient.

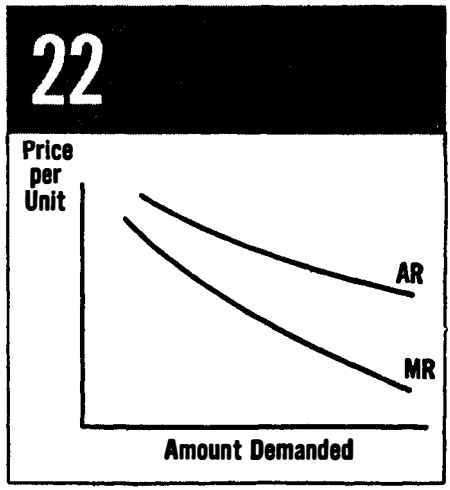

Figure 22. *Average and Marginal Revenue: Demand Elastic*

sales receipts from the sale of one unit. The total receipts for two units (160) equal the sum of the marginal revenue for the first unit (100) and the second (60). And so with three units, the total receipts (210) equal the sum of the marginal revenues corresponding to the first (100), second (60), and third (50) units.

This relation can be generalized. The total receipts for n units equal the sum of the marginal revenues for the first + the second + the third, and so on up to the nth unit. Table 25 will demonstrate the application of the rule.

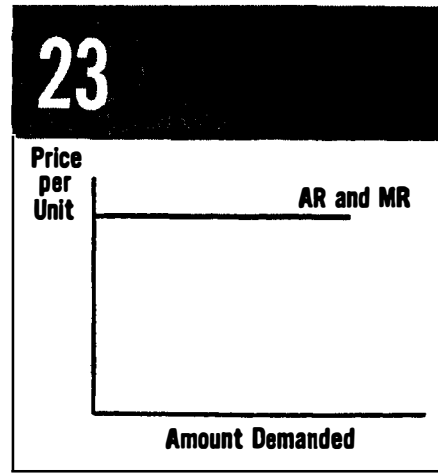

Figure 23. *Average and Marginal Revenue: Demand Perfectly Elastic*

TABLE 24

Relation Between Total Receipts and Marginal Revenue

I Volume of Sales	II Price per Unit	III Total Sales Receipts	IV Marginal Revenue
0	—	0	—
1	100	100	100 for 1st unit
2	80	160	60 " 2nd
3	70	210	50 " 3rd
4	62.5	250	40 " 4th
5	57	285	35 " 5th

We can thus express the total receipts from the sale of n units as the sum of the marginal revenue of the first, the second, the third, . . . and so on, up to and including the nth units.

It will be noticed that the relation between total revenue and the sum of the marginal revenues is essentially the same as the relation between total variable costs and the sum of the marginal costs.

The diagrammatic treatment of this equality should be clearly understood. Let us suppose that, as in the diagram we have previ-

TABLE 25

Relation Between Total Receipts and Marginal Revenue: General

Volume of Sales	Total Receipts	Marginal Revenue	Sum of Marginal Revenue from 1st Unit to This Output
0	0	—	—
1	R_1	R_1	R_1
2	R_2	$R_2 - R_1$	$R_1 + R_2 - R_1 = R_2$
3	R_3	$R_3 - R_2$	$R_1 + R_2 - R_1 + R_3 - R_2 = R_3$
4	R_4	$R_4 - R_3$	and so on, R_4
5	R_5	$R_5 - R_4$	R_5
6	R_6	$R_6 - R_5$	R_6
.	.	.	.
.	.	.	.
n−2	R_{n-2}	$R_{n-2} - R_{n-3}$	and so on, R_{n-2}
n−1	R_{n-1}	$R_{n-1} - R_{n-2}$	R_{n-1}
n	R_n	$R_n - R_{n-1}$	R_n

ously employed, we measure the number of units sold along the horizontal axis and the value of the marginal revenue for each unit along the vertical axis. Thus with reference to Figure 24, column A indi-

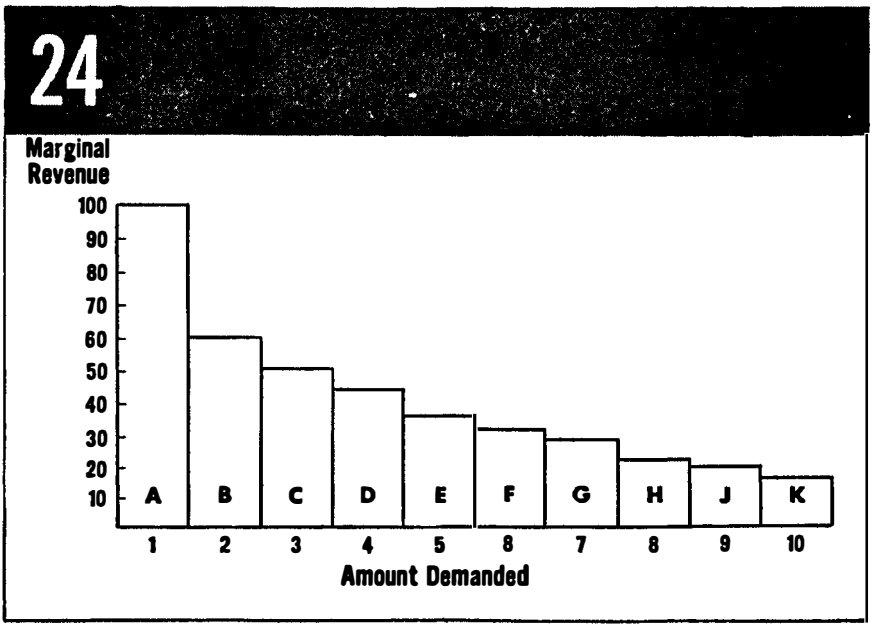

Figure 24. *Marginal Revenue and Total Revenue*

cates that the marginal revenue for the first unit is 100; column B shows that the second unit adds 60 to total receipts; and column C shows that the third unit increases sales receipts by 50 more. The three columns together are simply the sum of the marginal revenues for the first, second, and third units. If we had plotted values for the marginal revenue up to ten units, the ten columns we should have would equal the sum of the marginal revenues for all units up to and including the tenth. And, as we have seen, the sum of these columns is equal to the total receipts from the sale of ten units.

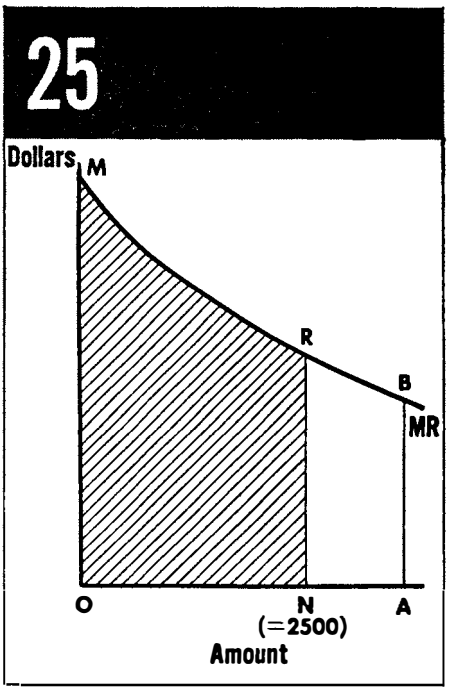

Figure 25. *Total Receipts Measured by Marginal Revenue Curve*

When we are concerned with a relatively small number of units, the marginal and average revenue curves will not be smooth but will be composed instead of a number of steps. But if the number of units is large, these steps will blend into a more or less smooth curve. (See Figure 25.) In that case, the sum of the respective marginal revenues for all units up to, let us say, 2500, would be represented by the area under the marginal revenue curve from unit 0 to unit 2500 corresponding to the shaded area *ONRM*. This area also measures the total receipts from the sale of 2500 units, since the total receipts from the sale of any output equal the sum of the marginal revenues for all units from the first to the last. To represent the total receipts from the sale of *A* units, we should have to draw a vertical line from *A* to the marginal revenue curve, cutting it at *B*. The area of the figure *OABM* is equal to the total receipts from *A* units. This method of showing total receipts will prove very useful to us in later chapters.

It must also be remembered that total receipts can be obtained by multiplying the volume of sales by the price per unit. Just as total

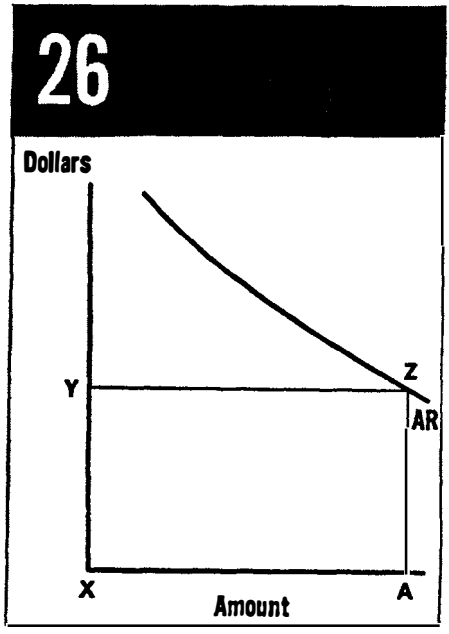

Figure 26. *Total Receipts Measured by Average Revenue Curve*

costs can be represented by using either a marginal or an average cost curve, so total receipts may be shown by using an average or marginal revenue curve. Such a construction based on the average revenue curve is illustrated in Figure 26. As we saw in Figure 25, total receipts for A units were represented by the area $OABM$. In Figure 26, total receipts for the same volume of sales are represented by the area $XAZY$, which should be equal to $OABM$. These are alternative ways of showing the total receipts, and of course they give the same results. Sometimes it will be more convenient to use the area under the marginal revenue curve, and at other times it will be more convenient to use the measurement based upon the average revenue curve.

Summary

The elasticity of demand measures the responsiveness of the amount demanded to a price change. The elasticity is high when a small change in price leads to a greater percentage change in the amount demanded. It is low when the price change leads to a smaller percentage change in the volume of sales. The dividing line between high and low elasticity comes where a certain price change, say a 2 per cent increase, leads to an *equal* change, a 2 per cent decrease in this case, in the volume of sales; the elasticity of such a demand is one. Since a firm's sales receipts equal the product of the price and the amount demanded, it follows that when the demand is elastic a reduction in price leads to an increase in sales receipts. Or to put this in a different way, when the demand is elastic, an increase in the volume of sales brings an increase in sales receipts. When the demand is inelastic, the result is just the reverse.

The marginal revenue is defined as the difference that a one-unit expansion in a firm's sales makes to its sales receipts. It is positive when the elasticity of demand is greater than one, for then an increase in sales leads to an increase in sales receipts; it is negative when the elasticity is less than one, for then as sales are increased there is a *decrease* in sales receipts.

Marginal revenue and total receipts are related in the same way as marginal cost and total variable cost. The total receipts of any output, say n units, equal the sum of the marginal revenue for the first, second, and third units, and so on up to and including the nth unit. Interpreting this result diagrammatically, it means that the total receipts from the sale of n units are measured by the area under the marginal revenue curve between the vertical axis and a line drawn to that curve from the nth unit. Alternatively, we can measure the total receipts by the area of the rectangle subtended by the average revenue curve at the output n.

12

Factors That Determine the Elasticity of Demand

UP TO THIS POINT we have simply assumed various values for the elasticity of the average revenue function. We shall now inquire into the circumstances that determine that elasticity. Why is the elasticity of demand for the product of one firm high and that for another firm low? We shall see that these factors can be divided into two broad groups: first, relevant characteristics of the commodity being sold; and second, characteristics of the industry of which the firm is a component part.

We must remember that in measuring the elasticity of demand or of the average revenue function, we are measuring the degree to which the volume of sales responds to a change in price — and that the greater the response, the higher is the elasticity. We assume, of course, that the demand determinants — for example, the income level and the prices of related goods — are fixed. We are thus concerned to discover why, given the demand determinants, the amount demanded sometimes changes markedly when the price is altered, and why at other times the change in the amount demanded resulting from an equal change in price is very small.

Characteristics of the Commodity and the Elasticity of Demand

Within rather wide limits, some commodities will be purchased no matter what the price may be. Buyers feel that certain articles are practically essential to their physical well-being, and they will pay almost any price in order to continue buying them. This is obviously true of food in general, though it is by no means true of individual

food items. If food prices should all increase by 15 or 20 per cent, the decline in the total volume of sales would be small, though the drop in the sale of certain items might be much greater. Some families might be willing to reduce their purchases slightly, and others would doubtless shift to the purchase of cheaper types of food. Thus, while there might be less call for caviar, pickles, or grade A beef, such items as bread, beans, and hamburger might be expected to sell in practically unchanged amounts. But most families would contrive — perhaps by skimping elsewhere, by saving less, or by borrowing — to maintain their food purchases. Since the sales drop off very slightly when the price rises, the demand for food in general is relatively inelastic. Naturally, families that continue to purchase the same amount of food as they did before prices rose, are forced to spend more money for what they get.

But if, instead of a general increase in food prices, there were an increase in the price of only one item — for example, canned corn — the sales of this one commodity would certainly decline more sharply. There are so many other kinds of food which satisfy practically the same need that many families would be tempted to substitute something else. And when a rise in price brings about a relatively large reduction in sales, the demand is elastic. Thus we may conclude that the elasticity of demand for essentials — as for food in general — is very low, while the elasticity of demand for any one commodity, such as a particular food item, is somewhat higher, since there are very few commodities for which good substitutes cannot be found.

Cigarettes are another commodity which many purchasers consider almost a necessity. So again we should expect only a small decline in sales if cigarette prices are raised, and only a small rise in sales if they are reduced. But as with food, although the demand for cigarettes, in general, is quite inelastic, the demand for any particular brand would be somewhat more elastic. The sales of things that are thought to be necessary will generally prove to be relatively unresponsive to a price change. For such commodities, the elasticity of demand is relatively low.

Likewise, sales may fail to show any great response to a price change, not because the article is considered a necessity, but because it is bought mainly by people who can easily disregard the price change. If the commodity is purchased chiefly by very wealthy people, a change in price will probably have little effect on volume

of sales, particularly if it is not extremely expensive. For to people of great wealth a 10 or 15 per cent increase in price for something which costs five or ten dollars will be of little consequence. If this commodity is bought by those with high incomes, the elasticity of demand will be very small.

Moreover, for some articles customarily purchased along with others that cost a great deal more, the price is of slight importance to the buyer because it represents so small a part of the total outlay. Take doorknobs, for instance. A certain number of doorknobs are needed in every new house that is built. But no architect would alter his plans in order to reduce the number of doors in the building just because the price of doorknobs had gone up by 15 or 20 per cent. Nor would any prospective house purchaser be deterred from buying a house on this account. The cost of a doorknob in the completed house can be very nearly disregarded. The elasticity of the demand for doorknobs must consequently be very low, and so would it be for a large number of other small items whose sale is contingent upon the sale of something much larger and more expensive.

We have seen that when there are a number of effective substitutes for a commodity — canned corn, for instance — a rise in its price will probably induce many purchasers to turn to one of the substitutes. Contrariwise, when there are no satisfactory substitutes, as perhaps with salt or shoes, buyers could not satisfy their needs by shifting their purchases elsewhere, and the change in the volume of sales resulting from a price change would be relatively slight.

Summary: Commodity Characteristics and the Elasticity of Demand

It would be possible to make an extensive list of the characteristics of commodities which affect the elasticity of demand, but the chief ones are: the degree to which it alone can satisfy a want, its importance to the standard of living, and the income groups to which it makes its chief appeal. It is enough to see the type of factor that helps determine whether the demand for the commodity is elastic or inelastic. The more essential it seems to be, the less easily substitutes for it can be found; and the more easily its purchasers can meet a change in price, the lower the elasticity of demand will be. And while undoubtedly other circumstances also affect elasticity, these, if not the principal ones, are at least representative and will suggest others.

The Structure of the Industry and the Elasticity of Demand

Certain features of the industry of which the firm is a part are also important in analyzing the elasticity of the demand for its products. By the term *industry* we mean that group of firms which sell roughly identical commodities. Thus we talk about the automobile industry because the individual firms that comprise it — Ford, General Motors, Chrysler, and others — sell an article which is basically the same. Likewise, the steel industry is made up of such firms as the United States Steel Corporation, Bethlehem Steel Corporation, Republic Steel Corporation, and several others, all of which sell practically identical products. The characteristics of the industry, and in particular the relation between the individual firms in it, are of critical importance in determining the elasticity of the demand for the product of each firm in the industry. Let us see more specifically how the structure of the industry affects the elasticity of demand.

There is a very wide range in the structure of American industries. In some there are literally thousands of firms each producing and selling a commodity that is almost or quite indistinguishable from that of its competitors. The wheat industry and the raw cotton industry are excellent examples. Wheat was raised on approximately 1,364,000 farms in the United States in 1935; cotton on even more. Thus each wheat or cotton grower had well over a million competitors. In contrast to these, certain other industries are composed of a very small number of firms. For instance, in 1940, the only producer of scientific precision glass in the country was the Bausch and Lomb Optical Company. Until quite recently the Aluminum Company of America produced and sold the entire output of aluminum in the United States. And for a long period of years the United Shoe Machinery Company has been the only American producer of shoe machinery. The individual firm in each of these industries, in sharp contrast to the situation that exists in the wheat and cotton growing industries, has practically no competitors. The single seller of one of these articles, although he may perhaps have other worries, need not concern himself about the actions of competitors. Thus, at one extreme are industries composed of vast numbers of firms in vigorous competition with each other, and at the other, industries made up of perhaps only one or two firms. In order to determine the importance of industries of these various structures in the national economy, we shall in the next chapter describe some of the chief industries in the United States.

Just as industries differ in the number and importance of their constituent firms, so they may differ as the products of the constituent firms are identical, clearly similar, or only vaguely similar. In some industries, such as wheat farming, the products of the individual firms are indistinguishable for any one grade and type of grain. In other industries, the products are more clearly differentiated. Any small boy can tell a Ford from a Chevrolet, or a Dodge from a Pontiac, with half a glance. In still other fields, the pharmaceutical industry, for instance, while the analytical chemist probably cannot distinguish between the products of the individual firms, the ordinary consumer can — or at any rate thinks he can — and acts accordingly. For these products are branded, and advertising is directed to persuading the buyer that the merits of one particular brand of a standardized commodity are beyond compare. If this effort is successful, the effective substitutes for the product of any one firm are few and unimportant. And as we shall see, whether buyers choose their products on rational grounds or on some quite irrelevant basis does not affect the elasticity of demand.

These differences between industries — in the number of constituent firms and the strength of buyers' preferences — are of decisive importance. We shall generally find that firms in one type of industry are faced with an extremely elastic demand function; that firms in another type of industry are faced with a less elastic demand, and so on. We shall now examine these various types of industry in more detail.

The Perfectly Competitive Industry and the Elasticity of Demand

Let us first consider an industry composed of an extremely large number of firms that produce an identical product. The fact that their product is identical, in the eyes of buyers, means that no one has any reason for preferring the product of one firm to that of any of the others except on grounds of price. The speculator in wheat, or the miller, will not choose Farmer Brown's wheat in preference to Farmer Jones's unless it is cheaper. If all the firms in the industry charge the same price, the buyer will be no less likely to buy from one than from any other. But if one firm should charge a price even slightly higher than the rest, no one would buy the higher-priced article. If, then, the firm raises its price above the prices charged by the others, the amount of its product which is demanded will fall to

zero. This means that for prices above the present market price, the demand is perfectly elastic.

Conversely, a firm which charged less than the market price would find that the amount of its product demanded was extremely large, for presumably all buyers prefer to purchase at the lower price. But actually, the firm will not be under any inducement to lower its price. For the kind of industry that is being considered is made up of a very large number of firms, and any one firm provides but a very small fraction of the total output of the industry. The individual wheat farmer, for instance, may supply 1/10,000 of 1 per cent of the total amount of wheat produced in any given year. Consequently, if he wants to sell not 1000 bushels, but 1250 bushels, he will not have to lower his price in order to induce buyers to take the additional 250 bushels. So small an amount would not be noticed in comparison with the one billion bushels grown each year in this country. And since it is not necessary to reduce price in order to expand sales, there would be no reason for doing so, for the lower the price, the lower are the sales receipts. It seems clear, then, that each firm would charge the market price. It could not charge more and find buyers; it would not need to charge less in order to sell as much as it had available. The demand for its product is therefore extremely elastic.

But we may also consider this situation from a slightly different point of view. We noted earlier that if there were good substitutes for a commodity, the elasticity of the demand for that article would be relatively high. The product of an individual firm in an industry such as we are now considering sells in competition with a very large number of perfect substitutes; *viz.*, the wheat of Farmers A, B, C, D, and so on. And we have seen above that when good substitutes are easily found, the demand for the product of any one firm in this industry should be extremely elastic. Indeed, the fact that sales can be expanded without reducing price, and that raising the price above the market level will cut sales to zero, shows that the demand that faces a firm in these circumstances is *perfectly elastic*. Accordingly, the demand curve for the product of each firm in such an industry is perfectly horizontal.

An industry having a very large number of firms producing identical goods is said to be *perfectly competitive*. And in a perfectly competitive industry the demand for the product of each firm is perfectly elastic. A brief survey of the important industries of the American economy

will demonstrate that very few of them meet these conditions for perfect competition in full. Some agricultural industries very nearly do so. But most industries, as we shall see, are composed of a relatively small number of firms; and in most of them strenuous efforts are made by the separate firms to differentiate their particular product from that of all the others.

From the phrase *perfect competition* we might infer that in industries which are not perfectly competitive, firms do not compete very vigorously. This inference, however, is quite unjustified. Competition in such industries may be vigorous enough, as the activities of the larger cigarette companies make quite plain, but it will frequently take a different form. Price competition will be less important than competition in advertising, product research, patent control, and so on. But this conclusion can hardly be developed at this stage. For the present, it is important only to notice that the phrase *perfect competition* does not denote anything ideal, or for that matter, anything extensive or complete. It has a purely technical meaning, and simply indicates that any firm in an industry which is perfectly competitive is able to sell its product in a manner described by an infinitely elastic demand curve; that is to say, at a price which does not depend upon the firm's own output.

The Demand for the Product of a Firm and the Demand for the Product of the Industry

The demand for the product of a firm in a perfectly competitive industry is perfectly elastic. The firm does not have to take into account the effect upon price of a change in the volume of its sales. But this does not mean that the demand for the product of the industry of which this firm is a part is perfectly, or even highly, elastic. Although the individual wheat producer is safe in assuming that the price of wheat will not drop because he dumps a thousand bushels on the market, this does not mean that the price of wheat is independent of the amount offered by wheat farmers generally. In fact, the demand for wheat is quite inelastic, for a relatively large reduction in its price is necessary in order to persuade buyers to increase the amount they buy by even a small figure. But these results are not contradictory. Any apparent contradiction is resolved when it is remembered that any one producer provides but a very small fraction of the total, and so he is able to expand his sales by what is to him a rela-

tively large amount, without affecting, except negligibly, the total amount put on the market by all producers. Thus while one producer does not find it necessary to lower the price at which he offers wheat in order to expand the volume of his sales by 50 per cent, this is true only because such an increase in his sales means an increase in the total sales of wheat so small that it is not noticed. So, although the demand for wheat, the product of the industry, may be very inelastic, this does not mean that the demand for the product of a single firm in that industry cannot be very, or even perfectly, elastic.

The Elasticity of Demand for a Monopoly

At the opposite extreme from a perfectly competitive industry is one composed of only a single firm. In such an industry the question of buyers' preferences is not relevant, because no alternative sources of supply are available. Before 1941, for example, one either purchased aluminum from the Aluminum Corporation of America (with very minor exceptions) or one did without. What about the elasticity of the average revenue function applicable to such a firm? It is not, as students frequently guess, perfectly inelastic (a vertical straight line), for that would mean that sales were not affected by price, and obviously this is not true. Although buyers could not at that time shift to another aluminum supplier in the event of a price increase by Alcoa, they could refrain from buying aluminum. They could either substitute other commodities for it — such as steel, copper, or plastics — or they could simply reduce their purchases of aluminum articles. Hence the demand for Alcoa products, and in general for the products of other monopolies, would be fairly elastic, but far from perfectly so. We shall later see reasons for believing, in any case, that the elasticity generally is greater than one.

The Elasticity of Demand in an Imperfectly Competitive Industry

Some industries, although they are composed of a very large number of firms, are not perfectly competitive because they do not satisfy the condition that consumers' preferences are absent. That is, in the minds of the buyers, there are differences between the products of the competing firms. The men's clothing industry illustrates this situation. There are so many firms in this industry that no one firm, even the largest of them, can claim a very large portion of the total production and sales. But the products of the individual firms are dissimilar,

and some buyers will want "suits by Acme" while others can be satisfied only with "suits by Apex." Whether the differences between these suits are substantial or not scarcely matters for our purpose. The mere fact that buyers entertain preferences for one brand over all others is sufficient. How elastic is the demand for the product of a firm in an industry like this?

Because each brand has its loyal supporters, the effect of a price change will be very different from what it is for a commodity like wheat. If the wheat farmer attempted to get a price higher than the ruling price for his crop, he would find that his sales would fall to zero, for no miller or grain speculator is so interested in Farmer Brown's wheat that he would pay one-eighth of a cent more a bushel for it than for any other wheat of the same grade. But the customer who strongly prefers to be clothed by Agile would be willing to pay an additional two dollars to avoid having to wear other kinds of suits. Thus the producers of Agile clothing can raise their price by two dollars, and some of their staunch customers will not reduce their purchases. Obviously, however, not all their customers will be equally loyal. Some were probably wondering, in any case, whether they should not buy a different brand next time, and the increase in price will persuade them. But notice the contrast between this situation and that in a perfectly competitive industry. In the latter, if one firm raised its price above that charged by the other firms, its sales would fall to zero. But where buyers' preferences are established, a firm may raise its price without suffering the loss of *all* its trade. Sales will decline to a degree that depends on the strength of customers' preference, but the decline will not be as great as in a perfectly competitive situation.

However, a firm in a situation in which buyers' preferences exist will find it necessary to lower its price in order to expand its sales. This situation is also quite different from that which rules in a perfectly competitive industry. Where preferences exist, an expansion in sales can come about only as a result of attracting customers from other firms or persuading people to buy clothing they had not planned to buy at that time. To succeed in either of these things without a vigorous advertising campaign requires a price reduction. Accordingly, the demand for the product of any one firm in such an industry is not perfectly elastic. The effect of consumers' preferences of the kind we have been discussing will be to reduce the elasticity of demand,

and the more strongly these preferences are felt, the greater will be the reduction in the elasticity. An industry in which such conditions exist is said to be *imperfectly competitive*, or *monopolistically competitive.*

The departure from perfect competition will be particularly marked if there are only a few firms in the industry, for then each individual firm plays an important role. Thus, even if consumers' preferences are not strongly felt, any one firm has fewer competitors from whom to draw customers if it attempts to expand its sales by lowering price. On the other hand, its customers will have fewer alternative choices if a rise in price makes them seek a new brand. Under these circumstances, sales will be even less responsive to price changes, assuming of course that other considerations regarding the commodity are the same, than in the cases we have previously studied. Therefore the elasticity of the average revenue function applicable to any one firm will be still lower. So this kind of industry also is said to be imperfectly or monopolistically competitive.

Some shift of customers between firms — away from the one that raises its price and to the one that lowers its price — may be confidently expected so long as competing firms leave their prices unchanged. We expect such behavior in an industry where there are more than a very few competing firms, for the actions of any one firm in such an industry would rarely affect its competitors enough to induce any of them to change their prices. Thus the firm regards the demand for its product as elastic in the above circumstances because it may plausibly suppose that if it does change its price, its competitors would not thereby be induced to alter theirs.

The Elasticity of Demand in an Oligopolistic Industry

There is one more situation to be considered before we conclude our examination of the structure of industry and the relation of that structure to the demand for the product of any firm. For industries of the type we have already analyzed, it has been possible to assume that what any one firm does about prices will go more or less unnoticed by its competitors, or at least that it need not fear retaliation. When there are a reasonably large number of firms in competition, or when buyers' preferences are very strong, this assumption is reasonable, for a price cut by any one firm, while it would attract some customers from the other firms, would probably not attract a sufficient number from any one of them to lead that firm to retaliate by reduc-

ing its price. The probability of provoking retaliation when the firm has a large number of rivals, or when products are well differentiated, is insignificant. Hence a firm in this situation, when estimating how its sales will respond to a change in price, is justified in supposing that its competitors will not reduce their prices in retaliation.

But the situation is very different when there are only a few firms, say three to ten, in the industry. In such a situation what one firm does is of decisive importance to the others, and there is always the chance that a price change by one will encourage its competitors to change their prices. One method which is open to any firm for regaining its sales, after a price cut by a competitor, is to reduce its own price. If most of the other firms follow suit, the advantage gained by the firm first to reduce its price will be wrested from it. And if its rivals should decide to reduce their prices, instead of a sizable increase in sales as a result of lowering its price, it will come out with only a slight expansion, or worse yet, an actual decrease.

Let us illustrate by supposing that there are four important firms in the industry — A, B, C, and D. As the story begins we assume that they are each charging the same price for their product — 18 cents. Firm A begins to wonder about the possible gains from reducing its price to 16 cents. So long as B, C, and D keep their prices at 18 cents, the gains to A at the lower price may be quite substantial, for many of the customers of B, C, and D will shift to A. In other words, the elasticity of demand for A's product is reasonably high as long as A's competitors do not change their prices. If the competitors numbered not three but three hundred, that might be the end of the story. For while B, C, D, and all the others would not welcome A's move, none of them would be hurt seriously enough to feel impelled to reduce its price to 16 cents too. If, however, there are only the three other firms, B, C, and D, it is likely that the reduction in their sales following A's price cut will be substantial. Hence they will each feel impelled to do something to win back customers. Suppose that to achieve this, B, C, and D all mark their prices down too — B to 16 cents, and C and D to 15 cents. Now A's position, *vis à vis* its competitors, is worse than it was when they were all charging 18 cents. Instead of a sizable expansion in sales, A would experience only a slight increase or possibly a decrease.

If A feels that cutting its price will lead, not to a large expansion in sales, but to retaliation and thus only to a very small expansion,

the elasticity of demand for its product is very low for prices below the current price. For if when it reduces price from, say, 18 cents to 16 cents, its sales increase by only 1 per cent, the elasticity of demand is only 1/100 ÷ 2/18, or .09.

But in such an industry, the response of sales to a price increase is quite different. When a large number of firms is involved, a price increase by any one of them will drive many of its customers to other sources of supply, but probably none of its competitors will gain enough to induce it to raise its price. And even when a fairly small number of firms is involved, a price increase by A would hardly induce B, C, and D to raise their prices If A should raise its price to 20 cents, it would lose many customers to B, C, and D, but unless the manager is extremely optimistic, or unless he has already come to an agreement with B, C, and D, he would have no reason to expect them to raise their prices to 20 cents too. For they naturally profit at A's expense by keeping their price at 18 cents. Of course, if they should raise their prices, the decline in A's sales would not be so great, but this is unlikely to happen unless they have decided to act in concert. Hence, since the firm would have little reason for expecting competing firms to raise their prices as a result of its actions, it would follow that the elasticity of demand for its product at prices *above* the market price would be rather high. When it raised its price, it would have to anticipate a sharp reduction in sales (elastic demand); when it lowered it, it could expect at most only a small increase in sales (inelastic demand).

In summary, we see that there is a certain lack of symmetry in what a firm could reasonably expect its rivals to do in response to increases and decreases in its own price. In an industry made up of only a few firms, if one firm lowers its price, it may expect its rivals to lower theirs, and hence it can count on a very small increase in sales. Therefore, the demand at any price below the one it is presently charging is quite inelastic. On the other hand, if it raises its price, it has no ground for expecting its rivals to raise theirs, and hence it may fear a very great decline in its sales. Therefore, the demand at prices above the market price is quite elastic. Such a demand curve is illustrated in Figure 27. Notice that at the point which represents the market price (we need not inquire at the moment how this price is established), there is a kink in the demand curve — that is, the curve, instead of being smooth and continuous, has an

angle in it. The marginal revenue curve appropriate to such a demand curve is also illustrated in this diagram. We showed in the preceding chapter that when the demand is very elastic, the marginal revenue curve is near the average revenue curve (being the same curve when the elasticity is infinite), and when the demand is inelastic, the marginal revenue curve is well below the average revenue curve. The elasticity of demand changes very suddenly at the output n, being higher for smaller outputs and very low for larger. Hence there is a gap in the marginal revenue curve at this output. The existence of the gap may easily be demonstrated by the use of the calculus.

The above considerations do not apply when the number of firms is fairly large, in which case we have ordinary imperfect competition — or when there is but one firm in the industry, for in that event we have monopoly and there are no rivals about which the firm need worry. Nor do they apply when buyers are very strongly attached to the products of individual firms, for then, too, retaliation need not be expected because the rivalry of the firms is tenuous and imperfect. But when there are several firms — the number needed depending on the kind of market served, the existence of preferences, understandings between the firms, and so on — these considerations can become important. An industry in which such a situation prevails is called *oligopolistic.* Oligopoly may perhaps be thought of as a special kind of imperfect competition, something between perfect competition and monopoly, with an additional ingredient, the danger of retaliation whenever a firm changes its price. We shall see in the next chapter that oligopolistic industries are common in the American economy.

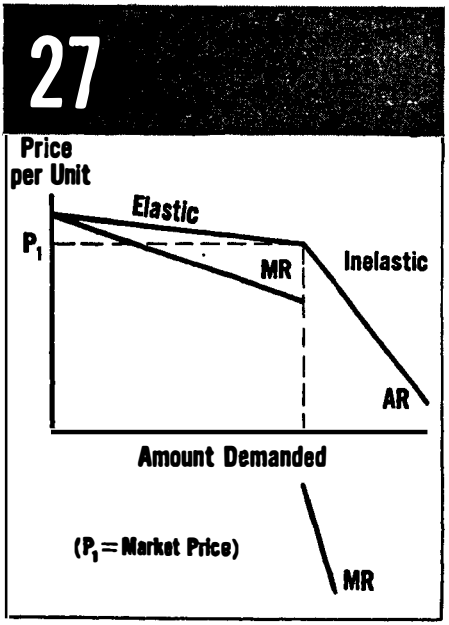

Figure 27. *Demand for Product of Oligopolistic Firm*

Summary

The question that we set ourselves to answer in this chapter had to do with the factors that determine the elasticity of demand for the product of a firm. We have seen that this elasticity has to do with the nature of the product itself — who wants it and why — and also with the nature of the industry of which the firm is a part. If the industry is perfectly competitive, the demand for the product of the firm will be infinitely elastic; if there is simple imperfect competition or monopoly, the elasticity will be less; and, finally, if the industry is oligopolistic, the elasticity is high for that part of the demand function concerned with prices above the ruling price, and low for that part of the demand function concerned with prices below the market price. We have now completed what we shall have to say, for the time being, about average revenue. We found it necessary to investigate the demand for the product of a firm, as in earlier chapters we analyzed costs, because it is one of the factors that determine profits. And in an economy such as ours, it is essential to understand how profits are determined if we are to understand the behavior of the business firm.

And as we saw earlier, in order to understand such problems as how in our economy employment and income are determined, or what causes inflation and how to check it, we must be fully informed about the way in which the individual firm determines its policy. An understanding of the nature of the cost and demand functions is an essential step in the analysis.

13

Competition and Monopoly in the United States

Is our economy competitive? Is the typical American industry competitive or monopolistic? The answer is hard to find in the confused chorus of slogans and war cries that fills the air. "The American competitive system will deliver the goods"; "The Free Enterprise System is on trial"; "The forces of competition will check inflation." Statements like these — and there are a host of them — imply that competition is the rule. In contrast, such slogans as "Monopoly capitalism brings inflation," and "Monopolies and Trusts cause unemployment," insinuate that competition is a rare phenomenon in our economy. The view we hold seems to depend entirely upon what newspapers we read. Let us see what the facts show.

The Meaning of the Terms

Unfortunately the terms *competition* and *monopoly* have gathered about them so many moral implications that it is difficult to use them precisely. We do not like monopolies, hence we brand with that title all institutions that we condemn. If somebody charges a price we think too high, we call it monopolistic. If we think a firm is unfairly resisting its employees' demands for higher wages, we call it a monopoly. Competition, on the other hand, is a word of approval. Competition is itself a virtue — although apparently there can be too much of a good thing, as when we talk about cutthroat competition. But in general, when we want to support an institution we call it competitive, and we refer with approval to a fair competitive price. Because of these emotional overtones it becomes doubly necessary to define words carefully.

Perfect competition, as we have already seen, has more to do with arithmetic than with morals.[1] We call an industry perfectly competitive if it is composed of a very large number of firms whose products are so similar that buyers have no preference as to the source of their supply. In this sense perfect competition has a technical meaning; and when it exists it has certain effects which are technical in nature also. The most important of these is that each seller is able to get rid of as much or as little as he chooses at the given market price and does not have to lower his price in order to expand his sales. Moreover, he cannot get a higher price by withholding a part of his output from the market. In other words, the average revenue curve for the product of any one seller in a perfectly competitive industry is perfectly elastic. Competition is so perfect that, in one sense, there is no real price competition, for each seller may charge the market price, and no one has any reason to charge less. But in another sense, there is very sharp price competition, as a seller would soon learn if he were to charge more than the market price, for in that case he would sell nothing at all, and his competitors would have all the business. When competition is perfect, it is somewhat impersonal; and when there are five million producers of a certain commodity, the seller in one spot feels no special rivalry with the seller down the road. Nevertheless, circumstances force each seller to adapt his price to that of his rivals; otherwise he loses everything. Price competition is intense with perfect competition.

And what about monopoly? Monopoly means simply there is only one seller in a market. When an industry is monopolistic, there is only one source from which the commodity may be obtained. This again is a matter of arithmetic rather than of morals. When there is only one seller, his average revenue curve, as we have seen, is not perfectly elastic, for his sales fall off if he raises his price, and increase if he lowers it. Since there are no other suppliers of the product, he does not have to face competitors who supply the same commodity. He may, however, have competitors who supply something which can be substituted for his product, or he may have to face potential competition from new producers. Thus his life is not free from business worry. He is not necessarily more grasping than other people, although he is perhaps in a position where he can grab more successfully. He will not even charge the highest price that the market can

[1] But not only with arithmetic, as we shall see later.

bear, as we shall see later; for this would not pay. His profits are not necessarily exorbitant, although his monopoly position gives him special advantages in earning profits. Because he is the sole producer, he does not have to use the weapons of competition, except very indirectly. Hence there is no price competition in a monopolistic industry.

Inspection of the structure of many industries shows that most of them are neither perfectly competitive nor completely monopolistic. Very few industries qualify as perfectly competitive, and not many more can qualify strictly as monopolies.[1] Most industries fall somewhere between these two extremes. In other words, most of them are industries where there is monopolistic competition in the simple sense or oligopoly.

The structure of those industries that are neither perfectly competitive nor perfectly monopolistic is extremely varied. In some of them, like the sulphur industry, only two firms provide the bulk of the output; in others, there are hundreds of relatively small firms. In some, one or two large firms supply most of the industry's output, and a large number of relatively small firms provide the remainder; in others, the firms are roughly equal in size. In some, the firms act in unison on matters of price, output, and labor relations, perhaps in pursuance of formal agreement and perhaps through an informal understanding; in others, the firms compete vigorously among themselves, using all the competitive weapons at their disposal. These industries that occupy the immense middle ground between perfect competition and strict monopoly differ widely from each other.

Competition, as we have implied, has two dimensions. The intensity of competition depends in part upon the number of competitors. If a firm has no competitors, there is of course no competition. If it has three million, competition *may* be intense. But the intensity of competition also depends upon the vigor with which firms compete. If there are 500 firms in an industry, and 499 of them follow the lead of the remaining one in setting price, there is in effect no price competition. But if there are only twenty, and they do not reach an agreement about price, competition may be quite severe. If there are 100 firms and they understand one another well enough to refrain from price-cutting, there is no price competition. But when there

[1] Nevertheless, as we shall see, some industries in which there is more than one firm are in effect monopolistic in regard to price.

are three, and they compete vigorously, using price as a weapon, we come much closer to perfect competition. Competition, then, is not merely a matter of numbers; it is partly also a matter of how much each firm in the industry is willing to use price adjustments as a method of improving its lot at the expense of its competitors.

Firms operating under perfect competition, as we have defined it, must be ready to make these price adjustments. Failure to do so may cause business failure. Hence, when there are many firms and identical products, there is not only a large number of competitors, but, of necessity, vigorous competition in addition. With perfect monopoly, no competitors exist, and so there is no question about the strength of competition. But industries that on the surface look to be somewhat competitive may in fact be perfectly monopolistic. The industry composed of 100 firms that act in concert on price does not have any price competition; in fact, such an industry is monopolistic, for price is determined by but one seller — the firms in consultation.

This should warn us that we must not apply our arithmetical criteria too rigidly in trying to determine whether we should file the industry in the monopoly drawer or the simple monopolistic competition drawer. Ideally, the answer depends not only on the number of firms in the industry, but also upon the force behind their competitive actions; or in other words, upon their price policies. Unfortunately, the amount of information on price policy now available is not very great. Little is known about many industries, and so it is not possible to classify them accurately.

Competition may take many forms. Firms can compete on the basis of price, or in any of a number of other ways. They may compete for buyers' favor through advertising, or through offering better quality, faster delivery, better repair facilities, or easier credit. Since the present analysis is particularly concerned with price, industries will have to be classified according to the nature of their price competition. But unless the terms are used carefully, this may be confusing. There are many industries where price competition is absent, and which therefore must be treated as monopolistic when their price policy is the determinant, but this does not mean that no competition or rivalry exists in such industries; the firms may compete strongly in other ways, as by the use of advertising. Although we shall call such industries monopolistic because price competition is absent, this will not imply that no competition of any kind exists.

Perfectly Competitive Industries

Strictly speaking, it appears that only a few agricultural industries can be classed as perfectly competitive. Of the approximately 7 million farms in the country, more than 4.5 million produced cotton in 1936. In spite of the fact that there are many different kinds of cotton, we may classify the cotton grower as a member of a perfectly competitive industry. The wheat industry may also be regarded as perfectly competitive. In 1936, 1.5 million farmers raised wheat, and since in addition the product is sufficiently standardized, the industry qualifies as perfectly competitive; it has a very large number of individual producers, and a product so homogeneous that consumers are indifferent to their source of supply. There were almost 4 million farms on which hogs were raised in 1936, and thus for similar reasons the hog-raising industry could be regarded as perfectly competitive. But these, and some other agricultural industries, are about the only ones which fall into this category. For though there are other industries, such as haircutting and domestic service, in which the number of producers is very large, the market in which each firm sells is very small. A barbershop in Cambridge, Massachusetts, for example, is one of perhaps five which near-by residents may patronize. Moreover, consumers are very much concerned about the source of their supply of haircuts or of domestic service. Thus such industries cannot be thought of as perfectly competitive.

Perfect Monopoly

Perfect monopolies are somewhat more common than perfectly competitive industries. Perhaps the most notable of them, although it is no longer perfectly monopolistic in the strict sense, is the aluminum industry, for the Aluminum Company of America until 1942 produced practically all the aluminum in this country. Another nearly perfect monopoly is the telephone industry, since the American Telephone and Telegraph Company *is* the telephone industry in most localities. It is an immense firm, too, with assets in 1935 of practically $4 billion. Again, the manufacture of shoe machinery in this country is for the most part in the hands of the United Shoe Machinery Company, which has no competition whatsoever in the sale of most items. The provision of electric light and power and the manufacture of gas in most communities are also monopolistic. The states generally allow only one firm to provide electric power in any one community,

and although there are thousands of public utilities firms throughout the country, there is only one in each market area. Thus each firm in the public utility industry is a practically perfect monopoly. Similarly, nylon, rail service between two points, express services, and certain other goods and services are provided by firms which are monopolistic in nature. But perfect monopoly is not much more prevalent than perfect competition.

Illustrations of Monopolistic Competition

Most industries are neither perfectly competitive nor perfectly monopolistic, because in most there is a relatively large number of firms. Some industries are more or less dominated by a few very big firms, while in others the buyers in any one market have an effective choice from as many as one hundred firms. Table 26 shows for some important industries the percentage of the total output produced by a few dominant firms. The table lists the number of important firms in the industry and the percentage of the total output these firms provided in a particular year. This list shows one striking fact: a large proportion of the total output in most of these industries is provided by three or four firms. Industries in which such a control of output is established are obviously not perfectly competitive. On the face of it, they do not appear to be perfectly monopolistic either, because there is more than one source of supply.

But the mere fact that there is more than one firm in the industry does not necessarily mean that there is no monopoly. For the dominant firms, instead of competing, may be acting in collusion; and if they are, then so far as the buyer is concerned, the industry is under monopolistic control. If the officers of the several companies sit around a table to discuss price, it may be assumed that the price is determined at monopoly level. Since such collusion to determine price is illegal, we are not likely to encounter evidences of it; but this, of course, does not mean that it does not exist. We do, however, find evidence of a situation which has the same effect as monopoly, but which does not imply conspiracy or collusion. In a number of industries in which only a few firms dominate production, price is determined by a single firm, generally the largest, and the others follow its lead.[1] It may be generally understood that they will do so, even

[1] The situation differs slightly from perfect monopoly when there is any doubt that price changes will be generally followed.

TABLE 26

Index of Concentration in the American Economy

Product	Number of Companies	Percentage of Output
Steel ingots	4	52 (1945)
Banana imports	2	90 (1939)
Tin cans	4	89 (1939)
Raw copper	4	82 (1944)
Refined copper	3	88 (1944)
Magnesium	1	100 (1938)
Electrical machinery	4	44 (1935)
Air brakes	2	100 (1940)
Agricultural implements	4	72 (1935)
Refrigerators	4	46 (1935)
Sewing machines	4	79 (1935)
Incandescent lamps	2 plus licensees	85 (1939)
Adding machines	10	100 (1944)
Dictating machines	5	100 (1944)
Automobiles	4	88 (1935)
Locomotives	4	85 (1945)
Refined oil	9	55 (1938)
Rubber tires and tubes	4	81 (1935)
Nitric acids	4	87 (1945)
Hydrogen	4	68 (1945)
Soap	4	72 (1935)
Copper sulphate	4	83 (1945)
Phosphorus	4	95 (1945)
Synthetic nitrogen	2	50 (1945)
Industrial alcohol	5	91 (1945)
Rayon	7	86 (1945)
Meat-packing	4	48 (1939)
Canned milk	3	44 (1937)
Chewing gum	4	92 (1935)
Bread	4	about 20 (1928)
Sugar refining	3	50 (1939)
Cigarettes	4	90 (1935)
Artificial leather	4	57 (1935)
Carpets and rugs	4	51 (1935)
Boots and shoes	4	28 (1935)
Glass, general	4	49 (1935)
Plate glass	2	95 (1935)
Men's clothing	4	5 (1935)
Lumber	4	5 (1935)

though no agreement is put into writing. When this happens, it is as though one firm determined price for the whole industry — or in other words, as though monopoly ruled in that industry.

Some examples of price leadership have come to light. Interesting

evidence on this point was presented to the Interstate Commerce Committee of the United States Senate for the steel industry. Mr. W. A. Irvin, president of the United States Steel Corporation in 1936, was questioned by the Chairman of the Committee as to how base prices in the steel industry were determined. The dialogue follows:

Mr. Irvin. I would say we generally make the prices.

The Chairman. You generally make the prices?

Mr. Irvin. Yes, Sir. We generally make the prices, unless some of the other members of the industry think that that price may be too high, and they make the price.

The Chairman. You lead off, then, with a price charged, either up or down, at Gary. Is that correct?

Mr. Irvin. Yes.

The Chairman. I notice the price on March 19 at Chicago for soft steel bars was $1.90, while in Pittsburgh the price was $1.85.

Mr. Irvin. Yes.

The Chairman. When *The Iron Age* printed that announcement, did they ask Bethlehem or did they ask you, or from what company did they receive that information?

Mr. Irvin. I do not think I can answer that.

The Chairman. At any rate, you were the one that fixed it?

Mr. Irvin. We fixed our own price at $1.90 and $1.85; we always notify the trade papers; I think our commercial people always notify the trade papers and others interested as to what our prices are.

The Chairman. Then the rest of them follow that?

Mr. Irvin. I think they do. That is, I say they do generally. They may quote the same prices, but maybe they need some business and make a better price. We do not always know that until it is over.

The Chairman. Those are the exceptions, are they not?

Mr. Irvin. Yes.

The Chairman. Those are looked upon as the price-cutters in the industry?

Mr. Irvin. Yes; and we have them with us always.

The Chairman. They represent a comparatively small percentage, do they not?

Mr. Irvin. It depends, Senator, on business conditions. I would say that when we are going at 30 or 40 per cent, we have more of them with us than when we are going at 60 or 70 per cent.[1]

Price leadership also seems to be practiced in the beryllium industry. Mr. H. L. Randall, president of the Riverside Metal Company, a

[1] *Hearings before the Temporary National Economic Committee*, Part 27, p. 14250.

relatively small firm engaged in the fabrication of beryllium alloys into rod wire sheets and strips, presented evidence before the Temporary National Economic Committee.[1] Mr. Hugh Cox, of the Department of Justice, who at this time conducted the hearings for the Committee, asked Mr. Randall what percentage of the industry his company controlled. Mr. Randall thought that it controlled less than 1½ per cent. He was then asked about prices:

> Mr. Cox. Mr. Randall, would it be correct to say that there is a well-crystallized practice of price leadership in the industry in which you are engaged?
>
> Mr. Randall. I would say so.
>
> Mr. Cox. And what company is the price leader?
>
> Mr. Randall. I would say The American Brass Company holds that position.
>
> Mr. Cox. And your company follows the prices which are announced by The American Brass?
>
> Mr. Randall. That is correct.
>
> Mr. Cox. So that when they reduce the price you have to reduce it, too; is that correct?
>
> Mr. Randall. Well, we don't have to, but we do.
>
> Mr. Cox. And when they raise the price you raise the price.
>
> Mr. Randall. That is correct.

Later Mr. Randall was asked why, when the price of one of his raw materials was reduced, the price of the fabricated product was increased. The questions and answers were as follows:

> Mr. Cox. I will put this question to you, Mr. Randall. Why didn't you reduce the price of the fabricated product following that decrease in the price of the master alloy?
>
> Mr. Randall. Well, of course I would not make a reduction in the base price of beryllium copper unless The American Brass made a price reduction in beryllium copper.
>
> Mr. Cox. And The American Brass Company made no reduction at that time?
>
> Mr. Randall. If they did, we did, as indicated on that sheet.
>
> Mr. Cox. Assuming you didn't make a price change then, the reason you didn't was because The American Brass Company didn't.
>
> Mr. Randall. That is correct.
>
> Mr. Arnold. You exercise no individual judgment as to the price you charged for your product, then, in a situation?

[1] *Ibid.*, Part 5, pp. 2085–87.

Mr. Randall. Well, I think that is about what it amounts to; yes, sir.

Mr. Arnold. And if the price of the raw material went down the price of the finished product actually went up due to that situation?

Mr. Randall. I don't know that that condition existed.

Mr. Arnold. In other words, the situation is such that you can't pay any attention to the price of the raw material in fixing the prices.

Mr. Randall. Of course, as Mr. Cox first stated, the industry is one of price leadership, and a small company like ours, making less than 1½ per cent of the total, we have to follow, and I think we have a statement of our price policy here which would perhaps clear that up a little.

Mr. Arnold. When you say you have to follow, you don't mean anybody told you you had to follow?

Mr. Randall. No, sir; I don't mean that at all.

Mr. Arnold. But you have a feeling that something might happen if you didn't.

Mr. Randall. I don't know what would happen.

Mr. Arnold. You don't want to find out, do you?

The Chairman. Well, as a matter of fact, Mr. Randall, if The American Brass Company raised the price, would the Brass Company consult you about raising it?

Mr. Randall. No, sir; not at all.

The Chairman. You would, however, follow them without exercising any independent judgment as to whether or not it was desirable.

Mr. Randall. That is correct.

. . . . . .

The Chairman. Why do you do it?

Mr. Randall. It is the custom of the industry, at least of the smaller companies, to do that.

The Chairman. And other small companies do the same thing?

Mr. Randall. Yes, sir.

The Chairman. Is there any reason outside of custom for it?

Mr. Randall. No, sir.

The Chairman. Isn't it likely to reduce the amount of business that you can obtain?

Mr. Randall. I don't think so.

. . . . . .

The Chairman. What, in your opinion, is the reason for this custom to follow the leader?

Mr. Randall. Well, of course, that is a question which has been prevalent, I think, in the industry, for many, many years prior to my entry into it.

The Chairman. Oh yes; we hear a lot about price leadership, but I am trying to get the picture of this practice as you see it, and why you follow it.

Mr. Randall. Well, I don't think I have ever given the matter very much consideration. We simply, when the new prices come out, print them just as they are. We don't give the matter any consideration. The prices are published and we print those prices.

The Chairman. Is there any sort of compulsion, moral or otherwise?

Mr. Randall. Absolutely none.

. . . . . .

Mr. Arnold. In other words, there is a situation here where there is a lot of competitors and no competition.

Mr. Randall. Well, we simply, as I said before, follow the prices that are published, and that is what we have been doing for a good many years.

. . . . . .

Mr. Cox. Mr. Randall, I think, just to make the record clear, we should perhaps explain that 1½ per cent of the business which you say your company controls, or less than 1½ per cent, is not merely the beryllium alloy, but all of the alloys which you sell.

Mr. Randall. That is correct.

Mr. Cox. And this system of price leadership which you have been describing is a system which applies not only to the price of the fabricated product made of the beryllium alloy, but all fabricated products?

Mr. Randall. Yes.

It is only fair to point out that the representative of The American Brass Company would not agree that the firm occupied the position of price leader in the industry. Yet this does not alter the fact that Mr. Randall followed the price policy of the larger firm. It is this fact which is important.

When a relatively small number of firms dominates an industry, obviously there cannot be perfect competition. If the firms have some understanding about price, then, as far as price goes, there is a monopoly. Even though there is no formal agreement, if they follow the lead of one firm in changing their prices, there is a monopoly. If the firms adopt neither of these practices, the industry is either one of simple monopolistic competition or, as pointed out earlier, of oligopoly. Table 26, which showed some of the industries in which a few firms occupy the dominant position, makes it clear that some

variant of monopolistic competition or of actual monopoly is characteristic of large areas of the American economy.

Extensive advertising is a sign of the absence of perfect competition. The purpose of advertising is to build up in the minds of the buyers a preference for a particular brand, to raise demand and to make it possible for the firm to raise its price without having to fear the loss of all of its customers. Obviously, when this can happen, price competition is not perfect. And advertising plays a very important role in the American economy. In many industries, for example the cigarette industry, the soap and drug industry, the cosmetics industry — and in fact in most consumers' goods industries — a large part of the total costs goes for advertising.[1]

In the American economy perfect competition is extremely rare, while perfect monopoly exists in some industries, and elements of monopoly are present in most. But to say that the American economy is not competitive is not to say that there is no competition between firms in the economy. Competition may, in fact, be bitter, as a look at the advertising carried on by business firms will indicate. But the competition is not necessarily in terms of price. Firms do not compete with one another for the trade of their customers by manipulating the price of their product. They compete through advertising, through developing new products which have a greater appeal to customers, or through improving the quality of their product. In many industries there is effective competition in a number of these fields. But this kind of competition is not price competition. This, of course, is not an indictment of American industry. We have not seen any reason to believe that perfect competition in respect to price is essential to the health of the American economy. Whether it is or not is something to be determined later; the fact of the matter is that it does not exist except in relatively unimportant sections of the economy.

[1] In 1940, 11.3 per cent of the total revenue of the cigarette industry was spent on advertising.

14

The Determination of Price and Output

THE COMPLICATED ANALYSIS of the last few chapters has a purpose of which we must not lose sight. It is not simply a new kind of parlor game in which we christen the pieces *marginal revenue*, *the elasticity of the average revenue function*, and *cost determinant number one*. There is certainly no virtue in making something appear mysterious just for the sake of being mysterious, although this practice may prove remunerative for the medicine man. What then is the reason for introducing these concepts? To what use can they be put? The answer is to be found in the complexity of the problem which we posed initially. A firm has to weigh a great many factors in determining how much to produce; the level of production is hard to decide upon. To see how the firm arrives at its decision, we need elaborate and refined tools of analysis. And in order for us to determine with any precision how, for example, changes in certain economic variables will affect the operations of the economy, we must make use of these strictly defined and somewhat abstract concepts. The justification, then, for what some readers of the preceding chapters may regard as so much mumbo-jumbo, is that it is useful; that without it the solution to problems we have set out would be much more difficult.

At this stage it is appropriate to restate the central problems of this section. Our fundamental purpose is simply this: we must understand how our economy functions because, lacking this understanding, we cannot know what measures to adopt to remedy any defects in its operation. Frequently it falls short of providing full employment, and at such times our standard of living is far below what it need be. The

distribution of income is very uneven, and it is questionable whether the maximum social well-being is achieved with this unequal division. Finally, it seems that the proportion in which certain commodities are found in the total output is far from ideal. To take an example, we sometimes appear to produce relatively too large a volume of agricultural commodities and too small a volume of such commodities as aluminum, magnesium, and electric power. Until we understand how our economy runs, we cannot, except by lucky chance, determine the policies to follow in order to remedy these defects, if they are defects.

In our economy, unlike such economies as the Russian, the individual profit-seeking firm is the basic unit. What happens in the economy is the composite of what a very large number of firms individually decide to do. Thus, our first task in working toward an understanding of the economy is to see how the single firm determines its policy. Our analysis of the nature of costs within the firm, and of the demand for the product of the firm, is essential to this understanding. The exact manner in which costs and demand determine that decision we shall investigate in this chapter.

We saw in our preliminary discussion of the firm that decisions about price and output are based upon considerations of profit, that the firm chooses that output which promises to be most profitable under the circumstances. In this chapter, with the aid of those tools of analysis which we have now developed, we shall attempt to find where, under given conditions of demand and cost, the most profitable output is to be found.

Review: Total and Marginal Revenue, Total and Marginal Cost

Before beginning our examination, however, let us recall two important relations which were described earlier — that between total receipts and marginal revenue, and that between total cost and marginal cost. Briefly, we saw that for any volume of sales — for example, n units — the area under the marginal revenue curve up to n (or alternatively, the sum of the marginal revenues for the first, second, third . . . and so on up to and including the nth units) was equal to the total revenue from the sale of n units. Likewise, we saw that the area under the marginal cost curve up to n was equal to the total *variable* cost of producing n units; in other words, to the total cost of producing n units minus the fixed costs.

Graphic Representation of Profits

Now let us see how, by building on these relations, we may represent profits at different output levels. In Figure 28 the curve *MC* represents the marginal cost and *MR* the marginal revenue. The

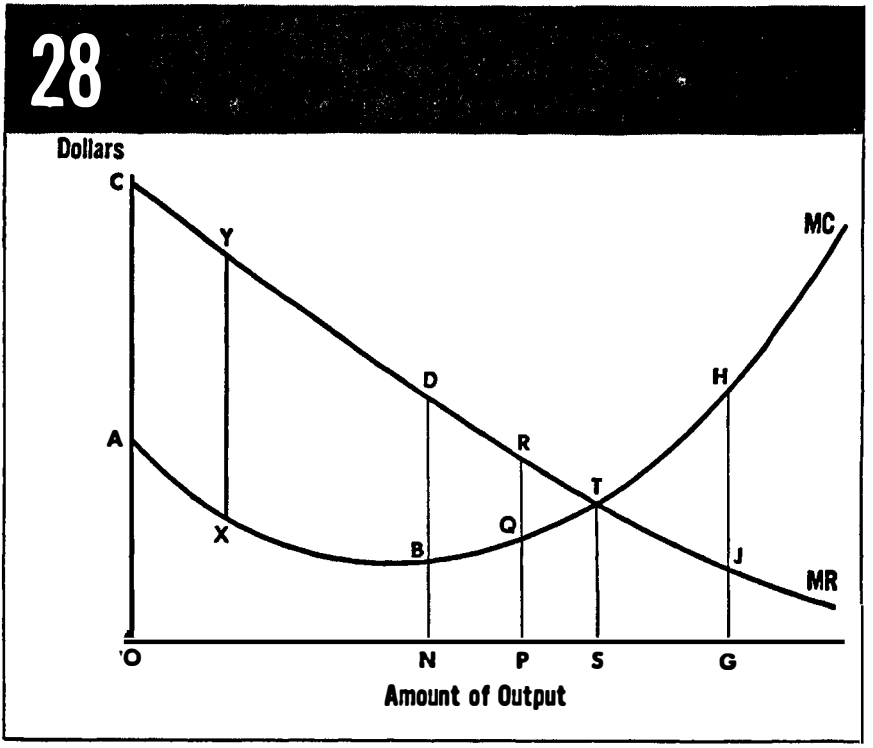

Figure 28. *Marginal Revenue, Marginal Cost and Profits*

total receipts from the sale of *ON* units may then be represented by the area *ONDC*.[1] The total cost of producing *ON* units can be measured by the area *ONBA* plus something to represent the fixed costs. The line *XY* is drawn so that the figure *AXYC* measures the fixed costs; consequently the total cost of producing *ON* units is *ONBA* + *AXYC*. Now, since total profits are equal to total receipts minus total costs, we have the total profits from the production of *ON* units represented on the diagram by *ONDC* minus (*ONBA* + *AXYC*), in other words, by *XBDY*.

[1] Compare Chapter 11.

Next let us suppose that instead of producing ON units, the firm produced OP. Total receipts would then be equal to $OPRC$ and total variable costs would be represented by $OPQA$. Since fixed costs do not vary with output, they are represented by the figure $AXYC$, just as they were when the output was ON units. Hence the total cost of producing OP units is $OPQA + AXYC$. The difference between the total receipts ($OPRC$) and total costs ($OPQA + AXYC$) is the profit earned from an output of OP. Therefore the profit from OP units of output is represented by the area $XQRY$, which is distinctly larger than the area $XBDY$, the profit earned in the production of ON units. And since the firm would prefer the more profitable to the less profitable output, it would accordingly produce OP rather than ON units, if these two production levels were the only alternatives.

The Most Profitable Output

But let us see whether there is a level of output even more profitable than OP — in fact, more profitable than any other. Such a level of output may indeed be found; it will be OS, the output at which marginal cost equals marginal revenue. If OS units are produced, the profit is measured by the area XTY. This area is of course greater than $XBDY$ — the profit from ON — and greater than $XQRY$ — the profit from OP units. Furthermore, we can easily see that the level of profit at OS is higher than it would be for any larger output, such as OG. For if OG units were produced, total revenue would be equal to $OGJC$, and total variable costs would equal $OGHA$. As before, fixed costs would equal $AXYC$, and since profits equal total revenue minus the sum of variable and fixed costs, they will in this case be XTY minus TJH. Since this is manifestly less than XTY — the profit on OS units — we can see by a process of elimination that profits are a maximum when output equals OS.

We have shown in Figure 28 that the output OS is more profitable than any other in the conditions of cost and demand there denoted. But we must not regard this demonstration as simply a page of geometry. We must be constantly aware of the meaning that lies behind these geometrical drawings. What characterizes the output OS? The fact that the marginal revenue of the Sth unit equals the marginal cost. In other words, the firm makes as large an addition to its receipts by producing that unit as it adds to its cost; thus the Sth unit neither adds to the profit of the firm nor subtracts from it. For the

*P*th unit, the addition to total receipts (*PR*) is greater than the addition to its costs (*PQ*); the firm would naturally want to produce the *P*th unit, since it adds something to its profits — and indeed to produce any unit which added more to its receipts than to its costs. But the *G*th unit adds *GH* to costs and only *GJ* to receipts. On that unit and, as can be seen from the diagram, on each unit in excess of the *S*th unit, the firm loses money. The firm wants to maximize its profit, and this means that it is willing to produce anything that at least pays for itself. Every unit up to *OS* adds more to receipts than to costs, and thus more than pays for itself;[1] each unit in excess of *OS* adds more to costs than to receipts, for the marginal cost beyond *OS* is everywhere above the marginal revenue, and thus it fails to pay for itself. The firm can make more profit by producing *OS* than from any other output. This is a slightly different way of showing that the output at which marginal cost equals marginal revenue is more profitable than any other.

The output that is most profitable does not depend on the level of fixed costs though of course the actual profits do. Fixed costs may be much higher than *AXYC*. Indeed, they could exceed *ATC*, and if so the firm would suffer a loss from producing *OS* units, or for that matter, from producing *ON*, *OP*, *OG*, or even zero units. The problem is then to determine the level of production at which the loss is minimized, and again the answer is *OS*, the point at which marginal cost and marginal revenue are the same. This also can be illustrated by reference to Figure 28. Thus, if the fixed costs equal not *AXYC* but *ATC*, the firm's loss is at a minimum when output is *OS*. For when *OS* units are produced, losses are zero, since the receipts are *OSTC*, and the costs are (*OSTA* + *ATC*), or *OSTC*; and the difference between receipts and costs is, of course, zero. When *ON* units are produced, losses are *BTD*, the difference between *ONDC* and (*ONBA* + *ATC*), a greater figure. When *OP* units are produced, the losses are *QTR*, and when output equals *OG*, losses are *TJH*. Thus, just as when fixed costs were *AXYC*, losses are least when production is set at *OS* units, the output at which marginal cost equals marginal revenue.

In general we may conclude (one exception will be noted later) that the output at which marginal revenue and marginal cost are equal gives maximum profits or minimum losses; and since the firm is in business to make profits, we may therefore conclude that it will pro-

[1] Except the *S*th unit itself, which just pays for itself.

duce the output at which its marginal cost and revenue are the same.[1] This output is the *optimum* output from the point of view of profits.

The Significance of Fixed Costs

It is instructive to analyze why the most profitable output is *OS*, where marginal cost and marginal revenue are equal, no matter whether fixed costs are equal to *AXYC*, are greater, or are less. If fixed costs are greater than *AXYC*, the profits earned in producing *OS* units will be somewhat less than *XTY*. Nevertheless, *OS* represents the most profitable (or the minimum loss) output. Fixed costs represent obligations that have been made in the past; currently, nothing can be done about them. They represent a lump sum that must be subtracted, no matter what the level of output, from the difference between total receipts and *variable* costs in order to ascertain profits. And whether this lump sum is great or small, the output that gives the maximum difference between total receipts and *variable* costs also gives the maximum figure for profits. The level of fixed costs does not affect the *current* decision as to which scale of output is most profitable, although it does, of course, play a role in determining just how large such maximum profits will be. This point must be kept clearly in mind when we analyze the effect upon output of a change in fixed costs.

The Most Profitable Price

As we have seen, the most profitable output will be *OS* units, or that at which marginal revenue equals marginal cost. Now, what about the price the firm will set? To answer this question we must first plot the demand or average revenue curve, for it records in the most convenient form the needed information, that is, the price that may be charged in order to market any given output.[2] The most profitable output is *OS*, and the average revenue curve (see Figure 29) indicates that the price must be *OL* in order to persuade purchasers to take *OS* units. The most profitable price is therefore *OL*. Since receipts equal the product of price and the number of units produced, total receipts for *OS* units can be expressed as the area *OSRL*.

[1] There is a second condition, namely, that for levels of output in excess of *OS*, marginal cost exceeds marginal revenue.

[2] Compare Chapter 10.

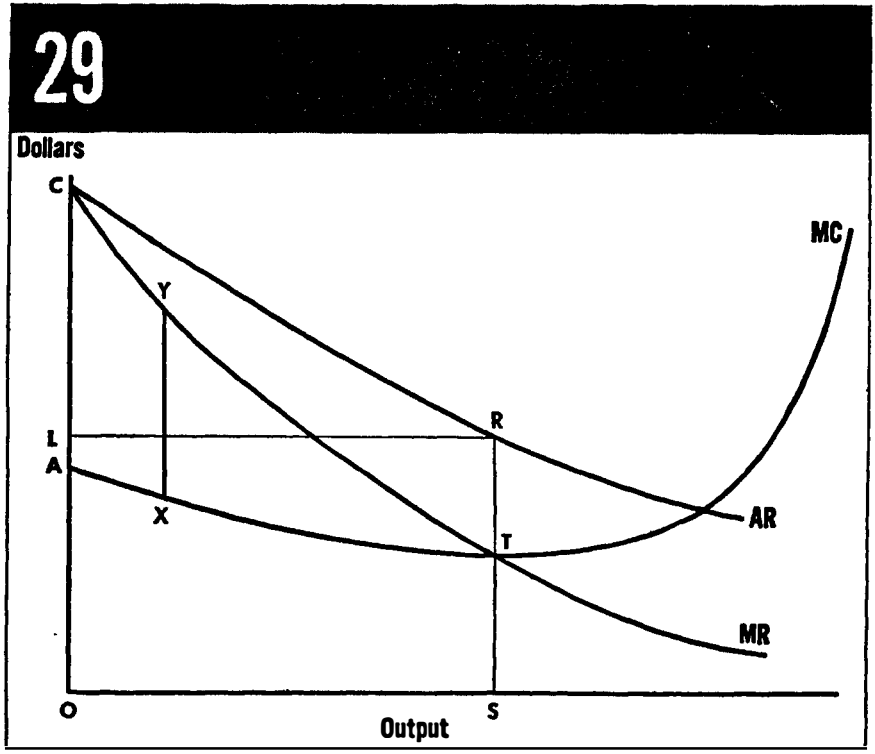

Figure 29. *Most Profitable Price*

Total Profit; Alternative Representation

If we plot the average total cost curve, we can represent total costs graphically, just as we showed total receipts. The firm incurs costs of *OC* for each unit of output when it produces *OS* units, as shown in Figure 30. Therefore the total cost of *OS* units is represented by the area *OSYC* (or *OS* times *OC*).

When you sell something for 10 cents a unit that costs only 6 cents to produce, you realize a 4-cent profit on each unit you sell, and if you sell 10,000 units a week your total profit is $400. If the unit cost of production is subtracted from the price, the difference equals the unit profit, or the profit margin (that is, *LC*). Hence we can conveniently represent the total profits earned in producing *OS* units, for it is the product of the profit per unit (*LC*) and the number of units produced (*OS*). The rectangle *CYRL* measures total profits. Or, to look at it

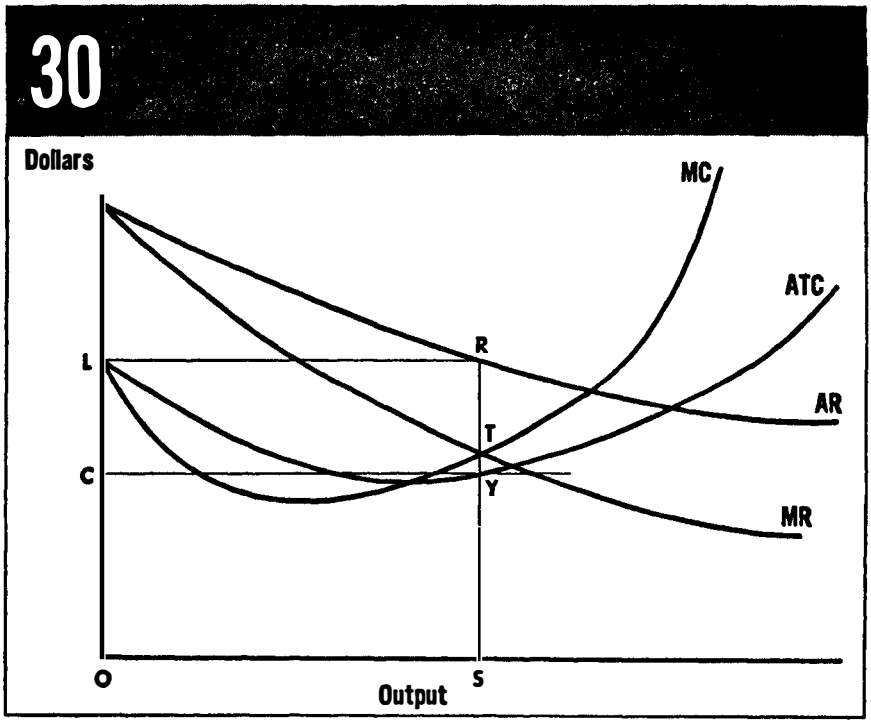

Figure 30. *Average Cost and Profit*

from a slightly different point of view, we may represent profits from an output OS in terms of the difference between total receipts and total costs. We have seen above that total receipts can be denoted by the rectangle $OSRL$ — or price per unit by the number of units produced; and that total costs can be shown by the rectangle $OSYC$ — or cost per unit by the number of units produced. The difference between these two rectangles, $CYRL$, thus represents the total profit.

It is worth pointing out that this method of representing profits is an alternative one to that employed in Figure 28, where XTY represented the total profits. The amount is no different, whether we represent it with a marginal or an average curve, but for some purposes the latter method has certain advantages. For although, as we shall see, it is a very lengthy task to determine the most profitable level of output by using the average measures, once the most profitable output has been established, it is easier to estimate profits by referring to the area

of a rectangle like *CYRL* than by trying to judge the area of an irregularly shaped figure such as *XYT* in Figure 28.

Using only the average revenue and average cost curves, we should have great difficulty in determining the output that promised the maximum profits. We can illustrate the problem by reference to Figure 31. The profit from any output, say *OA*, equals the product of

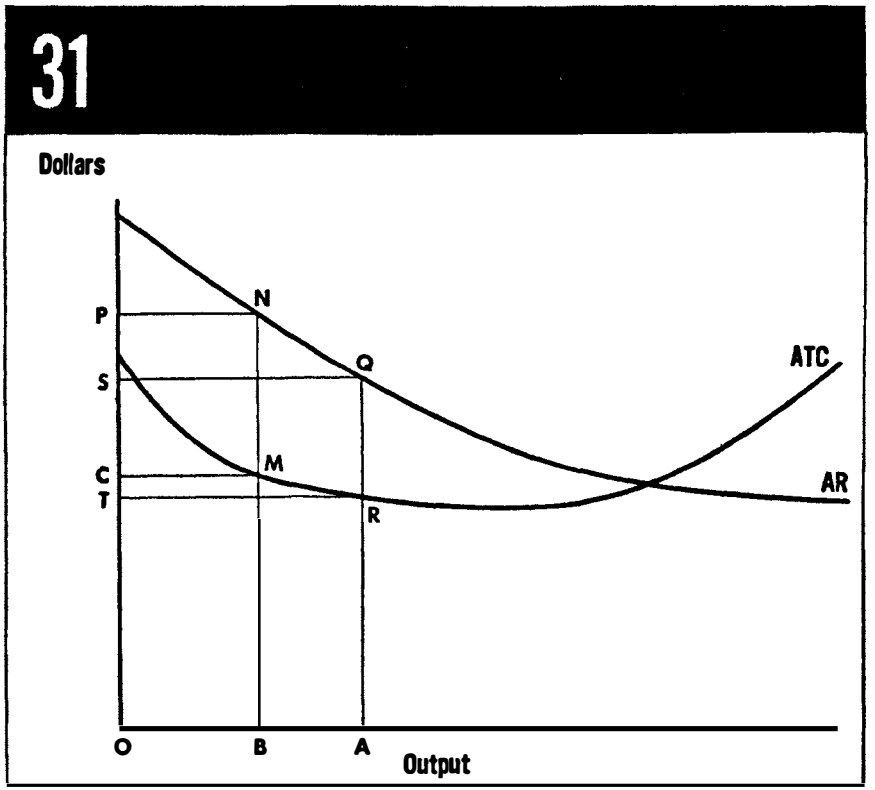

Figure 31. *Average Cost, Average Revenue and Most Profitable Output*

the profit per unit, in this case *ST*, and the number of units produced, in this case *OA* or *TR*. The rectangle *TRQS* thus represents the total profit at an output of *OA* units. Likewise the total profit at an output of *OB* units equals the rectangle *CMNP*. But which area is the larger? We have to compare two rectangles, one of which is lower and wider than the other. Thus, although the profit per unit is less

for *OA* units than for *OB* units, the output is greater. Does the increased output make up for the lower profit margin? This can hardly be answered by a casual inspection. Nor is this the only difficulty. The firm is not confined in making its decision to a choice between *OA* and *OB* units of output. We should have to construct comparable rectangles [1] for each of the thousands of alternatives, and then to compare the areas of these rectangles. But if we use the marginal revenue and marginal cost curves instead of the average curves, we can easily and quickly see where profit is highest. Once we have found the most profitable output, we can conveniently illustrate the amount of profit by constructing a rectangle similar to *CMNP*.

The Decision Not to Produce

We have already stated that there is one exception to the rule that a firm produces at the level at which marginal cost and revenue are equal. Instead of producing that output, the firm may be able to minimize its losses by not producing at all. It will do this if every output in excess of zero means a greater loss than no output does. If it produces nothing, its sales receipts will of course be zero, though its costs will still be positive, for so long as it continues to exist, it will have to meet certain fixed charges, such as salaries and taxes, depreciation, heat, light, power, and insurance. Hence, if it produces nothing at all, its losses will equal its fixed costs. But its losses at every positive output will exceed its fixed costs if at each such output it receives a price which is lower than the average variable cost; or in other words, if the total receipts for every output are less than the total *variable* cost of producing that amount. Figure 32 illustrates the situation in terms of marginal cost and marginal revenue.

The total revenue minus the total variable cost of producing *OS* units would in this case be *YATB* minus *CYR*, a negative figure.[2] Thus the firm would suffer a loss, no matter what positive output it produced, even before making an allowance for its fixed costs. Since to produce nothing would involve the firm in losses equal to its fixed costs only, it will be seen that the firm would minimize its losses by ceasing operations. Thus when price does not even cover the average variable cost, the firm suffers a smaller loss by producing nothing than

[1] Or, to the same effect, we should have to multiply profit per unit by the number of units produced for a host of alternative output figures.

[2] This, of course, implies that price is lower than average variable cost for every output.

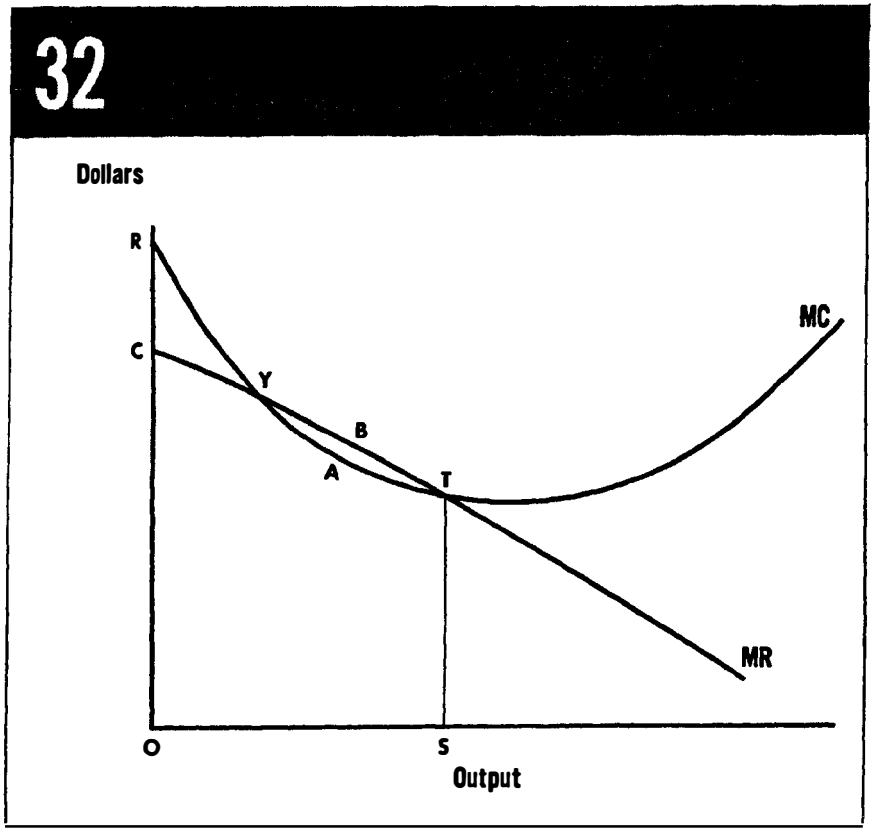

Figure 32. *Output with Total Receipts below Total Variable Costs*

by producing any positive output. In all other circumstances, the firm maximizes its profits or minimizes its losses by choosing the output at which marginal cost equals marginal revenue.

The Firm in Perfect Competition

A firm producing under conditions of perfect competition is in a special situation, for it can sell any output at all at a given price, that is, the market price. Unlike a firm in an industry where competition is imperfect, it does not need to reduce its price in order to expand its sales. Thus the average revenue curve is perfectly elastic or horizontal, and the marginal revenue curve is identical with it. For a firm in a perfectly competitive industry, one condition for maximum profit — that marginal cost equal marginal revenue — can be re-

stated as follows: profits are at a maximum when marginal cost equals price. The output is thus set at the level where the added cost of the last additional unit is equal to price.

Perhaps at this stage it may be asked how price itself is determined. When the firm is producing under conditions of monopoly or monopolistic competition, this question is apparently answered directly once the average revenue curve is drawn, for as we have seen, the average revenue curve illustrates the price that can be charged for each level of output. However, under perfect competition, the price a firm can charge is the same no matter what the level of its output, and each firm is practically forced to sell on the same terms as its competitors. However, we shall have to postpone until a later chapter our analysis of the determination of that price.

Figure 33 represents the situation of a firm in a perfectly competi-

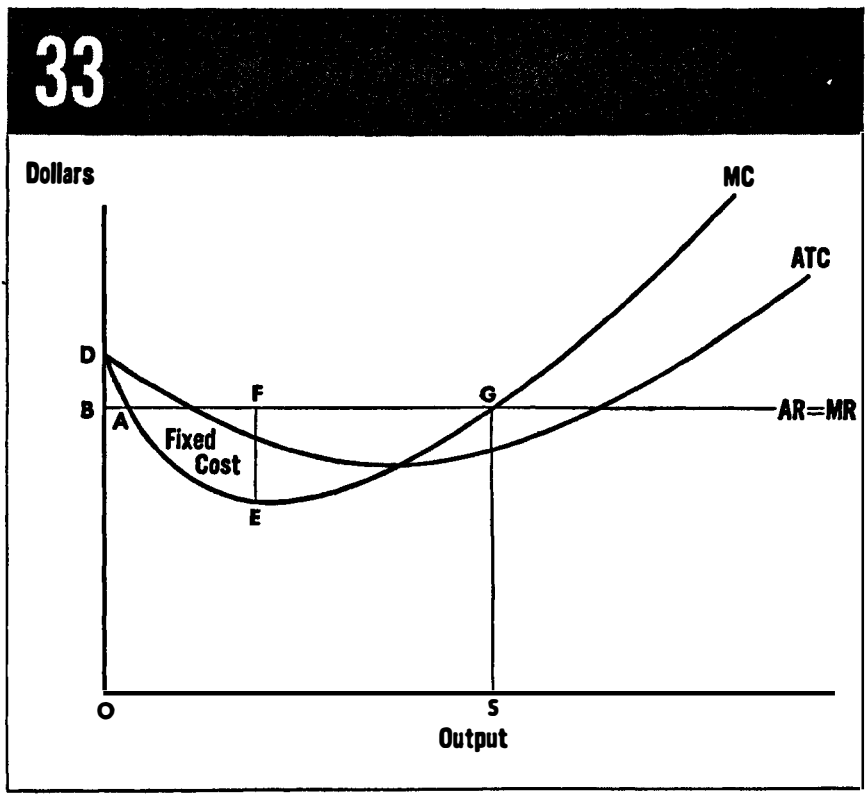

Figure 33. *Output Determination: Perfect Competition*

tive industry. The average revenue curve is infinitely elastic, and the marginal revenue curve accordingly coincides with it. The most profitable output is *OS*, where profits are equal to *EGF* minus *BAD*; *AEF* representing fixed cost. The price will naturally be *OB*, the price level for the whole industry. No other level of output would give such high profits.

The Output of an Oligopolistic Firm

We have seen that in many industries the firms sell under oligopolistic conditions. At what level will such a firm determine its output and its price? Will it also select the output where marginal cost equals marginal revenue? Profits, we can easily see, are maximized where marginal revenue and cost are equal, as with firms in imperfectly and perfectly competitive industries. But to say that the most profitable output is that at which marginal cost and revenue are the same is not in this situation as much of a solution as it appears to be at first glance.

First, we must recall that a firm in an oligopolistic industry believes itself to be faced with a peculiar kind of market demand. It knows how much it is selling at the price it is currently charging. At any price above that level it anticipates a much lower volume of sales, and for any price below it expects only a slight increase. Thus the average revenue function is quite elastic for increases in price, but quite inelastic for reductions. Corresponding to the kink in the average revenue curve at the current price, there is a gap in the marginal revenue curve. This situation is illustrated in Figure 34.

If the marginal cost curve passes through the gap in the marginal revenue curve, as it does in this diagram, profits are at a maximum with the output of *OS*. At this output they are equal to *XGFY*, since as in previous illustrations we measure the fixed costs by *AXYC*. At any other output, profits lower than *XGFY* can be expected. Hence the most profitable output would be that at which the marginal cost curve intersected, or more descriptively, passed through the gap in the marginal revenue curve. Price would of course be *OP* — the current price.

This statement about price perhaps illustrates most clearly the deficiency and incompleteness of this analysis. Until we know the current price, we do not know where the kink in the average revenue curve will come. But once we are given the current price, it follows that the gap in the marginal revenue curve will appear at the very

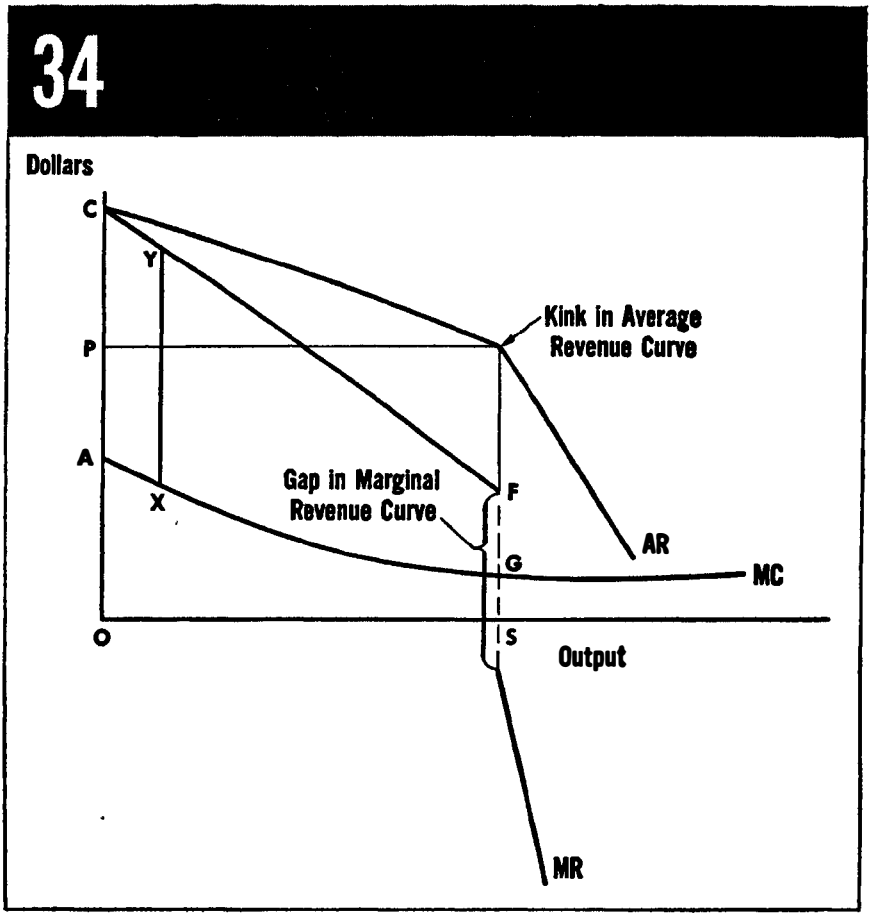

Figure 34. *Output Determination: Oligopoly*

output that can be sold at that price, and since the marginal cost curve would probably pass through the gap, the current output is the one at which maximum profits may be expected. Thus, in order to determine the price and output which an oligopolistic firm would set today, we have to know what price the firm inherited from yesterday.

More light will be thrown on the oligopolistic firm when we discuss changes in demand and costs. Meanwhile, it should be noted that we have not challenged the formal statement that profits are greatest at the output where marginal cost equals marginal revenue.

Summary

We have seen that a firm, whether it is producing under conditions of monopoly, of perfect competition, of ordinary imperfect competition, or of oligopoly, maximizes its profits by producing the output at which marginal cost and marginal revenue are equal, and will charge that price at which it can market such an output. The only exception to this rule is the firm which suffers losses no matter how much it produces, but which can minimize its losses by producing nothing at all. Only when it is unable to get a price that covers its average variable cost will it be induced to cease production. Otherwise, firms that seek maximum profits produce the output for which marginal cost equals marginal revenue.

It is sometimes helpful to look at the condition for maximizing profits from a slightly different point of view. Suppose we knew that if a firm produced 100 units of output it could realize profits of 1000. Under what conditions would it be tempted to produce a greater output? Only if a larger output, say 101 units, meant profits in excess of 1000. But 101 units of output would be more profitable than 100 units only if the 101st unit added more to receipts than to costs. If the 101st unit did add more to receipts than to costs, and hence if it added to profits, it would mean that the marginal revenue corresponding to the 101st unit exceeded the marginal cost for that output. Thus, so long as the marginal revenue exceeded the marginal cost, the firm could make greater profits by expanding output. But how far could this expansion be carried? Obviously not into that range of output where marginal cost exceeded marginal revenue, for this would mean a reduction in profits. The highest level of profits would be realized if output were increased from, say 100, to that level at which marginal revenue just ceased being higher than marginal cost; that is, if output reached that level at which marginal cost and marginal revenue were equal. This is an alternative way of looking at the conditions for maximizing profits, the determination of that output at which marginal cost equals marginal revenue.

To say that a firm's profits are highest at the output where the marginal cost and marginal revenue are equal is not to state a newfangled code of action for businessmen. Rather it simply describes in technical language how the management of a business firm determines its output — assuming that its motive is to maximize profit. It is thus a description [1] of business practice, not a formula for earning

[1] How accurately it describes actual business practices we shall consider in Chapter 17.

money. And what we as economists want is precisely that: an account of how business firms actually operate. This knowledge is needed as the foundation for an understanding of the way in which the economy functions.

15

Changes in Demand and the Effect on Output

WE ARE NOW IN A POSITION to analyze the way in which a change in the demand or in the cost function will affect the amount the firm would produce and the price it would charge.

Increase in Demand with Costs Unaltered

Let us consider first the effect of a change in demand, the conditions of cost being assumed constant. We shall suppose that the firm belongs either to an imperfectly competitive industry or that it is a monopoly. Only very minor adjustments are needed in the analysis when the firm is in a perfectly competitive industry, while under an oligopoly, the situation is somewhat different.[1]

Again we may most easily illustrate the situation by the use of a graph. In Figure 35, AR_1 represents the *original* average revenue curve, and AR_2 represents the *changed* (in this case, increased) demand. MR_1 is the marginal revenue curve drawn with reference to AR_1, and MR_2 is the marginal revenue curve based upon the increased demand. Since the conditions of cost are assumed to be unchanged, MC represents the marginal cost, both before and after the rise in demand. In the original situation the most profitable output is indicated by O_1, the output corresponding to which the marginal cost and the original marginal revenue curve intersect. To sell O_1, the firm will charge P_1, for undoubtedly it will not choose to sell at a lower price and thus voluntarily limit its profits. But when demand increases, the most

[1] The first of these two situations will be discussed in Chapter 18; the second later in the present chapter.

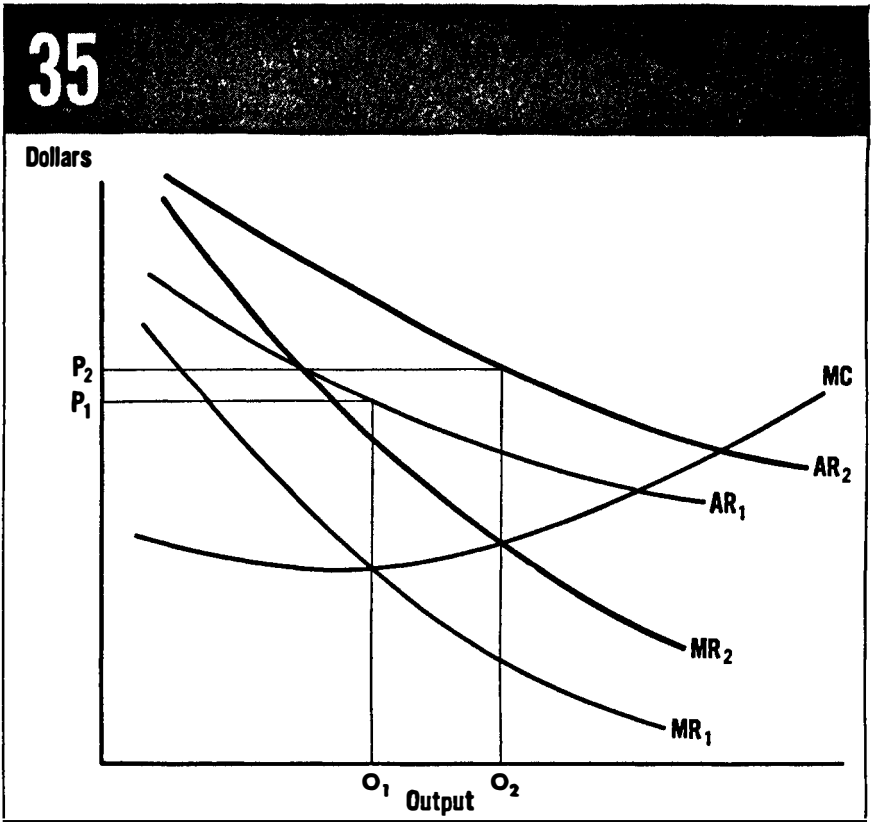

Figure 35. *Effect of Change in Demand*

profitable output is no longer O_1, for at that output the marginal revenue, as given by MR_2, is well above the marginal cost. Hence greater profits could be realized by expanding output to O_2, the level at which the new marginal revenue curve and the original marginal cost curve intersect. And the price which the firm is able to secure is now P_2.

Marginal Cost and Change in Price and Output

An increase in demand will ordinarily, as we have seen, lead the firm both to expand its output and to raise its price; that much is clear. To the extent that output is increased, it satisfies the increased demand. And the greater the increase in output (for a given increase in

demand), the more completely the market is satisfied. But the rise in price is not a response which the firm's customers would desire; in fact it is just the opposite. In a sense it is a confession of failure — of failure to meet the increased desire for the product. For as can be seen by an inspection of Figure 35, the greater the increase in output, the smaller is the increase in price; the smaller the rise in output, the greater is the rise in price. What the rise in price does is to turn away some of the customers, for naturally less is wanted at the higher price than would have been wanted if the price had not been raised. Obviously, if the response to an increase in demand is chiefly to expand output, the buyers gain much more than if the main effect is a rise in price. It is thus important to see what determines the relative increase in price and output.

As we shall see in our examination of Figures 36, 37, and 38, the size of the increase in price and output for a *given increase in demand* will depend upon two things: first, upon the character of the marginal cost curve, and second, upon the relation between the average and marginal revenue curves. The influence of the first factor is easily seen by reference to Figure 36. If the marginal cost curve sloped upward steeply instead of gently, the newly drawn marginal revenue curve would intersect it at a point only slightly to the right of O_1, and the increase in output would be much smaller while the increase in price would be much greater. If, on the other hand, the marginal cost curve were nearly horizontal, the new marginal revenue curve would intersect it at a point considerably beyond O_1, so that the increase in the output would be large, while the increase in price would be very slight or perhaps zero. In Figure 36, two marginal cost curves are drawn, one MC_1, which rises sharply, and the other MC_2, which rises only gradually; and the increase in demand is shown by the shift from AR_1 to AR_2. P_1 and O_1 are the original price and output, while P_2 and O_2 are the new price and output, after the increase in demand, when the firm is faced with only slowly rising marginal costs. For comparison, P_3 and O_3 are the new price and output, after the rise in demand, when the firm is faced with rapidly rising marginal costs. It should be noticed that the firm increases its output only slightly and its price considerably in response to the increase in demand when it is confronted with rapidly rising marginal costs, while it increases its output considerably and its price only slightly when its marginal costs rise only gradually as it raises its output.

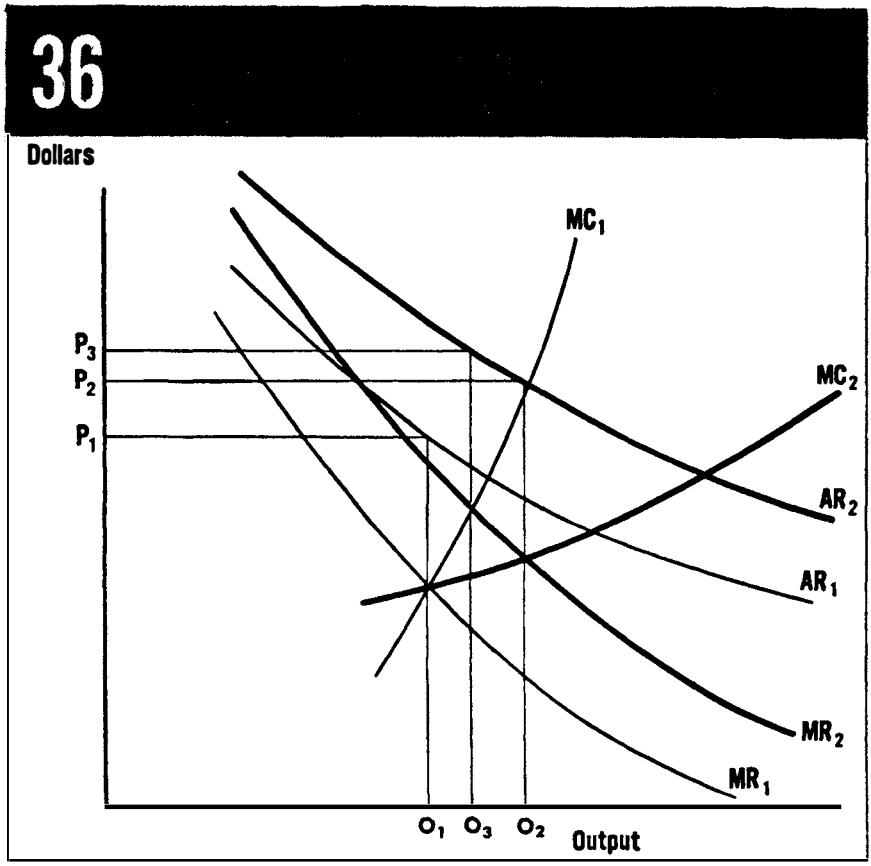

Figure 36. *Shape of Marginal Cost Curve and Effect of Increased Demand*

This conclusion could be expressed in less technical language somewhat like this: when it is easy and relatively inexpensive for the firm to expand its output (that is, when the marginal cost curve is approximately horizontal), an increased demand will lead the firm to expand output freely and to raise price only slightly. On the other hand, if the firm can expand its output only with very great difficulty and at a sharp increase in cost, it will, if demand increases, raise its price considerably and expand its output only slightly.

This conclusion is important in the understanding of depression, prosperity, and wartime boom. In a later chapter we shall discuss these implications at some length. For the moment, however, it is important to continue with our present analysis.

Change in Elasticity of Demand and Relative Change in Output and Price

The other factor determining the relative size of the increase in output and price in response to a given increase in demand is, as we pointed out above, the relation between the average and marginal revenue, and particularly any change in that relation. The point of intersection of the new marginal revenue curve with the original marginal cost curve shows where profit is maximized after the rise in demand, and thus indicates the new output and price. Hence, even though the average revenue curve were to shift far to the right, the increase in output would be very small, and that in price would be relatively large, if the new marginal revenue curve moved only slightly in that direction. But if, in contrast, the marginal revenue curve moved much farther to the right than the average revenue curve, the increase in output would be great while price might actually decline.

The nature of the change in the marginal revenue curve for a given change in the average revenue function is not erratic or indeterminate. Generally speaking, when there is no change in the elasticity of the average revenue function, the marginal revenue curve changes by as much (proportionally) as the average revenue curve. However, if the new demand is *less elastic* than the original one, the marginal revenue curve will move to the right by less than the average revenue curve. And if the new average revenue curve is considerably more elastic than the original curve, the shift to the right in the marginal revenue curve will be greater than the shift in the average revenue curve.

These results can be easily proved mathematically. They can also be demonstrated convincingly enough to the reader who knows little mathematics, if it is borne in mind that as demand becomes more elastic, the gap between the average and the marginal revenue curves decreases. Thus, if the average revenue curve becomes more elastic as it moves to the right, the gap between it and the marginal revenue curve will lessen. Hence the marginal revenue curve would move even more to the right than the average revenue curve. This is illustrated in Figure 37.

AR_1 and MR_1 are the original average and marginal revenue curves; AR_2 and MR_2 are the increased average and marginal revenue curves. The new average revenue is more elastic than the original one. When the average revenue curve moves to the right a distance AB, the marginal revenue curve moves CD, which is greater than AB. O_1 and P_1

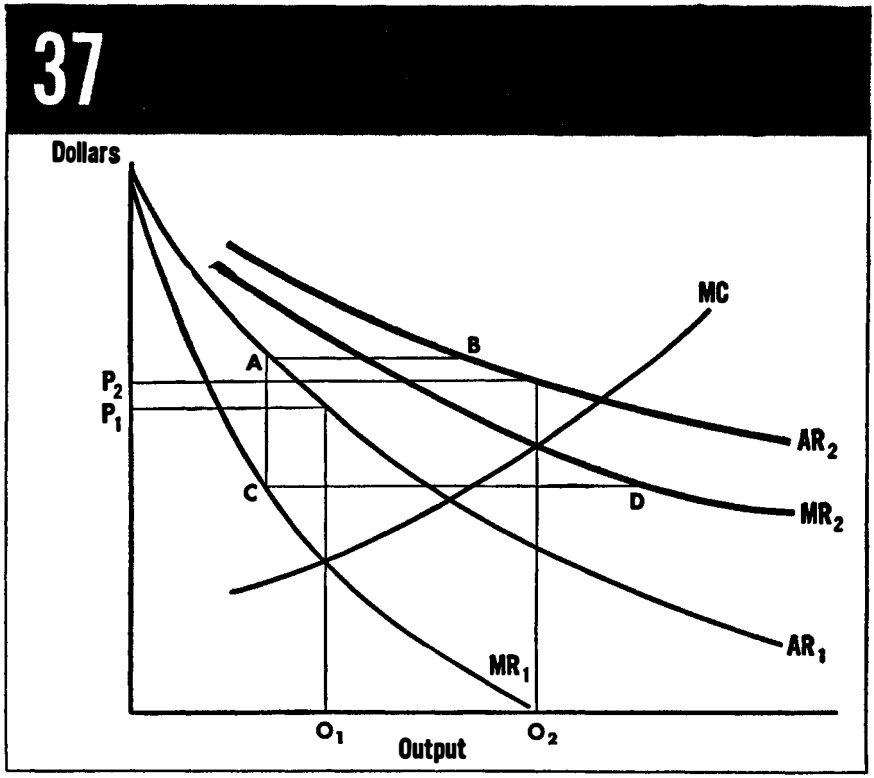

Figure 37. *Effect of Increased and More Elastic Demand*

are the output and price set by the firm in the initial situation. O_2 and P_2 are the firm's output and price after the increase in demand (which has become more elastic).

In the same way, we can show that if demand becomes less elastic as it increases, the marginal revenue curve will not move as far to the right as the average revenue curve. This follows from the fact that the gap between the average and the marginal revenue curves is greater, the lower the elasticity, as illustrated in Figure 38.

AR_1 is the original average revenue curve; AR_2 the increased and less elastic average revenue curve. MR_1 is the original marginal revenue curve and MR_2 is the new marginal revenue curve after the increase in demand. The increase in the average revenue may be measured by AB, that in the marginal revenue by CD. — Notice that AB is greater than CD. The original output is O_1, the output after the

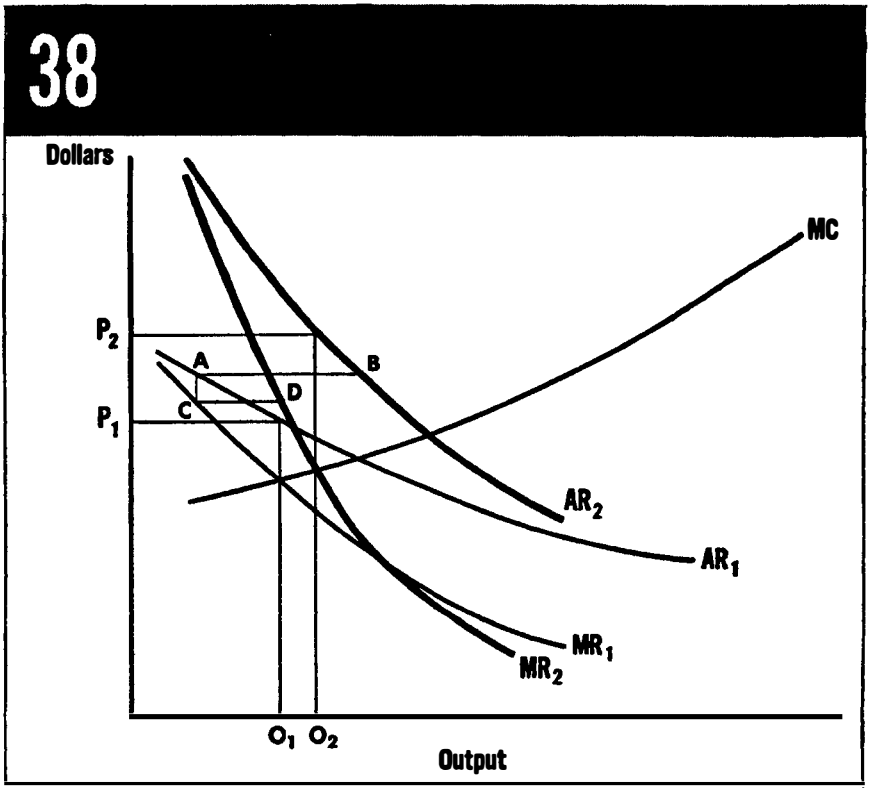

Figure 38. *Effect of Increased and Less Elastic Demand*

increase in demand is O_2, a relatively small increase. P_1 is the original price and P_2 the price after the increase in demand, a relatively large rise. This is in contrast to the situation illustrated in Figure 37, where output was raised by a large amount and price only slightly when demand increased and became more elastic.

We may conclude, therefore, that if demand becomes *less* elastic as it increases (and accordingly that the increase in marginal revenue is less than the increase in demand), then the increase in price will be relatively great and that in output relatively small. On the other hand, if demand becomes *more* elastic as it increases, the increase in price will be relatively small, and the increase in output will be great.

Effect of Reduced Demand on Price and Output

Now let us see what happens in the opposite circumstances; that is,

when demand falls. The effect of a decline in demand upon the price and output can be seen in Figure 35. In that diagram, let AR_2 and MR_2 now represent the original demand and marginal revenue curves, and AR_1 and MR_1 the average revenue and resultant marginal revenue curves after the decline in demand. Seeking to maximize profits, the firm would of course reduce its output from O_2 to O_1 and its price from P_2 to P_1.

The same considerations that were found to be important when demand increased will determine the relative size of the decline in output and in price when demand falls. If the marginal cost curve is approximately horizontal, the firm faced with a decline in demand will cut its output sharply and its price very little; while if the marginal cost curve slopes up to the right, the firm will reduce its price considerably and its output only slightly. Furthermore, when demand becomes more elastic as it falls, price will be cut by more, and output by less, than is the case when demand becomes less elastic as it falls.

Effect on Profits of a Change in Demand

As demand changes, profits will vary in the same direction, as long as the cost determinants remain constant. Under these circumstances an increase in demand will always bring an increase in profits, and a decrease in demand will necessarily do the opposite. This result does not depend on the shape of the marginal cost curve; nor, provided that the change in demand is not confined to only part of the range, does it depend on the change in the elasticity of demand. In other words, if demand increases, provided the new and old demand curves do not intersect, profits will be higher; and if demand falls, they will be lower. This result can be derived most easily by reference to Figure 39.

In this diagram, AR_1, MR_1, and MC represent the original average revenue, marginal revenue, and marginal cost. The increased demand is shown by AR_2, and the new marginal revenue curve is MR_2.[1] The area $BAXY$ measures fixed costs. Profits, when demand was at its original level, are shown by the area ACX. After the increase in demand, profits are represented by the area $ACDEYX$. Obviously, then, the higher the demand the higher are the profits.

In summary, when cost determinants are constant, an increase in

[1] If demand decreases, let AR_2 be the original, and AR_1 the new average revenue curve. The remaining steps will be obvious.

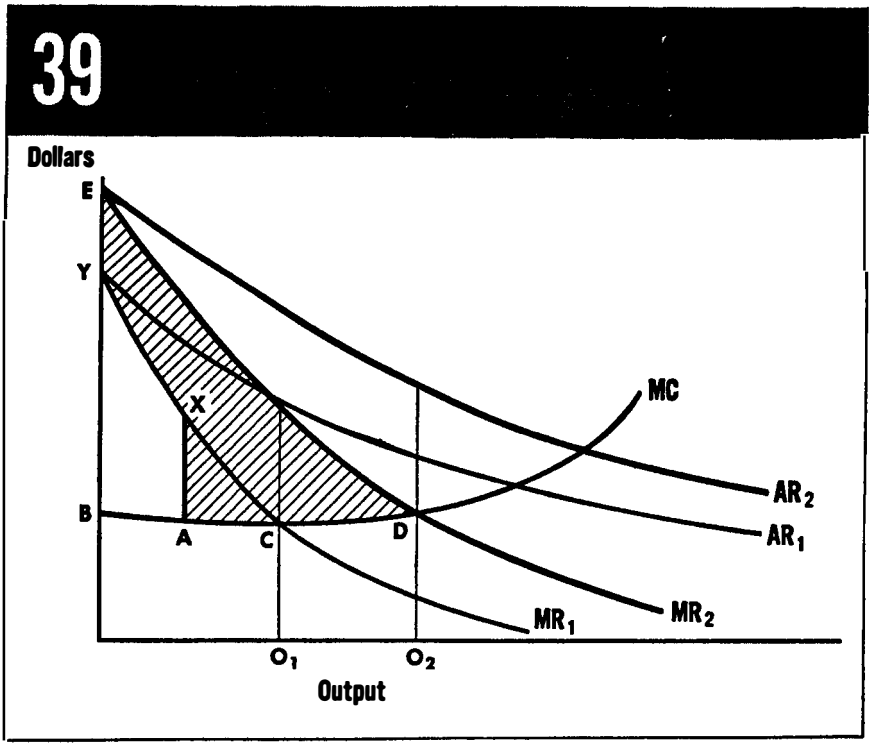

Figure 39. *Change in Demand and Effect on Profits*

demand will always bring higher profits and will usually lead to an increase in price and output, although occasionally to an increase in only price or output. A decrease in demand will have the opposite effect. But so far these results have been demonstrated only for a firm producing under conditions of simple imperfect competition or of monopoly. We have not yet examined the situation of a firm which is part of an oligopolistic industry.

Increase in Demand: Oligopoly

Under conditions of oligopoly, a price rise as the result of an increase in demand is much less probable than under imperfectly competitive conditions. In an oligopolistic industry, as we have seen, each firm is reluctant to take the initiative in raising its price because it fears that its competitors will not follow its lead. For, if its fear is justified, the firm will experience a sharp decline in sales after raising its price,

since many of its customers will transfer their trade to its competitors, who have not raised their prices.

In this situation an increase in demand may not persuade the firm to raise its price. Suppose that four firms, A, B, C, and D, make up the industry. Each of the four finds that at the existing price its sales have increased, and also that its profits have risen — assuming that they all have experienced an increase in the demand for their product. But firm A may be deterred from raising its price because of the high probability that B, C, and D would leave theirs unchanged. Thus the elasticity of demand for prices above the current price remains very high. This situation is represented in Figure 40.

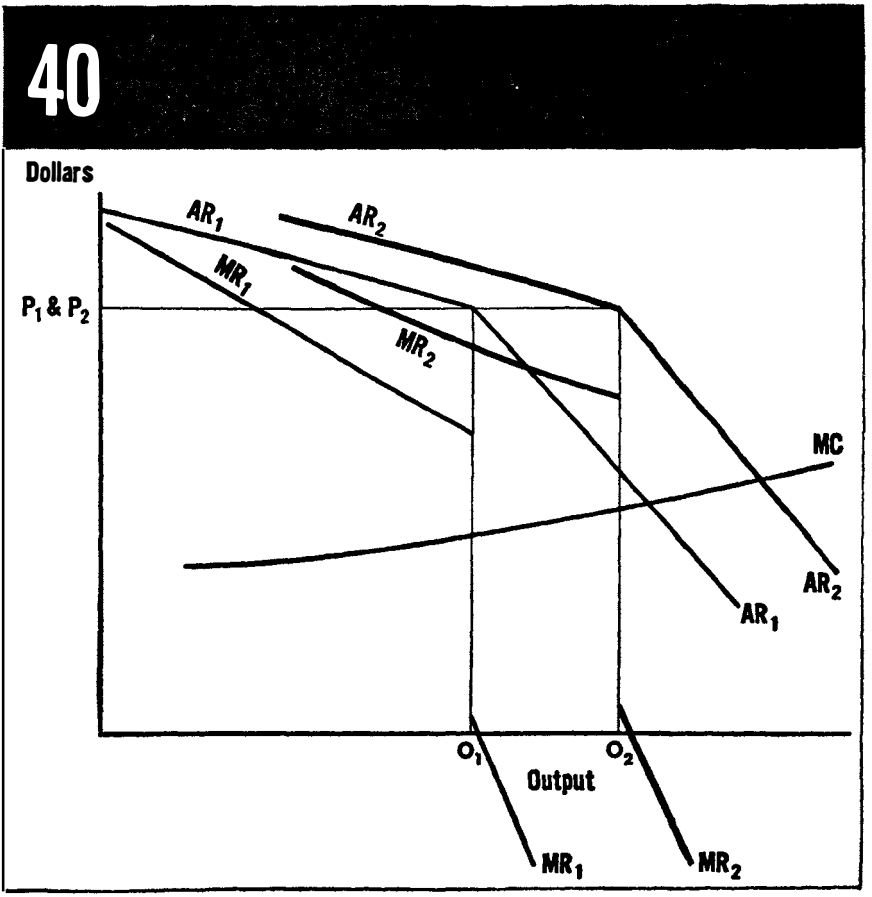

Figure 40. *Increase in Demand: Oligopoly*

The kinked demand curve AR_1 represents the demand that faced firm A originally, and MR_1 is the corresponding marginal revenue curve. Firm A, on experiencing the increase in demand, would find that at the current price, P_1, its sales were greater than before. But unless A felt considerably more hopeful that B, C, and D would follow its lead in raising prices than it did before the increase in demand, the new average revenue curve AR_2, would be no less elastic than AR_1, for prices above P_1. Hence the gap in the new marginal revenue curve would be about as large as in the original one. So long as the marginal cost curve passes, as before the change, through this gap, firm A would be unwilling to raise its price. Since the gap is likely to be quite large, and since the marginal cost curve is unlikely to slope steeply except when output is near capacity, a price rise following an increase in demand is improbable.

But firm A would raise its price if the marginal cost curve cut the new marginal revenue curve at an output smaller than that at which the gap occurred. This might happen if either the gap itself became smaller — as it would if firm A grew more hopeful that other firms would follow its lead in raising prices — or if the marginal cost curve sloped sharply upwards. This situation is illustrated in Figure 41.

Incidentally, once the price has been raised, the average revenue curve must be redrawn, first, so that it takes account of what the other firms actually do to their prices, and secondly, so that when firm A contemplates another change in price, the new curve embodies its revised expectations about the pricing policy of B, C, and D. But whether A raises its price or not, it seems clear that both profit and output will be higher following an increase in demand.

Decline in Demand: Oligopoly

The effect of a reduction in demand upon the price and output of a firm which is a member of an oligopolistic industry can also be worked out using this analysis. The reduction in demand will be experienced initially as a decline in sales at the current price. There will, however, be considerable resistance to lowering price because each firm would fear that if it did so its competitors would do likewise and it would not maintain the advantage which the lowered price would initially give it. This fear means, as was pointed out earlier, that for downward adjustments in price, the average revenue curve

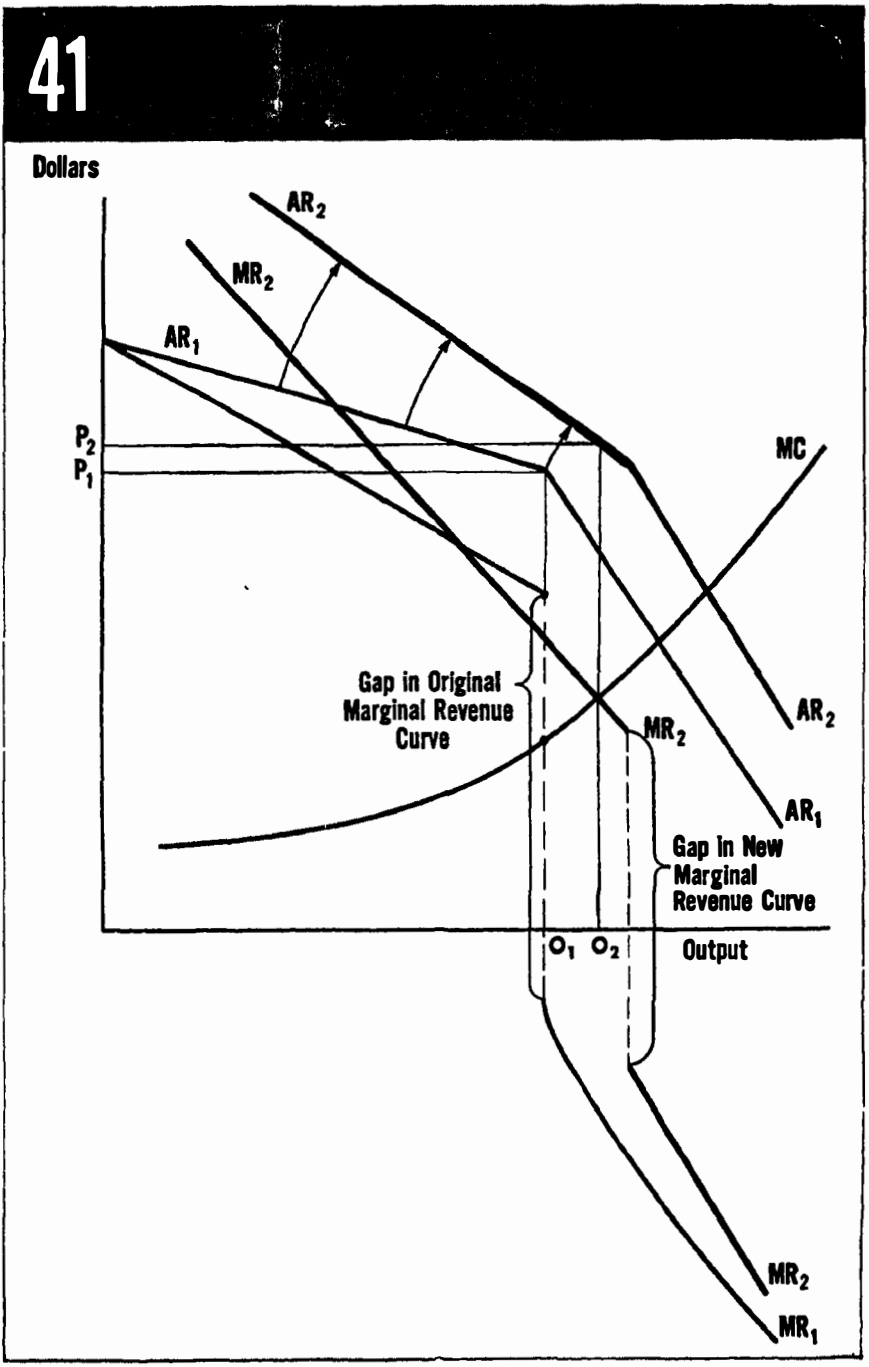

Figure 41. *Increase in Demand Leading to Price Change: Oligopoly*

is relatively inelastic. And as the decline in demand was experienced by all competing firms, the fear of beginning a price war would probably increase. This means that the elasticity of the average revenue curve grows even less for price reductions. Hence the gap in the marginal revenue curve would widen, thereby reducing the probability that the marginal cost curve would fail to pass through the gap. And so long as the marginal cost curve passes through the gap, the firm is not induced to change its price.

Oligopoly and Price Stability

It seems clear, then, that a firm in an oligopolistic industry is unlikely to change its price whether demand increases or decreases. For such a firm, price will tend to be stable. The geometric demonstration of this point should not obscure the common-sense argument in its favor. With oligopoly, the firm is deterred from raising its price because of the fear that its competitors will not do the same, while it is reluctant to lower its price for just the opposite reason: it fears that its competitors will also lower theirs and that it will then be in the same competitive position as before but with much lower revenue. Expressed this way, the reason for price stability in these conditions is sufficiently obvious.

Yet it must not be supposed that an oligopolistic firm will never change its price. Price may change in response to a change in demand, but it is less likely to do so under these circumstances than it would be if the firm were not part of an oligopolistic industry.

Summary

An increase in demand will probably induce a firm that is not producing under oligopolistic conditions to raise its price and output. Likewise, a decrease in demand will probably lead such a firm to reduce its price and its output. The change in price will be greater, and hence the change in output will be less, (*a*) as the upward slope of the marginal cost curve is greater, and (*b*) as the change in the elasticity of demand is greater, provided that this change is opposite in direction to the change in demand itself. But for a firm which is producing under conditions of oligopoly, price is not likely to vary with a change in demand. However, if the marginal cost curve has a steep enough slope, or if there is a drastic revision in the firm's expectations about its rivals, price may change.

There should be no confusion about *who* changes prices. It is easy to gain the impression that prices are changed by "the forces of competition," or "the laws of supply and demand," or "inflation," or "deflation." Indeed, a full-page advertisement, sponsored by an important business firm and carried by most of the American press during a period of controversy over price control legislation, informed its readers that "Due to inflation — caused by the reduction in the purchasing power of our money — most manufacturing costs and prices have risen substantially since the war." The implication is that prices are not made by man; that they are imposed by some economic juggernaut to whom man is a helpless slave. This is absurd, though it is easy to see why the firm which printed this advertisement was eager to shift the blame for higher prices away from itself to some acceptable scapegoat. Prices are raised or lowered because some men decide to raise or lower the price they charge. They may, of course, do so in response to the pressure of economic forces, as, for example, because of a change in demand, but they are no more *compelled* to do so than you are compelled to read this sentence. In reading later parts of this book it will be well worth while to remember this general truth.

16

Changes in Cost and Output

IN DETERMINING ITS PRICE AND OUTPUT, a firm must take into account not only the demand for its product, but also its production costs. Hence it may be induced to change its price and output if its costs alter, just as it may if the demand for its product shifts. Costs change, as we said in Chapter 9, when any of the cost determinants alter. They will change if the firm is increased in size, if the methods of production are modified, or if prices change for any of the factors of production, such as labor, raw materials, or land. In this chapter, then, we shall determine the effects of a change in any of the cost determinants upon price and output, following the lines of the analysis we employed to determine the effects of a change in demand.

In treating the effect upon price and output of a change in costs, we must first distinguish carefully between those changes which influence marginal cost and those which do not. We have previously seen that, even though average cost is altered by a change in a cost determinant, marginal cost may be unaffected. Thus, for example, a change in the wage rate (an element of *variable* cost) changes both average and marginal cost in the same direction, but a change in rent or in the salary of a high official does not affect the level of marginal cost because these are not items of variable but of fixed cost. As in Chapter 15, we shall consider first the effect of a change in costs on the output of a firm producing under conditions of imperfect competition or monopoly, and then the effect for a firm which is part of an oligopolistic industry.

Change in Cost Determinants Not Affecting Marginal Cost

Again it will be easiest to analyze the problem if we set it out in a diagram. Figure 42 presents the situation in which a change in the cost determinants does not affect marginal cost. In this case the

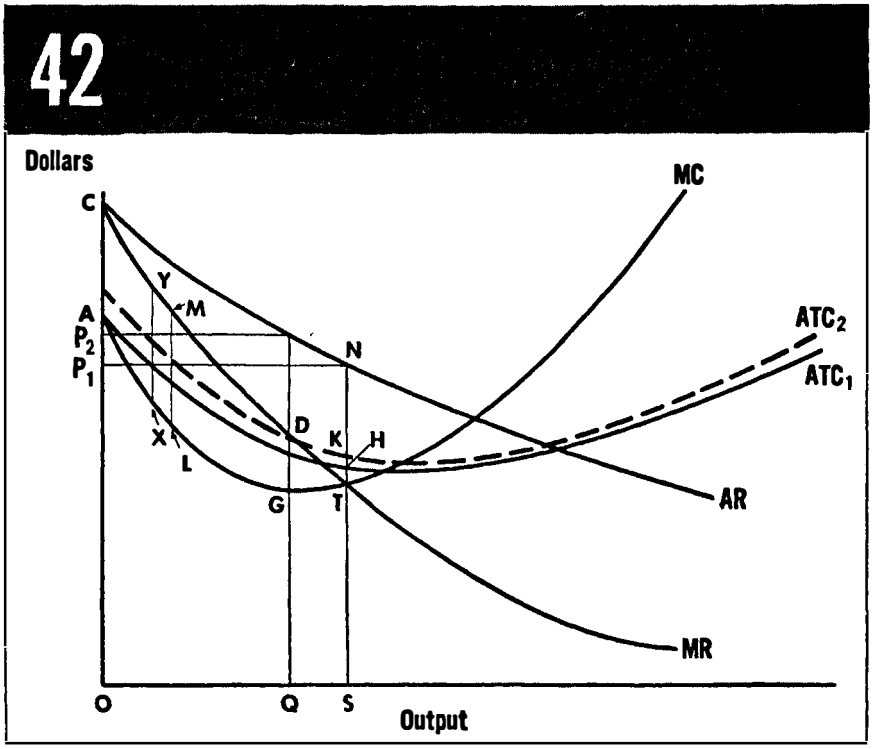

Figure 42. *Effect of Change in Fixed Cost on Price and Output*

intersection of the marginal revenue curve and the marginal cost curve is, of course, not changed. Hence we may conclude that the output which it is most profitable for the firm to produce is the same as before, even though *average* cost is altered. Furthermore, in this situation the firm does not profit by changing the price. This conclusion may seem paradoxical, since it implies that costs may rise and yet that the firm may not find it worth while to raise its price. Nevertheless, to give some examples, a rise in the property tax, a reduction in the interest charges on bonded indebtedness, or an increase in the salary of the vice-president of the firm would not induce

it to change its price or its output. Naturally any such change in cost would either raise or lower profits, and the firm would therefore welcome certain changes and resist others. But such changes would not, since its marginal costs would remain unchanged, lead the firm to alter its price or output, provided of course that the firm followed a policy directed to maximizing its profits. Thus, a rise in the property tax, for instance, would reduce profits but would not change the level of output at which profits were at a maximum. Profits would be lower than before the increase in tax at every level of output, but the output at which profits reached their peak would be no different.

As shown in Figure 42, profits are maximized at the output OS, both before and after the rise in costs. The fixed costs originally are denoted by $AXYC$, and after the tax increase, by $ALMC$. Profits before the increase in costs are equal to the area XTY; afterwards, to LTM. The profit margin falls from HN to KN; HN being measured against the original average cost curve ATC_1, and KN against the final one, ATC_2. But the firm earns maximum profits by producing OS both before and after the change in costs. Since the most profitable output is OS, the price which the firm would charge is OP_1. If the firm were to raise its price to, say OP_2, it would sell only OQ and its profits would be only $LGDM$, a figure less than LTM. Hence it would not pay the firm to raise its price if an element of fixed cost were increased.

Effect on Output and Price When Marginal Cost is Altered

Whereas a change in fixed cost does not change price or output, an increase in the price of an element of variable cost will have a much more direct effect. For in this case the marginal cost function will be changed in the same direction as the price of the productive factor. For example, if wage rates are increased, the marginal cost will also be increased, and the new marginal cost curve will intersect the marginal revenue curve at a point somewhat to the left of the original intersection, as Figure 43 shows.

MC_1 is the original marginal cost curve; MC_2 represents marginal costs after the rise in the wage rate. AX_1Y_1C and BX_2Y_2C represent fixed costs; the former measured with respect to MC_1, the latter with respect to MC_2. But since fixed costs are unchanged, the areas AX_1Y_1C and BX_2Y_2C are equal. Originally the output is O_1 and the price is P_1, while after the increase in the wage rate, the output falls to O_2 and the price is increased to P_2. Profits before the increase in

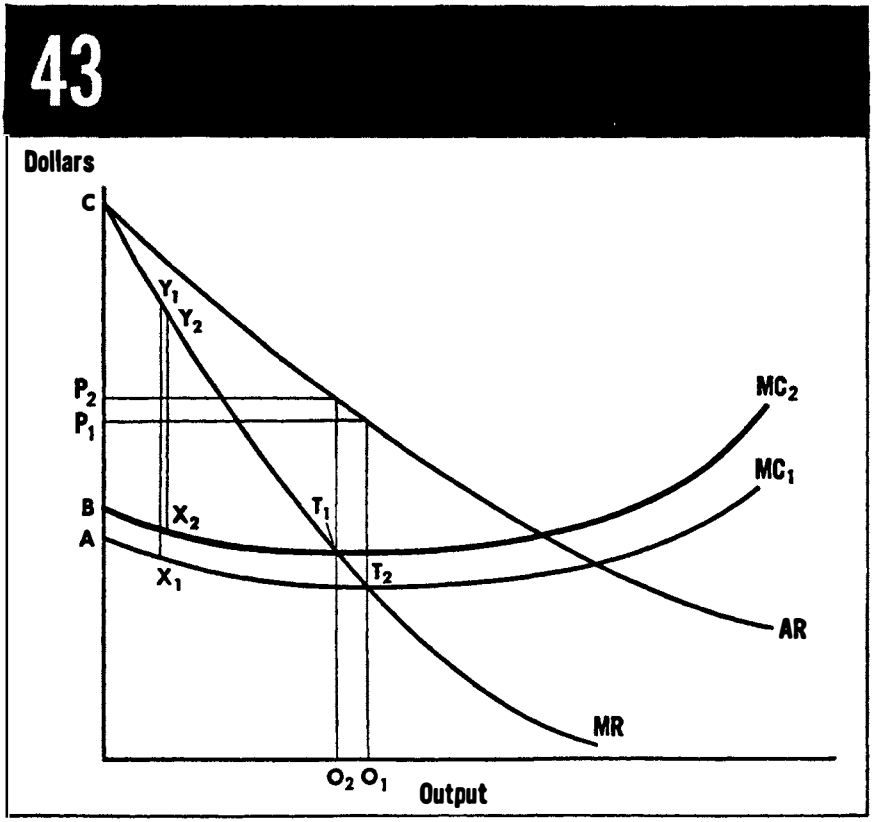

Figure 43. *Effect of Change in Variable Cost on Price and Output*

the wage rate are $X_1T_2Y_1$ and afterwards, $X_2T_1Y_2$, which is smaller.

In summary, then, the firm reduces its output after its costs rise in order to secure the highest level of profits, and it also charges a somewhat higher price. But it earns less profit than before in spite of the increase in price. Naturally, therefore, every firm resists an increase in costs even though it may be able to "pass on" a part of the increase to its customers.

We must be careful in interpreting the result just reached and in applying it to questions of public policy. It says that, *given the nature of the demand*, an increase in costs resulting from, let us say, an increase in wage rates, will induce the firm to produce less and to charge more. But there is a question whether, when wages change in the real world, it is justifiable to assume that the demand is *given*. It is quite possible,

as we shall see later, that a general change in wage rates will itself affect the demand, and if it does, the problem of determining the consequence of changing wages becomes very complicated indeed. One of the most important things to be learned in a preliminary study of economics is the nature of the conditions that must be satisfied if widely accepted conclusions are to be relied on. Thus, for instance, it is commonly believed that a rise in wage rates brings about reduced output and employment. As we have just seen, this conclusion holds, *provided we can assume that the demand is not affected by the change in wage rates.* What happens otherwise is another matter. However, we must postpone detailed consideration of this problem until we are in a position to study the consequences of such a change in wages upon demand.

There is one other kind of change in cost which it is desirable to explore, and that is the change that comes with the development of new methods of production. An improvement in the technique of production generally leads to a reduction in cost; otherwise the firm would have no particular inducement for adopting it. Such an improvement may mean that a given output can be produced with a smaller working force and therefore with lower payrolls and lower labor costs. While it is not certain that the marginal cost function would decline on this account, such a decline is normally to be expected. Or the new technique may not cut down the labor force, but may bring about a reduction in the amount of raw materials needed for a given volume of finished product, as in the case of recent developments in the refining of beet sugar. Therefore improvements of this kind also would lower average and marginal cost. Consequently, since technological improvements lower marginal cost, the firm will raise its output and lower its price, if it can be assumed that there will be no change in the demand for the product.

In conclusion, for a firm operating under conditions of imperfect competition or monopoly, price and output are generally sensitive to changes in the variable cost determinants, but not directly responsive to changes in fixed cost.

Cost Reductions in an Oligopolistic Firm

As may be expected, price and output for a firm in an oligopolistic industry will respond much less readily to changes in cost, just as is true for changes in demand. Thus, in contrast to the situation dis-

cussed above, there will often be no change in price and output by firms in such an industry even when there is an alteration in their costs. Suppose, for instance, that wage rates are reduced, and in consequence firms A, B, C, and D each enjoy lower costs of production. Naturally each firm would be inclined to lower its price if it only could be reasonably sure that its competitors would not also lower theirs. Indeed, if they were all in a perfectly competitive industry, or if each firm could feel that its own actions would not influence those of its competitors, they would all reduce prices and increase production when costs fell. But with oligopoly, that freedom of action would rarely be expected. Firm A is reluctant to lower its price because it is fully aware that doing so would give the signal to B, C, and D to lower theirs. Thus, if any one firm reduces its price, the others are quite likely to follow — all too likely in the opinion of the firm that contemplates the move. Since retaliatory price-cutting is so probable, a reduction in costs resulting from a cut in wages will probably not persuade a firm to lower its price. And, unless the demand is changed, a failure to reduce price means no change in output. Thus, in an oligopoly, neither price nor output will probably be altered as a result of general cost reductions.

This conclusion is not certain, but at best only very probable. It is given added support, however, by the following consideration: that if the reduction in wage rates affects not only one firm but all the firms in the industry, it will be all the more likely that if one firm cuts prices the others will follow suit; or so at least each firm will tend to feel. And thus the gain to any one firm from reducing its price will be less. Expressed diagrammatically, the reduction in wage rates, if general to the industry, would not only reduce the marginal cost curve, but it would also reduce the elasticity of that part of the average revenue curve that relates to price declines. This is illustrated in Figure 44.

MC_1 is the marginal cost curve before the reduction in wage rates; MC_2 is the same curve afterwards. AR_1 is the average revenue curve initially, and AR_2, which is less elastic than AR_1 for prices below the current price P_1, is the average revenue curve after the *general* reduction in wage rates in the industry. Notice that the gap in MR_2 is greater than in MR_1. Output is maintained at O_1 and price at P_1. It is clear, then, that a reduction in cost would be even less likely to bring reduced prices if all the firms in the industry experienced it than if only one firm felt it.

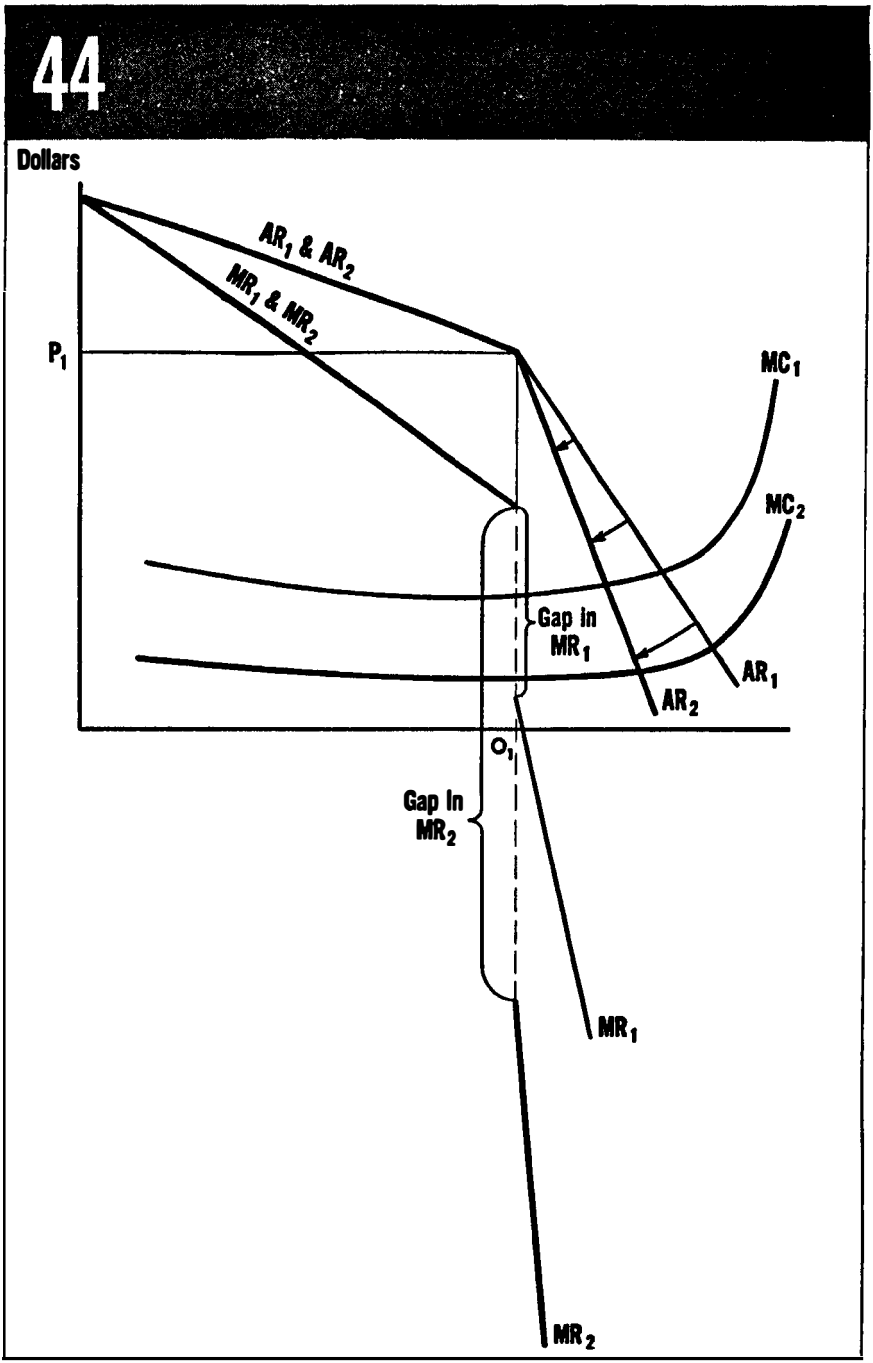

Figure 44. *Effect of Wage Reduction on Price in Oligopolistic Firm*

Cost Increases in the Oligopolistic Firm

The situation is somewhat different when wages or other costs are increased, for then it is much more probable that the firm will raise its price. In our discussion of demand and oligopoly, we saw that what deterred each firm from increasing its price was the fear that some of its competitors would hold back, and thus in effect undercut the firm that initiated the price increase. But if each firm knows its competitors also have had to pay higher wages, its fears on this score would be reduced. For example, if firm A knows that B, C, and D have also had to raise wages, he may suppose that they would be much more willing than formerly to raise their prices if he raises his. In these circumstances, an increase in wages or in other costs may lead to a general price increase, especially if the increase in costs is one that affects most of a group of related firms.

This amounts to saying that if most of the firms in an oligopolistic industry experience an increase in cost, each one of them will estimate that the elasticity of the average revenue function for price rises is lower than before the cost increase. Indeed, such a change in the evaluation of demand elasticity could be expected whether the affected costs were variable or fixed. The decreased elasticity of demand for price rises would make the gap in the marginal cost curve smaller, and thus would mean that a price increase was more probable. Thus, with oligopoly price increases may follow a general increase in cost, while price decreases are very unlikely after a general decrease in cost.

Summary

The effect of cost changes upon price and output is rather complicated. Except in an oligopolistic industry, we should expect price to vary in the same direction as cost, and output to vary in the opposite direction, provided marginal costs change. In an oligopoly, prices may not change when costs do. Indeed, it is quite unlikely that price will fall when costs are reduced, though they are much more likely to rise when costs increase. But even when costs rise, firms may not be willing to raise prices for fear that their competitors will not follow their lead.

An examination of the course of a large number of prices over the years shows that a significant proportion of them change only very rarely. This conclusion is, at first instance, surprising in view of the fact that demand and costs can be supposed to change frequently.

If firms all behaved in what we might call the conventional or un-oligopolistic way, this conclusion would be hard to explain. For we have already seen that price changes are to be expected when the cost or demand function of a non-oligopolistic firm changes. In the real world, however, oligopolistic considerations are an important factor, not only in price decisions but also in decisions on selling policy, wage policy, and other matters. Thus it is quite possible for demand and cost to change frequently and yet to produce no changes, or at any rate very few changes, in price. Thus the existence of oligopoly accounts for some of the price inflexibility that characterizes our economy.

17

Theory and Practice

ONCE, while this book was being written, the publishers asked for a rough estimate of its length, and within five seconds of hearing the estimate one of the editors said, "Well, let me see, that means we shall sell it for $\$X$." How could he determine the price so quickly? Can we suppose that he ground into some mental computing machine a marginal cost function based upon the length of the book, that he then ground in an estimate of the demand including the elasticity, and that the answer flashed back, "The most profitable price is $\$X$"? Do publishers, or for that matter other business firms, use the elegant apparatus we have constructed in the last few chapters? In other words, is business practice even remotely like theory? At first glance, it apparently is not, or so we might answer if we were to judge from this one incident alone.

But the answer to this question is a good deal more complicated than that. Not even the most confirmed armchair economist supposes that the manager of a firm actually prepares a chart on which he enters a marginal cost curve and a marginal revenue curve, and that he uses such a chart to determine his most profitable output. This does not imply, however, that the procedures described in the last few chapters give an unrealistic picture of what does happen. For the chances are that whether he thinks of them in these terms or not, the manager's calculations do for his immediate problem much the same thing that our analysis has attempted to do for similar but generalized problems.

Marginal Considerations in Business Practice

Thus, even without drawing marginal and average curves, it is obviously possible to weigh and balance all the considerations embodied in such curves before arriving at a decision about price and output. If, for example, the manager asks, "Would an expanded output which could be marketed only at a lower price be worth while?" his answer would almost necessarily involve the marginal concepts, whether he called them that or not. For if he decides that this expansion would not be profitable, he has actually decided that the additional cost needed to bring about the expansion will exceed the additional receipts. In other words, his answer means he has decided that the marginal cost of an output above the current level is higher than the marginal revenue. It would seem that if the question is cast in this form, the answer will inevitably be given along lines that could equally well be expressed in terms of marginal revenue and marginal cost.

The mere fact, then, that the businessman does not plot the marginal functions, does not mean that to use them in economic analysis akes us far from business practice. But there still remain other obections to consider.

Conventional Prices and the Marginal Analysis

For one thing, how shall we reconcile with the marginal analysis the fact that a few commodities always sell for the same price? We do not expect to pay seven cents for a chocolate bar or a package of gum. Such articles, and many others like them, sell at a standard or conventional price. Does this mean that the marginal revenue-marginal cost analysis is not applicable to such firms as Nestlé's or Wrigley's, or to the makers of penny matches and soft drinks? Probably not. For it is very likely that such a firm is usually unwilling to raise its price because it believes that the public would resent an increase in a price it had always taken for granted, and because the firm also believes that to provoke an adverse public reaction would mean lower profits. Now if we rephrase this thinking in terms of marginal analysis, we will say something like the following: If any departure from a conventional price would cause public resentment, then the demand would be very elastic for any price higher than the conventional price, and very inelastic for any price below the conventional price. So this kind of situation also falls easily into the pattern of the marginal analysis.

The Marginal Analysis and Sales Efforts

But there are other difficulties in applying the theory which are perhaps more important. One of these is that it does not seem to take into account the fact that the firm itself can influence the demand determinants; or at any rate we have not yet extended the analysis to cover this case. By putting on an extensive sales or advertising campaign, it can itself affect the demand for its product, as the leading cigarette and tooth paste manufacturers have proved again and again. The average revenue function is not then a constant, for if the firm can increase the demand for its product by spending more on advertising or by hiring more salesmen, it exercises some control over demand. Therefore in terms of the analysis, if the firm can do this, there is no unique or single intersection of the marginal cost and marginal revenue curves, since the position of the marginal revenue curve itself depends in part on the position of the marginal cost curve.

But this consideration, although it may complicate the task, does not mean that the marginal analysis is inapplicable. The management can still estimate profits at various levels of output, but it will also have to prepare each of these computations for different levels of advertising. And in choosing both the most profitable advertising level and the optimum amount to produce, it will have to deal with questions of this sort: "Will the spending of an additional $10,000 for selling be profitable or not? That is, will it add more to receipts than to costs?" and "Will the production of 10,000 more units be profitable or not?" As before, then, the marginal analysis applies to the problem whether it is formally worked out or not, even though the situation is a good deal more complex than those we have examined before.

Business Rules of Thumb

So far, then, we have seen no reason for rejecting the picture of business practice given by the marginal analysis. But now we must consider a common practice which is basically different from this analysis. If the manager of a firm is asked how he determines the price to charge, he very rarely answers that he estimates the profits to be realized at each price and then chooses the one which will be most profitable. Aside from such vague statements as "competition determines my price," the most usual answer is that he works out his average cost, or perhaps his variable cost per unit, and then adds a certain percentage for overhead and profit. For instance, the retailer

often claims to set his price at the wholesale price plus, let us say, 30 per cent. Price is so often said to equal variable cost per unit plus an allowance for overhead and profit that it is worth our while to examine this method of pricing with some care.

While it is probable that most business firms follow this practice and set their prices by estimating unit variable costs and then adding a certain percentage, it is impossible to describe the procedure more precisely because the percentage addition varies between one industry and another, and between periods of depression and prosperity. Thus, a firm in one industry may add 80 per cent in prosperity and only 60 per cent in depression, whereas a firm in a different industry may add only 50 per cent in prosperity and 30 per cent in depression. But whatever the mark-up, the point to be noted is that this procedure differs fundamentally from the marginal approach, for apparently no consideration is given to determining the most profitable output. But even though the mechanics of price determination are different in these two approaches, we may nevertheless get identical results. Hence we must determine whether a firm which follows this procedure will arrive at approximately the same price as it would set if it deliberately tried to maximize profits, calculating the optimum level by comparing marginal revenue and incremental cost; or whether the prices set by rule of thumb are radically different.

The Marginal Method versus Rule of Thumb

Before we examine this question it must be emphasized that if a price based on variable cost plus a fixed percentage differs from that set by the marginal method, the former would not be the price that promised the highest profits.[1] For undoubtedly the most profitable price is the one where the marginal cost and marginal revenue are equal. Any significant difference in the two prices would, of course, mean that the marginal method did not achieve the same results as the other method. But of equal significance, it would also show that business firms do not always charge the most profitable price. A divergence would imply that the analysis of the economists was faulty in that it was not describing business as it is. On the other hand, it would mean that business management was not doing its duty, which is to maximize profits. However, we can perhaps save face for both

[1] This means that the question really relates to the validity of our basic assumption that business firms choose the price and output that promises maximum profit.

economist and management if we can show that the price set by this rule of thumb is approximately the same as that set by using the marginal revenue and marginal cost functions. In that case, the conclusions derived from economic analysis would be correct, and at the same time management would be saving time in arriving at the optimum result by employing its convenient and simple rule of thumb.

And there is reason to suppose that these prices would be approximately the same. So that we may see why this is true, let us examine both methods of determining price, beginning with the marginal method. Following this approach, it is possible to demonstrate that the most profitable price is a simple function of marginal cost and the elasticity of the average revenue function, such that:

$$\begin{aligned} \text{if } p &= \text{the optimum price,} \\ m &= \text{the marginal cost,} \\ \text{and } e &= \text{the elasticity of the average revenue function,} \\ \text{then } p &= m\left(\frac{e}{e-1}\right). \end{aligned}$$

Hence if marginal cost and the elasticity of demand are known, the most profitable price can be easily determined. If, for instance, the elasticity is 5, the most profitable price, according to the formula, equals the marginal cost times 5/4, in other words, the marginal cost plus 25 per cent. If the elasticity of the demand is 2, the most profitable price equals the marginal cost plus 100 per cent, and so on.[1]

It will be recalled that for most firms the marginal cost curve is approximately horizontal over a fairly wide range of output. If the elasticity of the average revenue function is itself constant, it follows that the most profitable price is equal to the marginal cost plus a fixed percentage. If the marginal cost is \$2 no matter what the level of output within fairly wide limits, and if the elasticity of the demand for the product is 3, then the most profitable price is \$2 $\times$ 3/2 or \$2 + 50 per cent = \$3. Hence, within the range of output for which the marginal cost is approximately constant, the most profitable price

[1] It must be emphasized that the price, determined in this way, is simply the co-ordinate of the average revenue curve at the most profitable output — that is, at the output where marginal cost equals marginal revenue. In other words, price $= m\left(\frac{e}{e-1}\right)$ is the price represented in Figure 29.

equals the marginal cost plus a fixed percentage, given the elasticity of demand. At very high outputs, when we might expect the marginal cost to be higher than at normal outputs, the most profitable price would be higher than at other times.

We have also seen in Chapter 7 that the average *variable* cost curve is approximately horizontal over a wide range of output, and indeed corresponds with the marginal cost curve. In our example the unit variable cost would be $2 whether output was high or low. Hence if a certain percentage is added to the unit variable cost, again a price would be obtained which does not vary with output except at very large outputs. And if 50 per cent were added to the variable cost to determine price, it would be set at $3 a unit.

Thus, the price obtained by adding a certain percentage to the unit variable cost may actually coincide with the price determined by reference to the marginal functions. The exact percentage to add is of course critical, but experience gained though trial and error might in the long run insure that the choice was a good one. Thus, if a firm is normally faced with a demand of elasticity 4, the correct (most profitable) amount to add to the average variable cost would be 33 1/3 per cent, since marginal and average variable cost are approximately equal at normal outputs. Furthermore, the usual policy of business firms of adding a higher percentage at times of very high demand also tends to set the price at the most profitable level, since marginal cost rises more steeply for expansions in output as capacity is approached than does average variable cost. For comparison, we could set down in tabular form the results of the two methods of determining price, as in Table 27.

It will be seen that (*a*) if the allowance for overhead and profit is correctly computed, the firm can arrive at the same price as it would obtain by using the marginal analysis; and (*b*) the percentage allowance to be made at very high outputs must be greater than the allowance made normally. If the firm in this situation adds 25 per cent to its unit variable cost at normal times, and raises the percentage allowance to 45 per cent when its output is in the range in which marginal and average variable costs are rising, it will be seen that its price computed by rule of thumb corresponds almost exactly with the most profitable price.

To say that by following a simple rule for price determination the firm *can* get the right answer is not, of course, to say that firms always

TABLE 27

Comparative Results of Price Determination by Marginal Method and by Rule of Thumb

When Output Is	Results of Using Marginal Method if Elasticity Is 5		Results of Using Rule of Thumb		
	Marginal Cost	Most Profitable Price	Variable Cost	Percentage Allowance for Overhead and Profit	Computed Price
10	100	125	100	25%	125
15	100	125	100	25%	125
20	100	125	100	25%	125
25	100	125	100	25%	125
30	100	125	100	25%	125
35	100	125	100	25%	125
40	100	125	100	25%	125
45	100	125	100	25%	125
50	110	137.50	105	30%	136.50
55	125	156.25	115	35%	155.25
60	150	187.50	130	45%	188.50

do get the right answer. Where they should add, let us say, 40 per cent to their *average* variable cost, they may add 80 per cent or 20 per cent. Thus, it is not impossible that the economist gives an unrealistic picture and the firm manager chooses a price that is not the most profitable. But the economist can at least check his theory against reality to this extent: many business firms have been in existence for fifty to a hundred years, and the very fact of their survival argues that they have developed a rule for price setting which is not very far from the right one. If they do follow a rule that gives the correct answer, then the economists' analysis, while it may seem abstract and unrealistic, gives us a reasonably accurate picture of the results the businessman reaches, even if it presents an inaccurate picture of the procedure by which these results are obtained.

Our final conclusion about the two methods of price fixing may therefore be stated in some such way as the following. Because marginal and average variable costs are likely to be very nearly the same, it is possible to formulate a simple rule for determining the most profit-

able price. This rule of thumb is that the price should equal variable cost plus a certain percentage which is fixed except at times when demand is very high — then it should be increased. Business firms frequently claim to employ such a method of price determination, and as we have seen, provided that the percentage to be added is carefully chosen, the price thus set would be very close to the most profitable one.

Why Firms Use the Rule of Thumb

At best the rule of thumb can give an answer no better than the marginal analysis gives, and if the percentage addition is badly chosen, the answer will be much worse — that is to say, much less profitable. Why, then, is this method so commonly used? Presumably because it is easy, certainly much easier than the more complicated marginal analysis procedure, especially for firms that produce a wide variety of products. If it is also noted that even a careful estimate of demand, its elasticity, and marginal cost, when commodities of several kinds are involved, cannot command much confidence, then it is not surprising that the simple rule of thumb for price determination commends itself to the businessman.

Summary

At the beginning of this chapter, we asked whether economic theory gave a reasonably accurate account of business practice. On the whole we should now answer, with certain qualifications, that it does. While the economist does not attempt to duplicate the mechanics of price determination as employed by the business firm, there are grounds for believing that his conclusions as to how price is affected by changes in economic conditions are reasonably close to what actually happens. Moreover, the fact that most businessmen have never heard of marginal revenue and that they follow highly arbitrary rules in determining price does not mean that the methods of analysis which economists employ are unrealistic. These methods may, and we have reason to believe that they do, reach roughly the same goal the businessman reaches, though they may get there by a different route. In deliberately attempting to choose the most profitable output, and setting a price which makes it possible to sell that output, the businessman essentially follows marginal procedures, though he may not call his own method by so technical a name. If instead he uses a

simple rule of thumb to set price, as many obviously do, then again, though his methods of arriving at his decision are different, he may reach the very same result as would be obtained by following the marginal analysis.[1] What this means is that the economist who employs the elaborate technique illustrated in the three preceding chapters can reasonably expect that his description of the output and price set by a firm in given conditions of demand and cost will be realistic. Moreover, the response of the firm to a change in these conditions will, in the real world, be substantially the same as we should get if all price and output decisions were made with a view to maximizing profit (which implies that marginal procedures are employed). Hence to answer the question asked at the beginning of the chapter, economic theory gives a good description of business practice.

[1] It must be emphasized that if the firm reaches a price which is markedly different, as it would do if its percentage addition to cover overhead and profit was improperly chosen, then it would fail to make maximum profits.

18

Price and Output: The Industry

IN THE LAST FEW CHAPTERS we have concentrated nearly our whole attention on the firm and its problems — its costs, the demand for its product, and the way it determines its output and sets its prices. But this, as we can see, is far from being the whole of economics. In fact, as economists we are not primarily interested in what the single firm does, although an understanding of its major problems and the answers to them is an essential first step toward an understanding of the industry to which the firm belongs and finally of the economy as a whole.

Nevertheless, in the subsequent stages of our analysis we can, for a while at least, continue to make good use of the methods developed in the last few chapters, although we must elaborate and extend this analysis far beyond the present point before we can claim to understand, even imperfectly, how the whole economy functions. We have, for instance, developed the technique for finding out how much lumber a single mill will saw in given circumstances. But we have not yet investigated the factors that determine how much lumber is produced in the economy, and how changes in economic conditions can be expected to affect that output. We must therefore in the next two chapters extend our analysis of the firm and apply these results so far as we can to the *industry* — to the wheat industry, or the lumber industry, or the steel industry — as a whole. In other words, we must now answer questions for the industry similar to those we have already answered for the firm. We must determine, for example, what happens to the output of the industry when demand changes,

and also whether there is a difference between the immediate or short-run results and the eventual results of such a change. Or we can analyze the short-run and long-run effects on price of the development of cheaper sources of energy or of improved methods of production. These are samples of the problems to which we must now direct the analysis.

Here it must be noted that this change in emphasis from the firm to the industry does not break the continuity of our discussion. What happens to the industry is precisely the sum of what happens to the firms that make it up. For after all, an industry is merely a group of firms — anywhere from one to 4,000,000 — and it is the firms, not the industry, which determine price and output. In a sense, therefore, what we have to say about the industry is much the same as what we said about the firm, except that we now have a multiplication or an addition symbol to take into account.

It follows that if we want to know the effects of an increase in demand upon the output of an industry, we have to determine how such an increase affects the output of each of its constituent firms, and then to add the results. Therefore the question of how total output is set for the paper box industry or the publishing industry takes us back to the question already analyzed — how the individual firm in such an industry determines its output. Thus we shall build our analysis of the industry upon that of the firm.

The Output of a Perfectly Competitive Industry

Let us consider first a situation in which there is perfect competition. You will remember that competition is described as perfect when the number of firms producing a commodity is very large and, furthermore, when the buyers are quite indifferent to what firm they buy from, so long as its price is not above the price charged by others. As we have seen, every firm produces that output at which marginal cost equals marginal revenue; and since marginal and average revenue are, in this situation, the same, each firm produces the output at which marginal cost equals price. We may, therefore, ascertain the total output that the industry provides at any given price. It is equal to the *sum* of what each firm in the industry is prepared to produce at that price. In Figure 45 we have illustrated in the left-hand chart the situation for one firm in the industry, indicating the output O_1, which it will produce at the price P_1. The right-hand chart shows the situation

for the industry. The point X denotes the fact that all the firms in the industry will together produce M_1 at the price P_1. The figure M_1 is the sum of the outputs corresponding to O_1, made available by each of the many firms in that industry. The scale on the horizontal axis of the industry chart is of course very much greater than that used for the firm chart.

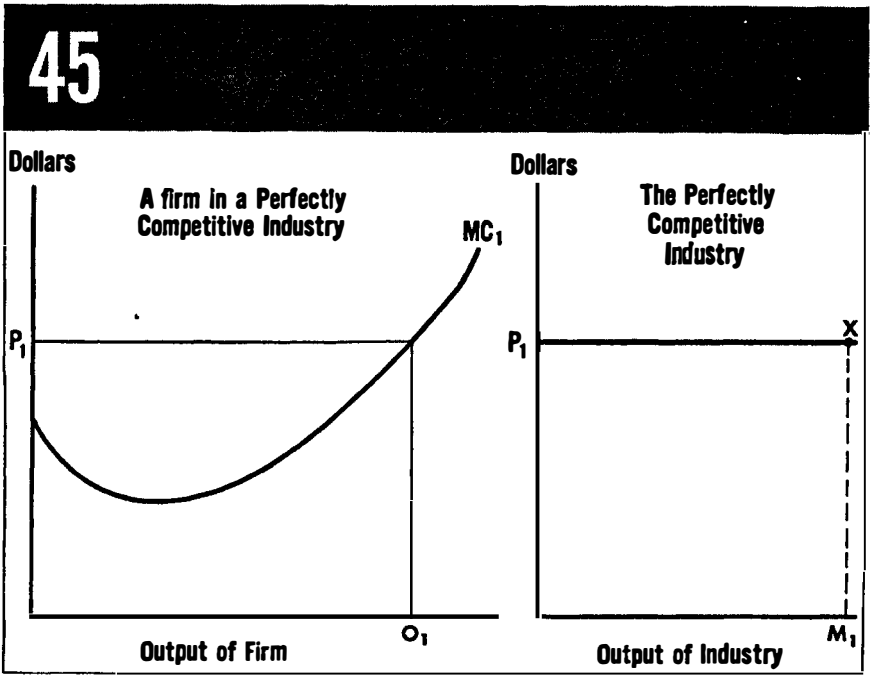

Figure 45. *The Output of a Firm and the Industry (Perfect Competition) at One Price*

Output When Price Changes

If the price were raised, to P_2 for instance, the output which the firm would be willing to produce would also increase to O_2. Hence the industry, composed of perhaps hundreds of thousands of firms, each of which would react in roughly the same way to the price increase, will provide at P_2, a larger amount, M_2. The point Y on Figure 46 illustrates this relation. And if the price were reduced from P_1 to P_3, we could conclude that, for reasons already set out, the industry would produce a smaller amount, M_3, as indicated by the point Z.

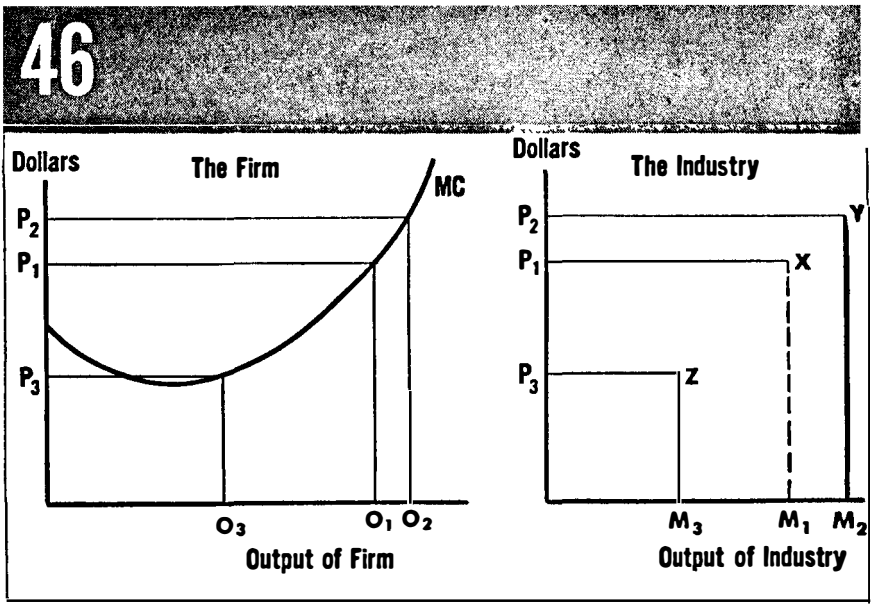

Figure 46. *Output of Perfectly Competitive Firm and Industry: Various Prices*

The Supply Schedule of a Perfectly Competitive Industry

We may draw up a complete schedule that relates each possible price to the corresponding output of the industry. This schedule or curve would be the locus of such points as X, Y, and Z in Figure 46. Other points could be determined in exactly the same way as X, Y, and Z. In general, at a low price the output of the industry will be small, and at a higher price the output will be greater. Finally, we can suppose that there is a price at which each firm in the industry will produce at its maximum capacity, and hence that at this price the industry itself will provide as large an output as it can. Obviously, then, the output of the industry can be expanded no further however great the rise in price beyond this point. Figure 47 illustrates this situation.

In the diagram, the curve SS_1 denotes the amount that the industry would supply at each possible price. At a very low price, P_3, the output of the industry would be very low, M_3. If the price, instead of being P_3, were somewhat higher, say P_1, output would be M_1, which is of course greater than M_3. Finally when the price is P_4, the output of the industry is at its maximum, and any further increase in price,

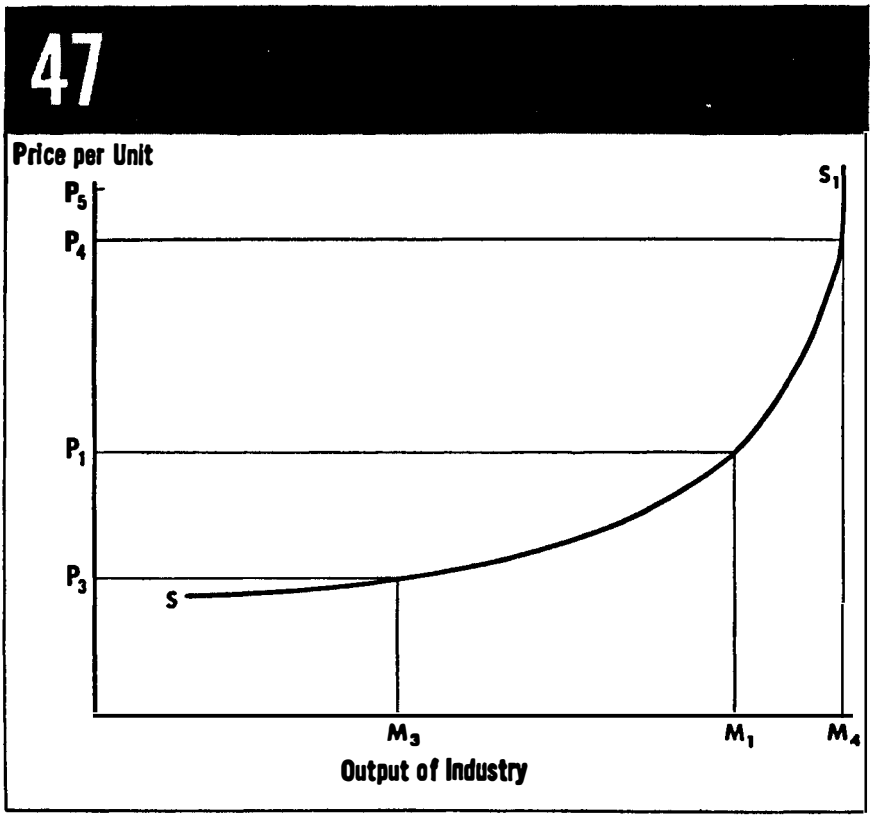

Figure 47. *The Supply Curve*

to P_5 for example, will have no further effect upon output. This relation and the curve which illustrates it are known as the *supply function* and the *supply curve*. These terms are generally used only with reference to the total output of an industry or, occasionally, of the whole economy. They are not ordinarily used in discussing a single firm.

The Supply Curve and the Marginal Cost Function

As we have seen, the supply function for a perfectly competitive industry shows that, up to a point, the amount supplied increases as the price rises. This rise in the amount supplied occurs because most or all of the firms in the industry find it profitable to produce more with every increase in price. The total increase in output, for example from M_3 to M_1 as the price is raised from P_3 to P_1, is simply the sum

of the increases in the output of each firm. But we have already seen that whether a firm increases its output by much or by little when the demand, or in this case the price, rises, depends upon the shape of the marginal cost curve. Thus, if the marginal cost curve rises very steeply when the price increases from P_3 to P_1, the firm will expand its output by only a little. If, on the other hand, the marginal cost curve is rather flat, the same price rise will bring about a much greater increase in output. Therefore the firm's reaction to a change in price is closely geared to the nature of its marginal cost function.

Now, since the supply curve for the industry simply combines the experience of all its constituent firms, this curve will reflect the marginal costs of all the firms. If each of them had exactly the same marginal cost function, the supply curve for the whole industry would mirror this marginal cost function with, of course, an output scale multiplied many times over compared with that for the firm.

As is shown in Figure 47, at low levels of price and output, the supply curve will be nearly horizontal, since each firm's marginal cost curve is nearly horizontal over a wide range of output up to near capacity production levels. As price increases and the output of the industry mounts, the supply curve also slopes upward more and more steeply, since the marginal cost curves for the constituent firms become nearly vertical as capacity output is approached. And finally, when each firm in the industry is producing at peak capacity and no possible rise in price can further increase the output, the supply curve for the industry as a whole becomes vertical, or perfectly inelastic.

Factors upon Which the Supply Curve is Based

The same care must be used in interpreting and using the supply curve as is necessary in using the demand curve. In the first place, the supply curve assumes a *given marginal cost function* for each firm in the industry. If the marginal cost function for any firm were to change, as would happen, for instance, if wage rates were altered, the supply curve built on the original marginal cost functions would not portray the new situation. In the second place, the supply function or curve is based on a *specified number of firms* in the industry at the time the curve is drawn. Hence if firms leave or enter the industry, the old supply function or curve will no longer apply. But as long as these conditions are satisfied — (*a*) that the cost conditions in each firm do not change, and (*b*) that the number of firms remains constant,

a supply curve provides important information. Briefly, it enables us to say exactly how much would be produced in that industry at any particular price. It is, therefore, rather similar conceptually to the demand function discussed earlier. But instead of dealing with the question of how much would be *purchased* at various prices, it tells how much would be *offered* or *made available for purchase* at various prices; hence, while the demand curve describes the reactions of buyers, the supply curve describes the reactions of sellers. And just as a demand curve can be drawn up only when the demand determinants are known, so a supply curve must be based upon known supply determinants. Thus one billion bushels of wheat may be forthcoming from the wheat industry when the price is $1.75 a bushel; but we should expect this relation between price and output to hold only so long as the number of wheat farmers remains about the same or the costs of production do not alter considerably. Any change in these determinants will bring about a change in supply; that is, a change in the amount supplied at any particular price.

The Effect of a Change in Costs on Supply

When there is a change in the marginal cost function of some firms in an industry, the amounts which these particular firms are willing to produce at a given price will of course change in consequence. As a result, the output which the industry as a whole would make available at that price will also change. If, for example, marginal costs in the firm are raised, it would pay the firm, as we saw earlier, to reduce its output at each price. The industry of which that firm is a part would consequently make less available than formerly at each price. Thus, if a large number of firms are affected by an increase in costs, the amount supplied by the industry as a whole will be greatly reduced. To interpret graphically this decline in the amount supplied, we shall have to draw a new supply curve, S_2S_3, as shown in Figure 48, which is above and to the left of the original curve. This new curve indicates that at any particular price, such as P_1, the amount which the industry now provides, M_3, is smaller than the amount it was willing to supply at the same price before marginal cost had risen. Such a change in the supply curve is known as a decrease in supply.

"Supply" and "Amount Supplied"

There is a distinction in meaning, in the language of economists,

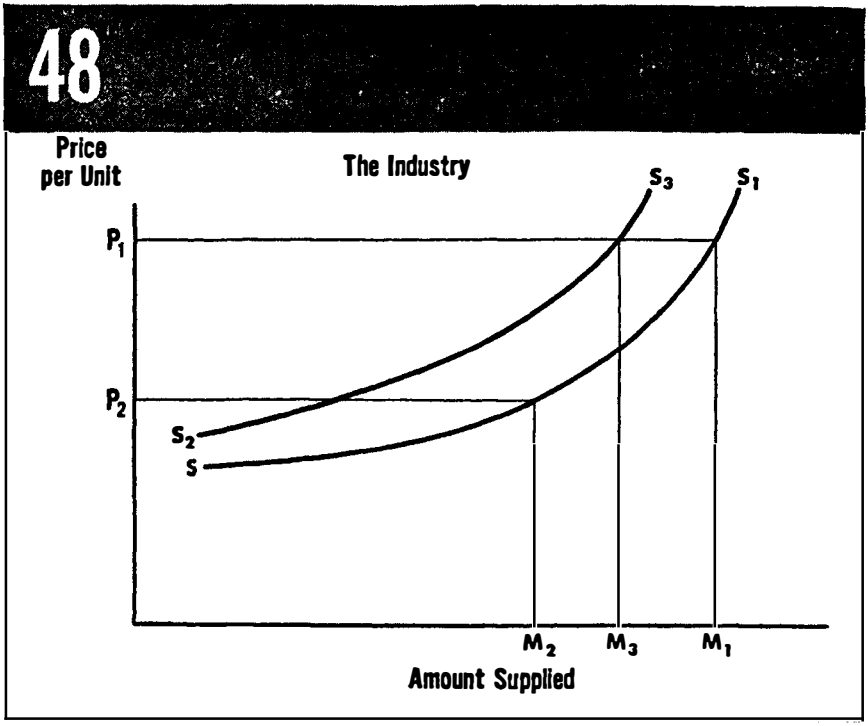

Figure 48. *A Reduction in Supply*

between the terms *supply* and *amount supplied* which closely parallels the difference between the terms *demand* and *amount demanded*. A decrease in *supply* refers to a change which occurs when the supply curve is altered so that at any particular price the amount supplied is reduced; in contrast, a reduction in the *amount supplied* is the change which occurs as the result of a change in price. Figure 48 illustrates a change, an increase as it happens, in the amount supplied, from M_2 to M_1 because of a change in price from P_2 to P_1 while the supply remains the same, SS_1. A change in supply is illustrated by the shift from SS_1 to S_2S_3. Strictly, then, it is inaccurate to say that "increasing prices will lead to increases in supply." Instead we should say that an increase in price leads to an increase in the amount supplied. In the language of the economist, supply changes only when the supply determinants change; as long as they are constant, a change in price brings about only a change in the amount supplied.

Price Determination in a Perfectly Competitive Industry

As we have seen, if we know the cost situation in the individual firms, and if we know the number of firms in an industry, we can draw a supply curve which shows the relation between the price offered and the amount supplied by the industry. We have also seen that the demand for the product of an industry shows how much is demanded at various prices. With these concepts as our tools, we are now ready to deal with the problem of price determination in a perfectly competitive industry.

Let us begin by plotting on a single graph both the demand and supply functions for a perfectly competitive industry. The demand curve shows that at price P_1, A_1 units are demanded; the supply curve

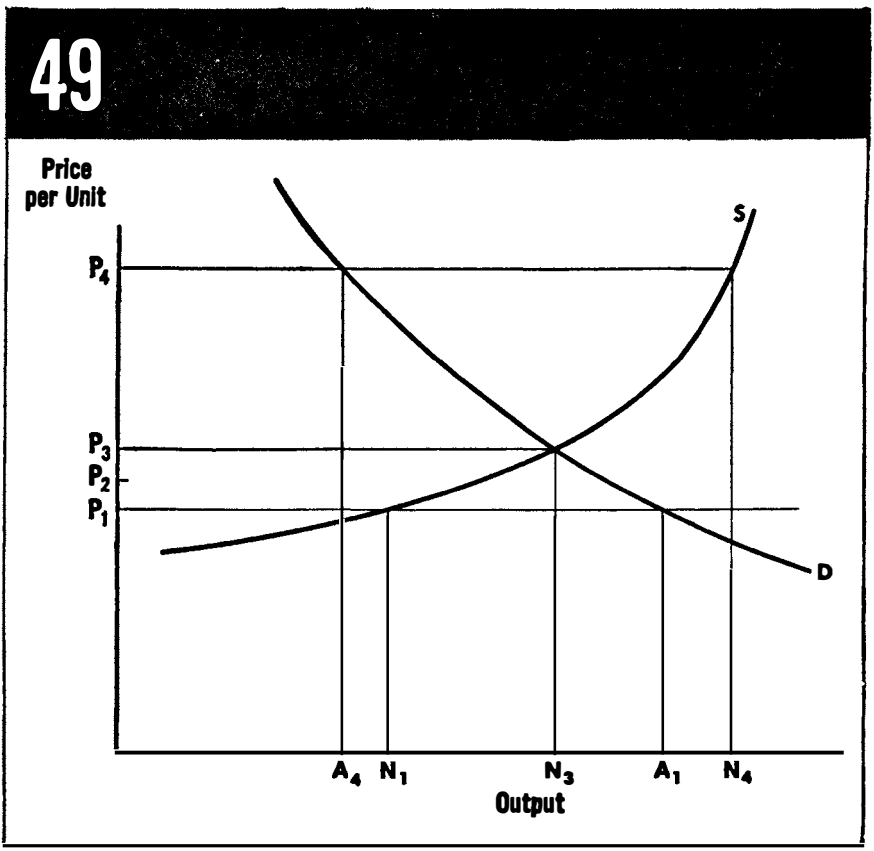

Figure 49. *Price Determination: Perfect Competition*

shows that at that price the industry is willing to supply N_1 units. It is obvious that A_1 exceeds N_1. That is, at price P_1, buyers are anxious to purchase a very much larger amount than suppliers are willing to sell. The question then arises, can the price remain at that figure? Hardly, for as buyers come to realize that they cannot purchase all they want at P_1, some show themselves willing to offer a higher price in order to satisfy their needs, and the price is bid up — perhaps to P_2. But even at this price, the amount demanded exceeds the amount the industry is willing to supply, and the pressure to raise prices is therefore maintained. But notice that the pressure has diminished in strength. Some who wanted to buy at P_1 will no longer be in the market, and others will reduce their requirements because of the higher price. At the same time the industry will increase its output because of the price rise. The size of the increase in output is shown by the shape of the supply curve. These two reactions will reduce the upward pressure on prices. Yet this pressure will exist so long as there is any positive difference between the amount which buyers wish to purchase and that which producers wish to sell.

When the amount supplied exceeds the amount demanded, the situation is exactly the reverse of what we have just seen. Thus, if the price were not as low as P_1, but very much higher, for example P_4, buyers would be willing to purchase only A_4, which is not nearly as much as sellers are willing to provide — that is, N_4. At the price P_4, a producer's sales are far below what he wants to sell, but he knows that if he cuts his price by only a small amount he can expand his sales just as far as he wishes, since customers buy where the price is least. Hence he has every reason, in this case, to lower his price somewhat, and so, as he well knows, have all his competitors. The pressure on prices would therefore be downward, and would continue so long as there was any positive difference between the amount offered for sale and the amount buyers wished to purchase at the current price. As is now clear, the situation will be stable only at the point where the demand and supply curves intersect. At the price P_3, the amount offered for sale is just equal to that which buyers are willing to purchase. It is in this sense that demand and supply determine price. No seller finds that his sales fall short of what he decides to sell at P_3, nor does any buyer find that he is not able to get all that he wants at that price. But this does not mean that either buyers or sellers are necessarily satisfied or delighted by the situation.

If the price is very low, the sellers will be dissatisfied; but they will nevertheless be able to dispose of all they wish to sell *at that price.* If the price is very high, the buyers will be discontented, but all are able to buy as much as they wish *at that price.* Yet, even though one group or the other may be dissatisfied with the situation, there will be no pressure to alter it.

The Equilibrium Price and the Ideal Price

Therefore, the *equilibrium price*, as we may call it, does not imply an ideal situation from either the buyer's or the seller's point of view. In December, 1932, the price of wheat was 49 cents a bushel, but it is doubtful that anyone would argue that this price was ideal. In 1945, wheat was $1.70 a bushel, and there is no reason to suppose that this price was perfect either. All that the equilibrium price does indicate is that in a perfectly competitive situation, such a price would be reached and maintained so long as there is no change in demand or supply.[1] The intersection of the demand and the supply curve, therefore, shows where the price is stable. It does not necessarily show where the price is perfect, or fair, or anything else.

A confusing belief prevails that the laws of supply and demand somehow indicate what *ought* to be done. We frequently see in the writings of people who enjoy airing their knowledge of economics that "we must not break the law of supply and demand." Perhaps the kindest thing to say about such statements is that they are meaningless. The law of supply and demand is only concerned with the question of where price settles under stated conditions of cost and market. To say that one should not break it is rather like urging an aviator to stay on the ground and not break the law of gravity. But the law of gravity does not make it illegal or foolish to fly, and sometimes it is convenient to travel that way. To apply this to the problem in hand: Up to this stage in our analysis, we have seen no reason why it is not expedient to interfere with the free workings of a perfectly competitive industry and to set prices at a level different from that which the law of supply and demand dictates, but which would be more acceptable to buyers or sellers. Yet, for that matter, neither have we seen any reason why it is desirable to set prices at a different figure. The point to be observed at this stage in our analysis is that the law of

[1] And so long as the long-period forces to be discussed in the next chapter do not come into play; forces which, incidentally, change the supply.

supply and demand merely describes what happens. It is not an economic decalogue or a moral code.

Effect of Changes in Demand and Supply

Changes in either demand or supply will, of course, bring about a new equilibrium price. If, for example, demand increases and more is demanded at the original equilibrium price than sellers are willing to put on the market, pressure will develop to raise the price and will be maintained until the price has increased to the point where the amount demanded is once again equal to the amount supplied. If demand falls, the equilibrium price will fall too. At the original price, the amount demanded, after the decline in demand, will be less than the amount supplied. Sellers, under pressure to sell more, will begin to cut prices and will continue to do so until once again a price is reached where the amount demanded equals the amount supplied. These reactions are illustrated in Figure 50.

The supply curve is shown by SS_1. The demand before the increase, or after the decrease, is denoted by D_1, and that after the increase, or before the decrease, is denoted by D_2. The equilibrium price before the demand has increased, or after it has fallen, is P_1. After the increase in demand or before the decrease, the equilibrium price is P_2. The change in price from P_1 to P_2, or vice versa, depends on the size of the change in demand and also upon the elasticity of the supply curve. If the supply function had been more elastic, that is, if the slope of the supply curve had been less steep, the change in price would have been smaller.[1]

So much, then, for what happens when demand changes. Now let us examine the results of changes in supply. When the supply increases, the amount which sellers want to dispose of at the original price exceeds the amount which buyers are willing to take, and as a result prices tend to fall. Conversely, when supply decreases, price rises. The effect of a change in supply is illustrated in Figure 51. It should be noticed that the less elastic the demand, the greater the change in price for a given change in supply. An analogous result was seen in connection with a change in demand; where the lower the elasticity of supply, the greater was the change in price following a given change in demand.

[1] The dotted supply curve S_2 is more elastic; notice that price would alter by much less when the supply curve is S_2 than when it is S_1.

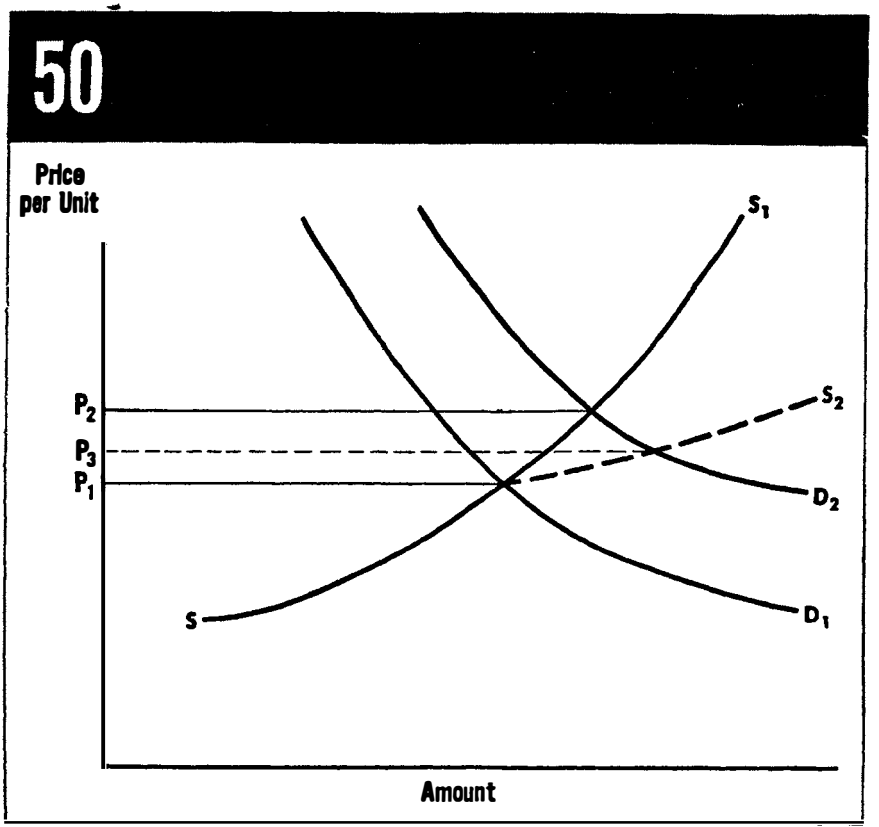

Figure 50. *Effect of Change in Demand on Price*

We may conclude that when competition is perfect, an increase in demand — supply remaining constant — brings higher prices, whereas an increase in supply — demand remaining constant — brings lower prices.

Now let us see how these conclusions about price fit into the picture of the firm in a perfectly competitive industry. If we recall that the supply curve for the industry reflects the marginal cost curve of the constituent firms, we can illustrate the situation by preparing two diagrams, one showing the cost situation for a typical firm in the industry, and the other the demand for the product of the industry as a whole. These are shown in Figure 52. The original supply curve, S_1, is based upon the marginal cost functions (one of which is denoted

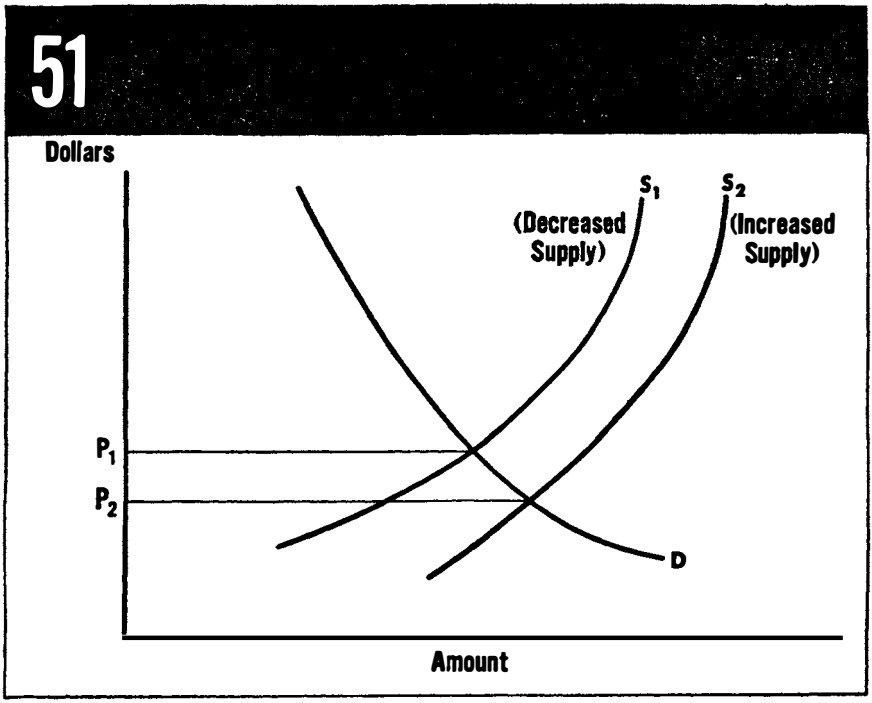

Figure 51. *Effect of Change in Supply on Price*

by MC_1), of the firms in the industry. Given the demand curve D_1, the equilibrium price will be P_3, at which level the firm is willing to produce M_3 units of output. M_3 is, of course, only a very small part of the total output of the industry, which is measured by R_3.

Now let us trace out the effects on the firm and the industry of a change in one of the economic variables — say in marginal cost. If the marginal cost function in each of the firms should fall, as it would if, for instance, an excise tax on the product were reduced, each firm would produce more than it did before, at any particular price. At P_3 the firm would raise its output from M_3 to N_3, since MC_2 and the price intersect at a higher output, and other firms would react in the same way. Hence the industry will put a larger amount on the market than it did before, when the price was P_3. That is, the total output will now be T_3 rather than R_3, as the diagram shows. Moreover, to illustrate this situation it will be necessary to draw a new supply curve, since it will be necessary to represent an increased supply. But with

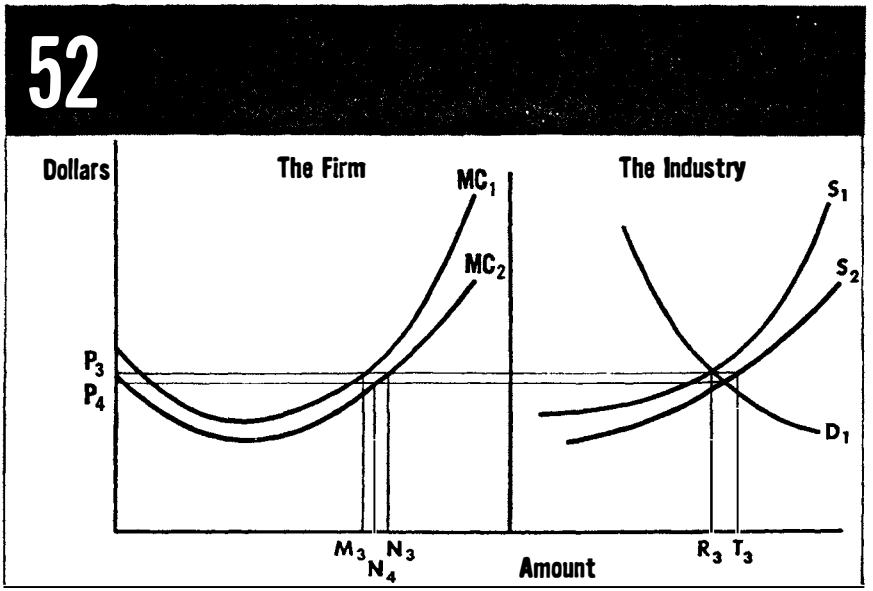

Figure 52. *Price Determination; Firm in Perfectly Competitive Industry*

the increased supply and the original demand, the price will not remain at P_3, for at that price sellers would be willing to put more on the market than buyers would take, and consequently price would be cut. In fact, price will eventually be stabilized at P_4, where the original demand curve and the new supply curve intersect. And at this price the firm will produce N_4 units.

The effect of a change in demand on the output of the firm and of the industry could be similarly analyzed. In brief, a change in either demand or supply will tend to bring about a change in the equilibrium price in an industry where perfect competition obtains.

Supply for an Industry Not Perfectly Competitive

Applied to an industry such as we have been discussing, the supply concept is clear-cut and therefore valuable. Unfortunately, it is somewhat confusing and therefore less useful when applied to the product of an industry which is not perfectly competitive, and hence it is rarely used for such industries. Let us see why this is true. In the first place, with perfect competition, there is no problem in defining the product of the industry, since the output of each producer is indis-

tinguishable from that of every other. But with an imperfectly competitive [1] industry there are likely to be marked differences between the products of the constituent firms. We shall discuss the significance of this fact below. The second difficulty that arises in applying the supply concept to an industry which is not perfectly competitive, is that the demand for the products of the various firms is not uniformly elastic. Under perfect competition, the supply function depends simply upon the marginal costs of the individual firms, as we have already seen; but under imperfect competition, this function depends also upon the elasticity of the average revenue function for each of the firms. It is the presence of this additional factor which causes the trouble. We shall now consider more carefully these two objections to applying the supply concept when competition is not perfect.

In discussing the product of a perfectly competitive industry, then, we are dealing with a uniform, standardized commodity, and there is no problem at all in adding the outputs of the various firms in order to secure the total output of the industry. But the product of a monopolistically or imperfectly competitive industry is not standardized. Each firm produces something which, at least in the minds of the buyers, differs from the product of every other firm in the industry. Chesterfield cigarettes differ profoundly from Camels — or at least so the advertisements proclaim. The owner of a Packard motor car knows just why his car is superior — or thinks he does. Sometimes the differences between such commodities are minor, sometimes considerable. The problem in this case is to determine just how these somewhat different items should be grouped. In dealing with the output of General Motors, Chrysler, Ford, and so on, should we lump all types together in the general category of automobiles, or should we group together only the low-priced cars, or should we include only low-priced two-door sedans? Since the products of the various firms differ among themselves, there are no clear-cut criteria to determine how many of these heterogeneous products should be lumped together and called a *commodity*. Lacking any precise determination of how much or how little the term *commodity* should include, we are unable to determine the appropriate supply curve until we make an arbitrary decision.

Furthermore, the well-known difficulty of adding two plums, three

[1] The reader is reminded that *imperfect competition* and *monopolistic competition* are used interchangeably.

apples, and four peaches arises at this point. If we cannot add apples and peaches, should we add Chryslers, Lincolns, and Studebakers? And if we do, is the total obtained a significant figure? This difficulty is not of great importance, however, for most problems. There is certainly more difference between an automobile and a motorcycle than between a Ford and a Chrysler. In practice there are clearly marked boundaries between any one product and the one most like it. Thus if the figure for automobile output is occasionally too crude, it is generally a perfectly acceptable measure.

The other objection to the use of the supply concept when competition is imperfect is, however, more substantial. The amount that a firm in an imperfectly competitive industry makes available depends upon its marginal costs and also upon the elasticity of the demand for its product. We can derive from a formula given in the last chapter [1] the generalization that the more elastic the average revenue, the lower the price for any given level of marginal cost; and, of course, the lower the elasticity, the higher is the price. Using this formula we can secure the following results:

When the elasticity of the average revenue function is:	3	4	5
If Output is.	10	10	10
Marginal cost is assumed to be:	10	10	10
And the price is consequently:	15 (= 10 × 3/2)	13.3 (= 10 × 4/3)	12.5 (= 10 × 5/4)
Or alternatively:			
If Output is:	10	12	15
Marginal cost is assumed to be:	10	11.25	12
The price is consequently:	15 (= 10 × 3/2)	15 (= 11.25 × 4/3)	15 (= 12 × 5/4)

[1] Price equals (marginal cost) $\times \left(\frac{e}{e-1}\right)$ where e is the elasticity of the average revenue curve. This formula is based upon two rules: first, that profit is maximized at the output where marginal cost and marginal revenue are equal, and secondly, that from the definition of elasticity we can deduce that the price at any output equals the marginal revenue times $\left(\frac{e}{e-1}\right)$.

The second formulation brings out most clearly the difficulty involved. The values chosen for marginal costs at these various outputs might well have been points on a single marginal cost curve. Yet if the elasticity of the average revenue function is 3, the output made available at a price of 15 is only 10; if the elasticity is 4, the output provided at that price is 12; and if the elasticity is 5, the output is even larger, or 15. Thus the amount produced at a certain price, say 15, depends not only upon the marginal cost function but also upon the elasticity of the demand.

This point is shown diagrammatically in Figure 53. The average revenue function for firm A is quite elastic, while that for firm B is less elastic. The marginal cost curve is the same for firms A and B. Firm A would produce X units at P_1, while B, as shown, would produce only Y units at P_1. Although the price and the marginal cost

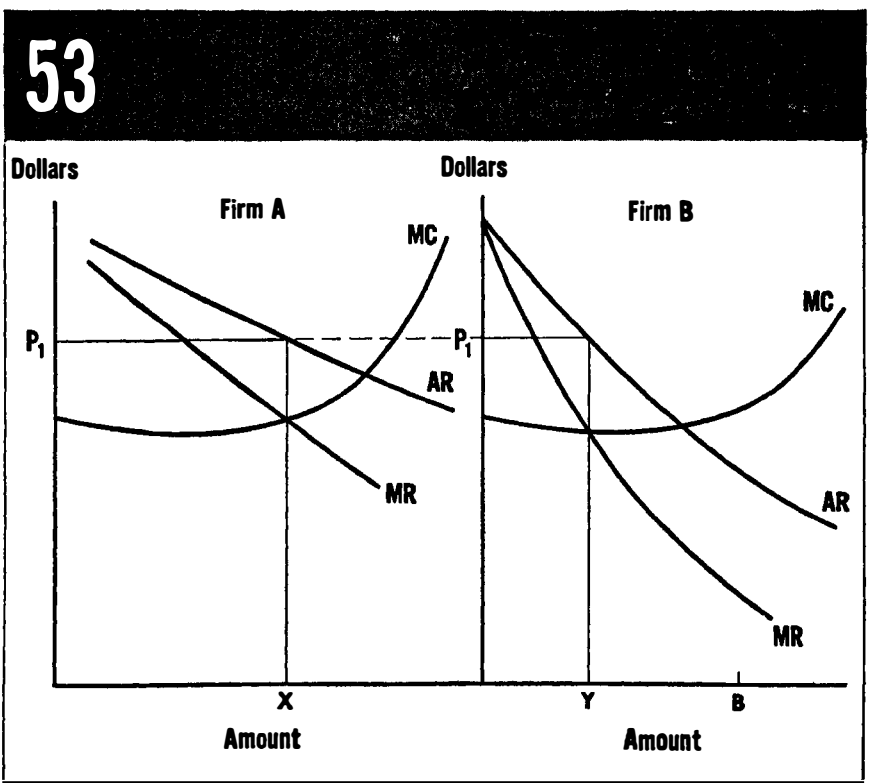

Figure 53. *The Elasticity of Demand and the Output of a Firm*

function are the same for the two firms, one firm would produce much more than the other.

This illustrates the fact that with imperfect competition, there is not a unique relation between price and the volume of output, as there is with perfect competition. The output of a firm depends not only on its price and marginal cost function, but also on the elasticity of the demand for its product. The supply function of a firm, therefore, must include this factor also. Hence if we wish to use this concept in a monopolistically competitive industry, we must be sure to base it on this additional determinant — the elasticity of demand. In a perfectly competitive industry we can overlook this determinant entirely, for with perfect competition, the elasticity of the average revenue curve for each firm is, by definition, infinite. But with monopolistic competition the elasticity can take any value less than infinity; and what is more important, a change in elasticity will bring about a change in the supply function.

As we have said, because of this added complexity, the supply concept is rarely used for monopolistic competition, nor is it needed in determining price. However, it is a convenient concept to use in dealing with questions of employment and certain problems of public policy towards monopoly. But in using it we must be careful to remember, in the first place, that it requires precise and arbitrary definition; and in the second place, that it depends not only on marginal costs, but also upon the elasticity of demand, or as we may call it, the degree of monopoly in the market in which the firms are selling.

Summary

The supply of the product of a perfectly competitive industry depends on the number of firms of which it is composed and on marginal costs in these firms. Any change in either of these factors affects the supply of the commodity. If for instance marginal costs decline, the supply increases, and so on. The price of the product of a perfectly competitive industry depends on demand and supply, and will stabilize at the level at which the amount demanded and the amount supplied are equal. Any change in demand or supply of course brings about a change in price.

The supply concept is most conveniently applied to a perfectly competitive industry. Because firms in such an industry are confronted by a perfectly elastic demand, we may overlook the depend-

ence of the supply function on elasticity. But when the industry is not perfectly competitive, the elasticity of the demand facing each firm can take any value less than infinity, and the supply function depends on the measure of elasticity. Because a change in the elasticity of demand affects the supply, the latter concept is not an especially useful one when the industry is not perfectly competitive. There is a further objection to using it, that the products of the firm that make up such an industry are not identical. We shall therefore apply the law of supply and demand only when the industry is perfectly competitive.

19

Long-Period Adjustments in Price and Output

UP TO THIS POINT we have been concerned with economic adjustments which take place over relatively short periods of time. Thus, when we inquired into the effects of an increase in demand on price and output, because we were dealing with fairly quick adjustments, we deliberately ruled out certain phenomena which would affect these adjustments but which could be expected to operate only rather slowly. Two such factors are that the number of firms in the industry may change, and that the size of the firms may change. Now if we want to make the analysis more complete, so that it covers not only the rapid adjustments involved in the original change, but also those which work themselves out much more slowly, we have to make an allowance for both the factors just mentioned. When the period of adjustment is long enough for factors of this kind to become fully operative, we talk about the *long-period effects* of a given change. This is the kind of adjustment with which the present chapter will deal.

Assumptions of the Short Period

When changes in economic conditions lead a firm to alter its output, there are a limited number of ways by which it can get quick results. To build a new plant and install new equipment will probably take a year or more. But if the firm wants to expand its output next month rather than next year, it will have to increase the number of men on its payroll and get more raw materials. It cannot increase output quickly enough by expanding plant and equipment or ad-

ministrative staff. The basis of the distinction between variable and fixed costs consists in this very fact.

Because rapid adjustments in output can best be made by varying the use of what we have called the variable factors, the short-period effects of any change in economic conditions differ from the long-period effects. Therefore, in dealing with the short-period effects of, let us say, a change in demand, it is proper to assume that while employment and the consumption of raw materials may vary, the size and number of the business firms are fixed. The marginal and average cost curves appropriate to the short period are constructed on these premises.

Assumptions of the Long Period

When time permits, it is often cheaper to expand output by building a larger plant than by trying to squeeze more from the existing plant. In fact, the additional output needed may be so large that this is the only solution. When, for example, a 500 per cent increase in aluminum production was needed for aircraft in the years 1941 to 1945, the only possible way to get so large an increase was to build new aluminum plants. Thus, when dealing with long-period adjustments, we do not exclude the possibility of expansions or contractions in the size of the firm. But even to vary the size of the plant does not exhaust the possible ways in which, given time, the output of an industry may be changed. An increase in output can come in two different ways: existing firms can employ more men, expand their plants, and use more raw materials; and, second, the number of firms in the industry can rise. Similarly, a decrease in output can result either from a reduction in the capacity and output of existing firms, or from an actual decline in the number of firms in the industry. Thus a long-period adjustment may include all those changes which ordinarily occur in the short period, plus those special ones, such as changes in the number and size of firms, which are unique to the long period.

Shor -Period and Long-Period Equilibrium

Short-period equilibrium is attained when each firm has made the best possible adjustment to prevailing conditions. In more concrete terms, this means that each firm has adopted that output and price policy which it finds most satisfactory in the light of such factors as

the demand for its product; the methods of production; the size of its plant, equipment, and administrative organization; and the wage rate it must pay. In still simpler terms, this means that the firm is producing the most profitable output. We say, then, that a firm is in short-period equilibrium if, with *existing plant* and related factors, it is producing the output it believes to be the most profitable.

While short-period equilibrium implies no tendency to change price or output by varying employment or the use of raw materials, as long as demand and related factors remain the same, long-period equilibrium implies no tendency to change the price or output of the industry in any way by varying either employment or the number and size of firms. If it were ever reached, long-period equilibrium could be recognized by the fact that the number of firms, their size, and the amount of labor they employ had become adjusted to the level of demand. While under these circumstances one firm might be growing and another contracting, there would be no tendency for the total output of the industry to alter.

Conditions of Long-Period Equilibrium

What possible circumstances can there be which would remove all incentive for firms either to expand or contract, or to enter or leave an industry? In other words, what are the conditions of long-period equilibrium? In a capitalist economy the conditions can be postulated simply enough, though perhaps not in a very helpful form. Briefly, if the businessman cannot anticipate additional profits by making the change, the change will not take place. If putting up an additional plant, installing more equipment, or permitting equipment on hand to wear out did not promise higher profits, we may assume that these things would not be done and the size of firms would not alter. Likewise if setting up a new firm or closing down an old one did not promise higher profits or smaller losses, the number of firms in an industry would not vary. But to say that the conditions for long-period equilibrium are satisfied when no additional profits can be anticipated from a change in the size or number of firms, while it is a formally correct statement, is not particularly helpful. And unfortunately, as soon as we try to cast more light on the matter, we necessarily give an impression that is imprecise. For the fact of the matter is that the forces in the real world which determine whether a new firm will be organized or an old one closed down, and whether

an existing firm will be expanded or contracted, are varied and complex.

Some of the factors that have a part in determining these decisions are listed below. The extent to which existing capacity is used is obviously important and anticipations about the future use of existing capacity are perhaps even more so, for firms will be expanded in size and new ones may enter an industry if it is widely believed that the present size of the industry will not be great enough to satisfy future demand economically. Another determining factor is the case and cost of expanding or organizing a new firm in the industry. The ease of entering other industries, and the inducements which they have to offer, are also important considerations. And the rate of interest, that is, the price that must be paid for borrowing the money to set up a new plant, will sometimes tip the balance. In short, the factors which may influence a businessman's decision to expand or contract a firm, or to open a new one or close down an old one, are numerous. In general, however, we may assume that if profits are exceptionally high in an industry, the conditions for an increase in the number and size of the constituent firms are satisfied. Thus when profits are especially high, we may expect that new firms will enter the industry and existing firms will enlarge their capacity. If, on the other hand, profits are very low or losses are unusually heavy, we should expect the number of firms in the industry to decline and the capacity of the remaining firms to fall.

These conditions for the expansion of an industry have been expressed with deliberate vagueness since there are so many exceptions and special cases that it is impossible to be more precise without misstatement. What are "exceptionally high profits," for instance, and when are losses "unusually heavy"? Profits were very high in the aluminum industry for a number of years, but the number of firms remained constant, at one. Through the same period profits in agriculture were very low, and many farmers lost heavily; yet the number of farms in the country grew rather steadily from 6,288,648 in 1930 to 6,812,350 in 1935. Therefore, we may use the term "abnormal profits" to mean profits high enough to induce expansion in the number and size of firms, but we must define the term with great care. It might not imply high profits, for example, when applied to retail trade or agriculture; on the other hand, when it is used in connection with some other industries it might imply an extraordinarily high

rate of profits. And precisely the same qualifications must be understood with reference to the phrase "subnormal profits or losses." Now let us examine the long-period effects of a change in demand and costs on industries of various types.

Increase in Demand in a Perfectly Competitive Industry

We shall begin by considering a perfectly competitive industry. Let us assume that before the increase in demand the industry is in long-period equilibrium. This means that the constituent firms are earning normal profits, since when long-period equilibrium rules, the number of firms and their size are constant. Now if demand increases, existing firms would produce more and prices would rise, as we showed in the previous chapter. Furthermore, profits would be higher, and if they were normal before, they would now be abnormally high — in the technical sense. So new firms would enter the industry and firms already in it would expand their capacity. Consequently, the supply of the product would increase — that is to say, at any price, the amount which this larger number of firms would make available would be greater than originally. Unless the demand rose still further, prices would fall and the profits earned by the firms now in the industry would be somewhat reduced. However, as long as profits remained high enough to induce new firms to enter the industry, there would be further increases in the supply of the commodity, and further reductions in price. Not until profits had once again fallen to normal would the expansion of the industry come to an end, though finally a new long-period equilibrium would be reached, with prices certainly lower than they had been shortly after the increase in demand.

Decrease in Demand in a Perfectly Competitive Industry

The situation would be just the opposite if demand were to fall. The first effect would be a reduction in price, output, and profits. In many of the firms in the industry, profits would be subnormal, and eventually some firms would close while others would probably let their plants deteriorate. Hence the supply of the commodity would fall, and the amount made available at any given price would be lower than before the number and capacity of firms was reduced. This decline in supply would cause a rise in price, as long as there was no further reduction in demand; and this in turn would reduce

the losses of the firms remaining in the industry. Firms would of course continue to leave so long as profits remained below the critical level, but with the exit of each firm, the situation for those remaining would become happier, and prices would tend to rise until profits were once again normal. Thus it is clear that the final position is one of long-period equilibrium, whether it results from an increase or a decrease in the number of firms in the industry. For this is the situation towards which the industry tends when the demand remains unchanged over a long period of time.

The accompanying diagram (Figure 54), which shows the situation for a perfectly competitive industry, and also for a firm in the industry, graphically shows these adjustments to changes in demand. We shall assume that the profits earned in the firm when the price is P_1 correspond to a situation of long-period equilibrium for the industry, and hence that in this situation the number of firms in the industry, and their capacity to produce, tend to remain the same. Now if the demand for the product increases from D_1 to D_2, and if supply remains

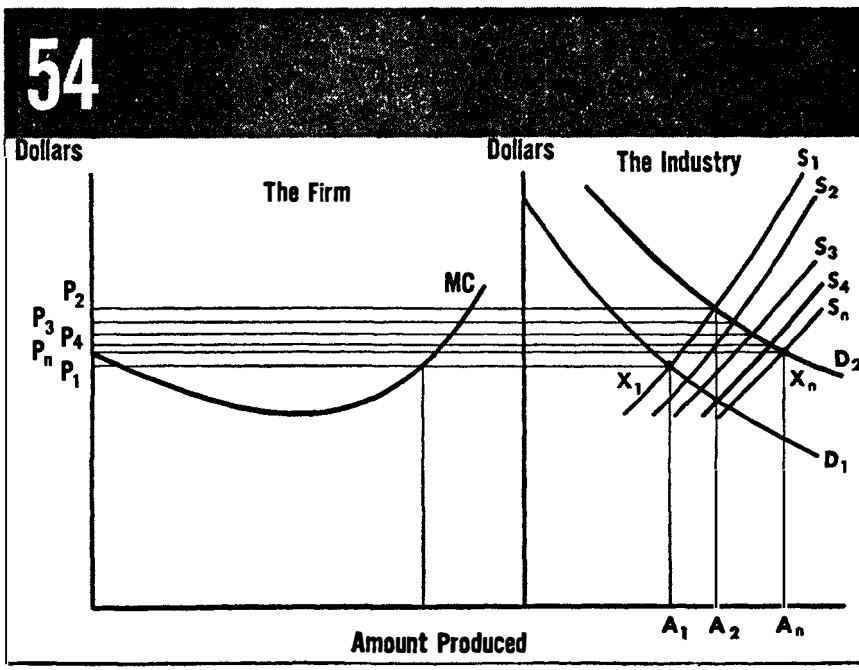

Figure 54. *Short- and Long-Period Adjustments to Change in Demand*

constant, price will increase and so will profits. Because of the now abnormally high profits, we should expect that after a time new firms would enter the industry and some of the original ones would expand. Consequently, supply will increase, and the supply curve S_1 is no longer apposite, since it was based upon the original capacity of the industry. As the number of firms is increased, it is necessary to redraw the supply curve in order to picture the changed situation. S_2 represents the supply function after the capacity in the industry has changed by a relatively small amount; S_3 after a further increase, and so on. With each increase in supply, price is brought lower, and probably profits fall at the same time, until the situation once again approaches a long-period equilibrium. Finally, when no further expansion is deemed profitable, long-period equilibrium would be restored.

Increasing-, Decreasing-, and Constant-Cost Industries

But would the price return to the original level? Not necessarily. The initial price P_1 would give way to P_2 after demand had gone up but before the industry's capacity had been altered. With the increase in capacity, the price would fall, as we have seen, to P_3, then P_4, then P_5, and ultimately to the level at which once again profits were normal. Whether the price would finally reach a level higher or lower than P_1 would depend on a number of factors. If costs began to rise as new firms entered the industry and others already in it expanded, then prices would not go all the way back to P_1. Such an increase in costs might occur if, for example, the union succeeded in winning a wage increase because of the increased employment of labor. But if the average cost curves for most firms fell, the final price would go even below P_1. Reductions in cost might be looked for if the expansion in the industry made it possible to improve the facilities for training labor or handling raw materials. Also, price would settle at a new level if the expansion of the industry changed the conditions of its long-period equilibrium. Thus, if much higher profits were required than formerly to keep the larger number and size of firms constant, we should not expect the final long-period equilibrium price to be as low as P_1.

In any case, whether price would eventually be stabilized at a point below its original level or above it, the increase in demand would bring about a much greater price increase in the short period than in

the long. Once capacity has increased in response to greater demand, we expect price to fall again. If demand remains at its new level for long enough, and there are no autonomous or spontaneous changes in the conditions of supply, a new long-period equilibrium will be reached. When that equilibrium is finally attained, the price may, as we have seen, be either higher or lower than the original price. Industries for which price reaches a new long-period equilibrium above the original equilibrium level, are known as increasing-cost industries; and those where the final equilibrium price is below the original level are known as decreasing-cost industries. Industries where in these circumstances price returns to its original level are known as constant-cost industries.

While price at first rises and then falls again, after the rise in demand, output can be expected to rise continuously, as a glance at Figure 54 will show. At first, the original firms will produce more, because at higher prices it pays them to do so. Then as these firms enlarge their plant, administrative staff, and so on, and as new firms are organized, the industry's output will rise still further. On the above diagram, the increase in output before the long-period forces have had time to operate is shown as A_2 to A_1. When these forces come into play, the output of the industry will rise still further, from A_2 to A_n.

Short- and Long-Period Supply Curves

Inspection of Figure 54 will show a number of supply curves, S_1, S_2, S_3, and so on. As we have seen, the supply curve S_1 was based on the costs and on the number of the firms originally in the industry. Hence it may be called a short-period supply curve. With the expansion of the industry, it became necessary to draw a new supply curve, S_2, which like S_1, was based upon the costs in the somewhat expanded and more numerous firms. Hence it too is a short-period supply curve. After further expansion of the industry, the supply curve S_2 gives way to S_3. Indeed, each of the supply curves, S_1, S_2, $S_3 \ldots S_n$, is a short-period supply curve appropriate to some given level of capacity.

Long-period equilibrium is represented in Figure 54 by the points X_1, which denotes the original position, and X_n, which denotes the final position for the industry. These points record the fact that the industry would in long-term equilibrium make available A_1 units of output at a price of P_1, and A_n units of output at a price of P_n. If

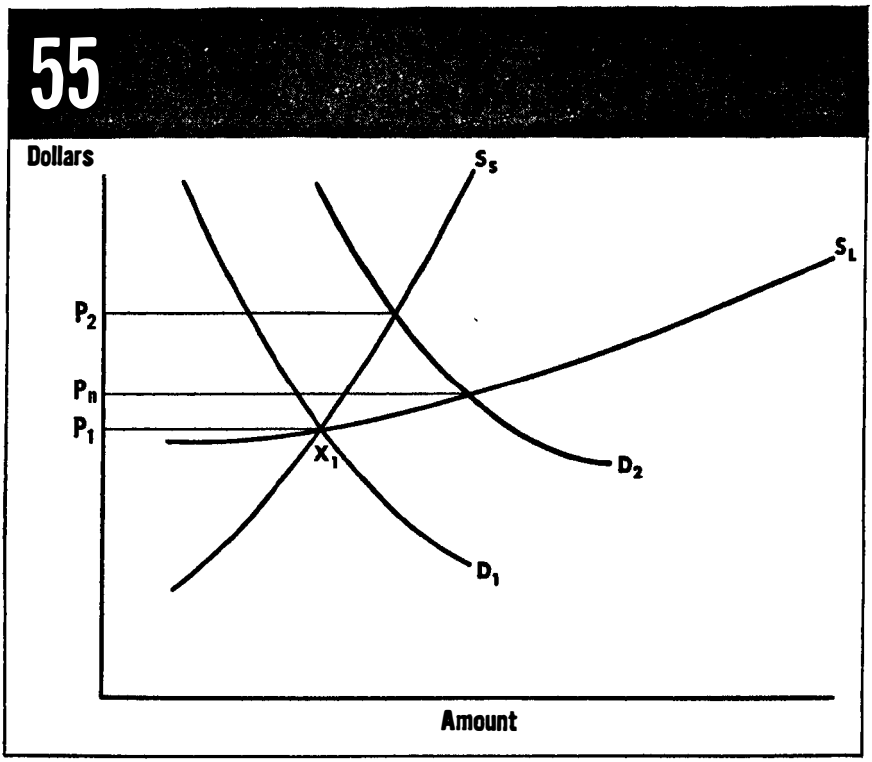

Figure 55. *Short- and Long-Period Supply Curves*

other points of this nature were determined, we could show the price at various levels of output A_m, A_s, A_t for an industry in long-period equilibrium by joining these points in a long-period supply curve, as in Figure 55. Since the long-period supply curve is the locus of points which describe long-period equilibrium positions, we may determine the final result of a change in demand upon price directly. The original demand is denoted by D_1. If the demand rises, let us say, to D_2, the price will at first rise quite far — the short-period supply curve S_s through X_1 shows the increase — but ultimately will fall again to P_n, given by the intersection of the new demand curve and the long-period supply curve *SL*. It should be clear that if the long-period supply curve rises to the right, it applies to an increasing-cost industry; if it declines to the right, it describes a decreasing-cost industry, and so on.

Changes in Costs and the Long-Period Adjustment

The method of analyzing the short- and long-period effects of a change in demand may also be applied to a change in cost. Therefore we shall simply sketch the procedure, for to trace it in detail would be to repeat much of what has already been said.

Suppose that costs for an industry in long-period equilibrium are reduced, perhaps because of a cut in wages or an improvement in the technique of production. The first effect is likely to be an increase in supply, although this would come about only if marginal costs in the firms were reduced as a result of the improvement. If demand remained constant, the increased supply would cause prices to fall. But whether prices fall or not — and as we have seen, they would not if marginal costs were not reduced — profits in the industry would rise. If they should rise enough to become abnormal in our sense of the term, new firms would be induced to enter the industry, and capacity would be increased. This in turn would give rise to a new and higher short-period supply curve, and price would fall still further. Long-period equilibrium would finally be restored when so many new firms had entered the industry, and so many old ones had expanded, that further expansion did not seem profitable to businessmen.

If costs were reduced, it is probable that price [1] would fall and output would rise. But when long-period adjustments begin to occur — namely, the entry of new firms — there would be a further fall in price and a further increase in output. The effect of a shift in costs is rather different from that of a change in demand. As we have seen, when demand changes, price and output at first move in the same direction, but in the long period, price moves back toward its original level while output continues to change in the same direction as the original shift in demand. When costs change, however, price moves in the same direction and output in the opposite one, for both the short and the long periods.

Long-Period Equilibrium: Monopolistic Competition

We have already seen that it is inconvenient to use the supply concept for an industry where competition is imperfect. It will therefore be more convenient to limit our discussion of long-period equilibrium

[1] If marginal costs were unaffected, as would happen if fixed costs were reduced, neither price nor output would be immediately affected. However, in the long period, profits would become abnormal, the capacity of the industry would increase, price would fall, and output would expand.

under monopolistic competition to the consideration of a single firm in such an industry. Let us suppose that the industry is initially in long-period equilibrium, and that one of the firms is earning normal profits, as represented by the rectangle *ABCD* in Figure 56. Under these circumstances our firm would respond to an increase in demand to AR_2 by raising both its price and its output. Profits would be higher than before, and hence, in our usage, abnormal. New firms would be attracted into the industry, and some old ones would be induced to expand. Because of this increased competition, the average revenue function for each firm would tend to fall, as indicated by AR_3. But the expansion of the industry would not come to an end until the demand for the product of each firm had dropped so far, to AR_n, say, that profits had once again been restored to normal. And when profits were once more normal, a position of long-period

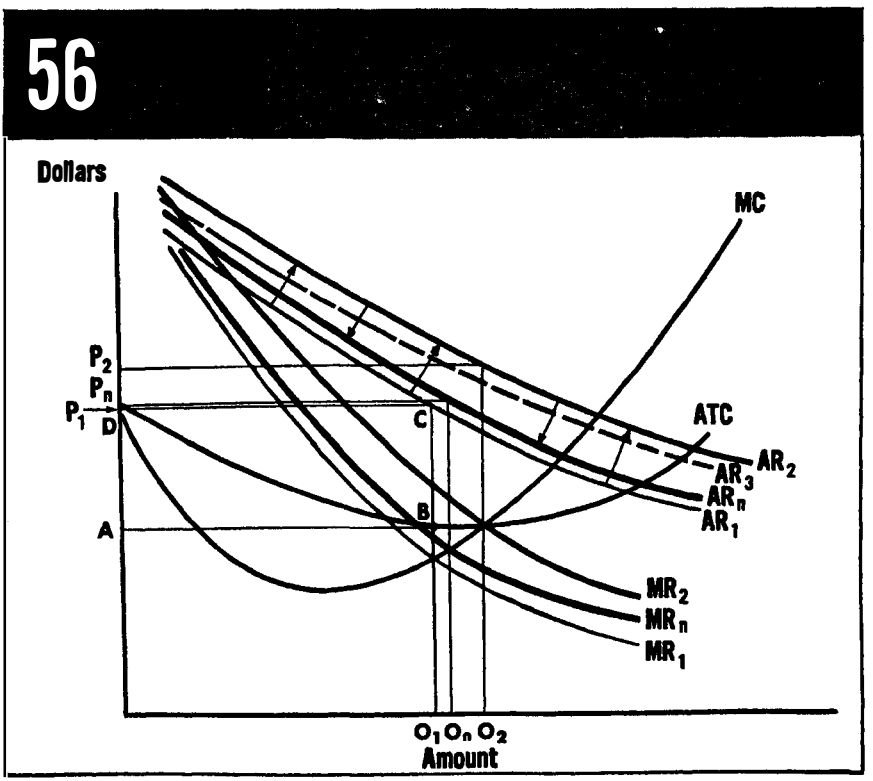

Figure 56. *Short- and Long-Period Adjustment: Monopolistic Competition*

equilibrium would again have been reached. It is possible, but not certain, that price would decline below the original level as increased competition reduced the demand for the product of any one firm. But any change in the elasticity of demand would be a critical factor. Whether the price would finally be established above or below the original level would depend not only on this factor but also on the change in cost induced by the entry of new firms into the industry or the expansion of old ones.

It is also possible to determine the long-period adjustment of price and output to a reduction in demand or an increase in cost. Generally, if demand falls, the long-period equilibrium price will be near the original level but above the price established shortly after the decrease in demand. Output, however, will be lower than at first, or than before firms began to leave the industry. If costs go up, price in the short period would go up, and output would decline; price in the long period would probably be set quite close to the original level, but output would decline still further. The reader should attempt to reach these conclusions independently, by reference to earlier demonstrations if necessary, since it is important to understand the reasoning which underlies them.

Long-Period Output Adjustments

While most of this discussion has been cast in terms of price, the output adjustments of an industry are perhaps even more important. They should, however, be quite clear from the preceding discussion. If because of the entry of new firms into an industry, the price should drop, the implication is that output is increasing; similarly, if price should rise because of a decline in the capacity of the industry, the implication is that output has been cut.

Summary

We have seen that the short-period adjustments of output and price to a change in demand or cost are likely to differ from the adjustments over the long period. A change in demand will usually affect price much less in the long period than in the short period, but it will usually affect output much more. The first effect of an increase in demand is a rise in price and output. Later, price is likely to fall and output to expand still further. If costs change, the longer the period under consideration, the greater the effect on both price and

output is likely to be. Thus in the short period, if costs fall, price will probably fall and output will probably increase; in the long period, we should expect a further fall in price and a further increase in output. Under imperfect competition, changes in the elasticity of the average revenue function can be a further complicating factor. But when no such change occurs, what happens under imperfect or monopolistic competition is just the same as what happens under perfect competition. The modifications of the theory for an oligopolistic industry should be worked out by the reader.

20

Monopoly, Competition, and the Satisfaction of Wants

OUR ECONOMY is made up of some ten million business firms which use labor and capital goods of all types to produce thousands of different commodities. The variety of their individual contributions to our total output is immense: needles and thread, socks and shoes, bread and cheese, automobiles and locomotives, permanent waves and performances of symphonies, and so on almost to infinity. In one year these firms may produce 80 million tons of steel, 500 million pairs of shoes, 1 billion pounds of aluminum, 5 million automobiles, and 650 million glass tumblers. A picture which showed in complete detail the output of these firms in any period would be enormously complicated. And a moving picture that tried to give some impression of the activities of these ten million firms, ranging from the boy who cleans shoes on the corner to the enormous corporation that supplies the nation's telephone service, would give an impression of bustling confusion and not much else. For the structure of our economy is complicated in the extreme; it seems to resemble an enormous jigsaw puzzle rather than a well-designed bridge.

Perhaps even the analogy with the jigsaw puzzle is too tame. Perhaps a jigsaw puzzle in which each piece determined its own size, color, and shape with no regard to the total design would seem to be nearer the truth. For each firm in the economy makes its own independent decisions on what and how much to produce. The United States Steel Corporation may decide to supply 30 million tons of steel ingots in a year. The Free Sewing Machine Company may decide to manufacture 35,500 sewing machines. The Marquette Cement Man-

ufacturing Company may decide to produce 6 million barrels of cement. And the corner grocery store may decide to keep $420 worth of canned goods on its shelves. In a capitalist economy, each firm arrives at its own decisions. Can we then hope that these individually determined pieces could ever be assembled into a coherent pattern? With perhaps 500 breweries separately determining how much beer to produce, with about 25 companies determining independently how many sewing machines and parts to manufacture, with about 4 million farmers determining how much cotton to grow, and so on for 10,000 commodities, can the result be anything but chaos? Or do these separate flows of output from the millions of firms in the economy somehow miraculously fill the need and leave nothing over to rot?

If there is a pattern, instead of a chaos of output, it exists by accident — or at any rate, not by design. In contrast to our planlessness, a socialist economy has its Central Planning Board to determine the economy's production of steel, shoes, wheat, and watches, and to insure that there is some sort of balance between the individual components of its total output. One of the jobs of this board is to see that the economy does not produce too much of one thing and too little of something else. The pieces of the jigsaw puzzle are all cut by plan, and while the pattern may not be perfect, at least it is intended to be. But in a capitalist economy no such over-all plan is drawn up, for there is no Central Planning Board that tells General Motors how many automobiles to produce or General Foods how much cereal to package. Each firm cuts one piece of the jigsaw puzzle in the color and size it wants. Do the pieces fit? Or do we run the danger of having far too many shoes and automobiles and not enough socks and gasoline?

This chapter will attempt to answer that question. We shall find that there is indeed a pattern, and a good one. Production is reasonably well-geared to meet buyers' wants, though not perfectly. One of the chief factors in distorting the pattern is the existence of varying degrees of monopoly in our economy. For unless the government overrules them, the more monopolistic firms always tend to cut their pieces a little small, and the more competitive firms tend instead to cut theirs a little too generously. The composition of the total flow of output for that reason does not exactly correspond with the wants of buyers. There are other reasons, too, why the correspondence is

not perfect. But even if it is not ideal, the correspondence is at least good. Our first task is to see the general picture; to see that the pattern made by the individual flows from the millions of firms in our economy is one that broadly reflects our wants.

The Ideal Pattern of Output

What pattern of output would best reflect our requirements? How can we determine whether a certain combination of commodities is better than any other, judged by reference to its ability to meet our wants? If society prefers one combination of goods to all others that cost the same amount, surely that combination is in this sense ideal, for obviously we prefer the combination that corresponds most closely to the pattern of our requirements. To say that the production of commodities is as closely related as possible to the wants of consumers and other buyers, is to say that buyers do not prefer a different combination of commodities to the one they are now getting, as long as the costs of the two combinations are the same. If they should prefer some other combination that costs no more, the implication is that their present combination is not as well-adjusted to their wants as it could be. By way of a simple illustration, let us suppose there are but two commodities, A and B, and that the same amount of labor and other resources is required to produce one unit of A as to produce one unit of B. Suppose, too, that our present combination consists of 100 units of A and 200 units of B. When we say that the present production of A and B conforms as closely as possible to the wants of buyers, we mean that buyers prefer a total output composed of 100 units of A and 200 units of B, to one made up of 101 units of A and 199 units of B. For it must be noted that it would be just as easy to produce one combination as the other; we could as well produce a combination of 101 units of A and 199 units of B, as one consisting of 100 units of A and 200 units of B. But so far as buyers are concerned, 100 units of A and 200 units of B is better than any other combination that can be produced with the same volume of resources.

How do they show that they prefer such a combination? Buyers do not mark ballots in our economy to indicate their preferences. If they did, we should know directly. As it is, the only indication is to be got from what they are willing to spend. They show that they prefer one combination to any other by their willingness to spend more to get it than to get any other. If one combination of goods

attracts more money from the pockets of buyers than any other, this is a sign that buyers prefer it.

It is not, however, a very accurate sign, for this reason: it weighs the importance of buyers by the amount of money each one has. It is as though each voter at an election for six members to a committee were assigned not one vote, but a number that varied with, let us say, the number of letters in his name. When one voter is given ten votes, another twenty, and still another thirty, does the election conform to the voters' preferences? No, for those with the largest number of votes do most of the electing, and the final choice corresponds more closely to their preferences than to the preferences of those who had few votes. Likewise, when one man has $500,000 to spend, and another has $500, can we say that the output that corresponds most closely to total buying corresponds most closely also to their preferences? No, for the combination that attracts the largest sum of money is obviously the one that conforms almost entirely to the wants of the rich man, who has many more votes than the poor one. The combination of goods for which buyers are willing to pay the most differs from the combination which reflects most closely the real wants of buyers because of inequalities of income.[1]

The "ideal" output has been defined as that which corresponds most closely to buyers' preferences. At first sight it seemed that we could determine whether any particular combination of goods satisfied this criterion by seeing whether buyers were willing to pay more for it than for any other combination. And so we can, but with important qualifications. If it were not for inequalities in income distribution, there would be fewer doubts,[2] and we could perhaps conclude that the combination of goods which attracted the largest amount of money from the pockets of buyers was indeed the best, or at least that it most closely conformed to buyers' preferences. But because of these inequalities, this conclusion is not acceptable. Instead we must rephrase our conclusion about as follows: buyers spend most for the output which they prefer, except for the effects of inequalities in income. Or we may say that the ideal pattern of output is that for which buyers spend the most — again except for the effects of the unequal distribution of income.

[1] This implies that the desires of each buyer are of equal importance, or at least that the importance of their desires is not measured by their ability to get money to spend.

[2] Except for relatively small difficulties because people themselves differ, so it is said, in their capacity to enjoy, and so on.

The Pattern of Production — Perfect Competition

Can we say anything about the combination of goods and services which the firms in the economy produce? Do we find, for example, that they provide an output which, out of all those that could be produced with the same total of resources, attracts the largest sum of money? Or is it widely different from this, and if so, why? For if it should be true that the output which firms produce attracts the largest sum of money from buyers, then we might conclude that it departs from the ideal only because of the unequal distribution of income. Perhaps the easiest way to test whether firms do in fact produce the ideal output is then to see what would happen if all firms sold in a perfectly competitive market. Is there any force that leads them to produce the output upon which buyers spend the maximum sum? When a firm sells in such a market, it is able to expand its sales without reducing price. Therefore, when it increases its sales from, let us say, 100 to 101 units, its receipts rise by the price of the unit; in other words, marginal and average revenue are the same. For convenience, we define our units of the commodities in such a way that the same volume of resources is required to produce one unit of each of them. In other words, the same volume of resources would turn out a combination of 1000 units of A and 1000 units of B, or a combination of 800 units of A and 1200 units of B. Let us suppose that there are only two kinds of goods produced, and that 800 units of A and 1200 units of B are the total output of the two industries when each of the firms in both industries is making the amount of goods which will maximize its profit.

Now, if buyers were willing to spend more for, let us say, 400 units of A and 1600 units of B than for the combination of 800 units of A and 1200 of B, would the firms continue to produce their present amounts? If consumers were willing to spend more to acquire the former output than the latter, it would signify a willingness to pay a higher price for the extra 400 units of B [1] than for the alternative 400 units of A,[2] perhaps $15 a unit for B and only $12 a unit for A. If this happened, the makers of B would find it profitable to bid resources away from firms that make A, since a certain amount of resources working for the A producers would add only $12 to their receipts, whereas the same amount working for the B producers would now add $15 to these firms' receipts. Hence the B producers could

[1] That is, the units between 1200 and 1600. [2] That is, the units between 400 and 800.

afford to bid more and more for these resources, and as they expanded their output and profits, the A producers would contract theirs. But as the output of B increased and that of A declined, the price of B would fall while that of A would rise. This bidding of resources away from A would go on until the price of A and B were again the same.[1] For until they are equal, it will continue to pay the B producers to get resources from the A producers. And all through this process, B's output will go up and its price will fall while A's will do the opposite, until eventually the production pattern yields equal prices so long as each firm is free to determine its output by considerations of profit. This would occur because, as long as prices differed, it would pay some firms to expand their output, and for others to contract theirs as the cost of a package of resources rose.

When the prices of A and of B are the same, it implies that consumers feel no preference as between the final unit of A and B. And this implies that they spend more upon such a combination of goods than on any other. To illustrate: Let us suppose that when the prices are equal, output is 600 A and 1400 B. With prices equal, buyers would not spend more for the 1401st unit of B than for the 600th unit of A, or for the 601st unit of A than for the 1400th unit of B. From this, it follows that they are willing to spend at least as much for 600 of A and 1400 of B as for any other combination of outputs. For if any other combination of outputs would attract more dollars from their pockets, it would signify that either the price of the 601st unit of A was above that of the 1401st unit of B, or the other way about. Now we have already seen that when competition is perfect, the pattern of production is such that the price per unit of each commodity is the same.[2] We may therefore conclude that with perfect competition, the pattern of production that firms create in their search for profits is such that buyers are willing to spend more upon it than upon any other combination.

Summary: Ideal Output and Perfect Competition

Up to this point the argument has consisted of these steps: (1) The

[1] It will be remembered that the units are defined so that the same package of resources is needed to produce one unit of A and one unit of B. We do not mean, of course, that the price of one orange would equal the price of one automobile.

[2] The careful reader will notice the implicit assumption that each firm would be able to acquire a certain package of labor and other resources at the same price. If not, this conclusion would have to be modified, and the pattern of production would depart still further from the ideal.

ideal output is the one on which buyers are willing to spend the greatest amount; or it would be ideal if we could neglect the effects of unequal distribution of income. (2) When competition is perfect, firms attempting to maximate profits, produce a combination of goods such that their prices are all equal. (3) When the prices of all goods are equal, it means that buyers are willing to spend the maximum amount upon the combination being produced.[1] (4) Hence, when competition is perfect, firms produce that combination of goods which attracts the largest sums from buyers. And therefore, (5) when competition is perfect, firms produce the ideal output, aside from the effects of the inequality of income.

Our economy is unplanned. Each firm, as we have emphasized, determines its own output, and there is no central authority to see that total production conforms to buyers' preferences. Yet the output which firms actually produce, determined only by their own desire for profit, is the one which, with certain qualifications, conforms most closely to buyers' wants. The first qualification is that inequalities of income are not taken into account; the second, that the results apply, so far as we have yet seen, only if all firms are perfectly competitive.[2]

The Pattern of Production and Monopoly

Do firms produce the ideal output when some industries are monopolistic and others are competitive? Does the existence of monopoly in a competitive economy distort the pattern? To answer this question, let us suppose that the B industry is monopolistic and the A industry is perfectly competitive. Each firm in the latter expands or contracts its output without regard to the effect such changes may have upon the price charged; that is, each firm in the A industry assumes that no matter what the level of its output, it can get $12 a unit for as many units as it produces. But the firm which produces B — and since it is a monopoly, there is only one — looks at things in a different way. If its output is 1200, it gets $15, but if its output is raised to, let us say, 1220, it may get only $14.95 a unit, since the average revenue function for a firm that is not perfectly competitive is not perfectly elastic.

Suppose that when the output of A is 800, its price is $12; and when

[1] The reader is again reminded of the way in which the unit for measuring the output of each good was established.

[2] Because of our way of measuring units of a commodity, we assume too that all industries operate under conditions of constant returns, as the economist will realize.

the output of B is 1200, its price is $15. There are two things to determine: is this an ideal output, as defined above; and is there any force which would lead to a change in the output of A and B as there is under perfect competition?

In the first place, the output composed of 800 units of A and 1200 units of B is not ideal. Since A sells for $12 a unit and B for $15, buyers would spend more in total if the output of B were raised and the output of A were lowered. If, for example, 799 units of A and 1201 units of B were produced, buyers would spend $3 more than when the combination is 800 A and 1200 B. Hence the combination of 800 A and 1200 B is not ideal, for the ideal output is that which draws maximum spending.

But if the output is not ideal, will it tend to change? In this case, it will not change at all, or at the most, very slightly. With perfect competition we saw that each B producer could offer up to $15 for additional packages of resources; and so he was able and willing to expand his output, since each added unit of output increased his receipts by $15, the price of the unit. With monopoly, the situation is quite different. The single B producer would raise his receipts by considerably less than $15 if he raised his output by one unit. For when he produces 1200 units and sells them at $15 each, his total receipts are $18,000. But when he produces 1220 units, he is able to sell them at only $14.95 each, and his total receipts for this larger output are only $18,239. Hence he adds only $239 to his total receipts when he expands output by 20 units, or approximately $12 when he expands output by 1 unit. Now if he adds only $12 to his receipts for the extra unit, it is obvious that he would not be willing to pay more than $12 for a package of labor and resources. Since the firms that produce A are also willing to pay this amount, it follows that B would be unable to attract resources from the A industry and so would be unwilling to expand his own output. Hence output would remain at the original level, 800 A and 1200 B, whereas the ideal output would be composed of fewer than 800 units of A and more than 1200 units of B — or as assumed earlier, 600 units of A and 1400 of B.

Hence, in an economy which has some competitive industries and some monopolies, output is not ideal from the point of view of consumers' wants: the output of the competitive industry — A in our example — tends to be too large for a proper balance, and the output

of the monopoly tends to be too small. The ideal output is not reached because monopolies produce, relatively speaking, too little, while competitive industries produce, relatively, too much. Or to express this conclusion in other terms which may be more familiar; the price of monopolistically-produced goods, in this case $15, is too high; while the price of competitively-produced goods, in this case $12, is too low.[1] When a price is too high, output is too small; when a price is too low, output is too large, so the two ways of describing the effects of monopoly in a competitive economy come to the same thing.

Likewise, the ideal output is not achieved when different degrees of monopoly exist in the firms involved, though as might be expected, the departure from the ideal is less extreme than when only perfect competition and strict monopoly exist. If, for example, the economy were made up of some industries which were strongly monopolistic and others only slightly so, the composition of the output would be unbalanced. The more monopolistic industries would produce too little, and those that were less monopolistic would produce too much. In fact, the output of each industry would be above or below the desired amount as the degree of monopoly in that industry was less or greater than average.

Strictly speaking, monopoly cannot be condemned for distorting the pattern of production and forcing it to deviate from the ideal. It would be just as accurate to say that the trouble was caused by the competitive industries whose production tends to be too high. The responsibility for creating the unbalanced output really lies neither with monopoly nor with competition, but rather with the fact that the degree of monopoly varies over the economy. If all industries were *equally* monopolistic, output in this sense would not be unbalanced.[2] Nevertheless, there can be no doubt that it is unbalanced now and that the output is far from ideal with respect to the amount produced by the various component industries.

Does our economy provide the ideal output judging it by its correspondence with consumers' wants? Do we produce automobiles,

[1] Strictly speaking, we should have to say the monopoly price is too high relative to the competitive price. There is no absolute standard by which to judge them.

[2] There may be other objections to monopoly, however; for one thing it is possible to show that monopoly emphasizes the inequality of income distribution. In addition, it is very likely that if all industries were monopolistic, the total output would be lower than if they were all perfectly competitive; the output might, with equally strong monopolies everywhere, be balanced if we consider the relative proportions of commodity A, B, C, and so on; but it might be too low when set against our capacity to produce.

gloves, bread, and so on in approximately the correct proportions? We have now developed two reasons for thinking not: income inequality means that the total vote of buyers does not reflect their real requirements; and the components of our total output do not reflect our total vote (which does not reflect our total requirements) when the degree of monopoly varies from one industry to another. It is as though we gave some people 100 votes, others 50, and still others only 1, and in addition permitted some tampering with the ballot-boxes.[1] Would the list of successful candidates in such an election represent the real desires of the voters? It would be very strange if it did.

Changes in Wants and the Ideal Output

When wants change, the nature of the ideal output changes too. Originally the ideal output may have been 600 units of A and 1400 of B, but if consumers now want more A and less B, the output that *was* ideal no longer conforms to buyers' preferences. If the community now prefers an output composed of 700 units of A and 1300 of B, how will production in our economy respond? Will the output of A be increased and that of B be reduced? We have already seen the answer to this question.[2] When the demand for a commodity increases, the firms that produce it are induced to expand their output. At first they will have to produce the larger output with the existing plant and equipment, but if demand is maintained at the higher level, they will find it profitable to expand, and new firms will enter the industry, so that the increase in output in the long run will be greater than the increase in the short period. Conversely, when demand declines, output will be reduced promptly; and if demand remains at the lower level, some firms will eventually leave the industry while others will not maintain all of their plant and equipment. With the shrinkage in capacity, output will be still further reduced. Thus, we should expect the output of A to be raised and that of B to be lowered if demand shifted, as we assumed it did, from B to A.

The adjustments in output would not, however, be instantaneous. The firms that wanted to expand might not be able to find the labor they needed, or their output might already be at plant capacity. The entry of new firms might be delayed because of inability to acquire

[1] Votes are removed for commodities produced by monopolies, and illegal ballots are cast for commodities produced competitively.

[2] Compare Chapters 15 and 19.

rights to use processes of production on which existing firms held patents. Firms might be reluctant to expand their plants until they felt certain that the increase in demand was to be long-lived. It might take many months before output could be raised. Likewise, it might take a long time before the output of an industry with lowered demand declined to its final position. Firms might be reluctant to let skilled labor go. They might continue in the industry in the hope that conditions would improve. Thus the adjustment of output to a change in demand would certainly not be instantaneous; indeed, it might be very slow.

There have been striking illustrations of the economy's inability to adjust rapidly to shifts in demand, both during the period between 1942 and 1944, when munitions production was being increased, and during the reconversion period after 1945. In the earlier period, more aircraft, tanks, and other munitions were badly needed. At that time output was raised with phenomenal speed, but even so we were not producing all the munitions needed in those years. Our output was unbalanced — there were too few munitions and too many consumers' goods.[1] Again, during the reconversion period we urgently wanted more automobiles and nylon stockings than were immediately available. Our output was unbalanced, and the adjustments needed to bring it into balance were relatively slow.

Just here it is worth while to digress from the main argument in order to emphasize a related point: these adjustments sometimes involve heavy social costs. Workers dismissed from an industry faced by a declining demand do not necessarily find jobs in the industries for whose product the demand is rising — or anywhere else. And even if jobs are available they may not fully use the skills of the displaced workers. After demand shifts, the shift in the composition of output that brings it into tolerable conformance with consumers' wants not only may be slow; it may be painful. And because the adjustment is slow during this long period, the pattern of the actual output of the economy does not correspond to the pattern of buyers' preferences.

[1] Strictly, we cannot say this in view of our definition. It will be remembered that we defined a unit of each commodity in such a way that its production required a package of labor and other resources of the same size as was required for the production of any other. For munitions outputs larger than we were producing in, let us say, December, 1942, the unit of output was a very small amount of munitions. And we were not willing to pay the very high price needed to secure that unit at that date.

Government Policy and the Ideal Output

The principal reasons, then, why the output of our economy may not approximate the ideal are the unequal distribution of income, the unequal degrees of monopoly which prevail in the various industries, and the slowness of adjustments to changes in demand or cost. And there are other reasons,[1] though these are the chief ones. Can anything be done to achieve a better balance? Indeed yes. It is possible, for instance, to give more buyers' votes — or more purchasing power — to the lower income groups by distributing income more equally. It is possible to achieve a more nearly ideal balance of output by making monopolies more competitive, and perhaps also by making competitive industries more monopolistic. And it is possible to provide for faster and smoother changes in output in response to shifts in demand and cost.

In fact, measures are frequently taken in our economy to remedy this situation. The inequality of income distribution is attacked directly by the imposition of high tax rates on high incomes and on the inheritance of great wealth. Less direct attacks are made through legislation designed to raise wages — for example, by laws that set minimum wages. Certain goods, such as public education, are made available in quantities that far exceed what buyers would purchase if a price were set by business firms. This in effect gives lower-income groups some additional buyers' votes.

The second distorting factor, the unequal extent of monopoly and competition, has been attacked in a variety of ways. The most obvious of these measures have been those designed to break up monopoly, such as the Sherman and Clayton Anti-Trust Acts. Moreover, where it was regarded as unwise to destroy monopoly, firms have been forced to charge less than the monopoly price. This has happened with the railways and the public utilities. Obviously the actions of state regulatory bodies like the Public Utilities Commissions, or, on a national scale, the Interstate Commerce Commission, in setting prices below the monopoly level have the same effect as forcing the regulated firms to produce more than the monopoly output. Some monopoly firms are thus forced to cut their individual contributions to the economic jigsaw puzzle more generously than they would have done if they had been free to decide by themselves. One more example of the kinds of steps which have been taken to correct the effects of this dis-

[1] See footnote 2, page 242, for further suggestions, but there are still others.

torting factor will suffice. Legislation for agriculture has brought about conditions in most agricultural industries which are less than perfectly competitive. When the individual producer is subsidized to limit his output, it is as though the demand function for his product were made less than perfectly elastic. These measures have the effect of reducing the inequality in the degree of monopoly throughout the economy.

Much has also been done to improve and speed up the adjustment to changes in wants. Employment agencies have been sponsored by various states and by the federal government. The allocation of raw materials for many key industries was strictly controlled during the years between 1942 and 1945, and a strict system of priorities for building materials was established in 1946.

In short, we, acting through our government, have not been disposed to accept as perfect the composition of output that would have been produced in a completely *laissez-faire* economy. By various kinds of government regulation, we have interfered with that result in an effort to make the pattern of our output conform more closely to our wants.

Summary

We see, then, that there is some order in the make-up of our total output, not chaos as might at first glance appear. Even though each firm produces as much as it wants, the pieces of the jigsaw puzzle fit together reasonably well. In spite of the absence of a central planning authority, the pattern of our output is not too far from what buyers' preferences dictate. This is true because buyers express themselves by spending, and firms seek to maximize their profits. The "invisible hand" postulated by Adam Smith, in accordance with which the interest of society is promoted by each person seeking his own gain, is herein revealed. But the result is not ideal. Although the pattern is reasonably coherent — indeed wonderfully so in view of the complexity of our economy — it does not correspond perfectly, or even closely enough, to the pattern of our requirements. Three main factors make the pattern of output diverge from the pattern of wants: unequal distribution of income, unequal degrees of monopoly among industries, and slow adjustments of output to changes in demand or cost. Because of these things, our output is composed, in the absence of social control, of too many goods for the well-to-do and of too

many produced by the more competitive industries. And the response to changes in the pattern of demand and cost is much too slow.

Government action aimed at reducing the inequality of income and the unequal degree of monopoly in various industries contributes to making the output conform more closely to the ideal. Similarly, measures to speed up the shifts in output needed to bring it into conformity with consumers' wants help to make it more acceptable.

21

The Value of Output and The National Income

WHEN YOU PAY A NICKEL for a chocolate bar, the seller receives a nickel. From this obvious fact we can derive an important conclusion which is somewhat less self-evident. To state the obvious fact in more general terms: corresponding to every sum of money paid by a purchaser there is an equivalent sum received by a seller. The not-so-obvious conclusion is that the total value of all the goods and services produced in a period of time equals the national income for that period. Let us consider the relation between sales and income, beginning once more with the basic unit in our economy, the individual business firm.

The Single Firm

The firm is both a buyer and a seller, and it is enlightening to follow its receipts and its expenditures in some detail. A firm's sales receipts for a period of time are equal to its costs and its profits for that period. This follows from the definition of the term profits. For if it has receipts of $10,000 and costs of $8000, then its profits are $2000; if it has receipts of $10,000 and costs of only $3000, then its profits are $7000. And if it has receipts of $10,000 and costs of $11,000, then its profits are minus $1000. In each case its receipts equal the sum of its costs and its profits.

A firm's sales receipts are, of course, equal to the sum paid by buyers for its products. The firm need not receive cash. The buyer may purchase on credit; that is, on a promise to pay cash in, let us say, sixty days. But the firm counts not only its cash receipts from

sales but also the promises to pay it accepts in calculating the value of its sales. And the total amount spent by buyers upon its products constitutes its sales receipts.

Let us suppose that a firm, which we shall call A, receives $10,000 from sales in a certain period. What happens to this sum? A part of it is used to pay wages and salaries, another part is used to buy raw materials, still another is used to pay rentals and interest on the money it owes, and some is used to purchase light, heat, power, and miscellaneous supplies. These are its costs of production.[1] Suppose that these items of costs are:

Wages and salaries	$4000
Rent and interest	500
Raw materials[2]	4000
Light, heat, power, and miscellaneous	300
Total costs	$8800

The firm's profit for the period is therefore $1200. It is clear that the $10,000 in receipts may be broken down into wages, salaries, rent, interest, costs of raw material and so on, and profit. And the sum of the cost items plus profit equals $10,000. Even if the sum of these cost items had exceeded $10,000, and the firm had therefore suffered a loss, the sum of the cost items and profits (negative, in this case) would still have been $10,000.

The cost items may be divided into two classes: those paid directly to individuals, and those paid to business firms. Wages and salaries are paid directly to individuals, the employees of the firm. Likewise, rents and interest may be paid to private individuals rather than to business firms. But the payments for raw materials, light, power, office supplies, and so on, are much more likely to be made to other business firms which supply these products. The payments to individuals constitute their incomes for the period.[3] If the wage earners

[1] In the interest of simplicity we have deliberately omitted taxes and depreciation charges. If these also had been included, we should have been compelled to draw a distinction between the national income and the gross national product. While the distinction is important, it is not necessary to be concerned about it at this stage.

[2] We shall assume at first that the amount of raw materials used up in production equals the purchases of raw materials. This implies that inventories are kept constant.

[3] Payments may not be made in cash in the period. The wages and salaries due for the period, but not yet paid in cash, or the cost of raw materials bought on credit, are treated nevertheless as current costs from the point of view of the paying firm, and from the point of view of the employee or supplying firm, as current receipts — just as the accountant would treat them.

and salaried employees of the firm receive $4000 in the period, their incomes amount to $4000. If landlords [1] and lenders receive $500 for rent and interest for the period, their incomes amount to $500. Part of the $10,000, then, can be traced directly to the incomes of the firm's employees, its creditors, and its landlords. The $1200 profit of the firm would also be regarded as the income of the firm's owner. Thus, of the $10,000 which the firm receives, $5700 is disbursed to individuals who regard the receipts as their income for the period. The other $4300 paid for raw materials, light, heat, power, and miscellaneous supplies, constitute part of the sales receipts of other business firms, which we shall lump together as B.

A similar accounting can be made of the sales receipts of the firms collectively called B. The B firms that receive the $4300 from A will have to make wage payments to their employees, will perhaps pay interest on money they have borrowed, and will purchase raw materials, light, heat, and so on. And whatever is left over is available to their owners as profits. If these firms pay $1000 for raw materials and other items which came from still other business firms, then it follows that their payments to individuals are $3300. This sum, going for wages, salaries, rent, interest, and profit, is regarded by those who receive it as their income for the period earned in producing the supplies bought by firm A for $4300. The $1000 paid by firms B to other firms, C, makes up, of course, their receipts from the sale of raw materials and other items to B. This sum too can be broken down in further detail. Part of it is paid out by firms C for wages, salaries, rent, and interest as income earned by laborer, landlord, and lender in producing the $1000 worth of raw materials sold by C to B. Another part is paid for raw materials to firms D, and the remainder is profit for the owners of firms C.

It should be clear that as we trace the original $10,000 step by step, we can account for every cent as an item of income for someone who played a part, perhaps very indirectly, in producing the goods that firm A sold for $10,000. Firm A directly generated $5700 worth of income. Firm B directly generated $3300 worth of income. And firms C, D, E, and so on generated $1000 worth of income. The total income generated from the production of these goods was $10,000 — the exact value of the sales of firm A. The sales receipts of a firm in any period are thus equal to the incomes earned by those associated

[1] Assume that landlords have no expenses.

with that firm and with its sources of supply. If it sells $50,000 worth of shoes in a month, the incomes of all those who had a hand in producing the shoes are $50,000; or more accurately, the $50,000 represents the incomes they earned because of the part they took in producing the shoes.

In conclusion, a firm's sales receipts in any period are equal to the sum of its costs and profits for that period. All items of cost can be identified with someone's income: wage and salary costs directly, and costs for raw materials after one or several steps in the analysis of the activities of firms that supply raw materials to our firm, or to the supplying firm, and so on. Likewise, the profits earned are, of course, the income of the firm's owners. Hence a firm's sales receipts in any period equal the incomes earned in producing what it sold.

A schematic presentation of this argument may make it clear.

Suppose that the block A_1 in Figure 57 represents the receipts of firm one in a certain period. It pays out W_1 in wages and salaries,

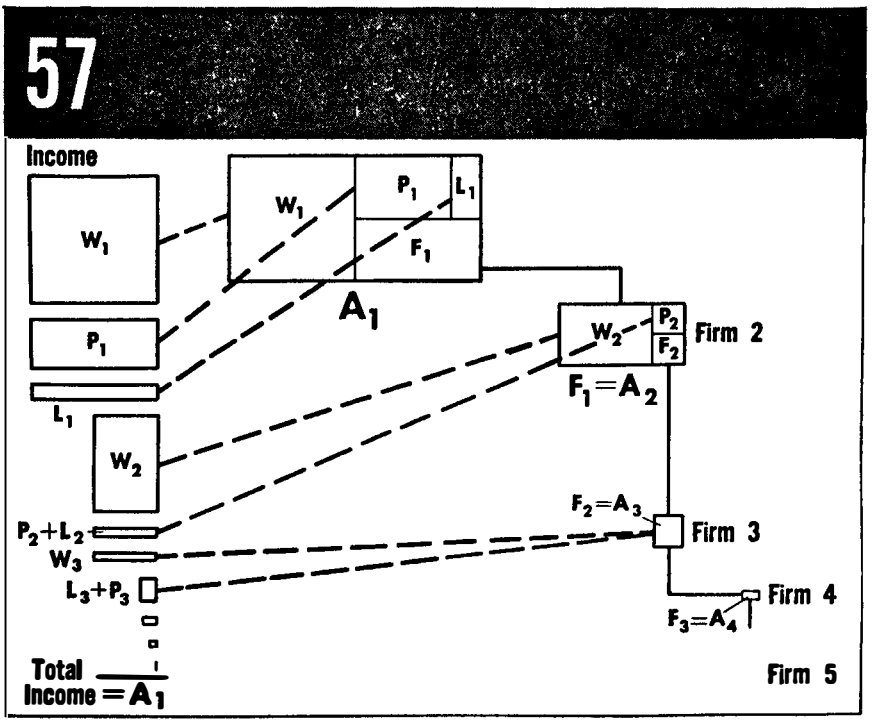

Figure 57. *Sales Receipts and the National Income*

L_1 in rent and interest, F_1 for raw materials purchased from other firms. Its profits are therefore P_1, or the difference between A_1 and $W_1 + L_1 + F_1$. Now W_1, L_1, and P_1 are all items of income. We can trace F_1 to several firms, but if we combine their accounts under the heading firm two, they receive F_1, which we may call A_2 — that is, the receipts of firm two. Now A_2 gives rise, as did A_1, to a series of individual incomes; to W_2 in wages and salaries, L_2 in rent and interest, F_2 in raw materials purchased, and the difference, P_2, in profits. As before, $A_2 = W_2 + L_2 + F_2 + P_2$. Next we could trace F_2 to the firms that we shall treat as a unit and call firm three. Here again the same analysis could be made: $A_3 = F_2 = W_3 + L_3 + F_3 + P_3$ and so on. Since $F_1 = A_2$, and $F_2 = A_3$, and so on, we may substitute for F_1 in the original expression, $W_2 + L_2 + P_2 +$ (instead of F_2) $W_3 + L_3 + P_3 +$ (instead of F_3) $W_4 + L_4 + P_4 + \ldots$ and so on. Hence the whole sum A_1 may be broken into $W_1 + W_2 + W_3 + W_4 + \ldots = W$; $L_1 + L_2 + L_3 + L_4 + \ldots = L$; and $P_1 + P_2 + P_3 + P_4 + \ldots = P$. Hence $W + L + P = A$.

A Number of Firms — Double Counting

Suppose that our economy were made up of only three firms: A, which sold to B, which in turn sold to C. This is an extreme simplification, of course, but the analysis is sufficiently complicated. Suppose further that in a certain period A sells $1 billion worth of goods to B, B sells $2 billion worth to C, and C sells $4 billion worth to consumers. What is the total income of the economy? Is it $7 billion or $4 billion? The answer depends on what happens to the goods bought by B and C. If they are entirely used up in the process of production, the economy's income, or, as it is usually called, the national income, is $4 billion. For if we were to trace back the $4 billion earned in sales by C, we should find that $2 billion of it (the sale from B to C) went for raw materials consumed in production, and that the other $2 billion went in the form of wages, salaries, rent, interest, and profit. Thus, firm C directly generates $2 billion in income. We should then find that firm B disbursed only $1 billion of income, including profits, the other $1 billion out of its total receipts being used to buy from A the raw materials which were used up in production. Finally, firm A had receipts of $1 billion, and since we assumed that it bought nothing from other firms, the whole of this sum was someone's income. For what was not paid out as wages, salaries, and so on, constituted the

profits of its owners. The national income would in this case be $4 billion — $2 billion disbursed by C, plus $1 billion disbursed by B, plus $1 billion disbursed by A.

This should warn us against the danger of double counting. The national income is not the sum of the sales receipts of all firms, when, as in our economy, certain firms supply others with goods that are used up in current production. The following data will give a rough estimate of the error we should commit if we were to forget this point. A crude measure of the total sales receipts of all business firms in the United States in a period of time can be secured by taking the figures for the total value of checks drawn outside New York City.[1] In 1945 total sales receipts corrected in this way came to about $570 billion. Our national income in 1945, adding up wages, salaries, rent, interest, profit, and so on,[2] came to about $160 billion, or less than one-third the amount of checks drawn. Thus the error from double counting would be enormous.

The National Income When Firms Invest

Sometimes, however, a firm buys materials from another firm which are not used up in production during the period. The purchasing firm may, for example, acquire machinery or raw materials from the supplying firm which it still has on hand at the end of the period. How does this process, adding to the stock of capital goods, which is called investment, affect our computations? To return to our original example, let us suppose as before that A sells $1 billion worth of goods to B, B sells $2 billion worth to C, and C sells $4 billion worth to consumers. But this time, instead of using up all the goods it buys from B, suppose C still has on hand $1 billion worth at the end of the period.[3] It therefore uses up only $1 billion worth of goods, and its costs for raw materials actually consumed are therefore only $1 billion. The rest of its spending for the products of B must not be regarded as an expense of current production, since the goods are still available for future production requirements. Firm C thus disburses sums that constitute income, not of $2 billion as before, but rather of $3 billion

[1] This measure is crude chiefly because some checks are drawn for the purchase of goods not currently produced, such as old houses or securities. Since most of the business in securities is carried on in New York City, we have omitted figures for that city.

[2] After subtracting business taxes and depreciation.

[3] Or, if it began the period with any, it ends the period with more; the additions to its stock being valued at $1 billion.

— the difference between its receipts and the costs for raw materials acquired from B. It pays out as wages, salaries, rent, and interest, and has left over for profits, a total of $3 billion. Firm B disburses as before $1 billion in the form of income payments, and likewise A generates $1 billion in income. The total income of the economy for this period is thus $5 billion: $3 billion from C, plus $1 billion from each of the other firms.

When all the products sold by A and B are entirely used up in the production of commodities sold by B and C, the national income is $4 billion. However, when $1 billion of the product supplied by B is not used up, but remains in existence, the national income is $5 billion; and the $1 billion worth of goods still existing at the end of the period are, within the boundaries of that period, to be regarded as *final products;* they are not used for further production *in the period.* Likewise, the goods sold by C to consumers are not used for further production, so they too are final products. But the $1 billion worth of goods sold by A to B are fully used up in the period in producing B's output; they are not final goods, that is, goods acquired for their own sake. Final products are thus goods added to the stocks of commodities kept by business firms or purchased by final buyers — that is to say, by consumers. When all of A's and B's output was fully used up in the period in question, the amount spent upon final products was only $4 billion, or the value of C's output; and that was the amount of the national income. But when C acquired $1 billion worth of goods to add to its stocks, then the sum spent upon final products was $5 billion. And when this occurred the national income was $5 billion. We may conclude, then, that the national income for any period equals the sum spent upon final products in that period.

The amount spent upon the products of any firm is equal to the value of that firm's output.[1] If a firm gets $1 million in sales receipts in a year, its output is valued at $1 million, and this figure is our only measure of the value of its output. Likewise, the sum spent upon all final products, not of one firm but of the whole economy, is equal to the value of all final products produced. Instead of saying that the national income equals the sums spent upon final products, we may therefore say that it equals the value of the output of final products in the economy.

[1] If a firm's product is added to its inventories, the value of its product includes the value of its addition to inventory, and that addition should be included in its sales receipts, as though it bought goods from itself. Essentially this is what the accountant does.

It will be noticed that the danger of double counting is avoided by directing attention to spending upon final products — those which are not used up in the period in facilitating the production of other goods.

Application to the American Economy

If we considered the national income in terms of final products, we would say that in a year when $100 billion is spent upon final products, the national income was $100 billion. But we may also consider the national income as the sum of the incomes earned in the year by wage earners, salaried employees, landlords, lenders, and owners of businesses; and this total also would amount to $100 billion. Thus the national income can be regarded from either of two points of view; it can be seen as the value of final products produced within the economy in a certain period, or as the amount of income earned by various kinds of income recipients. These two ways of measuring the national income can be illustrated as follows. In 1945 the total value of final products amounted to $197.3 billion. These final products were purchased by the following kinds of buyers: [1]

TABLE 28

Gross National Product by Types of Purchaser: 1945

(in billions of dollars)	
Consumers	104.9
Business firms	9.1
Foreigners	.3*
Government	83.0
Total	$197.3

* This figure is a net figure, our sales of final products to foreigners minus our purchases from them.

This, of course, does not mean that business firms bought only $9.1 billion worth of goods. They bought more nearly $300 billion worth, but they used up all but $9.1 billion worth in producing other goods. Thus they bought only $9.1 billion worth of goods that became final products for that year.

[1] As noted above, the present analysis disregards the distinction between the national income and the gross national product.

Since the value of the economy's final product for a year is equal to the income, we may break down the income into the nation's wages, salaries, rent, and so on, as in Table 29.

TABLE 29

Gross National Product by Distributive Shares: 1945

(in billions of dollars)	
Wages and salaries	114.5
Rents and interest	11.8
Profits (before taxes and charges for depreciation)	71.0
Total	$197.3

It will be noted that the above figures measure profits *before* payments of business taxes and allowance for depreciation. This sum, of course, does not measure what the businessman actually receives. His firm must pay business taxes to the government and make an allowance for the value of capital goods used up in the period, and these amounts must be subtracted from the profit figure above, in order to derive the net earnings of the owners of business firms. In 1945, the net figure came to $34.6 billion, the greater part of which was earned by farmers and other proprietors rather than by shareholders of corporations.

We may occasionally be interested not in the value of *all* final goods, but only in the value of certain final goods, for example those for government or for consumers. If so, the total sums spent on these final goods measure the income generated in producing them. If, for example, our economy produced $100 billion worth of final goods for consumers in one year, the incomes earned in producing consumers' goods would equal $100 billion.

Summary

In any period, the value of the output of final goods is equal to the economy's income. The argument by which this conclusion was derived may be restated in simple mathematical terms. Let us first consider a single firm which in a certain period of time has sales receipts amounting to A_1. In that same period it has to pay W_1 to wage earners, S_1 to salaried employees, L_1 to landlords and lenders

for rent and interest, and it pays X_1 to other firms for raw materials, equipment, and so on.[1] Suppose that the value of any changes in its stock of capital goods during the period is Q_1. Such changes could be brought about if it bought some new equipment or raw materials in excess of its current requirements, or if it reduced its stock of goods held in inventory. Then its profits for the period P_1 are $A_1 + Q_1 - (W_1 + S_1 + L_1 + X_1) = P_1$. Hence $A_1 + Q_1 = W_1 + S_1 + L_1 + X_1 + P_1$.

The payment of W_1 to wage earners represents income to them; likewise the payments of S_1 and L_1 represent income, and P_1 represents the income of the firm's owners. The total income directly generated by the firm's activities is thus $W_1 + S_1 + L_1 + P_1$. But since $A_1 + Q_1 = W_1 + S_1 + L_1 + X_1 + P_1$ we may as well write $(A_1 + Q_1 - X_1)$ for the total income directly generated by the firm.

We may proceed with the same analysis for the next firm in the chain, and the next, and so on. The income directly generated by the next firm is: $A_2 + Q_2 - X_2$; by the next is $A_3 + Q_3 - X_3$, and so on. Therefore the total income earned is: $A_1 + Q_1 - X_1 + A_2 + Q_2 - X_2 + A_3 + Q_3 - X_3 \ldots$ and so on. We may call this $A + Q - X$. In other words, $W + S + L + P = A + Q - X$.

Now let us examine the concept $A + Q - X$ more carefully. The first letter, A, represents all sales by business firms; X represents all sales by business firms *to other business firms*. Hence $A - X$ represents all sales by business firms *to non-business firms*. But the only buyers who are not business firms are consumers, the government, or foreigners. Hence $A - X$ represents all sales by business firms to consumers, government bodies, and foreigners. We have already defined Q as the value of the change in business firms' capital goods during the period; or in other words, as the value of inventory, equipment, and plant acquired by business. From this we may conclude that the sum of all wages, salaries, rent, interest, and gross profits equals the value of goods sold to consumers, foreigners, and government bodies, and in addition the change in the value of business firms' holdings of goods in inventory, plant, and equipment.

One more step must be taken. Goods sold to consumers, foreigners, and government bodies are final products, since they are not used during the period for further production in our economy but are

[1] X_1 is not necessarily the value of capital goods used in the period, but rather the total spending on the products of other firms.

bought by those whom we can regard as final buyers. Likewise, goods acquired by business firms in a period and not used up in the production process during that period, are, from the standpoint of that period, final goods. Regarded in relation to the period with which we are concerned, they too are bought by final buyers. Hence the sums spent on final products in a period equal the national income, including business taxes and depreciation. And since the amounts spent upon goods equal their value, we have shown that the value of the output of final goods equals the income earned in producing them.

If we are interested in one special class of final goods, such as goods for consumers or for government, the same equality holds, for in that case instead of adding all the A's for all firms, we should add them for those firms that sold goods to the government or to consumers. For a single sector of the economy, then, the value of the output of final goods equals the income earned in producing them.[1]

[1] Further study of the income data provided in this chapter may be helpful at this stage.

22

The Economics of the Firm: Summary

This chapter is intended especially for readers who do not require the detailed analysis of the firm presented in Chapters 5 to 21. While the following pages present the most important conclusions reached in those earlier discussions, obviously it is not possible to demonstrate in detail all of those conclusions within these few pages. The present chapter will therefore be useful in two ways: as a summary of Part Two for those readers who have studied Chapters 5 to 21; and as a brief discussion of the firm for those who have not.

A knowledge of the way in which the private business firm operates, and especially of the way in which it responds to changes in economic conditions, is necessary to an understanding of the capitalist economy. This is true because, as we have seen, the essence of capitalism is that the determination of price and output is in the hands of the individual business firm. To call our economy a "private enterprise economy" means simply that profit-seeking firms independently determine how much to produce and what price to charge. Each firm is free to produce as much or as little as it wants, and each firm is free to charge whatever price it wishes. Similarly, each firm may produce in whatever way it wishes, using the techniques of production that appeal to it, and so on. What happens in a capitalist economy, therefore, is no more than the sum of what happens in the individual firms of which it is composed.

The kinds of problems that confront us as economists relate to the total output of the economy, to the output of individual goods and services in the economy, and to the prices charged. We are not particularly interested as economists in whether books are bound in red

or green, but we are interested in whether the output of books is ten million or five hundred million a year. Therefore, we are invariably led back to the individual firm by such questions as the following: How many books are published? What is the output of the steel industry? What is the total value of goods and services produced in this country in a year? What is the relation between wages and prices? For, as we have seen, it is the individual firm that determines how much to produce and what price to charge. Hence in a capitalist economy it is essential, if we are to understand its operations, that we analyze the activities of the individual firm.

The firm in a capitalist economy is in business to earn profits. Under given conditions of demand and cost, it tries to do whatever is most profitable; and when these conditions change, it is frequently led to change its output and price, because by doing so it is able to keep its profits at a maximum. This search for profit is, in a sense, the highest common denominator of business firms. Apart from that common characteristic, there are enormous variations between firms. But the fact that they are all motivated by a desire for maximum profit permits us to generalize about the ways in which they change output and price in response to specified changes in economic conditions. The substance of Part Two of this book is simply an analysis of the way in which a business firm would alter its output and its price to meet a change in business conditions.

As we have indicated, firms are of all types. Some are very large, like the American Telephone and Telegraph Company; others are very, very small — so small, indeed, that they hire no labor at all, and their owner is employer and employee in one person. Some firms produce a commodity which meets with no competition from any other firm. Others produce a commodity which meets with the competition of perhaps five million producers. Some firms do not compete except in advertising; others compete in terms of price. Some firms produce commodities that buyers consider necessary; others produce only luxury articles. Firms are, therefore, extremely varied in nature.

But even though they vary so much among themselves, we can generalize about the ways in which they respond to changes in economic conditions. When there is an increase in the demand for their product, most of them react in much the same way. When they have to pay higher wages, most of them react in the same way.

When taxes are raised, when new methods of production are introduced, when competitors enter their industry, or when they have to pay more for raw materials, there is enough uniformity in their response to allow us to describe the results in terms of general laws. The generalizations at which we have arrived have been worked out on the assumption that the firm will determine precisely what course of action will be most profitable. We know that many firms do not make their calculations quite so scientifically, and that a large number of them simply follow conventional rules in adjusting to a change in demand or a change in the cost of raw materials. But there is every reason to believe that the results of precise calculation correspond very well with those reached in the real world by following such standard rules. Let us see some of these results. How would firms, seeking maximum profits, be expected to respond to changes in business conditions?

Changes in Demand: Short-Period Effects

It is not difficult in the real world to see how firms respond to changes in demand. The most striking change in the American economy between 1939 and 1945 was the enormous increase in the demand for almost every kind of goods. Demand more than doubled for many items, and the over-all figure for spending in 1945 was more than twice what it had been in 1939. What did firms do in response to this enormous increase in demand? They did two things. Almost all of them increased their output, and a large number of them raised their prices. Firms engaged in manufacturing and mining raised their production by 62.4 per cent between 1940 and 1945. Wholesale prices increased by approximately 35 per cent in that six-year period. Some firms were unable to raise their output because they could not find labor or raw materials. These firms were under even greater inducement to raise their prices, and the only reason they did not do so was government price control. There is no doubt, then, that an increase in demand will ordinarily lead a business firm to increase its output and to raise its price.

But the increases in price and output that follow an increase in demand are not always the same in size as those which occurred between 1939 and 1945. When firms are producing at a low level of output, and demand increases, they are likely to make their biggest response by increasing output; and the increase in price is likely to be

relatively small. It is only when their output is high that price is likely to rise significantly with increasing demand. The difference can be illustrated by comparing the change in output and price between 1933 and 1940 with the change in output and price between 1941 and 1945. Since wage rates increased at about the same rate in both periods, we can suppose that the effects of this change are comparable. The demand for goods increased very nearly as much between 1933 and 1940 as it did between 1940 and 1945. Thus, the change in economic conditions was about the same. What about the results? The increase in price was greater between 1940 and 1945 than between 1933 and 1940, and the increase in output was smaller. In the earlier period, wholesale prices increased by about 20 per cent, while between January, 1941 and 1945, they increased by about 33 per cent, in spite of the fact that price controls were in effect after 1942. In contrast, between 1933 and the end of 1940 the physical output of the economy just about doubled, while between 1940 and 1945 physical output increased by about 43 per cent. Thus, the demand for goods rose by about the same amount in the two periods, and so did wage rates; but in the earlier period output doubled, and wholesale prices increased by only 20 per cent, while in the later period output increased by only 40 per cent, and wholesale prices increased by 33 per cent. We must also remember that the increase in prices during the later period was held down by legislation. If price controls had been abolished, the increase in price would have been very much greater.[1] We may conclude, on the basis of this and other evidence, that a given increase in demand will lead to a greater increase in output and to a smaller increase in price when an industry is operating far below capacity than when the industry is operating near capacity.

But there are exceptions to this rule. For example, some prices did not change at all between 1932 and 1940, and others changed no more than once or twice in the whole period. Those articles whose prices remained steady were usually highly fabricated products, and they were frequently produced by industries that were dominated by a few large firms. In contrast, the prices of certain other commodities, such as wheat and cotton, changed almost daily.

[1] This is clearly shown by the fact that black-market prices were far above legal prices, and by the violent price increases during the few weeks in the summer of 1946 when price controls lapsed.

There are all sorts of reasons why, under certain circumstances, prices may not be altered. Sometimes firms simply abide by conventional prices, as for chocolate bars or cold drinks. Sometimes firms are reluctant to change prices because they know that their customers prefer price stability. Sometimes they are unwilling to change prices because they prefer not to introduce price competition into the industry or are unwilling to risk setting off a price war.

A full list of some of the commodities whose price moved only very rarely between 1926 and 1933 would be surprisingly large. Here are a few of them. The wholesale price of corn cereal was changed only once in this period of eight years, and that of ginger ale changed only twice. The wholesale price of a certain brand of dress shirts changed only twice, and that of certain kinds of stiff collars changed only once. The price of grain binders and cultivators, grain drills, three horsepower engines, thresher harvesters, hay mowers, corn planters, and indeed of most other agricultural implements changed only two or three times between 1926 and 1933. In the chemical industry, too, there were a number of commodities whose price did not change frequently. For example, liquid carbon dioxide sold at the same price throughout the period; hydrogen peroxide changed in price only four times; crude sulphur only three times; and sodium bicarbonate only three times between 1926 and 1933. It is obvious that the inflexibility of prices that this denotes existed in spite of the fact that demand changed often.

At the other extreme were such commodities as barley, corn, oats, eggs, apples, hops, cheese, lamb, coffee, gasoline, and crude rubber, whose prices changed in every one of the months between 1926 and 1933. Thus, our generalization about the effect of an increase in demand upon price and output must be interpreted with reference to the particular commodities in question. The price of some commodities does not change at all when demand increases, though the amount produced may rise sharply. For other commodities, prices increase along with increases in output. Thus, for most commodities, an increase in demand will lead to higher prices and increased output, but for some, price will not rise. The rise in price, if it occurs at all, will be greater and the increase in output will be less, for a given increase in demand, when output is close to capacity than when it is far below capacity.

The effects of a decrease in demand are just the reverse. When

demand falls, firms cut their output and frequently reduce their price. However, prices do not always fall, since the factors that bring about inflexible prices when demand increases are also operative when demand declines. In general, though, we should expect prices to decline when demand falls off. There is also good reason to expect the decline in price to be relatively great and the decline in output to be relatively small when demand falls from a very high level. When, in contrast, demand drops from a low level, we can expect the decline in output to be relatively large and the decline in price to be relatively small. To put this more concretely, let us suppose that the demand for a certain commodity declines severely. The firms that produce it will almost certainly be led to reduce their output, and — unless they operate in the special circumstances that cause inflexibility of prices — to reduce their price also. If originally they were operating close to capacity, they would reduce their price by a relatively large amount and their output by a relatively small figure. But if they were originally operating far below capacity, they would reduce output by much more than price. Finally, to the extent that prices are inflexible, a given change in demand will lead to the same change in output no matter what the original level of demand.

Changes in Costs: Short-Period Effects

Again, changes in cost conditions generally lead a firm to alter its output and to change its price. When firms have to pay higher wage rates or more for raw materials, a good number of them raise their prices. This situation was seen very clearly between 1933 and 1934, and again between 1936 and 1937, when wage rates increased substantially. For example, wholesale prices increased by about 15 per cent between 1933 and 1934. However, we should not attempt to account for the whole of this price change by the increase in wage rates in these months, since the demand for commodities also increased; and this, as we have already seen, would lead firms to raise their prices even if wage rates were unchanged. But there are grounds for expecting the rise in price to be greater when wages are rising than when they are constant. Perhaps no further demonstration is needed to convince the reader that prices are likely to move in the same direction as costs. When it raises the price of its product, a business firm generally explains this increase to its customers by pointing out that its costs have gone up. And its customers always look

forward to a reduction in price when they have reason to believe that costs have gone down.

Moreover, we should expect to find that firms will respond differently to changes in variable costs than to changes in fixed costs. When variable costs are increased, as happens when, for example, wage rates or the prices of raw materials increase, we expect prices to rise. But when fixed costs are increased, as happens when, for example, property taxes or the salaries of company officials are raised, then we do not expect increases in price. This conclusion may at first appear unrealistic, but more careful observation and analysis will show that it probably has a good basis in fact. There are two ways of looking at the matter. Either we can determine by strict profit calculation how a firm should respond to a change in fixed costs, or alternatively we can examine the procedures business firms normally use to determine price when they do not bother to go through these intricate calculations. But no matter how we analyze the problem, we arrive at the same result: a change in fixed costs does not lead the firm to vary its price.

Let us first see whether a firm that set price by calculating the level where its profits would be maximized would change its price in response to a change in fixed costs. Suppose that the results of its calculations are set down as follows:

TABLE 30

Relation Between Price and Profit Before Paying Fixed Charges

When Price Is	Profit, Before Paying Fixed Charges, Is
$ .65	$10,000
.85	10,700
1.00	11,600
1.10	11,800
1.15	11,850
1.20	11,820
1.25	11,760
1.40	11,600
1.50	10,300

Fixed costs by their very nature do not vary with price. If originally they are $5000 when the price is 65 cents, they are $5000 when the price is $1.50. Now the most profitable price is $1.15, since profit at

that level — $11,850 minus $5000, or $6850 — is higher than at any other price, and fixed costs have nothing to do with the case. If fixed costs are increased, as would occur when management's salaries were raised, would the price of $1.15 no longer be the most profitable? When fixed costs are increased to $6000, they remain at that level quite independently of what price the firm charges. If the price is $1.15, the profit is $11,850 minus $6000, or $5850. If the price is $1.20, the profit is $11,820 minus $6000, or $5820. The reader can easily satisfy himself that at no other price will profits be as high as at the price of $1.15. In other words, whether fixed costs are $1000, $5000, $10,000, or anything else, the most profitable price remains $1.15.

Apparently, however, many firms do not bother to make these calculations in order to determine their prices, but follow conventional rules instead. In order to see how a change in fixed costs will affect their prices, we have to examine these rules. The most common of them is this: Calculate the unit variable cost, then add a fixed percentage for overhead and a fixed percentage for profit, and the final figure is the price to charge. But here is the interesting fact: even though there should be a change in fixed costs, in many cases no change is made in the percentage added to variable costs for overheads.[1] If a firm is accustomed to adding 50 per cent for overhead to its variable costs per unit, it is likely to continue doing so, even though, let us say, its rent is increased, or, in the opposite direction, even though its bondholders have been persuaded to accept a lower interest return and the fixed costs of the firm are consequently reduced. This rule is by no means universal, but it appears to be common, and we may tentatively conclude that changes in fixed costs do not generally lead firms to change prices.

As we have seen, however, changes in variable costs usually do lead to changes in price, though if for one reason or another a firm tends to keep its price constant, then even a change in its variable costs will not lead to a price change. As an instance, in almost every industry there were large reductions in wage rates between 1929 and 1933, but great numbers of firms did not reduce their prices. And between 1933 and 1937, wage rises were common to almost every industry in

[1] The reader should not infer that this conclusion has been carefully confirmed by observation. The published evidence on the subject is very meager. A great deal remains to be done to fill in such gaps in our knowledge.

this country; yet there were many products whose prices did not rise. Nevertheless, there are many firms whose prices do respond to such changes. Consequently, for the economy as a whole, when costs alter we expect to find the price level changing in the same direction.

It is worth while to draw specific attention to the effect on price of improvements in technique. A firm does not adopt a new method of production unless it foresees a reduction in costs. But since, as we have seen, reductions in costs frequently lead to reductions in price, we may expect a price reduction to follow the adoption of improved methods of production. It is estimated that such improvements are made in our economy at a rate that leads to cost reductions of about 3 per cent per annum. On account of inventions we should therefore expect prices to decline by about 3 per cent per annum. But naturally such tendencies may be offset by the effects of changes in demand or in other costs.

Changes in costs also lead to changes in the opposite direction in output, as long as demand remains constant. If, for example, costs are reduced because of a reduction in wage rates, the firm usually is induced to lower its price. And unless the change in wage rates itself brings about a change in demand, this reduction in price will be associated with an increase in output. If costs are raised and in consequence the firm raises its price, it will also reduce its output unless the demand for the product goes up. But we must be careful in applying these generalizations to the real economy, because general changes in costs usually do not leave demand unaffected. If, for example, wage rates are increased, it is generally agreed that the demand also increases. For with higher wages, employees have more money to spend, and with increased spending power, we may be sure that they will increase their spending upon consumers' goods. The increase in costs thus brings about an increase in demand. And while an increase in costs would by itself lead to higher prices and *lower* output, the increase in demand would lead to higher prices and *higher* output. The total effect upon output and price thus depends upon the relation between the change in cost and the change in demand. If demand increases substantially, total output may be increased. If demand increases only slightly with the increase in wage rates, total output will probably fall. However, whether demand increased by much or little, price will certainly increase. But we must postpone our detailed consideration of this problem to Part Four.

The short-period effects upon price and output of a change in any of the economic variables can be summarized as follows: when demand changes, price and output move in the same direction. When there is an increase in the demand, most firms increase both output and price, though some increase only output. And when demand declines, most firms reduce price and output, though some firms leave price constant. When demand is constant, changes in cost usually lead to changes in price in the same direction and to changes in output in the opposite direction, though some firms may not vary their price in response to the cost change and hence would not vary output either. Thus, when costs rise, most prices rise, and if demand is constant, output falls. One exception to this rule should be noted: if the change in costs affects items that enter into fixed or overhead costs rather than items that constitute variable costs, firms may not be induced to alter price or output. We must note carefully, however, that our conclusion applies only when the change in cost does not affect demand. For that reason, it is generally not applicable in the real world, except when properly modified.

Changes in Demand and Cost: The Long Period

The long-period effects of a change in demand or costs also deserve consideration. When demand increases, profits rise. Higher profits may induce new firms to enter the industry and existing ones to expand. With the entry of new firms and the expansion of old ones, price will eventually fall toward its original level; and as it falls (if demand remains constant) output will rise. Thus, while the short-period effects of an increase in demand are to raise price and increase output, in the long run price tends to fall again toward its original level while output increases even further. A reduction in costs will have the same effect upon the profits of firms already in the industry, if the demand does not also fall. With increased profits, we should expect new firms to enter the industry and existing firms to expand their capacity. Consequently price should fall further and output should rise. Thus, the short-period effects of a reduction in costs are to reduce price; and, assuming that demand is unchanged, to increase output. The long-period effects are to reduce price and to raise output even more. However, a general reduction in costs would not necessarily bring about an increase in profits, since demand might also be reduced. If so, there would be no increase in the number of firms in

the industry; and in the long period, price and output would not be very different from what they were in the short period.

The Pattern of Production and Buyers' Wants

Industries in the United States are enormously varied, ranging all the way from those that are perfectly competitive to those that are monopolistic. Perfectly competitive industries are composed of an immense number of firms all producing a standardized product. Because it is standardized, consumers are completely indifferent to the source of this product; that is to say, they are as willing to buy from one producer as from any other, provided that the price is the same. A monopoly is an industry composed of only one firm. There are not many perfectly competitive industries in this economy, nor are there many where pure monopoly rules. Most industries fall somewhere between, in a category we have called monopolistic competition or oligopoly. The fact that our economy has this mixed character affects the pattern of production. If all our industries were perfectly competitive, or if all were monopolistically competitive (with the same elasticity of demand for the product of each), then the relative outputs of individual commodities would have this character; the greater the demand for any commodity, the larger would be its output.[1]

If we were to rank commodities in order of the size of the demand for them, then the commodity at the lowest end of the scale would be produced in the smallest quantity, the commodity for which the demand was next largest would be produced in a somewhat larger quantity, and so up the line — until finally, at the top, that commodity which was most wanted, in relation to the resources needed for producing it, would be produced in the largest quantity. But with monopoly, output tends to be smaller than with perfect competition under the same conditions of demand and cost; and with monopolistic competition, output is somewhere between these two extremes.[2] In other words, if the total demand for commodities A and B is the same, and if the conditions of cost are also the same, then the output of commodity A and commodity B would be equal only if the degree of monopoly in each industry were the same. But if commodity A is

[1] Allowance would also have to be made for differences in the costs of producing these goods. This could be done most conveniently by measuring them in units such that, no matter what the commodity, the cost of production was the same.

[2] It is assumed that generally the elasticity of demand is greater for the product of a firm in a monopolistically competitive industry than for the product of a monopoly.

produced by a monopoly and commodity B is produced by a perfectly competitive industry, then the output of commodity A will be less than that of commodity B. Hence, when some industries are monopolistic and others are competitive, equal demand does not yield equal output. The two outputs are somewhat unbalanced, for the output of A is less than the output of B, even though the demand for the two is the same.

If the demand for the products of all industries were exactly the same, the monopolistic industries would produce least, while the perfectly competitive industries would produce most. Naturally, the same forces operate when demand differs from industry to industry. The pattern of production is not closely adjusted to consumers' wants, as it would be if all industries were either perfectly competitive or equally monopolistic. Instead, the pattern of production reflects consumers' wants only in part; it also reflects the degree of monopoly in each industry. Other things being equal, the greater the want for the commodity, the larger the output; but, to distort this result, the greater the degree of monopoly in the industry, the smaller the output. Yet it is not logical to blame monopoly for so faulty a pattern of production; indeed, it would be just as sensible to blame competition. The maladjustment occurs, not because of either monopoly or competition, but rather because both exist in the same economy; that is, because competitive industries, oligopolies, and pure monopolies exist and function side by side. It is because industries differ, not because they are monopolistic or competitive, that our total output does not reflect precisely the relative strength of our desire for various commodities.

The unequal distribution of income is another most important factor in accounting for a divergence between the actual pattern of production and the pattern that would conform most closely to the real wants of consumers. A person whose income is $500 a year is not able to indicate on the market any but his most pressing needs; one with $500,000 a year can indicate his very slightest want, and it is this indication that matters to business firms. The pattern of production, then, is biased. It responds much more strongly to the least pressing wants of those in the high-income brackets than it does to the relatively urgent wants of those in the lower-income group.[1]

[1] This is not the only reason for the divergence. The interested reader should refer to Chapter 20 for a discussion of some others.

Since the pattern of production does not reflect the relative strength of our wants for different commodities, steps are taken in most economies to alter that pattern. Because the tendency is to produce too little of those goods offered by monopolies and too much of those produced by the more competitive industries, the controls should obviously expand the output of industries where the degree of monopoly is high and, if necessary, reduce that of the more competitive industries. Something like that has been done in this and in other economies. Certain monopolistic industries are subject to government regulation of price. For example, the electric light and power industry can charge only those rates which have been approved by the Public Service Commission of the state. As a consequence, power rates are lower, and the output of electric power is somewhat higher, than would otherwise be the case. Railway rates, telephone and telegraph charges, and the prices of certain other monopolies are also subject to state regulation. The regulation is almost always directed toward reducing prices, which means toward increasing output.

Steps are also taken to reduce the output of those industries in which the conditions of production are highly competitive. These can be most clearly seen in certain agricultural industries which have been subject in recent years to some government control. Under the program comprised in the Agricultural Adjustment Acts during the nineteen-thirties, measures were taken to reduce the output of most agricultural products. The motive for doing this may have been to raise the income of farmers through raising the price of their products. But one consequence of these actions was to reduce the output of many products that were produced under highly competitive conditions.[1] In an economy in which industries range from the perfectly competitive to the perfectly monopolistic, production does not conform closely to wants where there is no government regulation. With government control of some monopoly prices, and government measures to raise competitive prices, the pattern of production is made to conform very much more closely to the wants of buyers.

The Value of Output and Income

The amount of money spent on the new products of our economy in any period of time is equal to the amount of money earned in pro-

[1] Output did not actually fall in many cases, for demand increased; rather, the amount by which output rose was smaller than it would have been in the absence of these controls.

ducing them. This most important conclusion depends upon the following identities: the profit of a firm equals the difference between its receipts and its costs, or, to put this in another way, the receipts of a firm are equal to the sum of its costs and its profits. But the receipts of a firm are also equal to the amount of money spent upon its products. Hence the total amount spent upon the new products of our economy in any period is equal to the total costs plus the total profits of all firms in the economy. To develop these fundamental equalities further, it is necessary only to trace all the sums included as costs. Some of them very obviously enter directly into income. For example, all sums paid out as wages constitute the income of wage earners. Likewise, money paid as salaries, or interest, or rent to landlords is income. What remains of the cost payments and other expenditures are the sums paid to other firms for raw materials, new equipment, power and light, and so on. But the money paid to these firms equals their total receipts from sales in the period. They, too, have wages, salaries, rent, and interest to pay, and a relatively large part of their total receipts will be used to meet these expenses. And these expenses likewise constitute income for their recipients. Thus, a large part of the amount paid over for raw materials and new plant and equipment is seen, at the next step in the analysis, to flow out to wage earners, salaried employees, and so on, as income. The expenses that do not directly count as the income of the recipients are again expenditures for such things as raw materials, new equipment, and so on. Further, the receipts of the firms that sell raw materials, equipment, and power, can at this stage be broken down into certain dispersals that constitute income to those that receive them, together with profits for the owners of the firms, and expenditures for raw materials, new equipment, power and light, and so on. If every dollar spent on new products is followed down step by step, it will be found to end up as somebody's income.

We must be careful, however, not to count the value of the goods purchased by one firm from another both at the time of the purchase and again after they have been processed when they are purchased by still another user. Thus, if one firm spends $20,000 on tanned leather in a year and uses it to produce shoes which are sold to consumers, we must not count the tanned leather twice. If we count it as a part of the shoe purchase, we should not count it a second time when it is sold by the tannery to the shoe manufacturer. Making

the necessary allowance for such double counting, we conclude that the total value of the goods and services produced in a period of time is equal to the amount of money earned in producing them. Difficulties in connection with taxes and depreciation charges would take us rather more deeply into the concept of national income and gross national product than it seems advisable to go in an introductory course. Overlooking these difficulties, we conclude that when, for example, $200 billion worth of new goods and services are produced in our economy in the course of a year, the income of those engaged in their production is equal to the same sum — $200 billion a year. If only $60 billion worth of new goods and services are produced in the course of a year, then the incomes of those engaged in producing come to only $60 billion in the period.

Summary

The individual firm is the critical unit in a capitalist economy. Its reaction to changes in economic conditions determines the reaction of the whole economy. Part Two of this book has traced the firm's response to various changes. We have seen that when demand increases, both output and price are raised, and when demand declines, both output and price fall off. We have seen, too, that the results of changes in demand are not everywhere the same throughout the economy, but that in industries which are constituted in certain ways, price inflexibility is the rule. In such industries, changes in demand have no effect on price, but a large effect upon output. Furthermore, the response to changes in demand varies between depression and prosperity. We have also seen how individual firms react to changes in costs. Knowledge of these reactions, together with an understanding of the relation between income earned and the value of goods and services produced, enables us to tackle the problem which is perhaps the most interesting, and certainly the most important, in our economy — the problem of how the national income and the level of employment are determined. This we shall do in Part Four, but first it will be helpful briefly to survey the financial sector of the economy.

SUGGESTIONS FOR FURTHER READING

Many recent texts give clear analyses of the pricing policy of the firm. The more important discussions of the subject are quite technical. The following references are, however, clear and good.

Hall, R. L., and Hitch, C. J. *Price Theory and Business Behaviour.* (Oxford Economic Papers, No. 2; May, 1939.) Oxford (Eng.): Oxford University Press.

An important analysis of methods actually followed by firms in determining their prices. Not difficult reading.

Machlup, Fritz. "Evaluation of the Practical Significance of the Theory of Monopolistic Competition," *American Economic Review* (The Journal of the American Economic Association), XXIX (June, 1939), 227–36.

——— "Marginal Analysis and Empirical Research," *American Economic Review,* XXXVI (September, 1946), 519–54.

In the first article the author sets out very clearly the nature of monopolistic competition and oligopoly. In the second, he examines the problem of whether marginal analysis gives a realistic account of pricing policy.

Meade, J. E., and Hitch, C. J. *Introduction to Economic Analysis and Policy.* New York: Oxford University Press, 1938.

Part 2 of this excellent book is about the policies of the firm in conditions of competition and monopoly.

Sweezy, P. M. "Demand Under Conditions of Oligopoly," *Journal of Political Economy,* XLVII (August, 1939), 568–73.

The "kinked" demand curve was developed in this short essay.

United States Government. *Economic Concentration and World War II. Report of the Smaller War Plants Corporation to the Special Committee to Study Problems of American Small Business, United States Senate.* Washington: Government Printing Office, 1946.

This is a useful compilation of data on the structure of various industries.

United States Government: Temporary National Economic Committee. *Investigation of Concentration of Economic Power. Hearings before the Temporary National Economic Committee, Congress of the United States.* Washington: Government Printing Office, 1939–41.

——— *Investigation of Concentration of Economic Power. Monographs for the Temporary National Economic Committee, Congress of the United States.* Washington: Government Printing Office, 1939–41.

The TNEC Hearings and Monographs. Part 5 of the Hearings, "Monopolistic Practices in Industries," is a useful record, but there are many others. Monographs Number 1, "Price Behavior and Business Policy," and Number 21, "Competition and Monopoly in American Industry," may prove helpful in providing factual data.

PART THREE

Money and the Interest Rate

Introduction

THE NEXT THREE CHAPTERS will be devoted to an analysis of the workings of the financial sector of the economy — the banking system. The firms that make up that sector do not engage in production in the usual sense. Instead, their business is in making money, in the literal meaning of that phrase. And their peculiar significance in the economy arises from the fact that they play a very important role in determining the rate of interest and in setting the conditions on which money is lent. An understanding of the banking system requires an analysis of the activities of commercial banks and the Federal Reserve System. This analysis, together with that presented in Part Two, constitutes the foundation on which we are able, in Part Four, to build the Theory of Employment.

23

Money and Commercial Banks

LET US TRY to picture an economy in which goods are bartered directly for one another. The housewife who has too many eggs but wants a pound of butter has to find someone who has a surplus of butter and wants eggs. Once such a person is discovered, the exchange of butter for eggs can perhaps be arranged on satisfactory terms. But imagine the problem of getting a haircut under such a system. You would have to find a barber who wanted the very thing you were able to provide — which might be bananas. But if you knew only one barber who could cut your hair exactly the way you wanted it, and he did not want bananas, you would have to trade your surplus for something he would accept, or for something which in a later deal you could exchange for something he would accept. Each transaction would be like a gigantic version of the well-known parlor game in which you are asked, for instance, to identify a tropical fruit with a yellow rind that grows in clusters. By changing the first letter of the word to a C and rearranging the others, you form a word for an Oriental potentate; then by changing the second letter of that word to an A and again rearranging the others, you have a word for a malady of antarctic birds — and so on, until finally, after suitable transpositions and rearrangements, you end up with H–A–I–R–C–U–T. The problem of actually swapping bananas for a haircut would be hardly less complicated if we had to acquire goods by means of barter.

Obviously, such methods of carrying on trade would be workable only under extremely primitive conditions. In a modern economy

the range of commodities and the range of wants is so wide that if goods had to be exchanged by barter, it would be extremely difficult to locate two persons whose wants and resources precisely complemented each other. We do not trade goods directly for one another in a modern economy, and barter is of no significance. Instead we use money. When we want to buy anything, we find that its price is quoted in money and we must pay for it with money or a promise to provide it at a later date. Likewise, if we sell anything, we express the price in money and we receive money for it. We use money as a means of payment. Obviously we can exchange goods very much more swiftly and easily in this way than by bartering them.

Money transactions are all-pervasive in our economy. We purchase and sell all goods, borrow and lend, pay wages, salaries, and rents with money. But where does money come from? How is it produced? Does it affect the operations of our economy in any way except as it facilitates trade? These questions obviously demand an answer if we are to understand how the economy works. We shall find that the banks have a great deal to do with our money supply. Furthermore, we shall find that money is not simply a means of transmission — a device to facilitate exchange — but that it plays an important role in influencing the operations of the economy. The institutions that supply money help to determine the level of activity in the economy, chiefly through their influence upon the interest rate, which as we shall see later is an important factor in determining the national income. Thus the study of money, banks, and the interest rate is essential to an understanding of the economy.

Kinds of Money

First it is necessary to describe precisely what money is. Perhaps this seems unnecessary because the answers appear to be so obvious. But you can dispel any such belief by examining carefully the contents of your own pocketbook. Even when there is not much in it, the variety can be surprising. First look at the coins — the nickels, dimes, pennies, and quarters. You should notice that there is no statement on them that the money is "backed," and that there is no "promise to pay." Yet, in spite of the absence of such formulae, we know that these coins buy things just as readily as any other kind of money.

How much of this kind of money exists? If we had added together

the value of all the coins in all our pockets, in penny banks, the sugar jars, and in the tills of business firms, it would have amounted to a little over one billion dollars in March, 1946. This, of course, is more than enough for one pocket, but, as we shall see, it is a relatively insignificant sum in comparison with the amounts of other types of money.

Now let us examine the bills, which range in face value from one dollar to ten thousand dollars. They are of three main types: Silver Certificates, Federal Reserve Notes, and United States Notes. A small legend at the top center of the face side of the bill identifies it as being of one of these three kinds. There are certain other types of bills also, but their circulation is very restricted. Measured by the amount in circulation, the Federal Reserve Notes are much the most important kind of paper money. Those in pockets, jars, and tills (that is, in circulation generally), at the end of March, 1946, had a face value of over $23 billion. Notes of this kind are issued by the Federal Reserve Banks on conditions to be described later. The face of such a bill bears the legend: "This note is legal tender for all debts, public and private, and is redeemable in lawful money at the United States Treasury, or at any Federal Reserve Bank." In addition, the note states that "the United States of America will pay to the bearer on demand five dollars" (or the face value of the bill). These legends most certainly raise questions. For example, what is lawful money? In what form will the United States pay to the bearer the five dollars which he may demand? With another five-dollar bill?

Silver Certificates, of which there were about $2 billion worth in circulation in March, 1946, have a somewhat less ambiguous set of statements on their face. They say that "this certificate is legal tender for all debts, public and private," and "this certifies that there is on deposit in the Treasury of the United States of America five dollars [or the face value of the bill] in silver payable to the bearer on demand."

The third kind, United States Notes, were at one time the most important, as judged by the amount in circulation, though at present there are only about $300 million worth of these bills outstanding. The legends are rather like those on Federal Reserve Notes: "This note is legal tender at its face value for all debts, public and private" and "The United States of America will pay to the bearer on demand five dollars." There are also other kinds of bills in existence: for example, Federal Reserve Bank Notes and National Bank Notes. But

these types make up a very small part of the total supply of currency, and they are gradually disappearing from circulation.

The face value of all bills and coins in circulation came to about $28 billion at the end of March, 1946. In other words, we possessed about $28 billion worth of paper money and coins at that date, or about $200 worth for every person. This figure is very much higher than it was at the end of 1938, when it stood at about $7 billion, or $50 a person.

As consumers, and perhaps as wage earners, we do most of our trading with money of these kinds — coins and bills. But we should not be misled into supposing that most of the transactions in the economy are financed in this way. When a manufacturing concern buys raw materials, when an investment bank sells securities, or when the government pays for the construction of a large dam, payment is not effected by the transfer of bills. Instead, a check is drawn by the purchaser in favor of the seller. In terms of volume, the kind of money which is thereby transferred is very much more important than that transferred in the form of currency. For example, in March, 1946, in comparison with the $28 billion in currency outstanding, there was in existence almost $100 billion in what we may for the moment call check-money. And not only is the amount of check-money far in excess of the amount of currency, but check-money, because it circulates much more rapidly than currency, supports an even larger percentage of all transactions than might be supposed from a mere comparison of amounts.

Demand Deposits

Since the volume of check-money is so much greater than that of all other kinds of money combined, we must devote special attention to it. When someone writes a check, he instructs his bank to transfer a part of his demand deposit, or checking account, to another person. The check is simply an order to transfer funds from a demand deposit. Check-money, as we have called it, is nothing more than a demand deposit or a checking account in a commercial bank. But while we have called demand deposits a form of money, they are not always the same as currency money. The acceptability of currency is complete. The seller would never question a purchase financed by either a bill or a coin. An order to transfer a bank deposit is, however, less readily acceptable. When an ordinary individual writes a check, it

is usually necessary for him to identify himself before the seller is willing to accept it. But it is easy to overemphasize this difference. While it is sometimes awkward for those of us who carry on only a small amount of our business by check, there is no difficulty for a firm which habitually finances most of its business in this way. No one questions a check drawn by the United States Steel Corporation. The acceptability of an order to transfer demand deposits by a reputable business firm is virtually as good as the acceptability of Federal Reserve Notes.

There are other kinds of wealth, however, which are much less acceptable for ordinary transactions, and for that reason they are not treated as money. Government bonds or a personal savings account cannot be directly used to finance the purchase of, let us say, an automobile. Time deposits (savings accounts), unlike demand deposits, cannot be transferred by check, nor can they be easily withdrawn from the bank in the form of currency. Since savings deposits are not readily transferable, and consequently sellers are unwilling to accept them in payment, they are not usually counted as a form of money.[1] But there is, of course, no sharp break between the acceptability of certain kinds of assets and others. Savings deposits are different only in degree from demand deposits. Their acceptability is somewhat lower; that is all. We could include savings deposits as money; for that matter we could include government bonds. But the usual definition of money in this country draws the boundary between demand deposits and savings deposits. However, the line is not inflexible. In the immediate post-war period in Europe, American cigarettes were at least as acceptable as any other type of asset, and frequently they were more so than French francs, Italian lire, or German marks. Thus we might logically include cigarettes as a part of the money supply of those countries, at least temporarily; and if there were instances when Italians would not accept lire, we should have to exclude lire in computing the amount of "money" in Italy. In normal circumstances, however, we include only currency and demand deposits in the money supply.

Reserves Against Demand Deposits

A demand deposit in a bank is, of course, an asset or an item of

[1] Just where the line should be drawn between money and other assets is a problem which must be settled somewhat arbitrarily.

wealth for the depositor. From the point of view of the bank, however, the deposit is a liability, because the bank must pay it on demand. Unlike the promise to pay on the Federal Reserve Note, which does not specify the form in which payment is to be made, a demand deposit obligates the bank to make payment either by giving currency to the depositor or by transferring the deposit to another person.

Since the bank may at any time have to meet the claims of its creditors or depositors, it must keep sufficient assets on hand, and they must be in a form suitable for meeting these claims. The bank, if it were very prudent, might keep a reserve of liquid assets large enough to meet *all* of its claims at any one time; but long experience has demonstrated that such reserves would be unnecessarily large; for except in very unusual circumstances, only a small part of these claims is pressed each day. Only when there is a run on the bank do all depositors ask to withdraw all their deposits. Otherwise they draw out only a small part of the total in any one day. And to make matters even more convenient for the bank, as some depositors are making their withdrawals, others are adding to their accounts. Thus, a bank with deposit liabilities of $10 million may have to meet daily claims of only $150,000 — and while some of its depositors are demanding currency or writing checks on their accounts, others are likely to be in the process of depositing currency and checks drawn on other accounts. Consequently, the reserves which even a very prudent bank has to keep against its demand deposits are normally much lower than the deposits themselves. Generally the law requires reserves that are only a small fraction of the bank's deposits, and often even that fraction is above the level which prudence alone would dictate.

In this country the bulk of the banking business is carried on by banks which are members of the Federal Reserve System. Member banks, as they are called, must keep reserves that are usually a great deal higher than would be thought necessary on grounds of safety alone. In the last few years they have had to hold reserves [1] of from 14 to 26 per cent of their demand deposits, the exact figure depending on where the bank is located. But these figures are subject to change. For instance, between 1917 and 1936 member banks were required to hold reserves of only 7, 10, or 13 per cent of their demand deposits.

[1] The percentages here given are minimum values. The bank may hold reserves in excess of these figures.

Now, since these reserves have to be kept in order to meet possible withdrawals of demand deposits, one would expect that they would have to be in the form of currency. But strange though this may seem, the law does not recognize the bank's holdings of currency as a part of its required reserve. Instead, reserves have to be kept, for reasons which will become clear later, not in the form of currency, but rather as deposits which the bank itself maintains in a Federal Reserve Bank.

The Federal Reserve Banks — Preliminary

What, then, are these Federal Reserve Banks? Essentially they are banks *for* banks; they perform almost the same functions for a commercial bank that a commercial bank performs for the individual depositor. There are twelve of them in the country, situated in large centers of population such as New York, Boston, Philadelphia, Chicago, San Francisco, and so on. Every commercial bank which is a member of the Federal Reserve System is affiliated with one of these Federal Reserve Banks. The Federal Reserve Banks themselves have little direct contact with the public. Most of their business is carried on with commercial banks or with the government, rather than with ordinary business firms. Although they are owned by the commercial banks, they are public institutions and play an important part in controlling the activities of commercial banks.

By using facilities provided by the Federal Reserve Banks, the banking system can quickly transfer deposits between individuals who have accounts in different banks and in different places. We can illustrate the process in this way: If someone with an account in the Cambridge Trust Company draws a check for $1000 in favor of a person with an account in a bank in New Haven, the latter may either deposit the check in his bank or take payment in currency. If he deposits the check, the New Haven bank has the check for $1000, and it owes $1000 more to its depositor. It sends the check for collection to the Federal Reserve Bank of Boston. The Reserve Bank on receiving it reduces the Cambridge Trust Company's deposit with the Reserve Bank by $1000 and increases that of the New Haven bank by the same amount. Finally, the Cambridge Trust Company reduces the account on which the check was drawn by $1000. To summarize: When a check is written on Bank A and given to someone with an account in Bank B, changes are made in four accounts. The account

of the payor decreases; that of the payee rises; Bank A's deposit account in the Federal Reserve Bank declines, and Bank B's increases. The transfer of funds by check between individuals with accounts in different banks thus causes a transfer of the deposits kept by commercial banks in the Federal Reserve Banks.

It can thus be seen that when an individual writes a check on his account and gives it to an individual with an account in a different bank, the reserves of one bank are reduced while those of the other bank are increased. For, as we have seen, the transfer of this sum is carried out by reducing the deposit which one bank keeps in the Federal Reserve Bank and by increasing the deposit which the other bank keeps. These deposits, as we have already seen, constitute the reserves which commercial banks are required to hold against their deposit liabilities.

Thus, a commercial bank uses its deposit with the Federal Reserve Bank in the same way that an individual uses his deposit in a commercial bank. One bank can transfer funds to another by having the Federal Bank reduce its account and increase that of the other bank. Conversely, it can have the Federal Reserve Bank collect for it from another bank by adding the amount to its reserve deposit, subtracting it from that of the paying bank. Also, just as you may go to your bank and draw currency from your deposit, so may a commercial bank draw currency against its deposits in the Federal Reserve Bank.

The Activities of a Commercial Bank

With this sketch of the organization of the banking system, we are in a position to begin our analysis of how demand deposits are created. To do so we shall have to examine the operations of commercial banks. We may best do this by following the activities of one of them, paying careful attention to the effects of its actions upon its assets and its liabilities. We shall summarize these effects by drawing up a simple balance sheet, or statement of assets and liabilities, after analyzing each operation.

To begin with, let us suppose that a number of citizens of a certain town decide to organize a bank. We shall assume that they have a certain amount of cash in their possession with which to begin — let us say $500,000. After they have gone through the legal formalities, their bank is chartered and becomes, we shall suppose, a member of

the Federal Reserve System. At this stage the balance sheet shows assets of $500,000 in cash, and liabilities (or, properly, capital) consisting of the rights which the owners of the bank could exercise if it were to be liquidated (the value of these rights is called capital stock) of $500,000. Accordingly:

Balance Sheet 1

Assets		Liabilities and Capital Account	
Cash	$500,000	Capital Stock	$500,000

Almost everything the bank does will have an effect upon two or more items in the balance sheet. In order to facilitate comparison, the items that are changed by the step under discussion are marked with an asterisk.

The bank now acquires buildings and equipment with which to carry on its business. If $50,000 is paid for this property, the balance sheet then looks like this:

Balance Sheet 2

Assets		Liabilities, etc.	
Cash	$450,000 *		
Property	50,000 *	Capital Stock	$500,000

The bank is now open for business. Certain residents of the town decide to deposit $100,000 in the bank, and so it has more cash. But it is also liable to these depositors for the amount of their deposits. Hence, although its assets in the form of cash have increased, so have its liabilities in the form of demand deposits. After this operation, the balance sheet is as follows:

Balance Sheet 3

Assets		Liabilities, etc.	
Cash	$550,000 *	Demand Deposits	$100,000 *
Property	50,000	Capital Stock	500,000

Since the bank is a member of the Federal Reserve System, it has to maintain a deposit in the Federal Reserve Bank of its district equal, we shall suppose, to at least 20 per cent of its demand deposits — or, in other words, to at least $20,000 — as a reserve against its demand deposits. Probably, however, the bank would wish to maintain a larger reserve than this with the Federal Reserve Bank, for its deposits may increase further. For purposes of illustration, suppose it deposits $50,000 in cash in the Federal Reserve Bank as a reserve against its demand-deposit liabilities. Following this operation, the balance sheet will read:

Balance Sheet 4

Assets		Liabilities, etc.	
Cash	$500,000 *	Demand Deposits	$100,000
Property	50,000	Capital Stock	500,000
Deposits in Federal Reserve Bank	50,000 *		

Note that while the bank was legally required to have reserves of $20,000, or 20 per cent of its demand-deposit liabilities, it actually has deposits of $50,000, or excess reserves of $30,000.

Now, if the bank next purchases $450,000 worth of government bonds for cash, its balance sheet is changed to the following:

Balance Sheet 5

Assets		Liabilities, etc.	
Cash	$50,000 *	Demand Deposits	$100,000
Government Bonds	450,000 *		
Deposits in Federal Reserve Bank	50,000	Capital Stock	500,000
Property	50,000		

One of the depositors now writes a check for $10,000 in favor of a person with an account in a different bank. The check is collected, as we have already seen, through the Federal Reserve System, and the bank's reserves in the Federal Reserve Bank will be reduced by the amount of the check. The balance sheet is therefore like this:

Balance Sheet 6

Assets		Liabilities, etc.	
Cash	$50,000	Demand Deposits	$90,000 *
Government Bonds	450,000		
Deposits in Federal Reserve Bank	40,000 *		
Property	50,000	Capital Stock	500,000

Next, a depositor receives a check for $60,000, which he wants added to his account. The check is drawn on a different bank, and so our bank collects the sum from the other bank through the Federal Reserve Bank. The balance sheet is thus:

Balance Sheet 7

Assets		Liabilities, etc.	
Cash	$50,000	Demand Deposits	$150,000 *
Government Bonds	450,000		
Deposits in Federal Reserve Bank	100,000 *		
Property	50,000	Capital Stock	500,000

If one of our bank's depositors writes a check to a person with an account in the *same* bank, no transfer of funds between banks is required. Since the bank has merely to lower one of its demand-deposit liabilities and raise the other, no item in the balance sheet is affected.

Let us now assume that the bank, anticipating withdrawals of cash by some of its depositors, calls on the Federal Reserve Bank for $20,000 in cash. The Federal Reserve Bank delivers the sum in bills, and the accounts are now altered as follows:

Balance Sheet 8

Assets		Liabilities, etc.	
Cash	$70,000 *	Demand Deposits	$150,000
Government Bonds	450,000		
Deposits in Federal Reserve Bank	80,000 *	Capital Stock	500,000
Property	50,000		

It should be noticed at this point that the bank's reserves — its deposits in the Federal Reserve Bank — fall when the bank asks for cash or when one of its depositors writes a check payable to someone with an account in a different bank. Likewise, its reserves are raised when it deposits cash in the Federal Reserve Bank, or when one of its depositors receives a check drawn on a different bank and deposits it to his own account. There are still other ways, as we shall see, by which the bank's reserves may be affected, but for the present it is essential to mark only these.

Banks earn most of their money in two different ways: one by receiving interest payments on the government bonds or other securities they hold; the other by lending money at interest to private borrowers. The more securities the bank holds and the greater the amount it lends, the more it receives as interest. Hence a bank is generally anxious to increase its holdings of securities and to lend more — provided that it feels sure of being repaid. What, then, sets the limit on the amount it can lend or the amount of securities it can purchase? To answer this we shall have to see what happens when it lends.

When a bank lends money, the borrower does not ordinarily stuff his pockets with five-dollar bills, thank the manager, and leave. Rather, the bank usually accepts the borrower's note promising repayment at a certain date and gives him, not currency, but the right to draw up to the amount of the loan minus the charge for interest. In other words, the lending bank gives the borrower a demand deposit on which he may write checks, or, if he should wish, secure cash. It seems, then, that a bank can lend without limit, for, after all, it takes only a very small amount of work by a bookkeeper to set up an account for the borrower. But there is a limit, or rather there are two limits. The obvious one is that as the bank lends more and more its demand deposits increase. And since it is required by law to hold reserves equal to at least a certain percentage of its deposits, the amount of those reserves will ultimately set a limit on the amount it may lend. Thus, when its deposit liabilities are $150,000, as they were in our last balance sheet, it must have deposits in the Federal Reserve Bank of at least $30,000, if the reserve ratio is 20 per cent. Since its reserves are now $80,000, this limit is not very close, but it could be reached if the amount of lending were great enough. There is, however, another limit on lending that operates much more quickly, and this we must examine with very great care.

When a bank lends, it adds the amount of the loan to its demand-deposit liabilities, since the borrower is allowed to draw on his account to that amount (minus interest, which is paid in advance). Undoubtedly if he borrows, he will want to make full use of this privilege, for he has to pay interest on the amount of the loan and he would not want to borrow unless he had a use for the funds. Normally he will spend the borrowed funds by writing checks, most of which will go to people with accounts in other banks. Now we have already seen (compare Balance Sheet 6) that when a depositor writes a check on his account in favor of an individual whose account is in a different bank, the amount of the check is transferred through the Federal Reserve Bank, the reserves of the bank on which the check is drawn are reduced, and those of the receiving bank are increased. Thus, when the borrower writes checks on his newly created account, the lending bank's reserves fall, unless by a lucky chance the checks are given to persons who also keep their deposits in the lending bank. Therefore, if a bank lends, say, $100,000, it is very likely to lose reserves of about $100,000. But it should be noticed that when its reserves fall by that amount, those of other banks increase by the same sum.

The effect on the bank's reserves will be much the same, of course, if the lender takes currency as if he draws checks. For if he draws out any very large amount in cash, the bank may have to call on the Federal Reserve Bank to replenish its stock of currency (compare Balance Sheet 8). In this case, however, the reserves of other banks would increase only if they received currency from their depositors which they did not wish to keep in their tills and therefore transferred to the Federal Reserve Banks. In short, then, when a bank lends, it must be prepared to have its reserves fall by the full amount of the loan, and as its reserves decline, the reserves of other banks in the economy are likely to expand by an equal amount.[1]

When a bank lends, its reserves are likely to fall by the full amount of the loan. This sets a much more immediate limit to its lending powers than the rise in its demand-deposit liabilities. Let us see why this is so. To return to our example, if businessmen borrowed $60,000 from the bank, its balance sheet would look like this:

[1] The increase in the reserves of other banks may not be quite so large, for some additional currency may be wanted by the borrower or by those who gain by his spending. Thus, when loans are increased, banks may find that they have to call on the Federal Reserve Banks for more cash. This is a relatively unimportant factor which may be neglected in this analysis.

Balance Sheet 9a

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits[1]	$210,000 *
Government Bonds	450,000		
Loans	60,000 *		
Deposits in Federal Reserve Bank	80,000	Capital Stock	500,000
Property	50,000		

And so far, its reserves would be adequate, for it would need only $42,000 in reserves and it would have $80,000. But if the borrowers wrote checks for $60,000 on their accounts, as they might well do, the balance sheet would change alarmingly:

Balance Sheet 10a

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$150,000 *
Government Bonds	450,000		
Loans	60,000		
Deposits in Federal Reserve Bank	20,000 *	Capital Stock	500,000
Property	50,000		

The bank's reserves would now be inadequate, for it would need $30,000 against its deposit liabilities of $150,000, whereas it would have only $20,000. We may conclude, then, that the bank could not safely lend the $60,000 demanded.

Could it lend $50,000, the amount of its excess reserves, as shown in Balance Sheet 8? Let us see. The balance sheet immediately after it made the loan would be:

[1] It is convenient to assume, as we have done here, that the whole amount of the loan is added to the account of the borrower and that no discount is paid.

Balance Sheet 9b

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$200,000 *
Government Bonds	450,000		
Loans	50,000 *		
Deposits in Federal Reserve Bank	80,000	Capital Stock	500,000
Property	50,000		

At this stage its reserves would, of course, be adequate. And after borrowers had written checks for the full amount of the loan, its balance sheet would be:

Balance Sheet 10b

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$150,000 *
Government Bonds	450,000		
Loans	50,000		
Deposits in Federal Reserve Bank	30,000 *	Capital Stock	500,000
Property	50,000		

And its reserves would be just at the legal minimum. Hence we may conclude that a bank cannot safely lend an amount greater than its excess reserves, but that it can lend up to that amount without threatening its ability to meet its reserve requirements.

When a bank lends, it has to face the loss of its reserves: that fact, as we have seen, imposes a restriction on its lending ability. But — and this point is of critical importance for an understanding of the banking system — whatever it loses in reserves is gained by other banks. When, as in our illustration, its reserves decline by $50,000, those of other banks increase by a like amount. And when its lending ability is exhausted, as it is if its situation continues as in Balance Sheet 10b, the lending ability of the banks which now have larger reserves is increased.

Let us see how this works. Suppose that another bank has made

loans of $40,000 and the whole sum is transferred to our bank. The other bank was able to lend because it had excess reserves, but the transfer of funds reduces these reserves just as a similar transfer did for our bank. And now let us see what happens to the accounts of our bank after the $40,000 has been collected:

Balance Sheet 11

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$190,000 *
Government Bonds	450,000		
Loans	50,000		
Deposits in Federal Reserve Bank	70,000 *	Capital Stock	500,000
Property	50,000		

It has gained $40,000 in reserves. However, its excess reserves have not increased by that full amount, for its deposit liabilities have also increased by $40,000. It now needs reserves of $38,000, and since its reserves are $70,000, it has an excess of $32,000 and may therefore lend an additional $32,000.

The picture we have, then, is that a bank can lend up to the amount of its excess reserves. When it does so, some at least of the excess is transferred to other banks, which can then expand their loans. However, even though their reserves rise by the same amount as the reserves of the lending bank decline, they are not able to expand their loans by this full amount, for as their reserves rise, so do their demand-deposit liabilities.

Nevertheless, it is clear that the banking system, as distinguished from a single bank, can expand loans by much more than the amount of its excess reserves. How much may banks expand their loans when they have excess reserves? To illustrate, let us suppose that banks have excess reserves of $1 million; that — to simplify the analysis — all banks are required to keep reserves of at least 20 per cent against their deposit liabilities, and that the public does not want to increase its holdings of cash. At the first stage, the banks that hold the excess reserves may lend $1 million. When they do so, their demand deposits increase by $1 million. But since they now need to hold reserves

of $200,000 (20 per cent of $1 million) more than before, their excess reserves are now only $800,000. They may expect to lose to other banks, reserves equal to approximately $1 million; what they do not thus lose remains, of course, a part of their excess reserves. As other banks find their reserves increased by $1 million, so they find their demand deposits increased by the same sum. Hence their excess reserves are now only $800,000. The banks that made loans (compare Balance Sheet 10b) will have no excess reserves if the full amount of their loans is transferred to other banks. In any case, the total excess reserves of the banking system stand at $800,000. To put it differently, the total reserves are not affected because of the loans, but they are redistributed. However, demand deposits are now $1 million higher than formerly, and against this amount $200,000 additional reserves must be kept. The upshot is that while total reserves are the same, excess reserves fall from $1 million to $800,000, and they are held by different banks.

Successive stages in this process are easy to follow. The banks that now hold excess reserves of $800,000 can lend up to that amount, so that demand deposits increase by that figure, reserves required rise by an additional $160,000, and the excess reserves of the system fall to $640,000. But loans can again be increased, this time by $640,000. Demand deposits rise by this amount, and $128,000 more reserves are required against them. Hence excess reserves fall to $512,000. It will be seen that if the banking system possesses excess reserves of $1 million initially, it may lend originally $1 million and then $800,000 + $640,000 + $512,000 + $409,600, and so on, with each term only 80 per cent of the one preceding. The sum of all these terms is $5 million. Hence if the required reserve ratio is 20 per cent, the banking system may expand loans (or more accurately, demand deposits) by five times the amount of its excess reserves.[1]

Hence, we may conclude that, although no single bank can safely lend any more than its excess reserves (compare Balance Sheets 9a and 10a), the banking system as a whole can expand loans by a multiple of its excess reserves. The reason for this should be clear. As one bank lends, in a sense it exports excess reserves to other banks, which in turn are enabled to lend. And as they do so, they export

[1] The reader may be interested to develop the general rule. It is this: if the required reserve ratio is R, and excess reserves are S, the banking system may expand its loans by S/R. In the case discussed above, the ratio was .2; the excess reserves were $1 million; hence the total possible expansion in loans was $1 million/0.2, or $5 million.

reserves to still other banks. But as banks lend, their demand-deposit liabilities grow, and therefore the amount of excess reserves falls, even though the total reserves of the banking system do not change. This, of course, eventually brings a limit to the possibility of expansion.

If some banks with excess reserves refuse to expand their loans, they will not lose their reserves to other banks. Furthermore, other banks which do lend are likely to lose some of their reserves to the non-lending banks, so that these reserves become, in a sense, sterile. For as they grow in amount, the reserves available to the *lending banks* decline, and their power to expand loans is checked more quickly. Generally, the banking system can expand loans to the full amount set out above only if all the banks are willing to lend up to the full amount that their reserve position permits. If some are not willing, the ability of the whole banking system to expand loans is lowered.

A bank may also utilize its ability to create demand deposits for financing the purchase of securities. If our bank, in the situation indicated in Balance Sheet 11, wanted to buy government bonds, it could pay for them by opening an account for the sellers. If it bought $30,000 worth of government bonds, its balance sheet would be:

Balance Sheet 12

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$220,000 *
Government Bonds	480,000 *		
Loans	50,000		
Deposits in Federal Reserve Bank	70,000	Capital Stock	500,000
Property	50,000		

The purchase of securities (not financed by cash) thus affects the accounts in the same way as loans, and the same factors that limit the bank's ability to lend also limit its ability to purchase securities. Since, with increased deposit liabilities, it must face the likelihood of losing reserves, it cannot safely buy securities that exceed in amount its excess reserves. But since anything it loses from its own reserves goes to other banks, they in their turn are in a stronger position for making loans or buying securities.

When the banking system has excess reserves of a certain amount,

banks can lend [1] much more than this amount. And when they lend, they create demand deposits which, as we saw, are money. In summary, then, we see that banks create money by expanding their loans. Hence if the banks generally come into possession of increased reserves, their ability to create money is enhanced, whereas if their reserves decline, they are less able to create money. Indeed they may be compelled to destroy it.

A borrower normally pays back a loan by a check drawn on his own account, and the payment of this check reduces the bank's demand deposits by the amount repaid. To illustrate this, let us suppose that a borrower repays a $10,000 loan. The balance sheet will now be:

Balance Sheet 13

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$210,000 *
Government Bonds	480,000		
Loans	40,000 *		
Deposits in Federal Reserve Bank	70,000	Capital Stock	500,000
Property	50,000		

Demand deposits are reduced, and money has been destroyed.

If the bank suffers a large loss in reserves, it will be compelled to call in loans and sell securities, or if loans are maturing, it will refuse to make new ones. Suppose, for instance, that a depositor wrote a check for $50,000 on his account for someone who kept his account in a different bank:

Balance Sheet 14

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$160,000 *
Government Bonds	480,000		
Loans	40,000		
Deposits in Federal Reserve Bank	20,000 *	Capital Stock	500,000
Property	50,000		

[1] This term will henceforth include the purchase of securities.

Since the bank's reserves are now inadequate, it must either acquire more reserves or reduce its demand deposits. We shall discuss in the next chapter how it may add to its reserves. Let us assume for the present that it does not take that step, but instead undertakes to reduce its demand deposits. Because its present reserves are adequate to support only $100,000 in demand deposits, it has to adopt measures that bring its demand deposits down by at least $60,000. It may do this by reducing either its loans or its security holdings by that amount. Quite likely it will do a little of each. Let us suppose that it allows loans of $30,000 to mature, and sells $30,000 worth of securities to its own depositors. Its balance sheet is now:

Balance Sheet 15a

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$100,000 *
Government Bonds	450,000 *		
Loans	10,000 *		
Deposits in Federal Reserve Bank	20,000	Capital Stock	500,000
Property	50,000		

If instead it had sold $30,000 worth of securities to those who had accounts in other banks, and had let loans mature, the balance sheet would have been:

Balance Sheet 15b

Assets		Liabilities, etc.	
Cash	$70,000	Demand Deposits	$130,000 *
Government Bonds	450,000 *		
Loans	10,000 *		
Deposits in Federal Reserve Bank	50,000 *	Capital Stock	500,000
Property	50,000		

Here, though the selling bank gains reserves, it does so at the expense of other banks, for the buyers pay by checks against their

accounts, and the funds are transferred by the Federal Reserve Bank.

When reserves are inadequate, banks are compelled to reduce the amount of their demand-deposit liabilities. The required reduction will normally be a multiple of the amount by which their reserves are deficient. The analysis proceeds along the lines followed in determining the total lending ability of the banking system when it possesses excess reserves. But there is this difference: while banks are *enabled* to lend more freely when they hold excess reserves, they are *required* to reduce their demand deposits when they have inadequate reserves.

How the amount of their reserves is determined will be discussed in the next chapter.

Summary

Demand deposits make up most of the money of the economy, and almost all the economy's business is carried on by the transfer of this kind of money. Banks create demand deposits when they lend or buy securities, and destroy such deposits when their loans are repaid or when they sell securities. When they hold reserves that exceed the amounts required by law, they are able to lend more freely or to buy more securities. Because their interest return is likely to be greater if they expand their loans and security holdings, they are under inducement to do so. Although any bank, considered individually, may not safely expand its loans and "investments" by an amount greater than its excess reserves, the banking system may do so by several times the amount of its excess reserves. Hence, if banks acquire additional reserves, they are able to increase the amount of money by several times as much as the additional reserves.

When reserves are inadequate, banks have to reduce their demand deposits. To do so they are *compelled* to reduce loans or investments. Again, although no one bank need reduce its demand deposits by more than the deficiency in its reserves, the banking system is compelled to reduce total loans and investments by several times the amount of the deficiency in reserves.

Money is created and destroyed by commercial banks, but not without limit. Any increase in their reserves permits them to increase the supply of money by perhaps four or five times the amount of the increase.[1] Any decrease in their reserves has the opposite effect.

[1] The factor depends upon the reserve ratio required against demand deposits, the willingness of all banks to move together, and on the desire of the public to increase its holdings of currency as demand deposits increase.

Indeed, if reserves were just adequate before the reduction, the banks would be *compelled* to destroy money amounting to perhaps four or five times the amount of the reduction in reserves.

Power over reserves thus gives conditional power over the amount of money. If the Federal Reserve Banks can reduce them, as we shall see they can, they can effectively force the commercial banks to reduce the amount of money in existence. If they can increase the reserves, they can at least give an incentive to member banks to increase the amount of money.

24

The Federal Reserve System

THE FEDERAL RESERVE SYSTEM is responsible for the monetary health of the economy. It is able to discharge that responsibility because it can exercise some authority over the lending operations of member banks. This control has its basis in the fact that the commercial banks are required by law to hold reserves against their deposit liabilities in the form of deposits in the Federal Reserve Banks. Obviously, if we are to understand how the banking system operates, it is important to examine in some detail the nature of this control and the methods of exercising it. This is our objective in the present chapter.

Methods of Control

The Federal Reserve Banks exercise control over the system through their power to alter the excess reserves of member banks. Commercial banks may lend [1] more freely when their excess reserves are raised, but they must reduce their loans when their reserves become deficient. If the Federal Reserve Board decides that the amount of money should be increased, it can order the Federal Reserve Banks to raise the excess reserve of the member banks so that the latter are able to lend more freely. If, on the contrary, the Federal Reserve Board wants to reduce the amount of money, it can require the Federal Reserve Banks to drive member banks' reserves below the legal minimum and thereby compel them to call in loans. Power to alter the excess reserves of

[1] The terms *lend* and *loan* in this chapter will refer to both lending in the ordinary sense by commercial banks and their buying of securities.

the member banks thus means power to influence the amount of money. The Federal Reserve System has this power; for it is able to alter member bank excess reserves no matter what the member banks want. If it wants them to decline, the commercial banks can do nothing to prevent it; likewise, if it wants them to increase, the commercial banks inevitably find themselves with higher excess reserves. The Federal Reserve Banks can alter the size of member bank excess reserves in either of two ways: by changing the amount of reserves required against a given volume of demand deposits, or by changing the actual reserves which member banks hold.

Let us illustrate these two procedures. In March, 1946, the member banks of the Federal Reserve System had demand-deposit liabilities of about $71.5 billion, against which they had to keep reserves of nearly 20 per cent.[1] They also had time-deposit (savings account) liabilities of about $25 billion, against which they were required to hold reserves of only 6 per cent. The total reserves required were approximately $14.5 billion, while they actually had on hand reserves of $15.6 billion, and hence excess reserves of approximately $1.1 billion. Their reserve position may be presented as follows:

TABLE 31

Selected Items: All Member Banks, March, 1946
(in billions of dollars)

Deposits in Federal Reserve Banks	$15.6	Demand Deposits	$71.5
		Time Deposits	25.0

Within limits, the Federal Reserve Board may change the reserve ratios required.[2] Let us see what would have happened if the Board of Governors of the Federal Reserve System had wanted to induce the member banks to lend more freely and therefore had reduced the reserves required to, let us say, an average of 10 per cent for demand deposits and 3 per cent for time deposits. After such a move, member banks would have been required to hold reserves of $7.9 billion against

[1] Country banks were required to keep only 14 per cent reserves against their demand-deposit liabilities.

[2] The Federal Reserve Board may not set the required reserve ratios against demand deposits at less than 13 per cent, 10 per cent, and 7 per cent for Central Reserve cities banks, Reserve cities banks, and country banks, respectively, nor higher than 26 per cent, 20 per cent, and 14 per cent.

their deposits — the sum of 10 per cent of $71.5 billion and 3 per cent of $25 billion. Their excess reserves would then have risen from $1.1 billion to $7.7 billion. But even though their excess reserves had grown by so great an amount, they would not necessarily have expanded loans. Such an increase in reserves would only have increased their *ability* to do so, and not necessarily their *willingness*. If, in contrast, the Federal Reserve Board had wished to force the member banks to reduce the volume of outstanding loans, it would have raised the reserve requirements. If the requirements had been raised to 22 per cent for demand deposits and 6 per cent for time deposits, member banks would then have had to maintain reserves of about $16.9 billion; that is, 22 per cent of $71.5 billion plus 6 per cent of $25 billion. Since they had only $15.6 billion on deposit with the Federal Reserve Banks, it is clear that they would have had to reduce their demand-deposit liabilities. This they could have done, as we have seen, by allowing loans to mature without renewal and by selling securities.

The Federal Reserve Board has used this method of control only sparingly, though more frequently in the last few years. Reserve requirements were not altered once before August, 1936, and they were changed only eight times in the tempestuous period from August, 1936, to June, 1946. Obviously, the Board has been ready to use this method of control more freely in recent years, though even in this period not very often.

The Federal Reserve System may also take action to bring about a change in the actual amount of reserves which member banks possess. As we saw, member banks were required in March, 1946, to hold reserves of $14.5 billion against their deposit liability, but actually had $15.6 billion. If the Federal Reserve Banks had been able to increase these reserves to $20 billion, member banks would have held excess reserves of $5.5 billion instead of the $1.1 billion they had, and this growth in their excess reserves might have induced some of them to lend more freely. Or if, to consider the opposite situation, the Federal Reserve Banks had taken steps to reduce the reserves of the member banks by $3.6 billion to a total of $12 billion, their reserves would have been deficient by $2.5 billion and the banks would have been compelled to reduce their demand deposits. It is clear that the Federal Reserve Banks can affect member bank lending operations by changing the reserve ratios required and by changing the actual volume of reserves.

Open-Market Operations

We shall now consider the two methods by which the Federal Reserve System can alter the reserves of the member banks. First let us analyze "open-market operations." A commercial bank pays for the securities it purchases by opening a deposit in the name of the seller. When a Federal Reserve Bank buys securities, it does the same thing — it pays for them by a check drawn on itself. If it buys the securities directly from its member banks, their reserve deposits are increased, for they have the sums they receive added to their deposits in the Federal Reserve Bank. If the checks are for $1 billion, member bank deposits in the Federal Reserve Bank increase by a like amount. We can see this more clearly by setting up a simple balance sheet for all member banks which will present the situation both before and after the purchase.

TABLE 32

Balance Sheet *

All Member Banks — *before* Federal Reserve Bank purchase of securities
(in billions of dollars)

Loans and Investments	$100	Demand Deposits	$75
Deposits in Federal Reserve Bank	15		
(and so on)		(and so on)	

* It is not necessary to show other items: hence the balance sheet will not balance.

After the Federal Reserve Bank has purchased $1 billion worth of securities from the member banks, their combined accounts are as follows:

TABLE 33

Balance Sheet

All Member Banks — *after* Federal Reserve Bank purchase of securities
(in billions of dollars)

Loans and Investments	$99 *	Demand Deposits	$75
Deposits in Federal Reserve Bank	16 *		
(and so on)		(and so on)	

Since their demand-deposit liabilities are not changed, member banks do not require additional reserves, and yet their reserves have actually gone up by $1 billion. Hence, if they had $1 billion in excess reserves before, they now hold $2 billion in excess reserves.

If the Federal Reserve Bank buys securities from the public, the reserves of member banks will be affected in the same way, but less directly. Suppose as before that the Federal Reserve Bank purchases $1 billion worth of securities on the open market and pays for them by check. Private individuals and firms then receive checks totaling $1 billion drawn on the Federal Reserve Bank. These checks are deposited in commercial banks, and as a result the deposit liabilities of these banks increase by $1 billion. They in turn send the checks to the Federal Reserve Bank, which pays by raising member bank deposits in the Federal Reserve Bank by $1 billion. The resulting balance sheet for the member banks is shown in the following table:[1]

TABLE 34

Balance Sheet

All Member Banks — *after* Federal Reserve Bank purchase from public
(in billions of dollars)

Loans and Investments	$100	Demand Deposits	$76 *
Deposits in Federal Reserve Bank	16 *		
(and so on)		(and so on)	

Whether the Federal Reserve Banks buy directly from the member banks or from the public, the reserves of member banks rise by the full amount of the purchase. If they buy securities from member banks, the deposit liabilities of these banks are unaffected; whereas if they buy them from the public, member bank deposit liabilities also grow by the same amount as their reserves. This means that the *excess* reserves of member banks increase by less in the latter case than in the former, for even though their reserves rise, member banks must hold increased reserves against their now increased deposit liabilities.

The Federal Reserve Banks have purchased a large volume of securities in recent years. In 1938 they held about $2.6 billion worth of government securities, and in April, 1946, they held $22.3 billion.

[1] The balance sheet in Table 32 illustrates the original situation.

In other words, they bought about $20 billion worth of securities in eight years. These purchases would have brought about an equivalent increase in member bank reserves if their effects had not been offset by other factors, but as we shall see, reserves increased by only about $7 billion in that period. It is interesting to note that before 1938 the Federal Reserve Banks had never held more than $2.6 billion in government securities, and that during the nineteen-twenties their holdings were never above $1 billion.

When the Federal Reserve Banks sell securities, the reserves of member banks decline. If they sell to member banks, the buyers pay from their deposits in the Reserve Banks. If the Reserve Banks sell directly to the public, the buyers normally pay by checks drawn on commercial banks. When these checks are collected, the deposit liabilities of the member banks are reduced, and their reserves fall by the same amount. Hence the reserve ratio — actual reserves as a per cent of deposits — declines, and excess reserves are lowered.

The effectiveness of this measure for reducing member bank reserves will be limited when the Federal Reserve Banks hold only a small sum of government securities. In December, 1939, for example, member banks had excess reserves of $5.2 billion. But at that date the Federal Reserve Banks held only $2.5 billion worth of government bonds. Hence it is obvious that even if they had sold all their bonds, excess reserves would have remained very high.

When the Federal Reserve Banks buy or sell securities or, to give it its technical name, engage in open-market operations, they do so on their own initiative. Their motive is to raise or lower the reserves of member banks and in that way to influence member banks toward an easier or tighter credit policy. And it must be emphasized that the banks are unable to do anything to offset these measures. If the Federal Reserve Banks want to reduce the reserves of the member banks, they can do so by selling securities. And while any single bank may succeed in avoiding a reduction in its reserves by selling securities to other banks or to individuals who have accounts in other banks, if it does so, the reserves of other banks fall even more than they would have done otherwise. Total member bank reserves will change by the value of the securities bought or sold by the Federal Reserve Banks. Hence, open-market operations are one means by which the Federal Reserve Banks can control the size of member bank reserves.

Federal Reserve Rediscounting

The Federal Reserve Banks may also exercise control by varying the terms on which member banks rediscount. When a bank lends, it receives from the borrower a note that bears his promise to repay at a designated date in the future a larger sum than he actually receives. He may, for example, have to agree to give the bank $100,000 six months later in order to get $97,000. Such a note is said to be *discounted* by the bank at 6 per cent. The note is, of course, regarded by the bank as an asset, like other promises to pay, such as government bonds or industrial securities. And if the bank holds the note for six months, it can demand $100,000 from the lender. Obviously, it is not a piece of paper one would throw away.

The bank may also use the note to build up reserves. It can take it to a Federal Reserve Bank and ask to have it *rediscounted.* When this is done, the Federal Reserve Bank takes over the note and gives the member bank $100,000 minus the amount of the rediscount. If the note still has six months to run, and if the Federal Reserve Bank's rediscount rate is 1½ per cent, the member bank receives $99,250. Since it undertook to have the note rediscounted in order to build up its reserves, it will take payment by having that sum added to its deposit in the Federal Reserve Bank. On its balance sheet there will be a reduction in the item, Loans, and an increase in its deposits with the Federal Reserve Bank. Hence, obviously, its reserve position has become stronger.

When notes are rediscounted, the initiative comes from the member banks. Nevertheless, the Federal Reserve Banks exercise definite control over rediscounting, essentially through their control over the rediscount rate. The higher that rate, the smaller is the amount given to member banks for a discounted note. For instance, if the rate had not been 1½ per cent in the example above, but 6 per cent, the member bank would have received only $97,000 instead of $99,250. Obviously, then, when the rate is high, member banks are discouraged from rediscounting. Hence, by manipulating the rediscount rate, the Federal Reserve System is able to encourage or discourage the use of this method of raising reserves.

Member banks want to rediscount only when they seek higher reserves; otherwise they have no reason for doing so. When they have excess reserves, they are not tempted to rediscount, no matter how low the rediscount rate. Hence, control over the rediscount rate

can serve no purpose unless member banks chronically need larger reserves. During the nineteen-twenties, when commercial banks habitually rediscounted notes with the Federal Reserve Banks, changes in the rediscount rate had important effects on the size of member bank reserves. During the nineteen-thirties, when member banks generally had excess reserves and rediscounting was rarely used, changes in the rediscount rate were of minor importance. Rediscounting has been little used in recent years, as evidenced by the fact that in April, 1946, the total value of rediscounted notes in the possession of Federal Reserve Banks stood at only $.3 billion, whereas member bank reserves were about $15.5 billion. Thus it is clear that control over the rediscount rate was of no real importance at that date. Indeed, if the Federal Reserve Banks had raised their rediscount rate so high that rediscounting had ceased completely, member bank reserves would still have been $15.2 billion. Obviously control over the rediscount rate has not recently given much control over the banking system.

In conclusion, the Federal Reserve Banks can control member bank excess reserves, and hence their lending policy, by varying the reserve ratios required, by buying or selling securities, and by changing the rediscount rate. These methods are sometimes ineffective, as when member banks are not rediscounting, or when the Reserve Banks hold only a small amount of government securities. Moreover, as we shall now see, the Reserve Banks may not be able to employ them without limit to build up member bank reserves because of the restrictions which the Federal Reserve Act sets on their activities. We must now examine the rules by which the Reserve Banks are governed.

The Federal Reserve Banks

Just as member banks have to keep reserves against their demand- and time-deposit liabilities, so the Federal Reserve Banks have to keep reserves against certain of their liabilities. The chief of these liabilities has already been referred to: the deposits which constitute the reserves of the member banks. As your deposit in the bank is a liability from the bank's point of view, so your bank's deposit in the Federal Reserve Bank is a liability from the point of view of the latter. Therefore, the Federal Reserve Banks have to keep reserves against the deposits of their member banks. In 1946 the reserves required against these deposits were set at not less than 25 per cent.

The other important liability of the Federal Reserve Bank is the outstanding Federal Reserve Notes. We have already seen that these notes, which are issued by the Federal Reserve Banks, bear a "promise to pay to the bearer on demand" so many dollars in "lawful money." That statement marks them as a liability of the Federal Reserve Banks, although it is not one which should worry them very seriously. Nevertheless, the law requires reserves of at least 25 per cent of the amount of Federal Reserve Notes outstanding.

The reserves which the Federal Reserve Banks have to keep must be in the form of gold, or rather of gold certificates, which are essentially warehouse receipts for the gold. These reserve requirements impose limits on the actions of the Federal Reserve System, although not very rigid ones, since the Federal Reserve Board may suspend reserve requirements when it needs to do so.

We can see the function of these reserve requirements most clearly by analyzing the items in a balance sheet of the Federal Reserve Banks.

TABLE 35

Balance Sheet

All Federal Reserve Banks: end of April, 1946
(in billions of dollars)

Assets		Liabilities	
Gold Certificates	17.3	Federal Reserve Notes	23.9
U.S. Government Securities	22.7	Deposits	17.7

At that time, the Federal Reserve Banks had to have reserves of at least $5.975 billion against the Federal Reserve Notes outstanding, and of $4.425 billion against their deposit liabilities. Their total reserves, therefore, had to be not lower than $10.4 billion. Since actually the banks had $17.31 billion in gold certificates, they had excess reserves of about $6.9 billion.

Because their reserves were relatively high, the Federal Reserve Banks could buy many more government bonds without having to worry about their reserve position. If, for instance, they had purchased $20 billion in government securities, in order to add that sum to member bank reserves their balance sheet would have been:

TABLE 36

Balance Sheet

Federal Reserve Banks: after purchase of $20 billion in government bonds

(billions of dollars)

Assets		Liabilities	
Gold Certificates	17.3	Federal Reserve Notes	23.9
U.S. Government Securities	47.7 *	Deposits	37.7 *

They would have been compelled to hold reserves of $15.4 billion against such liabilities, or $1.9 billion less than they actually possessed. Hence they could have increased member bank reserves by $20 billion and still have satisfied the reserve requirements established by Congress. On top of that, they could have expanded their note issue considerably, for their reserve position would have been more than adequate even after such purchases.[1]

Federal Reserve Notes

When commercial banks need more currency, they get it by calling on the Federal Reserve Banks; that is, by drawing cash against their accounts with the Federal Reserve Banks, just as you or I might draw cash out of our own accounts. Where, then, does a Federal Reserve Bank get the currency? It prints it.[2] When it transfers the notes to the member banks, its liabilities are increased, for these notes in the hands of the public represent claims that can be made against it.

Commercial banks which have on hand more currency than they need send some back to the Federal Reserve Banks. When they do so, their accounts in the Federal Reserve Banks rise; that is to say, their reserves increase. From the point of view of the Federal Reserve Banks, such an increase means increased liabilities, for they now owe more to the member banks. However, they have received Federal

[1] In 1944 and earlier, reserve requirements for the Federal Reserve Banks were higher; 40 per cent against Federal Reserve Notes, and 35 per cent against member bank deposits in the Federal Reserve Banks. Hence, if these rates had been effective in April, 1946, the Reserve Banks would have been much more limited in their ability to increase member bank reserves or Federal Reserve Notes in circulation. The reader should apply these ratios to the data for April, 1946, given in Table 35.

[2] More accurately, it is printed by the Bureau of Engraving and Printing of the United States Treasury for the Federal Reserve Banks.

Reserve Notes, which incidentally are not assets to the Federal Reserve Banks; for if you have in your pocket a piece of paper on which you have written a promise to pay the bearer on demand one million dollars, you will not feel yourself a millionaire; and likewise, a Federal Reserve Bank which holds its own promise to pay does not on that account have any great wealth. These notes, in the possession of the Federal Reserve Banks, are regarded as so much paper. If they are not crumpled and worn, they are stored against the day when member banks may want them again. Otherwise they are destroyed. But when these notes come back to a Federal Reserve Bank, the reduction in its liability offsets the increased liability created by the increase in member bank reserves.

Let us trace the effects of an increase in the circulation of Federal Reserve Notes on the balance sheets of both member banks and the Federal Reserve Banks. The commercial banks know that with the approach of Christmas, currency withdrawals will be large. Let us say that they expect net withdrawals of a half-billion dollars in the four weeks before Christmas. How will these withdrawals affect the balance sheet of the member banks and of the Federal Reserve Banks? Let us suppose that initially their accounts are:

TABLE 37

a

Balance Sheet

All Member Banks, November 1

(in billions of dollars)

Assets		Liabilities, etc.	
Cash	$0.5	Demand Deposits	$75.0
Deposits in Federal Reserve Banks	16.0		
(and so on)		(and so on)	

b

Balance Sheet

Federal Reserve Banks, November 1

(in billions of dollars)

Assets		Liabilities, etc.	
Gold Certificates	$15.0	Federal Reserve Notes	$20.0
(and so on)		Member Bank Deposit	16.0

In anticipation of withdrawals during November, the commercial banks request $.5 billion in currency, and as they acquire it, the two balance sheets given above will alter thus:

TABLE 38

a

Balance Sheet

All Member Banks, November 30

(in billions of dollars)

Assets		Liabilities, etc.	
Cash	$1.0 *	Demand Deposits	$75.0
Deposits in Federal Reserve Banks	15.5 *		
(and so on)		(and so on)	

b

Balance Sheet

Federal Reserve Banks, November 30

Assets		Liabilities, etc.	
Gold Certificates	$15.0	Federal Reserve Notes	$20.5 *
		Member Bank Deposit	15.5 *
(and so on)		(and so on)	

Then, when the public withdraws the $0.5 billion from the commercial banks, the accounts of these banks are again changed:

TABLE 39

Balance Sheet

All Member Banks, December 20

Assets		Liabilities, etc.	
Cash	$0.5 *	Demand Deposits	$74.5 *
Deposits in Federal Reserve Banks	15.5		
(and so on)		(and so on)	

The balance sheet of the Federal Reserve Banks will not be altered by these withdrawals. The public now holds more currency, but its

demand deposits have fallen. The member banks have smaller reserves, but their deposit liabilities are lower. Hence the Federal Reserve Bank has exchanged one kind of liability, deposits of member banks, for another, Federal Reserve Notes.

After Christmas, the whole process is reversed. The public no longer wants to hold so much cash. Merchants normally decide at this time to deposit some of their surplus currency, and the commercial banks, since they do not need it all in their tills, send it back to the Federal Reserve Banks. Consequently, both balance sheets are altered, at each step, in the opposite direction to that indicated above.

This sort of thing happens not only at Christmas. Whenever business picks up, the public demands more currency. In 1946 it held about $24 billion in Federal Reserve Notes, whereas in April, 1938, when the national income was less than half as high, only $4.1 billion in Federal Reserve Notes was in circulation. As we can see by comparing the balance sheets above, this increase in the amount of Federal Reserve Notes outstanding tends to reduce member bank reserves, and at the same time to reduce their deposit liabilities. Furthermore, it alters the form of Federal Reserve Bank liabilities from "deposits of member banks" to "Federal Reserve Notes."

The Flow of Gold

The banking system is affected when gold enters or leaves the country or when gold, newly mined in this country, is sold to the government. When a Federal Reserve Bank — or, in the final analysis, the Treasury of the United States — acquires gold, member bank deposits and reserves are increased. And whether the gold comes from abroad or from domestic mines, the effects on the banking system are essentially the same. Let us suppose that a United States commercial bank acquires a claim in a foreign country through an American exporter who, having sold goods abroad, has been paid by a check drawn on a foreign bank. If he then deposits the check in his own bank, the bank has a claim on a foreign bank which can be met by the payment of gold. At the same time, of course, member bank deposit liabilities are raised. When the gold comes in, the commercial bank sends it to a Federal Reserve Bank and is paid by an increase in its deposits in the Federal Reserve Bank. The Reserve Bank then sends the gold to the United States Treasury and in exchange for it receives gold certificates. These gold certificates then become a part

of the reserve of the Federal Reserve Bank. Thus, the import of gold creates a chain of effects: deposit liabilities of member banks are increased, or rather, because they are increased, the banks acquire claims to be met in gold; member bank reserves are increased by the amount of the gold imported; and hence the deposit liabilities of the Federal Reserve Bank are increased by this amount; and finally, the reserves of the Federal Reserve Bank rise. When gold leaves this country, the effects are just the opposite: the reserves of both the Federal Reserve Banks and the member banks fall. The Federal Reserve Banks acquired about $6 billion in gold between October, 1938, and October, 1946, all of it before 1942. This acquisition would have led to a $6 billion increase in member bank reserves and deposits, as well as to an increase in Federal Reserve Bank reserves, if other developments had not occurred.

Survey, 1938–1945

It is advisable at this point to review the effects of various activities of the banking system on the items that enter into the balance sheets of both member banks and Federal Reserve Banks. When the Federal Reserve Banks buy securities from the public, member bank deposit liabilities and member bank reserves increase by the amount of the purchase. Likewise, the security holdings of the Reserve Banks and their deposit liabilities go up. When the public increases its holdings of Federal Reserve Notes, deposit liabilities of the member banks and their reserves both fall. Moreover, the deposit liabilities of the Federal Reserve Banks decline, while Federal Reserve Notes outstanding, which are liabilities of the Reserve Banks, increase. Finally, when gold is acquired by the Federal Reserve Banks, member bank reserves and deposit liabilities increase; the Federal Reserve Banks gain gold, and their deposit liabilities also rise. Naturally, when the Federal Reserve Banks sell securities, reduce Federal Reserve Notes outstanding, or lose gold, the effects are just the opposite.

Now let us make use of these results to account, so far as possible, for changes in member bank reserves and their deposit liabilities. The following balance sheet shows the changes in the Federal Reserve Accounts between 1938 and 1945.

The Federal Reserve Banks acquired $7 billion in gold and purchased $19.2 billion worth of government securities between 1938 and 1945. As we have seen, these operations would lead to an in-

TABLE 40

Balance Sheet

Federal Reserve Banks, 1938 and 1945

(in billions of dollars)

	1938	1945		1938	1945
Gold	11.0	18.0	Federal Reserve Notes	4.1	23.0
Government Bonds	2.6	21.8	Deposits of Member Banks	8.0	14.9

crease in member bank reserves of $7 billion plus $19.2 billion, or a total of $26.2 billion; and to a similar increase in member bank deposit liabilities, if not offset by other changes. But in the same period, the circulation of Federal Reserve Notes increased by $18.9 billion, and this growth in circulation would, if not offset, bring about an $18.9 billion reduction in member bank reserves and deposit liabilities. The combined result of these three operations, acquisition of gold, the acquisition of securities, and the issue of more Federal Reserve Notes, would thus lead to an increase in member bank reserves of $7 plus $19.2 minus $18.9 billion, or a total of $7.3 billion. Actually, their reserves rose by $6.9 billion.[1]

Now let us examine the accounts of the member banks:

TABLE 41

Balance Sheet

All Member Banks, 1938 and 1945

(in billions of dollars)

	1938	1945		1938	1945
Cash	0.7	1.1	Demand Deposits	23.9	67.6
Loans	11.9	20.6	Time Deposits	11.5	21.7
Investments	17.8	78.8			
Deposits in Federal Reserve Banks	8.0	14.9			
(and so on)				(and so on)	

[1] Other factors, principally the growth in Treasury deposits in the Federal Reserve Banks, account for this small discrepancy.

Member bank reserves grew, as we saw above, by $6.9 billion. The banks increased their loans and investments by the substantial total of $69.7 billion, and this led to a corresponding increase in their deposit liabilities. Moreover, the increase in Federal Reserve holdings of gold and government securities added another $26.2 billion. Thus, if only these changes had occurred, member bank deposits would have increased by $95.9 billion (= $69.7 + $26.2 billion). But the increase in Federal Reserve Notes subtracted $18.9 billion from this, leaving a net increase of $77.0 billion in deposits to be expected. Actually the deposits that are listed above increased by only $53.9 billion. The rest of the increase occurred in inter-bank and United States Government War Loan deposits (which grew by about $13.6 billion), against which reserves are not held.

The enormous increase in demand deposits and Federal Reserve Notes has already been indicated. The amount of money in circulation increased from $28 billion in June, 1938, to $90.6 billion in June, 1945. This increase was made possible because of the $7 billion increase in the gold stock of the Federal Reserve Bank and the $19 billion increase in their holdings of government securities. These increases permitted and brought about an increase in member bank reserves; and the increased reserves made possible an increase in member bank deposit liabilities. Finally, the greater part of that increase came about because member banks bought $60 billion in government securities in that period.

Summary

We should now have a clear picture of the operation of the Federal Reserve System. The commercial banks may extend credit by lending or buying securities on the strength of their reserves, that is to say, their deposits with the Federal Reserve Banks. The latter hold against these reserves, and against their Federal Reserve Note liabilities, a large stock of gold, the ultimate reserve in our banking system.[1] The Federal Reserve Banks can affect the willingness of member banks to lend or invest by changing the amount of member bank reserves or by changing the reserve requirements which the member banks must meet. They can change the actual reserves of the member banks, if their own reserves and security holdings permit, by buying or selling government bonds, or by changing the rediscount rate

[1] We shall consider in the next chapter whether such an arrangement is necessary.

charged member banks. And they can, within limits, change the reserves required by member banks. While the control of the Federal Reserve Banks over the member banks is not absolute, it can normally be made strong enough to force contraction; for the Reserve Banks, by selling government securities, raising the rediscount rate, and raising reserve requirements, can usually force member bank reserves below the level needed for current deposits. However, the control of the Federal Reserve Banks is not very strong in securing expansion. The most they can do, provided that the regulations about reserve ratios permit and their stock of gold is adequate, is to give member banks a larger volume of excess reserves. But this does not necessarily lead member banks to lend and invest more freely, for they must first find satisfactory borrowers and securities in increased volume. Hence the Federal Reserve System can either accommodate or check an increased demand for money, but it cannot do much to bring about such an increased demand. The monetary system is well provided with efficient brakes; but its accelerator is rather uncertain.

25

The Supply of Money, the Interest Rate, and Gold

BANKS PROVIDE many services to the economy. They transfer funds between individuals, and they provide convenient storing places for money. Most of us know them directly only when we use them in these ways. But they also have a more important role, one which affects most of us less directly, but is nonetheless much more vital to our well-being. For banks create our money. Of course, they do not do this as free agents. The government sets down the rules which regulate them, and the Federal Reserve Banks guide their money-supplying activities. But in accordance with these rules, the banking system determines how much money we shall have.

The fact that banks create money does not mean that they create income. Our supply of money is not our income. On March 31, 1946, we had a stock of $101 billion worth of currency in our pockets, tills, and bank accounts; whereas during March, 1946, we had a flow of income that amounted to about $15.2 billion. But though money and income are not the same, they are related, and a change in the amount of money is likely to bring about a change in income. We shall be able to understand this relation fully after we have mastered the analysis of Part Four; in this chapter, we shall advance part way toward that understanding. In Part Four we shall see that the interest rate helps to determine our income. In this chapter we shall take the first step in this direction by seeing how the amount of money and the interest rate are related.

The Interest Rate

The interest rate is the price we pay for borrowing. If we borrow $1000 from a bank to be repaid in a year, at an interest rate of 5 per cent, we are given $950 now, and the interest payment of $50 is the price we are charged for the use of the money. If we take out a mortgage at 4 per cent in order to buy a house, the interest payment computed at that rate is the price we have to pay for the money we borrow. When we buy for $75 a government Savings Bond which matures in ten years at $100, the $25 premium we receive is the price the government pays for the use of our money. To repeat, we have to pay interest when we choose to hold money, and this is true, in a sense, even when we hold our own money. For if we had not determined to keep it in our own possession we could have lent it to someone else and earned interest on it. Thus, whether the money is our own or someone else's, we may properly regard the interest rate as the price charged for holding it.

The Determination of the Interest Rate

It costs money to hold money. When the interest rate is low, the cost is relatively small, but when the interest rate is high, holding money is a more expensive pastime. To hold money is a costly way of holding wealth, for, instead of holding our wealth in the form of money, we could hold it in bonds, or stock, in a savings account, or in life insurance, in real estate, or in mortgages; and it is clear that wealth kept in these and other forms may provide an income yield.

Whatever our motive for wanting to hold money rather than other kinds of assets, it is reasonable to suppose that we will want to hold more money when the interest rate is low, and less when it is high. After all, if the interest rate were zero, there would be little reason to hold bonds or to lend money in other ways. We would rather keep our wealth in the form of money than let someone else have the use of it. But if the interest rate were 100 per cent, we would be very foolish, indeed, to hold much money, since we could improve our lot very rapidly by lending it, unless our motives for holding it were very compelling. When the interest rate is very low, we tend to hold a good deal of money, given our motives for keeping our wealth in this form; when it is very high, we tend to hold only a little.[1]

[1] The reasons for wanting to hold wealth in the form of money will be examined more carefully later in this chapter.

Let us assume that when the rate of interest is 3 per cent, we are willing to hold $80 billion in money, and that when the rate is 2½ per cent, we are willing to hold $85 billion. Then if the banks have created $85 billion in money, it is evident that the interest rate cannot be 3 per cent. It is important to see why this is true. Obviously, if the banks have created this sum, all of it must be somewhere and must have an owner, since it is most unlikely that $5 billion would be without a claimant. We have assumed, however, that when the interest rate is 3 per cent, we want to hold only $80 billion in money; hence at that rate some people must be holding more money than they want to hold. What then happens? Those who have more money than they want at 3 per cent will use their surplus money to purchase bonds and securities. But who can be persuaded to sell them? No one can, unless the price is increased. For at the current price or interest rate, holders of securities already have on hand all the money they want. Consequently, the price of securities begins to rise; hence, as those who hold more money than they want endeavor to convert some of it into securities, some people may be found who are willing to sell bonds and securities at the higher prices. But as bond prices rise the interest rate falls [1] and continues to do so until we are willing to hold not $80 billion but $85 billion in money. We have assumed that we should be willing to hold that amount if the interest rate were 2½ per cent. But until this happens there will be persistent pressure from those who seek to convert money into other assets — bonds, stocks, and I.O.U.'s — at the higher interest rate. These people will be eager to lend, but they will not be able to do so until they raise their offer price for bonds far enough. So bond prices will rise steadily, and interest rates will fall until they reach 2½ per cent.

Conversely, if the interest rate were very low — say only 1½ per cent — the situation would also be impossible, for with such a low

[1] To say that the interest rate falls when bond prices rise is arithmetic, not economics. Readers who have difficulty in seeing this should consider the following problems:

How much would you have to give today for a promise of $100 to be repaid in one year if the interest rate is 5 per cent? Answer: $95.

How much would you have to give today for a promise of $100 to be repaid in one year if the interest rate is 2 per cent? Answer: $98.

Now a bond is essentially a promise to pay; let us say, $5 a year for 20 years, and then $100. When the interest rate is 5 per cent, such a bond will cost only $100; when the interest rate is 2 per cent, it will cost more — about $144.05. Therefore, the higher the price of the bond, the lower is the interest rate, and vice versa.

yield on their securities, people would want to hold a great deal of money, perhaps as much as $100 billion. But if the banks have created only $85 billion, a great many people who will prefer to convert bonds into money because of the low rate of interest will not be able to find buyers for the bonds until bond prices are lowered. Then, as these prices fall and the interest rate rises, it naturally becomes easier to find buyers for bonds, since the amount of money people want to hold falls too. So once more the interest rate reaches 2½ per cent, the point of equilibrium.

We may illustrate the forces that affect the interest rate graphically as shown in Figure 58. Here SS_1 is the amount of money created by the banking system, and DD_1 represents the amount of money the economy wants to hold at various rates of interest — $A billion at 1 per cent, $B billion at 4 per cent, and so on. This curve is obviously a demand curve for money, just as the SS_1 curve is a supply curve.

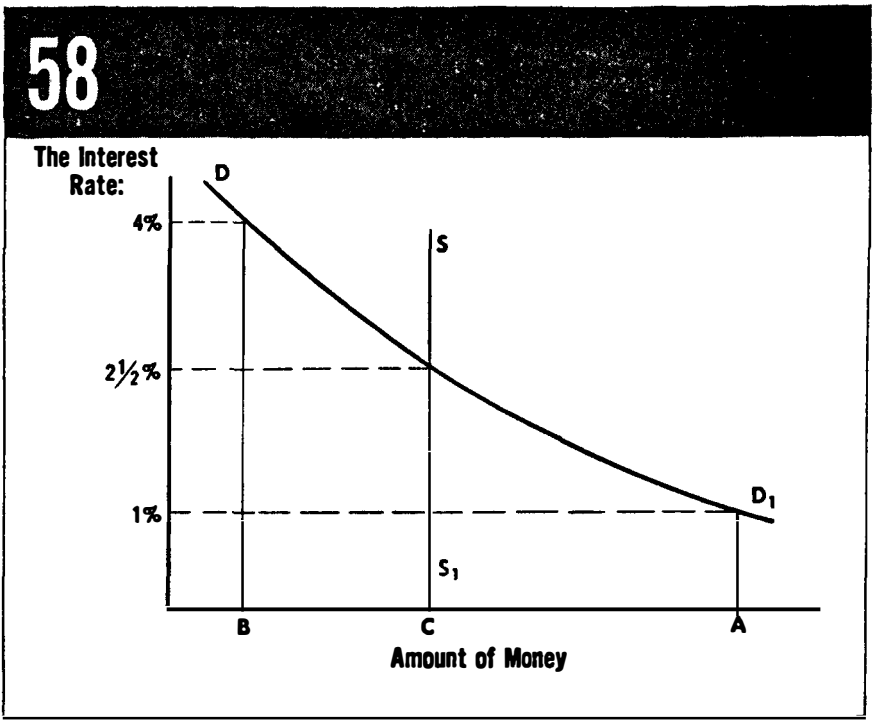

Figure 58. *The Determination of the Interest Rate*

The interest rate in this situation cannot be at 4 per cent, for the economy would want to hold only \$B billion, but would find itself with \$C billion — that being the amount in existence. In this situation, people will make every effort to take advantage of the interest rate by acquiring bonds and securities and by making loans. But they will be unable to persuade present bondholders to sell, or potential borrowers to borrow, until they offer a higher price for the bonds or a lower rate of interest on loans. Then, as the interest rate declines, the amount of money which the economy is willing to hold increases; and it continues to do so until the rate of interest reaches 2½ per cent — the equilibrium rate, as indicated by the intersection of the two curves DD_1 and SS_1.

A change in the supply of money would, of course, alter the rate of interest. If the supply were reduced, the rate would rise. The Federal Reserve Board could force such a reduction in the amount of money, as we saw in the preceding chapter. If it did so, banks would have to call in loans and sell securities in order to reduce their demand deposits. As the banks reduced the amount of money in circulation, it would no longer meet the wants of the public at the current rate of interest, and people would endeavor to get more by offering securities and bonds. But neither they nor the banks would find buyers at current prices, and so the prices would have to be lowered. And consequently, interest rates would rise.

Like a change in the supply of money, so also a change in the demand for money will affect the rate of interest, unless the supply is changed to meet it. Thus, if the demand for money rose,[1] interest rates would rise, and vice versa. The demand for money increases when the economy wants to hold more money than before at a given interest rate. If, when this happens, the banks are unwilling to accommodate this increased demand by expanding their loans and purchases of securities, those who want more money will try to sell securities and bonds. But since they will not be able to find buyers at current prices, they will have to ask less. As the prices fall, the desire to dispose of bonds is reduced and the willingness to accept them is increased, until finally equilibrium is again re-established. As bond and security prices fall, the interest rate rises, perhaps to 3½ per cent. Thus, because of the increase in the demand for money, there is an increase in the interest rate, unless the banks are willing

[1] The demand for money is often referred to as *the liquidity preference.*

to accommodate it. At the higher interest rate, the amount of money the economy is willing to hold is equal, as before, to the amount which the banks have created.

In summary, then, the rate of interest is determined by the supply of money and the desire to hold it. When the supply increases, the interest rate declines; when the demand increases, the rate rises.

The Demand for Money

We saw in Chapters 23 and 24 how the supply of money is determined. We must now consider how the demand for it is determined. A person who holds money rather than other kinds of assets is not the wealthier for it. A man with $500,000 in wealth can hold it all in the form of money, in government bonds, or in industrial securities, or in some combination of these and other assets. When he holds it in the form of money, either he has to pay an interest charge or he loses an interest income. If he held it all in government bonds, he would earn about $10,000 a year from them, whereas if he kept it all in money — including, of course, demand deposits — he would receive none. Why should he be willing to make this sacrifice? Why should anyone be willing to hold money in preference to income-yielding assets? The motives for holding money may be of three kinds, which Keynes has called the transactions motive, the precautionary motive, and the speculative motive. We shall examine these in turn.

We have seen that money is the most liquid type of asset. This means that the holder of money keeps some of his wealth in the form which is most readily acceptable to others. If he wants to buy anything, he can more conveniently do so by offering money than by offering other kinds of wealth. The housewife who has to pay bills cannot satisfy her creditors by giving them some of the books from the library or an old washing machine. Instead, she must pay by giving them money. To hold assets in their most liquid form — money — is, therefore, essential for those who want to buy. The ordinary consumer, therefore, always has to keep a part of his wealth in the form of money. The business firm is in exactly the same situation. Its employees would be very surprised if they were handed at the pay window on Saturday morning, not currency, but instead, several electric toasters. Payrolls must be met with money. Likewise, purchases of raw materials or equipment must be financed with money.

The business firm, then, like the consumer, must keep a part of its wealth in the form of money in order to transact ordinary business. Money so held is kept in accordance with the transactions motive.

The amounts of money held on account of the transactions motive will, of course, vary with the amount of buying. When the volume of purchasing in the economy is low, business firms and consumers will hold relatively small amounts of money. When a great deal of spending is being done, much more money will be held. Thus, the demand for money, or liquidity, for the transactions motive will normally be high when business is good and low when it is bad. The amount of spending also depends upon the price level, since the volume of spending is greater when prices are high, given the physical volume of business. The demand for liquidity to satisfy the transactions motive, therefore, depends mainly upon the volume of business and the price level, though it also depends upon such institutional factors as the frequency with which business and individuals receive and disburse money. If a man receives income once a month, he must keep a larger average amount of money in his possession than if he receives income once a week. The effects of these factors, which are not liable to great change, are obvious.

Since holding wealth in the form of money involves a sacrifice, the amount which business firms and individuals desire to hold also depends upon the interest rate. Because it costs dearly to hold money when the interest rate is high, consumers and firms cannot afford to hold it in great quantities, even though their requirements on account of the transactions motive are great. Conversely, because the loss suffered by holding money is small when the interest rate is low, business firms and consumers are generally willing, given the strength of their transactions motive, to hold relatively large amounts of their wealth in money at such interest rates. Hence, with the transactions motive given, the higher the interest rate, the smaller is the amount of money which the economy is willing to hold, and vice versa.

Thus, the transactions motive is one of the determinants of the demand for money. If its strength changes, the demand for money is altered. Thus, if the amount of buying increases because of a rise in the national income, the transactions motive grows stronger and the demand for money, other things being equal, increases. When the demand for money, or liquidity preference, increases, the amount of money which the economy wants to hold at each level of interest rate

becomes higher. Accordingly, the rate of interest is not a determinant of the demand for money, but rather is determined by it.[1]

As we have seen, the financing of current purchases is not the only motive for holding money. It may also be kept on hand for several other reasons. The first of these is to meet any contingency which may arise. This describes the precautionary motive. Like the hospital bed which must always be kept empty for an emergency case, funds are often kept on hand to meet unforeseen situations. The amount of money held for such purposes will also vary inversely with the interest rate.

The third reason for keeping wealth in the form of money is known as the speculative motive. Money is a measure of value in our economy, and despite the prevalence of the phrase "a fifty-nine-cent dollar," one dollar is always worth one hundred cents. However much the prices of other things may change in terms of money, a dollar is always worth ten dimes. Not so with other assets. Their value may diminish; their prices may fall; and if such falls are expected, it becomes worth while to hold money instead. Thus, a man with $1000 worth of securities who expects their price to fall can protect himself from loss by selling them and holding his wealth either in money or in other assets whose prices are not expected to decline. Thus money is often held as a protection against an anticipated reduction in the prices of other kinds of assets.

The strength of the speculative motive is subject to wide variation. At times it discourages the holding of money, as when people expect security prices to rise and so rush to purchase them rather than hold the money they have. At other times people will be encouraged to hold a great deal of money because of these speculative considerations. We have seen several instances of both extremes in our own economy, but none more spectacular than the shifts between 1927 and 1930. From 1927 to 1929, the steady rise in securities prices indicated that the desire for money on this account was very weak in some quarters. Many people were "bullish" about stock prices, and because they expected them to continue rising, they did not want to hold money, but preferred holding securities. At the same time, however, there was a growing "bearish" sentiment, a growing belief that security prices were due to fall. Acting on this belief, the "bears" turned their

[1] This conclusion is analogous to the one reached earlier, that the price of a good does not affect the demand for it.

securities and assets into money, with the rather paradoxical result that interest rates rose to very high levels, and at the same time securities prices also rose swiftly. A part of the market, at first the more important part, wanted to hold securities, but a growing part wanted to hold money. Suddenly the bearish views dominated. Stock prices began to fall as speculators sought to turn their wealth into money. Loans were called by the banks and the supply of money fell. At the same time, the fall in stock prices confirmed the position of the bears, and the expectation of further falls grew. As the speculative motive for holding money became stronger, stock prices fell further and further. Thus, in about three years' time, the speculative motive for holding money increased from an almost negligible to an absolutely overriding force. The desire to hold money on account of the speculative motive has, of course, fluctuated in many other periods of our history, though perhaps never more strikingly than at the end of the nineteen-twenties. This desire for liquidity is generally unstable, and its movements are relatively unpredictable, as anyone who has played the stock market knows only too well.

The amount of money held on account of the speculative motive is very sensitive to the rate of interest, especially at very high and very low rates. When the interest rate is very low, a large number of people believe that it will rise rather than continue to decline, since they feel that there is a limit below which it cannot go, and that it is certain to move upward once it reaches this lower limit. Whether this view is correct or not is unimportant; what counts is that commonly it is widely held. For when the interest rate falls, a large part of the market expects it to rise when it changes again, and is therefore induced to hold money rather than securities. The amount of money desired at very low interest rates is thus very high.

The same considerations operate for very high rates of interest. As the interest rate rises higher and higher, the belief gains strength that downward changes are bound to come. Since this implies that the prices of securities and bonds are expected to increase, people want to hold bonds and securities rather than money. Thus, at very high interest rates, the amount of money the economy wants to hold on account of the speculative motive is likely to be very low. Hence we may conclude that on account of the speculative motive, the economy will want to hold a great deal of money at very low rates of interest, and very little money at very high rates. But a rate that

may appear at one time to be very low may later appear quite normal. Views change on this matter, and the very low rate of 1929 may seem quite high in 1946.

The speculative motive for holding money is not only unstable; when it changes, it tends to induce other changes in the same direction. If the market anticipates higher security prices and so buys securities, their prices rise and consequently the market's belief that prices will rise finds confirmation. The belief may therefore grow stronger, eagerness to acquire securities may increase, and their prices may rise even faster. The process thus tends to become cumulative. Likewise, if the market expects security prices to fall, the decline also tends to become cumulative: lower prices lead to the expectation of still lower ones, and this in turn encourages selling, which puts more pressure on prices, and thereby confirms the market's views. Thus shifts in the strength of the speculative motive tend to set in motion further shifts in the same direction.

It is seen that, given the strength of the transactions, precautionary, and speculative motives, a rise in the interest rate discourages the holding of money, and a fall in the rate leads to an increased holding of money. The demand for money on account of the speculative motive tends to be very elastic at extremely high and low rates, but otherwise it is not particularly so. Hence, except at the extremes, the elasticity of the demand for money is like that of any other commodity.

So far, we have been talking as though there were but one rate of interest at any one time. In actual fact there are many. The rate on a loan for a very short time may be quite different from that on a loan for a long period. The rate for one borrower may be quite different from that for another, even though the two loans are for the same length of time. Thus, strictly speaking, we should not talk about *the* interest rate, but rather about the *structure* of interest rates. And it is interest rates in this sense that vary with the supply of money and the liquidity preference, or, as we may call it, the demand for money.

But though there is no single rate of interest, the various interest rates current at any one time are interrelated. Suppose there is an increase in the demand for loans of a certain maturity, for example, for loans due to be paid back in ten years. Under these circumstances, if there is no change in the supply of money, the rate of interest on

ten-year loans would rise, for the increased demand could not otherwise be satisfied. Because of the higher interest rate for these loans, borrowers would prefer to lend for this length of time than for other maturities, and would become less willing to hold securities that matured earlier or later. Therefore, the interest rate on other types of loans would also increase. Thus, a shift in one interest rate is very likely to affect others in the same direction. There are exceptions to this rule, but they may properly be left for the consideration of the more advanced student.

Summary: The Interest Rate

The interest rate is the price of money, and it is determined by the supply of money and the demand for it. The supply of money is set by the banks, through their willingness to lend and purchase securities. The demand for money varies with the three motives for liquidity that we have discussed. When the demand for money rises, the interest rate tends to rise; and when the supply rises, the interest rate falls.

Shifts in the demand for money are marked during the course of the business cycle. During the revival, the demand for money on account of the transactions motive grows, for in that period spending increases. But in the early stages of the upswing, the strength of the speculative motive for holding money may decline as speculators uniformly gain confidence. Hence the total demand for money may not increase sharply at this time. During the later phase of the revival, the strength of the speculative motive for holding money may increase as a feeling grows among speculators that the market has reached its peak. Hence at this stage the demand for money grows through both the transactions motive and the speculative motive. If the banks do not then add to the supply of money, interest rates will rise. During the early phase of a depression, the strength of the speculative motive may increase markedly, as it did in this country in 1930; and even though the transactions demand falls off, the total demand may rise. As the depression deepens, the speculative desire for money may decline as more and more speculators swing to the belief that the bottom has been reached, and at the same time business firms and consumers generally lessen their demand for money. Hence at such a time the interest rate may be expected to fall.

Through its control of the supply of money, the banking system plays an important part in determining the interest rate. When the

Federal Reserve Banks want lower interest rates, they provide member banks with larger excess reserves in order to encourage them to lend more freely and buy more securities — that is, to raise the supply of money. Federal Reserve Banks cannot require the member banks to expand; they can only provide them with the means for doing so. Hence the Federal Reserve Banks have relatively little power to lower the interest rate.[1] However, they can bring about an increase in the rate of interest because they are in a position to compel the member banks to reduce the amount of money outstanding. Thus the control of the Federal Reserve Banks over the interest rate is imperfect, being much stronger in raising the rate than in lowering it.

Gold and the Economy

At first glance there may seem to be little connection between the interest rate and gold. And true enough, there is no logical connection. But man-made institutions have created one, nonetheless. We ourselves have determined that gold should have a special significance in the economy, and thus indirectly that the above heading should be "Gold and the Economy." But it is sheer accident that we were not required to label the section, "Raisins and the Economy," or "Silver and the Economy," or "Pigs and the Economy." For any one of these commodities might just as well have been chosen for the special role we have given to gold.[2] But as it happens, we have endowed gold with certain unique characteristics, which somehow set it apart from everything else and give it its peculiar economic function. One of the characteristics, as you will notice, is that gold is somehow rather mysterious. We all feel that gold performs a peculiar function in the economy, though we find it hard to define that function with any degree of precision or certainty. But let us look into this a little.

Consider the following parable: Once upon a time the residents of a remote island in the South Seas had a very advanced monetary system. Although it lacked commercial banks and had no Federal Reserve System, it had a thing which many people consider much more important — a standard. But it was not a gold standard. It

[1] If they buy government securities directly, they can exert a strong pressure; but as we have seen, without a change in their regulations, their capacity to do this is limited.

[2] Gold has certain characteristics not found in pigs and raisins; you cannot produce gold unless you happen to own one of the few spots on this earth where it can be found, and it is very durable. However, these are not necessarily advantages when it comes to using it in our monetary system.

was a rock standard. The natives of this island had what economists call circulating media, the equivalent for all practical purposes of our nickels, dimes, and dollar bills. But this was not enough for them. They wanted to feel that their money had a backing. So they decided that their money should be backed by a large rock which stood on the shore. We may suppose that their dollar bills were inscribed, "Will Pay to the Bearer on Demand One Dollar in Rock," just as ours used to say, "Will Pay to the Bearer on Demand" a certain amount of gold. And for a long time all went well. But unhappily, an accident of nature dislodged their rock from its resting place, and it rolled into the sea. Imagine the consternation that there must have been! How could their money, now that it had lost its backing, be used to purchase anything? Luckily, they were saved from inflation — that is, from the loss of the value of their money — by a very fortunate observation. They discovered that on very clear days when the sun was at a certain angle, those of the tribe who had especially strong eyes could see the rock which had once served so tangibly as a backing for their currency, fathoms deep, under the water. And so their faith in their currency was restored, and once again it could be used to purchase whatever was available on the island. They had to call in their currency, it is true, in order to erase that part which read "One Dollar in Rock," and in its place they were now forced to write "Payable on Demand, One Dollar in Lawful Money." But their currency had a backing; confidence was restored; and their monetary system could function once again.

The monetary history of the United States since 1933 is in some respects similar to what happened to these islanders. Before 1933, currency was redeemable in gold; we could actually get gold for our money, and it even circulated in the form of gold coins. But once the private citizen could no longer obtain gold for his currency, our store of the precious metal performed for our domestic economy almost exactly the same function as their rock in the sea. We know that there is plenty of gold in this country, and we can examine it at our leisure. And we can then return to the outside world with renewed faith, if the examination comforts us, in our currency. But none of us can get gold in exchange for our currency unless we wish to engage in certain special transactions with foreign countries. Most of us have never seen the gold that serves as a backing for our money. Yet it is nonetheless generally supposed that our currency needs a backing,

that if somehow the gold in the Kentucky hills were spirited away, our money would no longer be capable of purchasing raisins and butter and milk. It is true, of course, that if the gold were not there we could not buy American gold with our money. But that would not be dissimilar to the situation we faced in 1944, when we were not able to buy new automobiles with our money because none were being produced — except that we could do without the gold more easily than without the new car. In any case, we are not able to buy the gold today, even though our national stock of it is bigger than that of all the rest of the world combined; for gold is sold only to special classes of buyers. Just what is the relation between gold and the value of money? Does money need to be backed? This question cannot be analyzed in detail until later in the discussion, but at least the nature of the relation between gold and money may now be made clear.

It is worth pointing out that through a good deal of human history, countries have not had a gold backing for their money, and most countries do not have it even today. Furthermore, although the United States has four times as much gold as it had in the nineteen-twenties, and about 25 per cent more than it had in 1939, its money has lost a good deal of its value since the latter year. In other words, prices have increased comparatively rapidly, though the backing for money is greater than it was a decade or so ago.

Although we are not able to get gold for our money, our monetary system is in some measure based upon gold. We have seen that the Federal Reserve Banks must keep reserves in gold. Hence, so long as their liabilities are limited by regulations which, incidentally, the Federal Reserve Board is empowered to change, the Federal Reserve Banks cannot print Federal Reserve Notes or create greater reserves for the member banks without restraint. The fact that the Reserve Banks are required to hold a certain percentage of gold as a reserve against their liabilities places a limit upon the size of these liabilities.

If gold flows into the country, as it did from 1934 to 1941, it provides the Federal Reserve Banks with greater reserves, which of course enable them to issue more Federal Reserve Notes and expand member bank reserves. Thus, if $100 million in gold enters the country in a certain period, the reserves of both the commercial banks and the Federal Reserve Banks rise by this amount, and this in turn permits the Federal Reserve Banks, since they are required to have reserves

of 25 per cent of their deposit liabilities, to add further to the reserves of member banks. Now, because the latter are required to have reserves of only about 20 per cent of their deposit liabilities, they too are in a better position to extend credit. A gold inflow, therefore, adds to the reserves of the Federal Reserve and commercial banks, thus permitting a large expansion in commercial bank credit. In the extreme case, an inflow of $100 million in gold enables commercial banks to expand their loans by almost $2 billion.

Shipment of gold out of the country has just the opposite effect upon the reserves of the Federal Reserve and commercial banks. When gold left this country before the war, these reserves fell by approximately the amount of the gold exported. Unless there are excess reserves in the banking system, such reductions in the reserves compel the banks to reduce deposits. If the reserves of the banking system are already being fully utilized, the export of $100 million in gold compels member banks to reduce their demand deposits by approximately $2 billion.

Gold flows affect the reserves of the banking system, hence the results of gold movements will be the same as those that follow an alteration of member bank and Federal Reserve Bank reserves brought about in any other way. We have already seen that when the reserves of the banking system are increased, the banks are enabled to lend more freely. Therefore, when gold comes into the country, the banks are enabled to expand their loans, or, in other words, to reduce their interest rates. But they are not compelled to do this, for banks have often been willing to hold excess reserves. When gold leaves the country, the reserves of the banking system are reduced. This, as we have seen, may compel the banks to reduce their demand deposits. Indeed, it will do so unless their excess reserves were high before gold was exported. Banks can reduce their demand deposits either by reducing their lending or by selling securities. In either event, we should expect the interest rate to rise. Thus, when gold comes into the country, banks are better able to reduce their interest rates. When gold leaves the country, they may be compelled to raise their interest rates — depending, as we have seen, on whether they held excess reserves originally. Gold flows, therefore, have sometimes been influential in determining the interest rate, although their effect recently has probably not been very great.

The connection between gold flows and the interest rate used to be

much closer than it is now. In the earlier years of the twentieth century, when the United States was still on the gold standard, bankers gave very careful attention to gold flows. When gold was leaving the country, this was felt to be a sign that the interest rate should be increased. When gold was entering the country, it was taken to mean that the interest rate should be reduced. The central bank of the country — the Federal Reserve Bank for the United States, the Bank of England for Great Britain, and so on — would then take steps to alter the interest rate in the appropriate direction as soon as it was convinced that the gold flow was important. The link between gold flows and the interest rate was therefore at that time much more intimate and direct than it was during the nineteen-thirties or is now.

It is interesting to speculate about what would happen if the stock of gold on hand were to disappear. Unless there were a change in the law, the Federal Reserve Banks would, of course, be required to reduce their deposit liabilities to zero. Obviously this would not be allowed to happen. Suppose that with the disappearance of the gold, the present law relating to Federal Reserve Bank reserves were repealed, and they were permitted to have any volume of liabilities that they thought desirable. How we believe the economy would be affected depends upon whether we suppose the Board of Governors of the Federal Reserve System is made up of men of intelligence or of lunatics. If the former, they would not be influenced in any way by their freedom to expand the liabilities of the Federal Reserve Banks. Their responsibility is to preserve satisfactory monetary conditions in the country. They would not want the liabilities of the Federal Reserve System to exceed a sensible, proper level, and this is determined, not by the gold supply, but with reference to the requirements of the economy. If the liabilities of the Federal Reserve Banks were not changed, member banks would not experience any change that could lead them to alter the volume of their deposits, and therefore they would have no reason for lending, or for buying or selling securities, more or less freely than before. In short, it is hard to see why any changes should result if gold reserves were done away with and if the Federal Reserve Bank requirements were repealed.

The use of gold in the present-day economy is somewhat absurd. The man-made law requires the Federal Reserve Banks to maintain gold reserves against their liabilities, and therefore our stock of gold fulfills a certain function. But the law seems to have been passed

only to provide us with an automatic check in case the governors of the Federal Reserve Board should all go mad. It appears to serve no other internal function.[1] If the law were repealed, our stock of gold would, of course, continue to be useful in industry or in purchasing commodities from other countries, provided they were willing to take gold. But it would have no function whatsoever in determining an internal monetary policy. It is difficult to modify inherited institutions. And certainly it seems true that since we have so much gold, the law requiring the Federal Reserve Banks to maintain gold reserves against their liabilities does not affect our economy adversely. But it is a restriction that we have imposed on our own freedom of action, and some day it may prove injurious.

SUGGESTIONS FOR FURTHER READING

The student should familiarize himself with the Federal Reserve Bulletin. In addition, he would find some of the following helpful:

Halm, George. *Monetary Theory.* Philadelphia: The Blakiston Company, 1946 (2nd edition).

Read especially Chapters 1, 3, 4, and 5 for a good account of the structure and workings of the banking system.

Machlup, Fritz. "Eight Questions on Gold: A Review," *Proceedings of the American Economic Association* (1940), XXX, No. 5, p. 30.

Robertson, D. H. *Money.* New York: Harcourt, Brace and Company, 1929 (2nd edition).

Very highly recommended, especially Chapters 1, 3, and 4.

United States Government: Federal Reserve Board. *Banking Studies.* Washington: Government Printing Office, 1941.

See particularly the following titles: "Operations of the Reserve Banks," "Monetary Controls," and "Instruments of Federal Reserve Policy."

——— *The Federal Reserve System — Its Purposes and Functions.* Washington: Government Printing Office, 1939.

A remarkably clear account.

[1] The role of gold in our dealings with other countries is discussed in Chapters 40–42.

PART FOUR

The National Income and Employment

Introduction

THE LEVEL OF EMPLOYMENT and the national income constitute the most important single factor in determining whether or not we enjoy a high standard of living. When the national income is low, economic misery is likely to be widespread; when it is high, most people live relatively well, in a material sense. The forces that determine whether we have prosperity or depression, inflation or deflation, are analyzed in Part Four.

26

The Theory of Employment: Introduction

NOW THAT WE HAVE HAD A BIRD'S-EYE VIEW of the whole economy and have examined in some detail the operation of its basic institutions — its business firms, and its banks — we are in a position to consider what is perhaps the most important subject to be discussed in this book: the theory of employment. This theory is in a sense the keystone of our whole analysis, and much of what we have learned in earlier chapters is significant because it is essential to a proper understanding of the analysis to follow. The theory of employment is important for two reasons. The central problem to which it is addressed is the most pressing and the most vital economic problem that confronts this country today. And toward the solution of this problem the economist has made in recent times his most important contributions.

Why should the problem of unemployment, the analysis of which makes up this theory, be so grave? After our experiences in the years between 1930 and 1940, it should hardly be necessary to ask that question. But memories are short, and the very special circumstances of the last few years may have obscured, for some, how great a problem unemployment is. How do unemployment and depression hurt us? Let us attempt to list and measure their adverse effects.

One criterion by which to judge an economic system is its ability to turn out a very large output of goods and services. When it produces a great deal, we agree that it is functioning satisfactorily, or at any rate that it is passing its first test; and when it fails to produce

large quantities, we conclude that its performance is unsatisfactory. When we judge our economy by this yardstick, we find that its grossest failures have occurred because of unemployment and depression. Unemployment was our key economic problem for years, and a large number of economists are convinced that it will become so again. Our country is not alone, of course, in having to wrestle with it. Advanced capitalist economies throughout the world — Germany, France, Canada, Britain, Australia, and others — have all been plagued as we have.

Let us attempt to measure the size of the difficulties which unemployment creates. First we shall do so in terms of the output loss resulting from depression and unemployment. Later we shall mention the other social evils which unemployment has brought about.

When there is unemployment in an economy, less is produced than when there is full employment. The difference between the actual output and the peak level of which the economy is capable represents the loss in output because of unemployment. For instance, when our whole labor force was working in 1944 and 1945, we were able to produce about $200 billion worth of goods and services a year. This represented an immense output — the equivalent of about $5700 a year for every family in the country. True, in the war years almost half of it was in the form of munitions, but it could as well have been automobiles, houses, clothing, and food. In short, the figure of $200 billion under wartime pressures at least gives a good indication of what we could produce with full employment in normal times. For though our peak might be somewhat lower because we would not normally work the long hours that were common during the war, it must be remembered that many of our most productive workers, who would in normal times be in factories and offices or on the farms, were otherwise engaged in 1944 and 1945. Let us assume, then, that with full employment we are able to produce $200 billion worth of goods and services in one year. That is our peak or capacity output.

Whenever there is unemployment, we produce less than this peak output. In terms of 1944 prices, we may produce as little as $70 billion worth of goods and services, as we did in 1932, when unemployment was very heavy. Or we may produce $120 billion worth, as we did in 1940, when unemployment was comparatively mild.[1] But

[1] Approximately 7.5 million men, but compared to most of the years in the preceding decade, this is a low figure.

whenever we have unemployment, we produce at some level below capacity; our economy is not running in highest gear. When we produce a $120 billion annual output, we of course produce a lot of goods, quite enough, certainly, to keep us from starving. But we produce less than we are capable of producing, since if everyone who wanted a job were employed, we should be able to produce a $200 billion annual output.[1] Hence, we should lose $80 billion worth of goods if, with relatively mild unemployment, our production should fall to $120 billion in, say, 1947. And obviously if $120 billion worth of goods and services is enough to provide us with a high standard of living, the loss of $80 billion worth is far from negligible.

It is hard to visualize the size of $80 billion, indeed, such an amount of money is almost beyond imagination's power to grasp. A line of $80 billion worth of pennies side by side would stretch from here to the sun and wind around it a few times for good measure. In terms of family income, a loss of $80 billion from the peak annual output would mean an over-all reduction of about $2300 worth of goods for every family in the country.[2] We may conclude, then, that each family is annually deprived of goods and services worth $2300 when we produce a 1940 output instead of a capacity output. Obviously, to be deprived of the many goods that we could have enjoyed for $2300 — of the clothes, automobiles, better food, and new furnishings — is not a welcome experience. To have it all happen in one year, and then to have it repeated year after year, seems almost more than one can bear. Worst of all — or perhaps best of all — it is avoidable.

Our failure to produce a capacity output has been almost entirely the result of unemployment, of our failure to have jobs for all who want to work. Natural disasters sometimes cause a little trouble, but in comparison to unemployment they are almost negligible. The New England hurricane of 1938 is said to have done about $1.5 billion in damage to property. In that year our loss in output because of unemployment may be estimated at about $34 billion. Unemployment has been by far the most important reason why our output has so often fallen below the peak level.

Measured only in terms of lost output, the cost of unemployment was staggeringly high during the nineteen-thirties. The estimates of

[1] Of course, our capacity to produce in 1932 or 1940 was less than it is today. Hence we should not set the 1932 output against the potential output of 1946, but rather against the potential of 1932.

[2] Omitting single individuals living alone, institutionalized persons, and so on.

the dollar value of this loss are widely varied, but they are all very high. The reason they vary is that we are uncertain about the value of the goods we could have produced if jobs had been available for all who wanted to work. For one thing, we do not even know, with certainty, how many people wanted to work. The data given in the following paragraphs are therefore not exact; they are only a rough estimate. But they are conservative, as we shall see.

In the period between 1930 and 1940, inclusive, we produced goods and services worth approximately $650 billion, measured in 1939 prices. Through much of this period we had depression. Unemployment was heavy, and consequently production was far below capacity. If we had produced, not at the depression level which prevailed for most of this period, but at capacity levels, the value of our output would have been about $1025 billion in 1939 prices. We could have produced $1025 billion worth of goods; we actually produced only $650 billion worth. The difference is $375 billion — and this is the value of our loss in output directly due to unemployment. But even this figure is not big enough to tell the whole story. It measures what we could have produced if everyone had been employed in the factories and with the equipment actually available from 1930 to 1940; but it does not take into account how much more productive our economy would have been if we had made a real effort during those years to build new factories, to expand old ones, and to introduce the most efficient machinery.

How much difference might this have made? If we had built new plants and added new equipment throughout the period at just the 1929 rate, we would have had about $60 billion more in plant and equipment by the end of 1940 than we did have. There is no doubt that with so much more plant and equipment in existence, our labor force could have produced very much more than the figures given above suggest. So, to say that we fell short of capacity output in this eleven-year period by something like $375 billion is to make a most conservative estimate. If we had had full employment throughout the period, and if (as certainly would have happened) some of that additional employment had been directed toward expanding plant and installing new equipment, thus adding to our capacity to produce, there is no doubt that we should have been able to produce a good deal more than $1025 billion worth of goods between 1930 and 1940. So actually we lost a good deal more than $375 billion in those years.

But let us see what even this cautious figure of $375 billion implies. Obviously this is a great deal of money, and even when spread over an eleven-year period it represents an enormous volume of goods and services. Since there were roughly 30 million families in the United States during those years, this sum would have given each family on the average more than $12,000 additional to dispose of in the period — $12,000 more for food, clothing, housing, automobiles, life insurance policies, government or industrial bonds, or what you will. At 1939 prices, $375 billion would have bought a new, comfortable house for every family in the country and a couple of automobiles besides. Our failure to rid ourselves of unemployment obviously deprived us all of much material well-being. Another way of viewing the direct cost of unemployment in terms of goods not produced is to compare the loss with the value of munitions produced during the war. Measured in 1939 prices, we produced about $240 billion worth of munitions between 1940 and 1945. We know that that represented an enormous quantity of ammunition, tanks, aircraft, and guns. And yet, if we had been able to avoid unemployment, we could have turned out about 50 per cent more than this in civilian goods during the period from 1930 to 1940.

Unemployment and depression (of which it is the most tangible sign) affect all of us. These are not things that concern only the unemployed or the economist. In a depression, corporations earn less money. They actually earned about $30 billion between 1930 and 1940, whereas, with peak prosperity and full employment, their profits would have been over $100 billion, according to estimates. Thus depression cost owners of corporations about $72 billion. Obviously, then, depressions are not good for the investor. Nor are they good for the farmer. Farmers earned about $48 billion between 1931 and 1941. They would have earned about $75 billion if we had maintained full prosperity during those years. The $27 billion difference is a measure of what unemployment and depression cost the farmer in that period. The small businessman suffered too. Not counting his losses through bankruptcy, we can estimate that the failure to maintain peak production between 1931 and 1941 cost him $39 billion. As it was, he earned only $85 billion; the additional $39 billion would have been a welcome supplement. Obviously the small businessman does not profit from depressions. The employee, of course, suffers the biggest loss of all, in absolute terms. Wage earners and salaried

employees actually received incomes of $475 billion between 1931 and 1941. If they had all been able to find jobs during those years, their income would have been $685 billion. Their loss, for which the economy's failure to run at full speed was to blame, came to about $185 billion. The depression cost us dearly as an economy. And it cost each class in the economy very heavily. Depression does not pay; we all lose from it.

Judged by its direct effects upon our economic well-being, depression is obviously harmful. The indirect losses we suffer because of our inability or unwillingness to solve the problem of unemployment are more elusive, though no less important. The loss in morale and psychological well-being for those who must suffer long periods of idleness is not the less because it cannot be measured precisely. Many of these people were used to working; they had been employed during most of the nineteen-twenties. They had families and homes to keep up, and they had self-respect. When unemployment hit them, the effects were bound to be serious. Because they had no opportunity to work, many of them lost skills which had taken years to acquire, for a skilled man long out of work loses his special abilities, and he must reacquire them before he is again fully productive. Part of our strength as a nation lies in the fact that our labor force is more skilled than that of other countries. Obviously, then, serious unemployment endangers that superiority. And just as our labor force loses its skills in periods of heavy unemployment, so, as we shall see later, our equipment becomes less efficient, our factories deteriorate, and our productive capacity is reduced generally.

If we leave coal in the earth, we do not lose the coal; we can mine it another day, and the energy it contains will still be there to use. Unfortunately, this is not true of labor. Labor unused today is not additional labor available tomorrow. Worse than that, the man who does not work today is even less capable tomorrow. The losses we suffer because of unemployment therefore cannot be made up. Likewise the factories or equipment which lie idle do not as a result become more productive tomorrow, but like labor grow less productive. The losses due to unemployment and depression cannot be made good later; they are permanent losses of productive energy.

A world in which depression prevails is not likely to be a world in which peace is secure. We must avoid depression, not only because of what it does to our pocketbooks, but also because of what it may

do to our lives. Whatever the causes of World War II, they were certainly reinforced by the dreadful economic condition of Europe during the nineteen-thirties. And unfortunately for our self-esteem, a most important cause of Europe's depression was our own. Beginning in 1930, we exported unemployment to Europe on a lavish scale. For unemployment, like influenza, spreads without regard to customs barriers and national boundaries.[1] In the nineteen-thirties, it spread from this country to Germany and Italy, to France, Great Britain, Australia, and Canada, and to most of the rest of the world. Nothing that these countries could do, or at least would do, was able to protect them from the effects of our severe depression. As the depression spread, so economic misery grew in Europe. With that came the Hitlers and the Francos, and with them, intensification of those economic policies that now look like the preliminary skirmishes of the fighting war.

Thus, avoiding depression is important, not only because of its direct effect on our pocketbooks. True, we can be richer if there are jobs for all who want to work. But in addition, avoiding a depression in this country is of vital importance in the maintenance of a skilled, efficient labor force at home, and in creating those conditions abroad in which war is less probable. To repeat: the cost of depression is immeasurable. In the ten years between 1930 and 1940, depression cost us at least $375 billion in goods; and in addition, such less easily calculable items as the immense sums not realized because our productivity did not increase at a satisfactory rate; the billions lost by deterioration in the skills of our labor force; the immense sums (if we can measure them in money at all) in damage to human personality — and on top of all that, a part of the expense of the war. These are costs beyond calculation. Depression is simply too expensive, even for us! When we are tempted to say that a certain measure to solve the unemployment problem costs too much, we should remember what we have to pay for not solving it.

Unemployment has been our most serious economic problem. Moreover, it has been perhaps the most serious economic problem of all the advanced capitalist economies. It is hardly surprising, then, that many of the most prominent economists of the last few decades have concerned themselves with this problem. With chronic depression in England and Germany in the nineteen-twenties, and severe

[1] The mechanism by which it spreads will be discussed in Part Five of this book.

depression almost everywhere from 1930 to 1935, it is only to be expected that much of the scientific progress in economics in recent years has been made in the analysis of unemployment and depression. The attention of most economists has in the last few decades been focused on the problem of depression. Incidentally, it is interesting to note that very little attention was given to this problem in the nineteenth century, simply because it was seldom a pressing one in the real world. In fact, few books on economics written in that century make any reference at all to unemployment or depression. In recent years, however, the advances which have been made in the theory of employment — or the theory of the national income, as it may be called — have been enormous.

The most influential work in this field among English-speaking economists has undoubtedly been accomplished by the late John Maynard Keynes. His classic book, *The General Theory of Employment, Interest and Money*, was published in England in 1936. The following chapters on the theory of employment are an attempt to give a simple account of Keynes's theory.[1] The book is difficult even for professional economists, nevertheless, the fundamental ideas — somewhat garbled at times — are passing more and more widely into circulation.

The central structure of Keynes's theory is simple, though it may not for a while appear so. A paragraph from the preface to his book may explain the reason for this:

> The composition of this book has been for the author a long struggle of escape, and so must the reading of it be for most readers if the author's assault upon them is to be successful, — a struggle to escape from habitual modes of thought and expression. The ideas which are here expressed so laboriously are extremely simple and should be obvious. The difficulty lies, not in the new ideas, but in escaping from the old ones, which ramify, for those brought up as most of us have been, into every corner of our minds.

The prime difficulty, then, is that the ideas are novel and therefore must compete against some mental furnishings which are pretty well established and hard to displace. It is not easy to change our ideas about anything — and especially about something like unemployment

[1] Keynes's book itself is difficult reading, although Chapters 18 and 24 of *The General Theory* can be read with interest and advantage after the study of the next twelve chapters in this book has been completed.

— when we are so certain of the correctness of our original views. And in the social sciences it is extremely difficult, first, to be fully aware of our preconceptions, and second, to examine them objectively. If we could bring to the task the same scientific attitude that guides the student of chemistry or biology, there would be no difficulty. But, as we have already seen, it is much harder to give up the belief that "money needs a gold backing" than it is, let us say, to accept a new model of the structure of the atom. It may prove especially difficult to do so in the field that we are now going to study; for very naturally, opinions about these matters constitute part of the ideology of our political parties. Our magazines and newspapers, our editorial writers, cartoonists, and columnists, all deal with these matters — perhaps crudely and even erroneously — but certainly not without assurance. And we cannot help accepting, sometimes unconsciously, what we hear and read in such quantity.

A word must be said, before we begin our analysis, about the political implications of the Keynesian theory. This is necessary because there is so much misinformation on the subject. The truth is simple. The Keynesian theory no more supports the New Deal stand or the Republican stand than do the newest data on atomic fission. This does not mean that the Keynesian theory cannot be used by supporters of either political party; for it can be, and if it is properly used, it should be. The theory of employment we are going to study is simply an attempt to account for variations in the level of employment in a capitalist economy. It is possible, as we shall see later, to frame either the Republican or the Democratic economic dogma in terms of the theory. After all, both good Republicans and good Democrats can analyze the causes of mental illness or of faulty timing in an automobile engine. And so the following chapters are neither an attack upon, nor a defense of, the beliefs of individual political parties. Rather, they are intended to show how a good many modern economists analyze this primary economic problem.

The importance of avoiding unemployment cannot be overstressed. The depression of the nineteen-thirties cost us very heavily in terms of output, in terms of morale and skill, and, in all likelihood, in terms of peace. There is every reason to believe that if we are not able to avoid unemployment in the future, it will cost us even more heavily. On so important a matter, it is obviously of very great importance that we approach the problem in as objective and scientific a spirit

as possible. We must for our own sakes, individually and collectively, be capable of viewing the evidence without prejudice, for if we approach the problem with our minds already made up, we can learn nothing.

27

The Determinants of Income

OUR ECONOMY in the years since the First World War has acted rather like a small boat on a stormy sea. It rose during the nineteen-twenties as if to the crest of a great wave, and then plunged into the trough so rapidly that it seemed doubtful whether its fall would ever cease. Then once again, in the middle nineteen-thirties, it began to climb, and, with relatively minor interruptions, had by the end of the decade reached a height about equal to that of ten years before. But instead of falling away this time as it did after 1929, it rose to new heights between 1940 and 1945, propelled upward by the Defense Program, Lend-Lease, the war, and reconversion. By the spring of 1947 it was shuddering at the top. What about the future? Is our economy again headed for a terrifying dive into depression? Can we depend upon "natural forces" to maintain prosperity? If not, can anything be done? Or is our economy to be subject in the future to the cycles of prosperity and depression that have been so characteristic during the last three decades? Do we have to face this succession of booms and slumps, or can the storm be calmed?

Let us first set out some bench-marks. In the early nineteen-thirties there were about seventeen million people unemployed in the United States, and in the early nineteen-forties almost none. Our total labor force, to put these figures in their proper setting, amounted to between 50 and 60 million workers, with roughly 37 million employees in non-agricultural industries. At the depth of the depression, (1932) farm national income totaled about $2.4 billion a year, and

in 1945 it reached $14.8 billion, a more than 500 per cent increase. In 1933 the average employed worker earned about 45 cents an hour; by 1945 his wage for an hour's work was almost $1. In 1932 profits of incorporated business firms before payment of taxes were *minus* $3.1 billion; by 1945 the profits of such firms, before taxes, amounted to approximately $20.9 billion. In 1932 the gross national product was about $55 billion, having fallen from almost $100 billion in 1929. Then, in 1944 and 1945, the gross national product was valued at the astounding figure of $200 billion. It is clear that the journey of our economy through the years has not been smooth or calm.

Moreover, our own country has not been the only one to suffer from these alternating seizures of economic chills and fever, although our attacks have been unusually severe. The United Kingdom, for instance, has also had its depressions and its prosperity. Unemployment, for example, was recorded at 1,176,000 in July, 1929; at 2,723,000 in December, 1932; and by April, 1946, it had fallen to 380,581. Similarly, Canada, Germany, France, and other important capitalist economies have experienced very great fluctuations of economic activity.

These rises and falls — rather, these soars and dives — in income, wages, employment, prices, and so on, are important not only to the statistician; they affect the welfare and even the life of every one of us. In the following chapters we shall analyze in some detail this interesting and important problem in an attempt to discover why employment and the national income fluctuate as they do. Obviously this question does not arise from idle curiosity; indeed the answer to it is a matter of very real importance to our future.

The Critical Role of the Firm

These fluctuations in production, income, employment, prices, and so on do not just happen. While a description of their underlying causes cannot be set out at this stage, it should be understood that in a capitalist economy the proximate or most immediate cause of these fluctuations is to be found in the actions taken by business firms. If production rises, it does so because these firms have decided to produce more; if employment falls, it does so because they have dismissed some of their employees. And while it would be absurd, or at least superficial, to say that business firms *cause* these cycles in income and employment, it is perfectly correct to say that in a capitalist economy,

the causative forces operate *through* business firms. The individual firm decides how much to produce, what price to charge, and how many men to employ. Each firm is free to produce as much or as little as it chooses — or, to put this in another way, each firm may hire as many or as few men as it wishes. Generally speaking, of course, it chooses to employ the number of men which in the circumstances appears to be most profitable. If by employing ten men it can anticipate profits of $2500 in the year, and if by employing nine men it can expect profits of $2600, it will ordinarily give jobs to nine rather than ten men. If by producing an output of 100 units, the firm can expect to earn greater profits than from any other output, it will produce 100 units. Hence, while the firm is a free agent in the sense that civil laws do not compel it to produce a certain predetermined amount, it normally seeks to maximize profits. In this attempt, it is subject to all kinds of economic forces. What we must analyze, then, is the nature of the forces that impel firms at times to give more jobs, at other times to dismiss employees; at certain times to raise prices, at other times to lower them — in short, to make all the various decisions it makes which affect the functioning of the economy.

Variations in Demand and Employment

In our analysis of the firm we saw that the greater the demand for its product, the larger is the output it would choose to produce, and therefore the greater its working force would be. An increase in demand would lead the firm to expand employment, to increase its output, and possibly to raise its price; whereas a decline in demand would induce the firm to dismiss some of its employees, to reduce its output, and possibly to lower its price. What is true of the single firm is obviously true of the agg egate of firms. If an increase in the demand for the product of one firm leads it to hire more men, and to produce a larger output, so an increase in the aggregate demand for the products of the whole economy leads firms in general to employ more men and to raise output. Variations in output and in employment are brought about by variations in demand. In seeking to explain such variations in employment and production, we must discover the factors that cause variations in demand. The analysis of these factors constitutes the core of what is now called the theory of employment.

The Level of Spending and Employment

To say that output and employment vary with the *demand* for the products of the economy is true, but it is not the formulation which is most convenient for further analysis. Let us see whether it is not possible to recast it in more usable terms by substituting a simpler notion for the demand concept. We must recall that when the demand changes, a firm normally changes its output and price in the same direction. But the product of the price times the amount produced equals the sales receipts. Hence when the demand for a product increases, the seller's receipts also increase; and when the demand declines, receipts fall. The relation between employment or output and demand can therefore be expressed instead as one between employment or output on the one hand, and sales receipts on the other. But even this formulation may be improved. Every cent a firm receives is a cent paid by a buyer. If buyers spend $10,000 a week on the product of a firm, its receipts are $10,000 a week. Sales receipts are thus equal to the amount spent. Hence, we may reach this tentative formulation: output and employment vary directly with the amount spent; when the amount is great, the amount produced and the number of jobs filled are both high; when the amount spent is small, the amount produced and the number of jobs are low. The level of spending is thus critical in determining whether output and employment are high or low; or to put this in other words, the level of spending determines whether we have prosperity or depression.

Although the relation between the amount of spending and the amount of employment is direct, it is not unique. That is to say, the level of employment is not determined by the amount of spending alone. Other variables, such as the level of wage rates, labor productivity, and the degree of monopoly, also influence the amount of employment.[1] Yet though employment does not depend on the volume of spending alone, a table in which these two variables are set out side by side for each year is revealing. It will at once be clear from data in Table 42 that employment has generally increased in periods when the volume of spending has increased, and fallen when the volume of spending has fallen. However, for reasons that have just been set out, the correspondence is not perfect.

[1] Or to put this in the language employed in Part Two, the amount produced depends not only upon the level of demand, but also upon its elasticity and upon marginal costs.

TABLE 42

The Relation Between Amount of Spending and Amount of Employment

Year	Volume of Spending * (in billions of dollars)	Employment † (1929 = 100)
1929	99.4	100.
1930	88.2	89.2
1931	72.1	73.4
1932	55.4	55.2
1933	54.8	57.1
1934	63.8	61.3
1935	70.8	68.5
1936	81.7	80.0
1937	87.7	81.9
1938	80.6	69.0
1939	88.6	79.2
1940	97.1	86.1
1941	120.2	101.4
1942	152.3	115.7
1943	187.4	120.9
1944	197.6	119.6
1945	197.3	110.5

* Gross national product in current dollars: estimated by United States Department of Commerce.

† A rough allowance has been made for changes in hours of work per week. Hence the employment series represents approximately man-hours of work.

Spending and the Gross National Product

We have already developed an identity of which we shall now make use. In Chapter 21, we showed that if we add the amounts of money spent in a certain period on the products of all firms, we reach, after allowing for duplication, a total exactly equal to the gross national product of the economy; that is to say, if $100 billion worth of new products are purchased in the course of the year (not including such duplications as occur when one firm buys raw materials from a second and uses them up in producing a commodity purchased by consumers), the total gross income of the economy in that period equals $100 billion.

Before going any further with our inquiry, let us summarize the points already made. First, in a capitalist economy, firms respond directly to changes in demand: the higher the demand, the greater is

the amount they produce and the larger the number of jobs they provide. Next, since changes in demand lead to changes in the same direction in spending, the condition for a large output and high employment is that spending be high. More generally, the level of spending determines output and employment. We can also equate the amount of spending with the gross national product, or after adjustments for taxes and depreciation, with the national income. Hence, in order to explain variations in employment and in the national income, we must account for changes in the level of spending.

Consumers' Goods and Investment Goods

The problem that we must solve is this: We must explain why the amount of spending has varied recently between $50 billion and $200 billion a year. At first glance, this looks like a hopelessly complex task. The variety of things purchased is so enormous that to discover any simple explanation for changes in the amount of spending seems to be impossible. After all, the spending with which we are concerned covers such various things as shoe polish, canned soups, dress suits, books, raw cotton for England, automobile parts for Canada, a highway, new houses, a plant for fabricating magnesium, and new equipment for a research laboratory. The only hope is to try to classify these items, for perhaps by so doing we can divide the main spending stream into a few smaller streams, each sufficiently homogeneous to permit a simple analysis of variations in its rate of flow.

In order to classify spending in this way, we must first examine the sources of spending. When a certain sum is spent on the products of our economy, who are the buyers? From whose pocketbooks and checkbooks does this money come? The greatest amount of purchasing is done by ordinary consumers. When we buy clothes, or a car, groceries, or meals in a restaurant, a ride on the bus, or theater tickets, we are buying as consumers. The common feature of all these purchases is that they are made not with a view to resale, but rather so that the purchaser or his family may use directly whatever has been bought. Such commodities are classified as *consumers' goods*. Indeed, the classification is somewhat narrower than this. Because of certain characteristics of international trade, we mean by the term consumers' goods, those goods that are purchased by consumers who are residents of this country. The amount spent on consumers' goods has recently varied from about $40 to $135 billion a year.

As we have said, most of our spending is for consumers' goods, but there are other spenders in the economy besides consumers. After all, many goods and services are bought for reasons other than those that motivate the consumer. Whoever buys a piledriver, for example, or a large power dam, a bomber, or a rolling mill, does not do so in the expectation of using it directly to satisfy a need. The president of the railway company does not buy locomotives simply to make his own commuting easy. All these goods which are not bought by consumers are called *investment goods*. Since investment goods (and services) comprise all that is not bought by consumers, it follows that total spending equals the amount spent on consumers' goods plus that spent on investment goods.

A sub-classification of investment goods is frequently helpful in analyzing changes in spending on those goods. Such goods bought by business firms are called *private investment goods*. When a firm buys investment goods, it does so because it expects to earn profits from their use. It may intend to use them in the production of other commodities which it can sell at a profit, or it may simply wish to store them until later, and then sell them at a higher price. This kind of purchase is motivated by the hope that one may "buy cheap and sell dear." Investment goods bought by governments are called *public investment goods*. A government does not have wants and needs like those of an ordinary consumer. When it lets a contract for the construction of a highway or a battleship, it does so for reasons entirely different from those that prompt us to buy a package of cigarettes or a shirt. Finally, investment goods bought by foreign purchasers are called *foreign investment goods*. In this category are placed all goods that are exported, whether eggs for the housewife, raw cotton for the manufacturer, books ordered by a private individual, or fuel oil for industrial use.

For convenience we shall often write *consumption* in place of *the amount spent on consumers' goods*, *private investment* in place of *the amount spent on private investment goods*, and so on. Now, since all goods can be classed as either consumers' goods or investment goods, and all investment goods as either private, public, or foreign investment goods, we see that total spending equals the sum of consumption, private investment, public investment, and foreign investment.

Perhaps it will now be clear why we have classified spending as we have. We set ourselves the problem of determining why spending

varies, which means that we have to analyze the factors that determine spending. Consumers' motives are so different from those of business firms and other buyers that we should hardly expect to find that consumption and private investment varied for similar reasons. Whatever the reasons that persuade consumers to buy more vegetables, they are not likely to be the same as those which impel a city to hire more school teachers, or induce British purchasers to buy more wheat, or persuade the General Motors Corporation to build a new assembly plant. Hence some classification is necessary. But since business firms are generally subject to much the same influences when they determine how much to spend, the attempt to explain in simple terms the variations in their spending is more promising. Likewise, consumers are a relatively homogeneous group, at least in the way their spending responds to economic influences; and accordingly we may hope to explain the changes in their spending. Thus the classification will help solve the basic problem — which is, how to explain fluctuations in total spending.

Problems in Classifying Goods

The classification of most commodities as either consumers' goods or investment goods is simple enough. There is no question, for example, that food in the kitchen of a private house is a consumers' good. Just as obviously, food in the kitchen of a restaurant is an investment good, since it will be used in preparing meals to be sold for a profit. In the same way, coal burned to heat a private house is a consumers' good, while coal used to heat a factory building is an investment good. The looms owned by a textile firm are investment goods. The stocks or inventories in the possession of a retailer are also investment goods, even though they may shortly be sold to consumers and become consumers' goods. The same good, therefore, may at one time be an investment good, and at another, a consumers' good. These cases are all quite obvious as long as we keep clearly in mind the basis upon which the classification is made.

But there are other cases where it becomes rather more difficult to apply the criterion. What, for example, should we call a house? If it is used by its owner, it should strictly be classed as a consumers' good; whereas if it is rented to someone else, it is more appropriately classified as an investment good. Generally, however, whether it serves as the residence of the owner or not, a house is treated as an

investment good. Again, how shall we classify a private automobile? If it is used simply for pleasure driving, it should be classified as a consumers' good. On the other hand, when a doctor uses his car to visit his patients, it should be regarded as an investment good. A car used several days a week for professional or business purposes and on other days for pleasure would have to be classified partly as a consumers' good and partly as an investment good. But while classification is occasionally difficult, in the majority of cases it presents no problem.

Spending on Consumers' Goods: The Level of Income

As we have already seen, the level of spending determines the size of the gross national product and the amount of employment. Whatever determines spending, therefore, determines whether we have depression or prosperity. The objective of our analysis in the next few chapters is to isolate the factors that affect spending. Let us begin by treating the spending of consumers. Why is it sometimes as high as $125 billion and at other times as low as $50 billion a year? This is obviously a very important question — indeed a $125 billion question.

In the summer of 1946, consumers' expenditures were higher than they had ever been before. Statistics for April, 1946, showed that in that month retail sales were $7.7 billion compared with $5.6 billion in April, 1945, and $3.2 billion a month in the period 1935 to 1939. In the summer of 1946, department store sales were about two and one-half times as great (in dollar terms) as from 1935 to 1939. Chain stores and mail-order houses were selling more than twice as much as in the period from 1935 to 1939 — almost four times as much women's wear, one and one-half times as much jewelry, and two and one-half times as much food.[1] Consumers in 1946 were a big market. The most important reason for this is that they had the money to spend. Their incomes were very high compared with those of, say, 1939, or 1935, or 1932, or even 1929. Consumers spend a great deal when they are earning a great deal; when they are unemployed, or earning low wages, or when their wheat is selling for fifty cents a bushel, they spend very little.

There is a good deal of statistical information which will help us to

[1] All these comparisons are of amounts spent. Because of the rise in prices, the increase in physical units purchased was somewhat less, though generally substantial.

determine with greater precision the relation between consumption and income. Here are some relevant data: The gross national product in 1932 was about $55.4 billion, and consumption was $43 billion. In 1935 the gross national product was about $70.8 billion, or $15.4 billion higher than in 1932, while consumption stood at $52 billion, or $9 billion higher. By 1937 the gross national product was $87.7 billion, and consumers' spending equaled $62.5 billion. Thus betweeen 1932 and 1937 there was a rise of $32.3 billion in the gross national product, and an increase of $19.5 billion in consumption. By 1941 the gross national product had reached $120.2 billion, $64.8 billion more than in 1932, and consumption had climbed to $74.6 billion, or $31.6 billion more than in 1932. Apparently, the higher the gross national product, the higher is consumption. This conclusion is hardly surprising. When the gross national product is high, most people have jobs that pay well, and many families have a good deal to spend. When the gross national product is low, there is a good deal of unemployment, wages are low, and many families are able to buy none but the most essential commodities.

It is also evident from the data that although consumption increases when income does, it does not increase by as much. Thus, for example, the gross national product increased by $32.3 billion (and the national income by $31.5 billion) between 1932 and 1937, but consumption increased by only $19.5 billion. Consumption responds to increases in income, but it appears that the response is less in amount than the increase in income which brings it about.

The results are similar when income or the gross national product is falling. Between 1929 and 1932, the gross national product fell by $44 billion, but consumption fell by only $27.8 billion. Again between 1937 and 1938 there was a decline in the gross national product of $7.1 billion, but the fall in consumption was only $4.0 billion. Generally, then, we may conclude that when income changes, consumption changes in the same direction but by a smaller amount. This characteristic of our spending habits is a very important one.

The Propensity to Consume

To indicate the relation between consumption and income (or the gross national product) we may prepare a graph, as in Figure 59. It will be noticed that the curve which shows the relation between consumption and income is similar in concept to a demand curve,

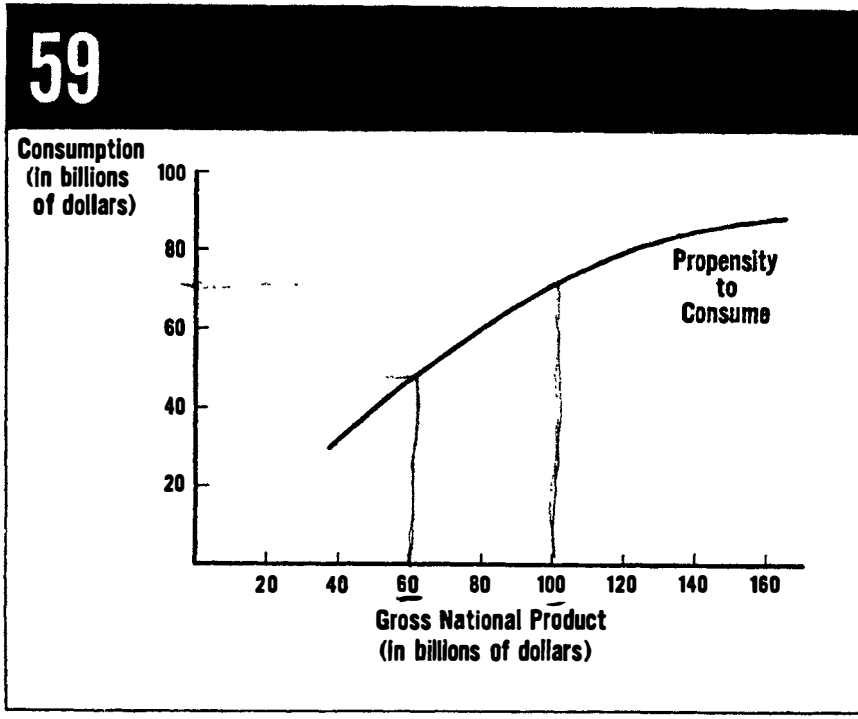

Figure 59. *The Propensity to Consume*

since it describes the aggregate demand for consumers' goods at various levels of income,[1] though in the ordinary demand curve, of course, the independent variable is not income, as it is here, but price. In a graph showing the relation between consumption and income, we conventionally represent the gross national product on the horizontal (OX) axis and consumption on the vertical (OY) axis. In accordance with our conclusion that changes in consumption are less in amount than the changes in income which bring them about, we have a curve whose slope is never as great as forty-five degrees, assuming that the same scale is used for the OX and the OY axis. This function or curve which is known as the *propensity to consume*, simply indicates the level of consumption forthcoming at each level of the national income.

The propensity to consume expresses the relation between the gross

[1] However, a demand curve as usually drawn shows not the total amount spent upon the product, but instead the number of units demanded. For this reason it slopes down to the right, while the consumption curve slopes up to the right.

national product and consumption. If we know its nature, we can determine consumption when we are given the gross national product. Reading from the propensity to consume curve in Figure 59, we see, for instance, that when the gross national product is $100 billion, consumption is $70 billion; when the gross national product is $60 billion, consumption is $45 billion; and so on. But we are able to determine consumption from the propensity to consume only if we know the gross national product. Hence the propensity to consume does not, by itself, determine consumption. All it does, to repeat, is to show consumption at each level of the gross national product.

The Propensity to Consume, Investment, and the National Income

The cornerstone of the theory of employment is that the national income (or gross national product)[1] depends upon the propensity to consume and investment. Given the economy's spending habits — that is, the propensity to consume — the amount of investment determines the national income and the level of employment. We shall endeavor first to show the truth of this statement arithmetically and, later, to show the sense of it in more concrete terms.

An analysis that seeks to determine causal relation should be carried far enough to give meaningful results. It may of course be expressed in the most superficial terms, but such an expression, while correct, will probably not be very useful. Thus, as we have already pointed out, it is formally correct to say that depressions are caused by the actions of business firms, but this is not a useful statement, for it says nothing about the causes of these actions. Or we could go one step further, and say that depressions are caused by inadequate demand. While this is a more useful statement, it still fails to explain depressions satisfactorily, for it does not make clear why demand should ever be inadequate. It may well be asked whether it is possible to explain in simple terms why the demand should ever be deficient for such diverse things as tooth brushes, canned tomatoes, warehouse buildings, steel rails, aircraft, naval vessels, and raw cotton for export. The very complexity of the list suggests that it is not possible to explain all these things simply. Hence the analysis must be extended.

[1] The national income differs from the gross national product by the amount of business taxes, allowances for depreciation, and a few minor factors. Although the two concepts are not exactly the same, the terms may frequently be used synonymously. We shall henceforth employ them interchangeably except in those special circumstances when it is important to maintain a distinction. The reader's attention will be drawn to those instances.

After additional steps we arrive, as we have seen, at a new formulation: that depressions occur when the sum of consumers' spending and investment spending is too low. This statement is, in an arithmetical sense, as correct as any, and in addition, it is more useful than the preceding ones because it directs our attention to factors which are causally more fundamental and significant. But it suffers from one great defect. We have seen that consumers' spending itself depends upon whether we have depression or not, or as we have expressed it, consumption varies with the national income. Hence, to say that depressions occur when consumers' spending is very low is rather like saying that depression occurs when we have depression, which sounds like a perfect circle instead of an analysis. But the argument is not circular, because in that statement, investment spending is also given as a determinant. Yet we should like a reformulation of this statement that saves its content but avoids the appearance of circularity. Perhaps the relation we have developed between consumption, the propensity to consume, and the national income will suggest a method.

Let us then test the following formulation: the national income depends upon investment and the propensity to consume. It is not conspicuously circular, as the previous one was, for the propensity to consume does not depend upon the level of income. But is it true? And is it useful? The second question can be answered only by trying it, by putting it to the test. This will be done in the following chapters. But what about its truth? Is the national income determined by the propensity to consume and by investment? Let us first see what this statement looks like in terms of arithmetic. What we must do is to see whether there is one value of the national income, and only one, which is possible when investment and the propensity to consume each take a certain value.

Assume that investment is 45 and that the propensity to consume is such that for a part of its range,

when the national income is:	100	125	150	175
consumption is:	80	92.5	105	117.5

Can we, on the basis of these figures, determine the level of income? Is there a single level of income which is compatible with these assumed values? First, let us try to determine income by trial and error. Can it equal 100 in this situation? If it were 100, consumption would be 80. Since investment is set at 45, we discover that income (which,

as we have already shown, equals consumption plus investment) is not the 100 we have assumed, but 125. Hence our first guess, that income equaled 100, is incorrect. Let us now see what happens if we take 125 as the figure for income. From the propensity to consume function, we read that consumption at this level of income would be 92.5. Hence with investment of 45, income, which is the sum of investment and consumption, would be 137.5. So we must discard this guess also.

Let us try again, this time assuming that income is 150. From the data given for the propensity to consume, we determine that consumption at this level of income is 105, and therefore that the sum of consumption spending and investment spending is 105 + 45, or 150. Our hypothesis that income would be 150 is therefore not inconsistent with the facts. We do not have to discard this guess. But all other guesses must be discarded, for the reader will see that the income could not be 175; nor interpolating for intermediate values of consumption and income, could it be at any figure other than the one at which we have already arrived, that is, 150.

Readers who are familiar with elementary algebra will probably recognize that what we are dealing with is essentially a problem involving two unknowns and two equations. We may set up these equations as follows: we designate income by Y, consumption by C, and investment by I, then we have:

(1) $Y = C + I$

and (2) $C = \phi(Y)$ where $C = \phi(Y)$ is the propensity to consume. Let us make the propensity to consume function more concrete, so that it reflects the figures we used for illustration. And let us also substitute 45 for investment. If this is done, we have:

$$Y = C + 45$$

and

$$C = \tfrac{1}{2}Y + 30$$ [1]

There is but one solution for these equations: $Y = 150$ and $C = 105$. And even though investment should be different, say, 25, and the propensity to consume function should be more complex, for example, $C = Y^{3/5} + 15.8$, the equations can be solved for Y and C. In a mathematical sense we may conclude that the national income depends upon investment and the propensity to consume.

[1] This formula for the propensity to consume gives the values that were set out above. For instance, when $Y = 100$, $C = 50 + 30 = 80$; when $Y = 150$, $C = 75 + 30 = 105$, and so on.

Now let us see how this dependence actually works out in the economy. As before, we shall assume a propensity to consume such that:

when income is:	100	125	150	175	200
consumption is:	80	92.5	105	117.5	130

And this time, for variety, we shall assume that investment, instead of being 45, is 32.5. Now, in order to test whether the income itself could be at the level of 100, let us further suppose that business firms decide on the output level that corresponds to this volume of spending on their products. Businessmen, their employees, and their families would then, in accordance with the assumed propensity to consume, spend 80 on consumers' goods. But, by assumption, 32.5 is being spent on investment goods. Hence businessmen are receiving 112.5 instead of the 100 which they anticipated. What this means is that they have underestimated their demand. Naturally, as soon as they realize their error, they will revise their production plans; and since demand is higher than they supposed, they will produce more. But the national income, even before they do so, is not 100, but somewhat higher; for since firms are receiving more than they expected, they are earning more profits than they expected. That is, the sum of wages paid out, profits actually earned, rents, interest payments, and salaries exceeds 100. We can summarize what actually happens as follows:

Assumed income:	100
Consumption based on assumed income:	80
Assumed investment:	32.5
Therefore actual income:	112.5 (*if* consumption is 80)

Hence income is not 100,
And therefore consumption is not 80.

Now let us see what would have happened if, on the basis of their demand estimates, firms had determined on an output and employment level that corresponded to a national income of 150. At this level of income, 105 would be spent on consumers' goods. But since only 32.5 go to investment goods, it is clear that firms have receipts of only 137.5. Consequently, profits will be lower than expected and income will of course be below 150. Indeed, it will be less than 137.5, for with profits actually lower than expected, consumers' spending will be less than 105. And as production plans are revised, men lose

their jobs and consumption sinks still further. Obviously, therefore, with the assumed propensity to consume and with investment at 32.5, the national income is neither 100 nor 150. Actually it is 125, for with incomes being earned at this rate, consumption will be 92.5; and since spending on investment goods is 32.5, total spending is 125. Thus income is generated at the rate of 125, which is consistent with the assumed income level. Business firms sell at the rate they anticipated, and they have no reason to revise their plans. In summary:

Assumed income:	125.0	
Consumption based on assumed income:	92.5	
Assumed investment:	32.5	
Hence actual income:	125.0	(if consumption 92.5)

Therefore income is 125.

We have seen that income could not be 100 with this propensity to consume if investment is 32.5. Under what conditions, then, would income be established at 100? This would occur if investment were 20. Consumption would then be 80, and total spending, and hence income, would be 100.

It is sometimes convenient to express the determinants of the national income, not in terms of money, but in terms of employment. Looking at the problem in these terms may help to clarify some difficulties. Suppose we undertake to determine the conditions that must prevail if the economy is to provide jobs for 50 million men. Let us assume that the propensity to consume is such that when this number are working, they and their employers choose to spend on consumers' goods just enough to make it profitable to hire 38 million men for the production of such goods. Thus, of the 50 million men presumed to be working, 38 million are employed in consumers' goods industries. The other 12 million, if they are to be at work, must then be engaged in producing either goods for government units, goods for business firms, or goods for foreigners — in other words, investment goods. If the demand for investment goods were not enough to require the labor of 12 million men, employment could not remain at the assumed level. Employers who anticipated that the demand for their products would be at a level at which they could profitably hire 50 million men would find that they had been unduly optimistic. And their expectations would be even further disappointed, for if fewer than 12 million men were employed in producing investment goods, the de-

mand for consumers' goods would be below the level at which it would be profitable to employ 38 million men in consumers' goods industries. Employers would therefore find, not that it was *necessary*, but instead that it *paid*, to reduce employment and output below the assumed figure.

On the other hand, if the estimate had erred in the opposite direction, and more than 12 million men were needed to produce investment goods at the desired rate, employers would find that sales exceeded expectations. They would therefore be induced to expand employment in the consumers' goods industries as well as in those producing investment goods. Hence the employment figure would rise above the 50 million mark.

To summarize: the national income can be said to depend: upon (*a*) the activities of business firms; or more helpfully, upon (*b*) the total demand for commodities; or still more usefully, upon (*c*) total spending; or still more helpfully, upon (*d*) the sum of consumption and investment spending; or finally, upon (*e*) the propensity to consume and investment. We have shown that this final statement is true, since it is merely another way of putting the essence of statements (*b*), (*c*), and (*d*). After clearing away some possible misconceptions, we shall further show that this final formulation is also useful in analysis.

The Meaning of Saving

First of all, we must be quite clear on the significance of saving in our economy. Saving is defined as the difference between the gross national product (not national income) and consumption.[1] The amount of one's gross income, including profits before the subtraction of business taxes and depreciation allowances, which is not devoted to the purchase of consumers' goods, is classified as saving, no matter how it is used. For instance, if out of an income of $2500 an individual spends $2000 on consumers' goods, the $500 difference is treated as saving. This $500 can be put into the bank, kept under the mattress, or used to purchase government bonds, corporation securities, a new machine, or a life insurance policy; it is still saving. Provided that it is not spent on consumers' goods, it is treated as saving no matter how else it is used. Thus, in a year in which the gross national

[1] This definition is not the usual one. It is adopted in this treatment because it is more convenient, in introducing the theory, to use the gross concepts such as gross investment and the gross national product. After the reader has become acquainted with the outline, it is an easy matter to substitute the more commonly used concepts.

product is $100 billion and consumption is $65 billion, saving amounts to $35 billion. For certain purposes we may be interested, of course, in knowing the form in which the saving is made. We may characterize that part of it which is kept under the mattress as "hoarding," and perhaps that part directed to the stock market as a "flow of speculative funds." Economists have made valuable contributions to our understanding of the economy by employing these distinctions. But for our purpose at this stage, it is sufficient to lump together all these various methods of saving.

As so defined, it will be clear that saving is a resultant; it depends upon the gross national product and the propensity to consume. In this respect saving is like consumption. In fact, where consumption is a positive disposal of income, saving is in essence simply a failure to consume income.

Analysis of the propensity to consume will show that saving generally varies directly with income. When the national income or the gross national product rises, saving increases, too, as long as the propensity to consume remains the same. When the national income falls, saving falls. This follows from that characteristic of spending habits to which attention has already been drawn — that when income changes, consumption changes in the same direction but by a smaller amount. This means that when income changes, the gap between income (or, more accurately, the gross national product) and consumption changes in the same direction; or, in other words, saving goes up when the gross national product rises, and vice versa. A numerical example will demonstrate this point. Suppose that:

When income is	100	
Consumption is	80	Then saving is 20.
And when income is	120	
Consumption is	90	Then saving is 30.

Hence saving rises with the increase in income and will, of course, do so because, when income changes, consumption changes too, but by a smaller amount.

Saving and Investment

Saving and investment are completely different. Investment is something active which helps determine income, while saving is merely a resultant, determined in part by the level of income. Invest-

ment measures the total amount spent by business firms, governments, and foreigners upon the products of the economy (after allowing for duplication); saving measures the amount of income (before taxes, and so on) not spent on consumers' goods. Hence while investment represents a positive activity, saving is a mere absence of activity.[1] Furthermore, saving is done — if a failure to act can be described as doing — by people in their capacity as income recipients, while investment is carried on by business and government units and by foreign buyers. Thus, saving and investment must be sharply distinguished.

Funds received as income which are not spent on consumers' goods may, of course, be used to finance the purchase of investment goods. Thus, out of an income of $3500, a man *may* spend $2000 on consumers' goods and use the other $1500 to purchase investment goods directly. Or he may instead buy some bonds that a corporation has just issued, and the corporation may itself use the $1500 to help finance the construction of a new plant. The actual dollar bills saved may be used in the purchase of investment goods. This is possible and it undoubtedly happens, but the direct and immediate transfer of money saved into investment purchases is of quite minor importance in the economy. Much of the actual money which is saved is used for the purchase of life insurance, or of bonds or securities that have long been outstanding, and much of it is deposited in banks. On the other hand, much of the money used for the purchase of investment goods has been borrowed (not necessarily from those who are currently saving) or has been accumulated from saving done in the past. In short, *the money* that is being saved currently rarely finances current investment. Thus, not only are saving and investment done by different people or institutions, and not only are they quite different in nature; they are not even linked by the fact that most of the money currently saved is used to pay for investment goods.

The Equality of Saving and Investment

It is necessary to be quite clear about the distinction between saving and investment because otherwise we may draw erroneous conclusions from an equality which will now be demonstrated. It can

[1] The sums saved may, of course, be used to buy securities, to build up a bank balance or even to buy consumers' goods in a subsequent period. But such purchases are *not* saving, nor do they even require a prior act of saving.

be shown that, despite the profound differences between the act of saving and the act of purchasing investment goods, saving and investment are equal in amount; that is to say, although saving (the non-consuming of income) and investment (the purchasing of non-consumers' goods) are quite distinct activities, the amount of saving is equal in any period to the amount of investment. Again it is emphasized that to show that they are equal in amount is not to imply that they are the same thing.

Let us assume a level of income or gross national product, say of Y. Let C represent consumption, and I represent investment. We have seen that Y is equal to total spending or that

$$Y = C + I,$$

$$\text{But Saving (by definition)} = Y - C$$

$$\text{Hence} \qquad C + I = Y = C + \text{Saving}$$

$$\text{And therefore} \qquad I = \text{Saving}$$

Or, in numbers, assume that

Consumption is	60
And investment is	30
Then the gross national product is	90

But saving, by definition, equals the gross national product minus consumption.

So Saving is 90–60, or 30.

Hence saving equals investment in amount.

As you can show with any examples you choose, this equality between saving and the amount spent on investment goods is one which is always maintained, no matter how short or how long the period within which incomes, consumption, saving and investment are measured, and no matter whether the economy is enjoying prosperity or suffering the deepest depression. The fact that saving equals investment is perhaps not of great importance in understanding the forces that determine the national income. However, it often proves a valuable check on an analysis, since we can be sure that we have made a logical error if, with the same definitions of saving, investment, and income, we imply in the course of the argument an inequality between saving and investment.

One point in connection with saving and investment requires very special emphasis. As perhaps it is a first step toward an understanding of the theory of employment to see that saving and investment are equal, it is a second step to see how that equality is maintained.

There is a temptation to conclude that the two are kept equal because a certain amount of saving necessarily brings about an equivalent amount of investment. We shall see later that this connection cannot be supported. In fact, we shall discover that saving and investment are kept equal, not so much because investment accommodates itself to the level of saving, as because saving accommodates itself to the level of investment. It is the level of investment which determines how much saving there will be, and not the other way about. We shall justify this statement at a later stage.

Summary

The argument of this chapter has been long and complex, but it must be understood thoroughly because it constitutes the essence of the theory of employment. To recapitulate it briefly: The national income and employment depend upon the amount of spending. When total spending is high, national income, output, and employment are all high; when total spending is low, they are all low. The amount of spending in any period is the sum of consumers' spending and investment spending: investment spending may be either private, public, or foreign. Consumers' spending depends upon the national income and the propensity to consume; this, of course, defines the propensity to consume. From the above relations we may derive this basic one: *the national income depends on the propensity to consume and investment;* hence, when these two fundamental determinants are set, the national income, and therefore the amount of employment, are also set.

Saving is defined as income (strictly, the gross national product) minus consumption. It follows from this definition that saving and investment, although they are completely distinct as acts, are nonetheless always equal in amount. But we have seen reason to believe that the amount of saving is a resultant; it does not determine the national income or investment, but rather it depends upon other factors. In contrast, investment is active, as we should already begin to see. It is a determinant of the other elements in the economy — of income, and through that, of saving.

28

The Determinants of Investment

INVESTMENT has been very unstable in the last few decades. It averaged $23.8 billion a year during the nineteen-twenties, and $19.8 billion during the decade from 1930 to 1939; but these averages obscure the extreme sharpness of the fluctuations. In 1922 it stood at $18.8 billion; in 1929 it had increased to $28.6 billion, a rise of more than 50 per cent. By 1932, at the depth of the depression, it had dropped to only $12.4 billion, or only 43 per cent of the 1929 level. Then in 1939 it had climbed again to $26.9 billion, a figure quite close to the 1929 high. But these figures were all made to seem lilliputian by what happened during the war, for in 1941 investment stood at $45.9 billion, and by 1944 it had reached the dizzying figure of $101.2 billion. Thus over the last several decades, investment has fluctuated just as wildly as the national income.

Investment has not only fluctuated tremendously in amount; in addition, in recent years the relative proportions of private, public, and foreign investment have varied almost as much. Over the whole period from 1921 to 1941, public investment was of the greatest importance, and constituted about 51.2 per cent of the total. During that period private investment accounted for almost as much, or approximately 45.9 per cent of the total, and foreign investment made up the difference, or 2.9 per cent. The pattern changed abruptly during the war years, 1942 to 1945. Public investment made up about 94 per cent of the total for that period and private investment about 6 per cent, while foreign investment was slightly

negative.[1] Throughout the period from 1921 to 1945 there have been very marked changes both in the total amount of investment and in the make-up of the total.

As we shall see later, the fluctuations in investment account for most of the variations in income and the instability in investment explains the instability of our economy. Hence we must discover why investment varies in order to understand why the national income varies and in order to see what can be done to stabilize it at a high level. The national income depends on investment and the propensity to consume. But what do these things depend on? We must now extend the analysis so that we can see how these determinants are themselves determined. What are the factors that affect investment? And what are the factors that set the propensity to consume? In this chapter and that following, we shall investigate the first of those two questions; and in Chapter 30 we shall investigate the second.

Composition of Private Investment

As might be expected, the series of private investment shows very great instability. The chart on page 372 (Figure 60) pictures its course from 1921 to 1945. From this figure it will be seen that in 1929, and again in 1941, private investment amounted to over $17 billion, while in 1932 it came to less than $2 billion. Over the whole period from 1921 to 1941 it averaged about $11.4 billion a year.

Private investment represents the total business spending of firms, including individual businessmen, farmers, and house purchasers, upon the products of the economy, after allowing for duplication. The allowance for duplication is made by subtracting from the total those expenditures for raw materials and other items of working capital which simply maintain the firms' inventories of such goods at a constant level. The items that make up private investment are of many kinds. The amount that a firm spends upon plant and equipment is included as a part of private investment. Thus, if a steel company spends $80 million in modernizing its rolling mills, or if a motor-car manufacturer spends $40 million in installing machine tools, we should count these sums as a part of the private investment of the period. When the railroads spend $280 million on rolling

[1] The method of measuring foreign investment employed by the United States Department of Commerce does not measure the concept used in this introductory account, but since the figures are relatively small, this difference may be overlooked.

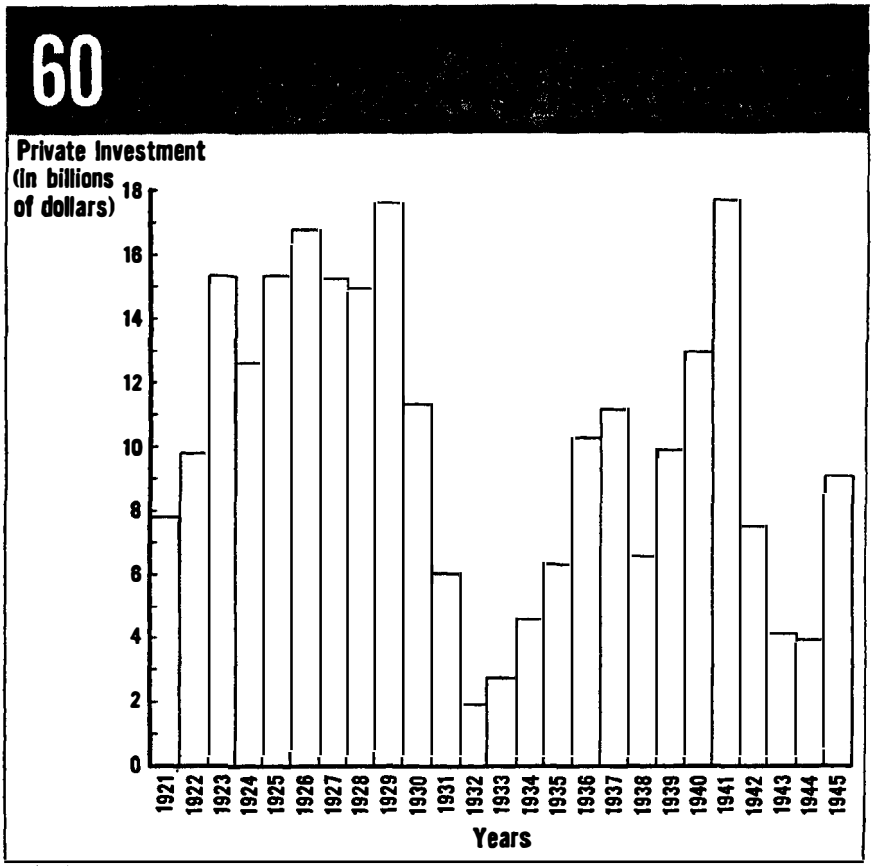

Figure 60. *Private Investment, 1921–1945*

stock, new lines, and other capital goods, as they did in 1939, this too constitutes private investment. The amount spent upon new houses is also included in the private investment figure. Likewise, spending by business firms which increases their stocks or inventories of raw materials, semi-finished, or finished goods, counts as private investment. Thus, when retailers build up their inventories by $260 million, as they did between April, 1945, and April, 1946, this figure is included with the other items of private investment. In concrete terms, then, private investment is the value of the output for private firms and individuals of such items as machinery, factory buildings, houses, goods added to inventories, dams, transmission lines, moving-picture

houses, apartment buildings, buses, ships, farm equipment, and so on. The list is complex and varied, but there is one characteristic common to all the items on it, and it is this feature which permits us to make useful generalizations about private investment.

The Profit Motive

The common feature of all private investment projects is that, since they are undertaken by business firms, they are made for profit. If a firm sees an opportunity to increase its profits by investing, it will do so. Otherwise it will not. A steel company will willingly build new blast furnaces or rolling mills if it believes that its profits will be increased as a result. A grocer will stock up if doing so promises greater profits. But the expansion will take place only when the firm believes that total profits will be higher because of it. Incidentally, "total profits" in this sense covers not the profits of a single year only, but the profits to be earned over the whole lifetime of the enterprise. Thus, an expansion of plant which promises high profits for only a few years would not necessarily lead a firm to invest. And it is clear that additional capacity may promise high profits for a short period but reduced earnings afterward. In fact, a great many firms faced just this problem in the years from 1941 to 1944, when they could see that expansion would give them much higher profits for a few years. But in certain industries there was a feeling that when munitions production declined, there would be no profitable way to use these new plants; some fixed costs would still have to be met, while there would be no receipts to offset them. Under these circumstances, the reluctance of firms to finance wartime expansion *with their own funds* is not hard to understand. Private investment, then, takes place only when it looks to be the profitable course of action.

To understand the forces that determine private investment, we must determine the circumstances which make expansion along various lines profitable. More accurately, since the firm can never *know* beforehand whether expansion will actually be profitable, we shall have to observe the conditions that normally would lead businessmen to *anticipate* profits from investment.

Investment, the Interest Rate, and the Marginal Efficiency of Capital

In attempting to decide whether or not an investment project will be profitable, a firm must weigh two considerations: the net yield it

can expect from this project over its lifetime, and the rate of interest. The net yield from the project, or, as Keynes called it, "the marginal efficiency of capital of that type," represents the rate of return on the original cost. If, before subtracting interest charges but after allowing for all other costs, an investment project promised a yearly net profit of $60,000 on a $1,000,000 outlay,[1] we should say that its anticipated net yield was 6 per cent. Anticipated net yield and "marginal efficiency of capital" are simply the technical names for a measure which is actually in common use. Thus, when someone says, "You should be able to earn 8 per cent on this, before paying interest," he is employing the concept without using the term.

If a firm can borrow money at an interest rate below the anticipated net yield on the project, it can expect to profit by undertaking the investment project. To complete the illustration introduced above, if a firm could borrow the million dollars needed for the project at 5 per cent a year, and if the added capacity were expected to yield 6 per cent over its lifetime, the firm would add $10,000 to its profits each year by making such an outlay. On the other hand, if the firm found it necessary to pay 6 per cent to borrow this sum, there would be no net profits, and hence no reason at all for undertaking the project. And if the rate of interest were as high as 7 per cent, the investment could be made only with the clear expectation of loss, which implies that it would not be made.

Thus, the relation between the rate of interest and the marginal efficiency of capital of the type, or the anticipated net yield, is decisive in determining whether a particular investment project will be undertaken. If we list together the anticipated net yields from all possible projects, then we may in summary say that private investment depends upon (*a*) the rate of interest, and (*b*) the marginal efficiency of capital.[2] We may anticipate far enough to say that the higher the marginal efficiency of capital, the greater the amount of private investment at any given rate of interest; and the higher the rate of interest, the smaller the amount of investment at any given marginal efficiency of capital.

[1] The allowance for depreciation looks after the return of the principal.

[2] The marginal efficiency of capital used in this general sense refers to the complete list of expected yields for all projects under consideration.

The Rate of Interest and Private Investment

First, we shall investigate more carefully the effect of the interest rate on investment. We have already seen that a high rate of interest discourages investment because some projects whose anticipated yield is not particularly high are not worth undertaking at high rates. A project with an estimated marginal efficiency of 5 per cent would be carried out if the rate of interest were 4 per cent; but if the rate of interest were 6 per cent, it could not be profitably undertaken. Thus, other things being equal, the higher the rate of interest, the fewer are the investment outlets which business firms can exploit to advantage.

This is obvious when firms have to borrow in order to purchase investment goods. But what about firms that already have funds on hand? After all, a good deal of spending on plant, equipment, and inventory is financed, not by borrowing or by selling securities, but with funds retained by the firm from an earlier period. For example, the United States Steel Corporation between 1921 and 1938 spent on plant and equipment about $1.2 billion. More than 75 per cent of this amount came from depreciation allowances, and another 15 per cent was retained profits. In the same period the General Motors Corporation spent approximately $1 billion on new plant and equipment. A spokesman for the firm said before a committee of the United States Senate that "In the eighteen-year period there has been substantially no outside financing. It is largely financed within ourselves." About half the funds were obtained by General Motors from retained profits and the rest from depreciation allowances. So it will be clear that a firm may finance an investment purchase without having to borrow. The question, then, naturally arises whether the rate of interest helps to determine the amount of investment on such occasions, as it so obviously does when firms must borrow.

And the answer, briefly, is that even in these cases the rate of interest is effective. A firm that has accumulated funds can use them in various ways. It can buy investment goods with them or it can lend them. If the rate of interest is high, compared to the anticipated net yield from the investment goods, the firm will make larger profits by lending than by investing. A firm that seeks to maximize profits will not invest if the marginal efficiency of capital of that type is below the rate of interest.[1] Thus the lower the rate of interest, the more

[1] There is, of course, no single rate of interest; some firms can borrow at low rates, others

numerous are the investment projects which under given circumstances of demand and cost will be worth undertaking. As the rate of interest increases, *other things being equal*, the number of such projects falls, and hence so does the amount of investment. A report of the British Iron and Steel Federation to the British government illustrates the significance of the rate of interest. In discussing the scope for major reconstruction in the plate branch of the iron and steel industry, the report concludes: "The balance in favour of the new plants will be increased by every rise in the cost of fuel and labour *and by every reduction in the rate of interest.*" [1] We may conclude, then, that, other things being equal, a high interest rate discourages investment, and a low interest rate encourages it.

The effect of changes in the interest rate on investment relates to what was said earlier about the banking system, which, as we have seen, plays a most important role in determining the rate of interest. During depression it is sensible economic policy to reduce the rate of interest. The reason for this should be clear; it is intended as a stimulus to investment. But it is not certain that the step is a very effective one. Many economists, reviewing our experience in the nineteen-thirties, are inclined to believe that a change in the rate of interest has only a minor effect upon investment. While they agree that as the interest rate goes up, the amount of investment goes down, they feel that in depression, at any rate, the response to a change in the interest rate is slight. On this matter, opinions have changed considerably in the last fifteen or twenty years. During the nineteen-twenties it was generally believed that proper manipulation of the interest rate could stabilize the economy almost by itself. Present views are quite different, for the interest rate is now looked upon as a relatively ineffective weapon for securing stability.

The Marginal Efficiency of Capital and Investment

Now we must analyze in detail how the marginal efficiency of capital affects investment, and the factors upon which it depends. Given the rate of interest, the higher the yield expected on new capital projects, the larger would be the number of projects undertaken, and hence the greater would be the amount of investment. This conclu-

must pay more. A loan made for a number of years generally costs more in annual interest charges than a loan made for a very short period. The rate of interest which must be set against the anticipated net yield is that which must be paid for a loan of the type needed.

[1] The italics were not in the original.

sion is based upon the consideration advanced in analyzing the relation between the rate of interest and investment. The dependence of investment upon the marginal efficiency of capital is clear. But what does the marginal efficiency itself depend upon?

It must be emphasized that the marginal efficiency of capital is ordinarily not quoted on the market like the price of tomatoes. The plans for an apartment building are not tagged with a figure giving the yield which the market expects it to earn. Rather, the marginal efficiency represents the yield expected by the firm which contemplates carrying out the investment project. Thus, if the project is to add $30,000 worth of merchandise to the stock carried by a retailer, it is he who determines the marginal efficiency of capital of that type. In doing so he must estimate the state of the market for his product and his costs during the expected life of the capital assets. The uncertainty of such an estimate is, of course, obvious. Since it has so little firm support, it follows that the current "mood" of the businessman will be quite important. When he is optimistic, he will reach a much more favorable estimate of the yield than when he is pessimistic, even though the objective circumstances may be identical. Thus, the mood, or temper, or "confidence" of the business community colors the results, and is itself a determinant of the marginal efficiency. There are, of course, more objective factors than this, as we shall shortly see.

The Marginal Efficiency of Capital and the Market

The yield anticipated from any project depends in part upon the demand for the product of the firm, both present and future. If the firm can sell a larger output at a satisfactory price, it has an opportunity for profitable expansion. Let us consider an actual case. In 1922 the retail value of automobile sales was about $1.9 billion. By 1926 it had increased to about $3 billion, and in 1929 it amounted to $3.5 billion. This growth in the size of the market created very favorable opportunities for investment, and General Motors alone spent over $440 million on plant and equipment in this period. Then, between 1929 and 1933, automobile sales at retail declined very sharply, and during these years General Motors invested only $59 million in plant and equipment, or annually one-quarter of the earlier amount. Further interesting evidence for the view that the marginal efficiency of capital depends partly on demand comes from the elec-

tric power industry. Between 1922 and 1930 the total output of electric power more than doubled, indicating a very rapid expansion in demand. Hence it was profitable to expand the facilities for power production, and accordingly the total investment in plant and equipment in this industry came to about $6.5 billion in this period. Then with the depression, output fell sharply and did not again reach the 1930 level until 1935. And during these years investment in plant and equipment came to only $1.2 billion.

Again, essentially the same thing happened in retail trade. Much of the investment in that industry is in the form of inventory accumulation. Retailers may be expected to anticipate a relatively high yield from inventory investment when their sales are expanding rapidly, and a very low yield when sales are increasing slowly or declining. Hence we should expect to find investment in inventories high when retail sales are increasing quickly, and low or even negative when they are falling. Let us see what actually happened. Between January, 1940, and January, 1942, retail sales increased from an annual rate of $38.1 billion to one of $52.1 billion. In the same period, retailers' inventories rose from $5 billion to $6.6 billion; in other words, investment was positive and high, $1.6 billion. In contrast for December, 1936, and again for December, 1938, retail sales stood at $4.1 billion. Instead of expanding inventories quickly between 1936 and 1938, as they later did between 1940 and 1942, retailers increased them by only $170 million, or by one-tenth as much as in the later period. What this appears to show is that investment is high when demand is climbing rapidly, and low when demand is increasing slowly or not at all. And though we must not equate investment with the marginal efficiency of capital, since the rate of interest also exerts its influence, that influence can be discounted for the period from 1936 to 1942, since interest rates did not change appreciably.

It is now clear that the marginal efficiency of capital depends on the state of the market and that investment is high when demand is growing. Nevertheless, we should not expect a *precise* relation between changes in *current* demand and investment. For one thing, other factors, such as changes in the interest rate, may influence investment. For another, the size of current demand, and the rate at which it is changing, are not critical in determining the expected yield on new projects, since such projects are intended to exploit not the present

market but a future one. A large demand for steel today does not guarantee a large demand over the next two decades. Instead of current demand, it is the demand over the life of the plant that determines whether its construction will be profitable. Only to the extent that expectations about future demand are based upon the current market should we expect investment to be closely related to the present level of sales and the rate of change in that level. Yet, since an estimate of the demand at various future dates is bound to be considerably influenced by current experience, we should expect to find, as apparently we do, some connection, though not a precise one.

In sum, then, a large and especially a rapidly growing market is favorable to the marginal efficiency of capital and hence to investment. And a market which is growing very slowly or declining is, in the same way, unfavorable to investment. This conclusion, incidentally, helps to explain why prosperity often goes so high and depression so low. During prosperity, consumers have high and rising incomes, and consequently the demand for many kinds of goods and services is great and increasing. We have seen that a growing market favors private investment. And when private investment is increasing, the national income is also growing. Thus when there is a revival, the market for all kinds of goods is growing, and this in turn helps to maintain a large volume of private investment. Revival, in other words, because it means that the demand for goods is growing, tends in this way to feed on itself. For much the same reasons, depression also tends to perpetuate itself. Demand for almost all products is low and falling during depression, and for this reason investment prospects are likely to be unfavorable.

Population Growth and Investment

A rapidly growing population is commonly supposed to favor private investment, and a slowing down in its rate of growth is believed to do just the opposite. This factor operates through its effect upon market prospects, and so should be considered at this stage. When population is growing rapidly, it is regarded as offering some promise of growing markets, and this expectation leads to favorable forecasts of yield on investment projects. When population is growing slowly or not at all, there is less reason to suppose that the markets for the products of business firms will grow rapidly. In such general terms, the argument may seem unreal and unconvincing. But if we con-

sider the rate of growth of the population of a metropolitan area, and investment in that area in housing and retail stores, we can see its pertinence at once. If the population of a city and its suburban areas increases quickly, and if that increase is believed to be permanent, real estate interests will be quite likely to sponsor housing developments, to build apartment buildings, and to erect retail outlets, whereas if population in the area is not changing they will be very reluctant to do these things. An increase in population persuades manufacturers, builders, and merchants that the demand for their products will be considerably higher a decade from now than it is today. Consequently, they will easily be persuaded to see opportunities for profitable expansion. Real estate groups may buy up the outlying portions of a community and build a number of houses. Railway lines may be extended into these rapidly developing areas. Manufacturers of consumers' goods would hesitate less about buying new machinery and putting up branch plants if they were confident that the market would continue to grow. Seen in this way, there seems to be an obvious connection between private investment and the rate of increase in population. But a rapidly growing population is not a *necessary* condition for high investment, though it is certainly a favorable factor.

Certain other population effects may also encourage investment. For instance, as population growth slackens, the population becomes somewhat older in composition, and the size of the average family tends to decline. Changes like these may lead to a much higher demand for some kinds of consumers' goods and for housing. Older people may spend more freely because they have already accumulated adequate savings. The rapid growth in the number of families, which, for a while at least, is not ruled out by the reduced rate of population growth, may create a very lively demand for housing. These effects may even offset the adverse and direct effect of the slowing down in the rate of increase of the population, an effect which ordinarily makes market forecasts somewhat less favorable. Thus certain results of a decline in the rate of population growth may actually encourage investment, though these stimuli appear to be much weaker than the direct and obvious stimulus that comes into being when population growth is rapid.

The Stock of Capital Goods and Marginal Efficiency

Yet another determinant of the marginal efficiency of capital is the amount of capital goods already on hand. For, other things being equal, the greater the stock of capital goods, the lower the marginal efficiency of capital will be. When enough plant and equipment to meet current and expected future demand already exist, the anticipated yield from new projects could only be very low. This conclusion has already been implied in our discussion of demand and investment. For obviously the critical factor in determining the expected profit of an investment project is not the demand for the product considered by itself, but rather the demand in relation to the capacity of existing equipment to meet it. The greater the existing capacity, or, in other words, the greater the existing stock of capital goods, the fewer the new investment projects that could be expected to yield a high rate of return when demand is at any particular level. Thus, just as high demand would be expected to encourage investment in an industry, so a large stock of capital goods in the industry would be expected to dampen private investment. Evidence already presented for the electric power industry supports this conclusion. As we have seen, the consumption of electric power was about as high in 1935 as it was in 1929, but capacity in 1935 was about 20 per cent above the 1929 level. We should therefore expect to find investment in this industry much lower in the later year than in 1929, and this is precisely what the data show. In 1929 almost $800 million was spent on new plant and equipment; in 1935 the figure was below $200 million.

In addition to plant and equipment, the stock of capital goods includes inventories on hand, and the conclusions which we have developed appear to be equally valid when applied to such goods. Thus, we should expect that when inventories are high in relation to sales, further investment in inventories would be low, and vice versa. And there is evidence to support this in available data on total trade. Wholesale and retail sales were almost exactly the same in 1928 and in 1936. In the earlier year they amounted to $41.8 billion, and in the later year to $41.6 billion. Trade inventories on hand in December, 1928, were valued at $10.8 billion, and in December, 1936, at $8.9 billion. We should therefore expect that the lower level of inventories in relation to sales would prove more favorable to inventory investment in 1937 than in 1929. This is what we find: in the year

1929, inventory investment by firms engaged in trade was almost nothing, whereas in 1937, inventory investment by trading firms came to $300 million. It must be pointed out that such statistical demonstrations are not to be regarded as proofs of the hypotheses set out above, but rather as illustrations. It is possible, indeed, to find instances where the data apparently disprove the hypotheses. The difficulty in analyzing the statistics and determining their significance is very great, because the number of variables that determine private investment is very large. Hence to reduce so complex a problem to a simple one by selecting a few instances that illustrate a conclusion should not be confused with proving that conclusion. But even so, there is abundant evidence that, other things being equal, the greater the stock of capital goods, the lower is the level of private investment.

The Stock of Capital Goods and the Business Cycle

This relation between the existing stock of capital goods and the amount of private investment is extremely important in accounting for prosperity and depression. In a country as advanced economically as our own, capital goods accumulate during prosperity at an enormous rate. Even after providing for all necessary replacements and repairs, it is possible for us to produce over $50 billion worth of capital goods for private firms a year, although we have never achieved anything like that rate.[1] Unfortunately, it is difficult to get a very clear idea of how such a level of investment would compare with the total amount of capital equipment available in the United States, for there is no accurate figure that measures the total stock of capital goods.[2] We do know, however, that growth at the rate of even $20 billion a year is very rapid, for at that rate it would take only ten or fifteen years to *double* the amount of capital equipment in the country. There is no doubt that if private investment were made at the rate of $20 to $40 billion a year, we would be adding to privately owned factories and equipment very rapidly indeed.

Some facts about how quickly it is possible to expand capacity in specific industries will demonstrate this point. In the electric power industry, where capacity was more than doubled between 1921 and 1929, it took only $5.8 billion worth of investment to attain this result.

[1] Government investment would have to be quite low if we were to do so.

[2] Using data supplied by the National Industrial Conference Board, I have estimated the value of capital goods in this country (including housing) in 1937 at about $260 billion.

Capacity in the steel industry was increased from about 82 million tons a year in 1938 to 96 million in 1945 — a 17 per cent expansion with an investment expenditure of only $2.5 billion. The aluminum industry could supply about 450 million pounds of aluminum a year in 1939, and by 1944 its annual capacity had increased to about 2.25 billion pounds. Thus, capacity in this important industry increased about five times over, and yet the amount of investment needed was less than $1 billion. There is no doubt that the stock of capital goods grows quickly when private investment is high.

Therefore, after a relatively few years of high private investment, the existing stock of factories, machinery, and other kinds of capital goods will have been appreciably increased. For as we have seen, if we are to enjoy prosperity, investment must be very high. Therefore, to the extent that prosperity is based on private investment, the stock of privately owned capital goods grows very rapidly. Just here is where the trouble starts. The rapid growth of this stock sets up increasingly serious obstacles to a *continuation of the expansion.* It may seem profitable to business firms to invest $20 billion a year for two or three or even four years, but to continue investing at this rate becomes less and less attractive, simply because each year finds more machines and more factories on hand and less need, therefore, for additional new ones. Thus, prosperity, unless other forces come to the rescue, tends to create the very conditions that bring it to an end, though luckily, of course, other forces frequently do intervene.

But it is not only prosperity that tends to cut its own throat; depression too creates the conditions that ultimately bring it to an end. In a period of depression, investment is very low. In fact, when the depression is particularly deep, as it was from 1931 to 1933, investment is so low that even the depreciation and wearing-out of plant and machinery is not made good. When this occurs, there is a gradual decline in the amount of effective capital equipment. Factories and machines deteriorate, and after several years the amount of efficient capital equipment on hand is less than it was before the depression began. For instance, in 1921 there were about 2,008,000 serviceable freight cars in the hands of railroad companies. Expenditures on railroad equipment between that date and 1931 were about $3.8 billion, and even though the greater part of this amount was for freight cars, the railroads had only 2,053,000 serviceable freight cars by 1931. Then in the next seven years, the annual rate of expendi-

ture was reduced to about 30 per cent of the 1921–31 level, with the result that by 1938 the number of serviceable cars had fallen to only 1,406,000, a substantial decline. When private investment is very low, the stock of privately owned capital goods actually declines.

Now with such a reduction in the stock of capital goods on hand, profitable investment opportunities become more numerous, other things being equal, and investment is thereby stimulated so that the revival begins. This does not mean that it is necessarily advisable to *wait* for the revival. The wait may be long and disastrous. The fire that consumes a house dies out eventually, after everything combustible has been burned, but that is no reason for not calling the fire department as quickly as possible. In economic affairs most of us are strongly predisposed toward "natural" remedies for such difficulties as depression. We do not like the idea of "tinkering," "meddling," or "interfering," with what we call "the operation of natural economic laws." But this preference for a "natural" cure may force us to suffer through a long and deep depression before our plant and equipment wear out and reduce our productive capacity enough to stimulate private investment, so that we may revel in a natural recovery. Luckily, we do not treat pneumonia that way. And surely if we do insist upon overcoming depression by "natural" means, we may be permitted to accelerate the process by speeding up the depreciation with a few suitably planted bombs. In the absence of other factors, prosperity would eventually give way to depression, and this in its turn would, after a long wait, turn into revival. This tendency toward cycles is, of course, present in the economy, but that does not mean that we can do nothing to iron out this periodicity.

Technological Development and the Marginal Efficiency

Still other factors must be considered in our analysis of private investment. Inventions and technological developments tend to raise the marginal efficiency of capital and thus to encourage private investment. When the continuous-strip rolling mill was adopted by the steel industry, when new methods of spot welding were worked out, when the radio and automobile were developed, the marginal efficiency of capital in certain industries was temporarily increased, and private investment was thus for a time stimulated. It is convenient in considering these effects to treat improvements in existing methods of production separately from the development of new products.

The development of improved methods of production stimulates investment in an industry, but only for a limited time. Suppose that the equipment in a certain industry has been designed to exploit a particular process, and that firms in the industry are earning a normal profit and producing enough to meet the demand at a reasonable price. In this stable condition investment would be relatively small — just sufficient to cover maintenance and repairs, with perhaps a small addition to bring about a slow growth in capacity.[1] If there should then be an improvement in the technique of production which enabled a firm to produce at a reduced cost, we should expect investment in that industry to rise to a rather high level for a time. Since any firm that could secure the more efficient equipment could undersell its rivals, every firm would be under pressure to install the improved equipment, and many of the firms in the industry would scrap their old equipment and order the new. This, of course, would mean a high level of investment in that industry, but only for a time. Eventually the capacity of the industry would be built up to the level suited to the size of the market, and when this happened, investment would decline again to a figure dictated by the rate of depreciation, the rate of growth in demand, and so on.

There are circumstances, however, when inventions, even cost-reducing improvements in technique, do not stimulate investment. Firms have been known to buy up a new process in order to withhold it from other firms, even though they themselves do not intend to exploit it. This is most likely to happen when competition is very imperfect, or when there are monopolistic agreements between firms ostensibly in competition. It is much less likely to occur when a large number of firms are in actual competition. Thus in general the effect of new developments in existing methods will tend to stimulate investment until the capacity of the industry is once more adjusted to the size of the market.

The development of new products is also likely to encourage investment. In a sense a new product creates its own demand, and the first few producers in the field, if the commodity is an important one, are likely to do very well. But in order to exploit the new product, firms will have to put up plant, install equipment, and build up inventories: that is, they will have to make the necessary investment.

[1] Strictly speaking, the investment that builds up capacity would be ruled out by our assumption of normal profits.

There have been some spectacular instances of the way in which the development of new products has stimulated investment. In the nineteenth century the most striking example of this was the growth of the railroads. The invention of the steam railroad not only encouraged investment in railway lines, locomotives, cars, and stations; it also made it profitable to expand the capacity of the iron and steel industry and to develop coal mining — to mention two industries which were enormously stimulated by railroad construction and operation. Finally, because rapid settlement of the West became possible once adequate transportation could be provided, the development of the railroad was also indirectly responsible for a heavy investment in housing and in the new industries associated with the growth of the West.

In the twentieth century the development of the automobile gives us perhaps the most striking example of this process. Here again we have far more to consider than the investment in the new plants where automobiles are assembled. Hundreds of factories were erected in which automobile parts were fabricated. The rubber tire industry had to be built up. The glass industry was stimulated. Petroleum refineries were constructed, and the iron and steel industry was given an enormous boost. But the stimulus to investment went even further than that. It was necessary to build extensive highways and to pave city streets, and consequently all the industries that had a part in road building were encouraged to expand. Suburban living became feasible, and this meant a considerable boom in housing, public building and the construction of retail stores. Other industries have recently experienced similar growth, though the effects have been less important than those associated with the growth of the automobile industry. The development of the electric equipment industry has provided another important source of investment demand, and the construction of factories for the manufacture of radios, vacuum cleaners, electric light bulbs, and so on, has been extensive. As a result of this expansion, the mining and refining of copper has been greatly increased; and other metal industries have also been encouraged to expand. The consequent development of the electric power industry has required the erection of dams, power plants, transmission lines, and transformers, all of which have required heavy investment.

Such important inventions as the railroad, the automobile, and electrical equipment have not been frequent. Each of them in turn

became a stimulus to investment which lasted over many years. It took a long time to build up all the industries that in any way depended upon the railway, the automobile, and the use of electricity. By now, however, these important industries have been developed to a point where they can comfortably meet normal demand. Hence we have to rely upon still newer products and processes if investment based upon invention is to be as high as it was in earlier decades. Whatever the future may hold, it is clear that the relation between technology and investment must be kept in mind.

Determination of Private Investment: Summary

It is time now to summarize what has been said about private investment up to this point. Firms invest when they expect to earn a profit by doing so. Or in economic terms, if the marginal efficiency of an investment project exceeds the rate of interest which the firms must pay on borrowed funds, the project can be expected to yield a profit. Therefore, other things being equal, the lower the rate of interest, the larger the number of projects which would be undertaken; and likewise, the higher the marginal efficiency of capital, the higher the investment would be. The marginal efficiency of an investment project is partly based upon expectations, and is thus subject to sharp changes. While an exhaustive list of the factors affecting the marginal efficiency is out of the question, it is possible to indicate some of the important ones. Among these the size and the rate of growth of the market is obviously important, since the more rapidly the demand for the product grows, the higher will be the marginal efficiency of an investment in the industry which produces it, provided, of course, that a later fall in demand is not expected. The level of capacity is an equally important determinant. For, assuming that the demand for its product is given, the greater the existing stock of plant, equipment, and inventories in an industry, the lower is the marginal efficiency of new investment projects, and vice versa. Technological developments, whether in the form of new methods or of new products, temporarily raise the marginal efficiency of capital. Wars, changes in tax rates, labor legislation, court decisions, the stock market, and numerous other factors also help to determine the marginal efficiency of capital and thus the level of private investment. How investment varies from period to period thus depends on a series of tugs, some having a depressant effect, like the accumulation of capital goods,

which goes on constantly except through very deep depression; and others having a positive effect, like a rapid growth in population or a series of important inventions.

Public Investment[1]

In the last few decades another type of investment has been increasingly important in amount. Between 1921 and 1942 public, or government, expenditures made up slightly more than half of the total investment done in the United States. This will not be surprising to those who remember the federal relief and public works expenditures during the depression of the nineteen-thirties. What should be more surprising is the great importance of government investment during the nineteen-twenties. That decade, it will be recalled, is commonly identified with "normal prosperity"; that is to say with prosperity based upon the investment activities of private business firms. And yet even in those years public investment accounted for about 40 per cent of the total investment. And during the war years, 1942 to 1945, public investment grew enormously, both in absolute and in relative importance.

In general, public investment has been a relatively unstable series, although before 1942 it was much less so than private investment. During the nineteen-twenties, public investment averaged about $10 billion a year; it fell during the years from 1930 to 1934, but then began to climb until by 1939 it stood at $16 billion. In the years from 1941 to 1945, the figures were very much higher, reaching about $100 billion in 1944.

As was made clear in the previous chapter, public investment covers the purchase of all goods and services by government units — federal, state, and local. It includes the carrying-on of such normal, peacetime activities as the construction of highways and the paving and repairing of streets; the building and maintenance of public hospitals, schools, and postoffices; the erection of public-owned dams for flood control and electric power; the organization of the police force, of the schools system, including, of course, the hiring of teachers; and the employing of other kinds of government employees. All such

[1] The definition of public investment employed here is different from that used in some other treatments of the problem. A decision had to be made on whether certain government expenditures should be treated as consumers' expenditures or investment expenditures. In the interest of simplicity, it was decided to treat them all (except for transfer expenditures) as purchases of investment goods (and services).

expenditures rank as government investment. We should also have to include special wartime expenses. These would include expenditures for aircraft, naval vessels, uniforms, and ammunition. The value of goods lend-leased to other governments, and expenditures by the government on new facilities such as steel mills, aluminum plants, aircraft assembly plants, and so on, must also be counted in this category.

What can an economist say about the determinants of this kind of investment? Unfortunately, not much. We do know that the decision to build a new schoolhouse, or to extend a highway, or to hire more postmen, or to build a battleship, is not made because it is *profitable*, for after all government agencies are not business concerns. These local, state, and federal government bodies are supposed instead to base their decisions upon the *needs* and *resources* of the community. But these are vague words which should be further analyzed, not by the economist, but by the political scientist. There are, however, a few very general comments that are worth making. During wars, or periods of preparation for war, government investment will always be very high. During depression, too, government or public investment is likely to be quite high. This latter tendency has recently undergone an interesting development. In the past, public investment during depressions was chiefly for relief payments. More recently in this country there has been growing support for a policy of expanding government investment when private investment is low; in other words, for making government investment counter-cyclical. Some state governments attempt to follow such a policy, and the federal government did precisely this during the depression of the nineteen-thirties. Other countries — Sweden, for instance — follow such a procedure quite closely in determining the level of government investment.

Another interesting kind of government investment in this country, although a relatively unimportant one, is the purchase of domestically mined gold. The amounts thus spent are very small — about $100 million a year compared with total government investment which ranges from $10 to $100 billion. However, this investment is especially interesting because the government finances the purchase neither from tax receipts nor by borrowing, but simply by printing money. As we have already seen in Chapter 25, when the government purchases the product of a domestic gold mine, gold certificates

are issued by the Treasury against the gold it receives. These gold certificates are given to the Federal Reserve Banks in payment for the gold, and the Federal Reserve Bank may pay the mine by giving it either Federal Reserve Notes or a deposit. When the government purchases anything else besides gold — whether it is a school teacher's labor or a mile of highway — it must either tax or borrow in order to make payment.

As we have seen, then, some public investment is necessitated by war and some is made deliberately as a remedy for depression. For the rest, it is much more difficult to say. Why was this schoolhouse built? What about that public park? Why were a thousand more people given jobs with the Post-Office Department? Perhaps the simplest answer in a democracy is that it happened because we wanted it to happen. Only a political scientist is capable of going behind this statement.

Foreign Investment

The sums spent by foreigners upon goods and services produced in this country constitute yet a third type of investment. In recent history, however, this type has been unimportant compared with private and public investment expenditures. For the whole period from 1921 to 1941, foreign investment accounted for about 3 per cent of the total. A detailed analysis of the factors that determine foreign investment must be postponed until we have discussed foreign trade. We can, however, describe in general terms how foreign investment is determined. Foreigners buy our goods when they want them and can pay for them. In this respect they are no different from Americans, for they will want our goods in preference to the products of other countries, including their own, if they can get them more cheaply. Whether they can do this or not depends upon the prices of our goods in our own currency, on the cost of transportation, and on the price which must be paid for our currency. The prices of our goods and currency, then, are one factor, but there are others. Like us, foreigners want more goods when their incomes are high, for when they are prosperous they consume more freely. Some of what they consume they will prefer to buy from us, some from other countries, and some from their domestic sources of supply. Obviously, then, our foreign investment tends to be high when foreign countries are prosperous, and low when they are not.

But, unfortunately, foreigners with wants are not enough to constitute foreign markets. If they were, we should always have full employment in this country, for in the simplest sense of the term the rest of the world wants an almost limitless supply of what the United States produces. But a Neapolitan housewife whose rooms need cleaning is not necessarily a customer for an Electrolux vacuum cleaner. She only becomes one if she is able to get American dollars for her lire. And since we do not empower Italian banks or the Italian government to print American dollars, she can get this currency only from us. We would not give it up unless we were willing to take lire for it, or possibly to lend dollars against a promise to repay later. And we would accept lire only if we could find Americans who wanted them in order to purchase something produced in Italy. Thus, when we are willing to import, we create one of the conditions in which foreign investment is high — we provide American dollars to foreigners. This point, and others concerning the availability of our currency to foreigners, will be considered in Chapters 40–42. At present we must be content with the general considerations mentioned above.

To sum up, then, foreign investment depends on how attractive our goods are to foreigners and on the supplies of American currency they can secure. If we could only export one of the printing presses used for the manufacture of Federal Reserve Notes to, let us say, China, our foreign investment would be enormously higher. Failing that, as we shall see later, an upper limit is set for it by the amounts of foreign goods, gold, and securities which we are willing to accept. Within this limit the actual level of foreign investment will be established by the appeal of our goods to foreign markets in the light of their price and shipping costs, and the level of income and tariffs in foreign countries.

The reader cannot escape a feeling that this analysis of the determinants of investment is enormously complicated. And so it is. But after all, the economy which the analysis describes is itself extremely complicated. In the real world very many factors affect investment, and an analysis simple enough to write on the corner of a handkerchief would be far too simple to give a useful account of the real world.

29

The National Income and Changes in Investment

NOW THAT WE HAVE HAD A GLIMPSE of the factors which determine the level of private, public, and foreign investment, we are in a position to return to the broader picture and to view investment in relation to the national income. As we saw in Chapter 27, investment and the propensity to consume are the two factors which determine the national income. At that stage, our demonstration of the fundamental relation was necessarily mathematical or logical. We may now achieve a clearer and more realistic view of the way in which the national income is fixed by tracing in detail the effects of a change in these determinants. In this chapter we shall analyze the effects upon income and employment of a change in investment when the propensity to consume is constant. Then in Chapter 30 we shall examine the determinants of the propensity to consume, and in Chapter 31 we shall consider the effects of a change in this factor when investment is constant.

The Initial Situation

We shall begin our analysis of what a change in investment will do to the national income by describing the situation as it might exist before investment is altered. We assume that the propensity to consume is such that when the national income is $100 billion, consumers spend $75 billion, and that initially investment is running at the rate of $25 billion a year. Given this propensity to consume and this level of investment, it is clear that the national income is $100 billion a year

— the sum of consumers' and investment spending. To summarize, we have originally:

Investment spending:	$25 billion
Consumers' spending:	75 billion
The national income:	$100 billion

Under these circumstances, saving will also be $25 billion a year, since it is defined as the difference between national income and consumption; that is, 100 minus 75. Now, so long as investment and the propensity to consume remain constant, the national income must remain fixed, for without a change in either of these two determinants, total spending will remain the same. On the other hand, if either of the determinants is altered, the national income will also vary.

Increased Investment and the National Income

But the national income would obviously not remain at $100 billion a year if investment rose above $25 billion. With the higher level of investment spending, total spending would increase and income would rise. The problem which concerns us in this chapter is to determine the size of the increase in income for any change in investment.

Let us assume that investment is raised from a rate of $25 billion a year to $26 billion. This means that purchasers of investment goods — private firms, government bodies, or foreigners — have increased their orders. In response to this increase, the firms that produce such goods will expand their output. More will be produced in machine-tool plants, building and highway construction will pick up, and there will be an increase in the output of shirts, socks, canned foods, and other items which are to be added to retailers' and wholesalers' inventories, and which therefore are also to be regarded as investment goods. In short, in at least some of the industries which produce investment goods, there will be increased activity. In order to visualize this more clearly, suppose that most of the increased investment is in the construction of houses. In order to produce houses more rapidly, it will be necessary not only for builders themselves to put up more houses, but also for the manufacturers of plumbing supplies, lumber, bricks, cement, nails, glass, roofing material, and so on, to expand their output. All this will mean, of course, that there will be an increased level of employment. More bricklayers will have to be hired, as well as more carpenters, plumbers, electricians, general laborers, lumbermen, and so on. For, generally, in order to produce

a larger output, firms will have to take on more men. Where will these men come from? Since according to our assumption, total investment is being raised, we shall have to assume that firms in the investment goods industries which have not felt an increased demand will continue to produce at the original level. Hence employment in these firms does not decline. Therefore the newly hired men in the expanding industries can only come from the ranks of the unemployed. Or if they were formerly employed elsewhere, their replacements in the firms they have left must have been unemployed before. In short, total employment will increase, and unemployment will decline.

Now when purchases of investment goods increase from $25 billion to $26 billion, there will be a $1 billion increase in the incomes earned in these industries. This is in accordance with the identity established in Chapter 21 between the value of goods produced and the income earned in producing them. Who will share this increased income? The firms which produce the investment goods now in greater demand will take on more men. They may also increase work hours, and thus be compelled to pay out more in wages. In addition, these firms will earn higher profits, since profits rise when demand increases. Most of the increase in income will go to workers formerly unemployed or now working longer hours, and to the owners of the firms involved. Thus, we could summarize to this stage as follows: the $1 billion increase in investment will encourage firms to expand employment. Hence, there will be a $1 billion *primary* increase in incomes earned by those connected with industries producing investment goods. This increase will be divided between labor and the owners of the firms. Men formerly on relief will now receive a wage income; men who had been receiving pay for perhaps thirty-five hours a week may now be paid for forty hours; and the higher profits earned by the firms will go to their owners. If this higher level of investment is maintained for a year, the sum of all these increments of income will be $1 billion.

Increased Income and Increased Consumption

But this is by no means the end of the cycle, for, as we have seen, when people receive higher incomes they spend more on consumers' goods. The amount of the increase in spending will, of course, depend on the circumstances in which the higher income is earned. Men who have been out of work and have been living on relief or on savings will spend a good deal more than formerly as soon as they begin to receive regular wages. Those who have been receiving good profits all along

will spend somewhat more on consumers' goods, but the increase in their spending will be much less than for those who have recently been unemployed. Hence there is no doubt that when incomes grow, spending on consumers' goods will also rise to some extent. But as we saw when discussing the propensity to consume, it is unlikely that spending on consumers' goods will go up by as much as the increase in income; for certainly those in the higher income brackets will not spend the whole of their increase on consumers' goods. And even the wage earner who has just left the ranks of the unemployed will probably not increase his consumption by the whole amount of his wage.

The following statistical data, compiled by the National Resources Committee from their study of expenditures on consumers' goods for the year from July, 1935, to June, 1936, have a bearing on this matter. It was found that about 4 million families had a total income for the year of less than $500, and the average income of families in this lowest income bracket was $312. The average expenditure on consumers' goods for these families came to $466. There were nearly 4 million families with incomes between $1000 and $1250 a year. The average income in this group was $1120, and they spent $1127 on consumers' goods. Let us suppose that a family which had to live on unemployment relief received an income of only $312, and that when the wage earner in the family got a job, he cleared something like $1120 a year. If we may in addition suppose that a family which enjoyed such an increase in income would spend on consumers' goods about as much as most families in the same income group, we can derive the following result: since the difference in income is $808 (from $312 to $1120 a year) and the increase in consumption is $661, we may reasonably conclude that increased income in the lower income brackets will bring an increase in consumption smaller than the increase in income. For those in the higher income brackets, the result is similar, except that consumption responds much less strongly, as Table 43 shows:

TABLE 43

Changes in Income and Consumption

Income Bracket	Average Income	Average Consumption
$10,000–$15,000	$11,353	$6097
$15,000–$20,000	$17,331	$9134

Thus, corresponding to a $6000 increase in family income (from $11,353 to $17,331), we find an increase in consumption of only $3000 (from $6097 to $9134).[1] We may, therefore, conclude that in general when people receive more income, they spend more upon consumers' goods; but that the increase in their consumption is less than the increase in their income.

The Marginal Propensity to Consume

We must now examine in greater detail the way in which consumers respond to changes in their income. The ratio of the change in consumption to the change in income is known as the *marginal propensity to consume*. Let us illustrate the method by which it is computed. If consumers' expenditures increase by $600 million when the national income rises by $1 billion, the marginal propensity to consume is

$$\frac{\$600,000,000}{\$1,000,000,000}$$

or 0.6, or 60 per cent. If consumers' expenditures fall by $1.4 billion when the national income falls by $2 billion, the marginal propensity to consume is

$$\frac{-\$1,400,000,000}{-\$2,000,000,000}$$

or 0.7, or 70 per cent. Thus, the marginal propensity to consume summarizes in simple terms the way consumers react to changes in their income. When the marginal propensity to consume is high, a change in income brings about a change in consumption which is almost as great. When it is low, a change in income brings about a much smaller change in consumers' spending.

It follows that if we know the value of the marginal propensity to consume, we are able to determine the change in consumption for any change in income. If, for example, the marginal propensity to consume is 50 per cent, a $10 billion increase in income would bring a $5 billion ($=10 \times 0.5$) increase in consumption. If the marginal propensity to consume is 66⅔ per cent, a $6 billion decrease in income would bring a $4 billion ($=6 \times \frac{2}{3}$) decrease in consumption. Finally, if the marginal propensity to consume is 65 per cent, a $1 billion

[1] While differences in spending habits are perhaps most important in accounting for this, the personal income tax which takes a good deal more from the higher than from the lower income is also important.

increase in income will cause a $650 million increase in consumption. The change in consumption is equal to the change in income times the marginal propensity to consume. Or, we may express this relation in symbols:

Let R = the marginal propensity to consume,
ΔY = the change in income, and
ΔC = the change in consumption.

Then by definition

$$R = \frac{\Delta C}{\Delta Y},$$

and hence,

$$\Delta C = \Delta Y \times R.$$

Or the change in consumption equals the change in income multiplied by the marginal propensity to consume.

The Marginal Propensity to Consume in the United States

Fortunately, statistics give some evidence on the reaction of consumers' spending to income changes in the United States. Generalizing from these data, we find that for every dollar of additional income distributed in the normal way, consumption will increase by about sixty-five cents. The results are not regular; so many other forces may influence consumption that we should hardly expect them to be. But they are quite good. For instance, between 1930 and 1931, the gross national product fell by $16.1 billion, and in the same period consumers' expenditures fell by $10.7 billion. The marginal propensity to consume, measuring it as $\frac{10.7}{16.1}$, was thus 66.5 per cent. Between 1934 and 1935, the gross national product rose by $7 billion, while consumers' spending rose by $4.5 billion. The ratio of the increase in consumption to the increase in income, $\frac{4.5}{7.0}$, is thus 64.3 per cent. However, in some other periods, the ratio differs from 65 per cent by a larger amount. Thus, between 1928 and 1929, the gross national product rose by $3.5 billion, but consumption increased by only $0.3 billion. The obvious explanation in this instance is not that the marginal propensity to consume was only 9 per cent, but that other factors — for example, the stock-market crash — forced consumption and the propensity to consume to a level much lower than normal.

We may, then, conclude from the experiences of the American

economy between 1921 and 1941 that the marginal propensity to consume was roughly 65 per cent.[1] This does not mean that we should expect the same value for other periods or other economies. Nor does it mean that the marginal propensity to consume would be the same in depression and in prosperity. Indeed, when we come to analyze the factors upon which the marginal propensity to consume depends, we shall find that it is liable to vary; that, for example, it is probably lower in England than in the United States, and that it may be lower in prosperity than in depression. Nonetheless, we can use the figure of 65 per cent as a bench-mark and a point of departure.

The Increase in Investment, Income, and Consumption

The digression on the marginal propensity to consume was required because we needed to know how to allow for the response of consumers to an increase in their incomes. Since, as we have seen, a $1 billion increase in investment yields a $1 billion increase in income, those whose incomes rise when investment increases spend more on consumers' goods. And in order to tell how much consumption will increase, we need to know the marginal propensity to consume. We then have these links in the chain of events: (*a*) an initial $1 billion increase in investment; which brings (*b*) a $1 billion increase in income; which, in accordance with the assumed 65 per cent marginal propensity to consume, causes (*c*) a $650 million increase in consumption.

But the chain does not stop at this point. The last link, the increased spending upon consumers' goods, exerts its influence, too. With increased purchases of food, clothing, services, flowers, and so on, incomes earned in producing these goods will also increase. Grocers, butchers, tailors, bootblacks, florists, and taxi-drivers will all begin to earn more. The number of employees in textile mills, clothing factories, food processing plants, shoe factories, and other firms that produce goods for consumers will increase. So the payrolls of such firms will go up, and at the same time profits will be higher. In accordance with the identity referred to above between spending and the income earned in production, when consumers spend $650 million more, the incomes earned in producing the additional goods increase by $650 million. This increase must be added to the $1 billion in-

[1] The *average* value was about 65 per cent (computed in this crude way), but there were, as noted, instances when the value found was quite different.

crease which was the primary result of the rise in investment. If we could neglect any further consequence of the $1 billion increase in investment, we would have a total increase in the national income of $1.65 billion; $1 billion earned by those who are connected with firms that produce investment goods, and $650 million for those associated as employers or employees with firms that produce consumers' goods. Thus we can summarize:

Increase in Investment:	Increase in Income:	Increase in Consumption:
$1 billion	$1 billion + $650 million	$650 million

But even that is not all. For this $650 million increase in the incomes earned in the consumers' goods industries will give rise to a still further increase in consumption, one which we may suppose will be 65 per cent of $650 million, or $422.5 million. For as we have seen, men who have been unemployed and are now given jobs will certainly buy more food and clothing — and this is true whether they work in the newly stimulated consumers' goods industries or in those producing investment goods, where the original rise took place. This constitutes one more link in the chain of events. Hence, up to this stage in our analysis, the increase in consumption is $650 million + $422.5 million. This last increase in expenditures will raise incomes in consumers' goods industries by precisely the same amount, $422.5 million, and will cause an additional increase in employment. The total increase in income that results from the $1 billion rise in investment now appears to be $2072.5 million; that is, the sum of $1 billion plus $650 million plus $422.5 million. But the chain of events does not stop even there. The newest addition to income will cause a further expansion in consumption expenditures — at this stage, of $274.625 million, if our rule still holds that for every dollar of increased income, an additional 65 cents will be spent upon consumers' goods. Because of this additional spending, we must take account of yet a further rise in income in the consumers' goods industries, this time of $274.625 million. Therefore the total increase in income now stands at $2,347,125,000, or the sum of $1 billion + $650 million + $422.5 million + $274.625 million.

It appears that this progression has no limit. Every increase in income brings an increase in consumption, which itself constitutes a

further increase in income leading to a still further increase in consumption, and so on, for what looks like infinity. Actually the series does effectively come to an end, since each additional term is only 65 per cent of the preceding term. The total increase in income, if we had patience enough to work it out to the final nickel, would be $2,857,142,857.15. And all this is brought about by a rise in investment of only $1 billion. It is worth emphasizing that in this example the greater part of the increase in income is earned, not by those who produced investment goods, though they are the first to benefit, but by the producers of consumers' goods. This is because, in accordance with the value assumed — 65 per cent — for the marginal propensity to consume, more activity is generated in the production of food, clothing, radios, automobiles, and the other consumers' goods upon which almost two-thirds of the increase in income is spent, than in the building trades and related investment goods industries. There can be no doubt that the effect of the original increase in investment grows something like a snowball. The effect is multiplied throughout the economy.

We can summarize the process as follows:

TABLE 44

Effect of Increased Investment on Income and Consumption

The Increase in: (in millions of dollars)				
Investment	Income (Marginal Propensity to Consume, 65 per cent)			Consumption
$1000	generates a	$1000	which leads to	$650
	which generates	650	which leads to	422.5
	" "	422.5	" " "	274.625
	" "	274.625	" " "	205.96875
	" "	205.96875	" " "	134.08566
	" "	134.08566	" " "	87.15568
	" "	87.15568	" " "	65.36676
	" "	65.36676	" " "	42.68390
	" "	42.68390	" " "	27.53945
	" "	27.53945	" " "	and so on
	" "	and so on		
Total $1,000,000,000		$2,857,142,857.15		$1,857,142,857.15

We assumed that originally the national income was $100 billion and investment was $25 billion. Hence the national income increases to $102.857+ billion when investment increases to $26 billion. Or, comparing the final with the original situation, we have (in billions of dollars):

	Original	Final
Investment	$25	$26
Consumption	75	76.857,142
National Income	$100	$102.857,142

It must be remembered that this result was computed when the marginal propensity to consume was assumed to be 65 per cent.

Increased Investment: Marginal Propensity to Consume of 50 Per Cent

It is clear that the size of the multiplying effect depends upon consumers' reactions to the changes in their income, that is, on the marginal propensity to consume. In fact, it is this response by consumers that enables the process — the change in income, which leads to a change in consumption, which in turn generates a further change in income which leads in its turn to a further change in consumption, and so on — to be carried on. If people did not modify their consumption when their incomes changed, the process would stop after the first step. The nature of the response obviously influences the total effect.

Let us see what happens when consumers react less strongly to increases in income. We shall assume that instead of spending 65 cents out of every dollar of increase in their income, they spend only 50 cents, indicating a marginal propensity to consume of 50 per cent. With this marginal propensity to consume, an investment of $1 billion produces a total rise in income of only $2 billion. It is worth taking a moment to see how this conclusion is reached. The $1 billion of additional investment will bring about a $1 billion expansion in income. This increase will cause a $0.5 billion rise in consumption, which in turn will occasion an additional $0.5 billion increase in incomes earned in the consumers' goods industries. This rise in income will bring about another increase in consumption and of income, but this time of only $0.25 billion, and with this increase in income, consumers' expenditures will be further increased by $0.125 billion. And so the process goes, with income rising in the successive stages thus: $1 billion + $0.5 billion + $0.25 billion + $0.125 billion + $0.0625 billion,

and so on out to the end of the series, when the total increase in income amounts to $2 billion. The results can perhaps be more easily seen in a table:

TABLE 45

Effect of Change in Investment on Income and Consumption

The Increase in: (in millions of dollars) Marginal Propensity to Consume, 50 per cent		
Investment	**Income**	**Consumption**
1000	1000	500
	500	250
	250	125
	125	62.5
	62.5	31.25
	31.25	15.625
	15.625	7.8125
	7.8125	etc.
	etc.	
Total $1000 (or $1 billion)	$2000 (or $2 billion)	$1000 (or $1 billion)
Summarizing we have:	**Original**	**Final**
Investment	25	26
Consumption	75	76
Income	100	102

In this case the rise in income is twice as great as the expansion in investment which initiated it. In the earlier situation the rise in income was 2.857 times as great as the rise in investment. Thus, in both cases the rise in income is considerably greater than the change in investment which brought it about. As we have already said, so long as consumers increase their expenditures when their incomes increase, the total increase in income is much greater than the increase in investment.

It should be noticed that the change in income is greater the more strongly consumers react to a change in their income. Given a marginal propensity to consume of 65 per cent, a $1 billion increase in investment raised the national income by about $2.86 billion; given a marginal propensity to consume of 50 per cent, the national income increased only $2 billion. The greater is the consumers' response to

changes in their income, the greater is the resultant increase in income from any increase in investment.

Marginal Propensity to Consume of 100 Per Cent and Increase in Investment

A marginal propensity to consume of 1.0, or 100 per cent, would produce what could quite properly be called a chain reaction. A small increase in investment would lead to an infinite increase in spending, and hence in money income. When the marginal propensity to consume is 100 per cent, for every additional dollar of income consumers increase their spending by a dollar. When the national income rises by $1 billion, consumers raise their expenditures by the same amount. Let us trace out what would happen if, with such a marginal propensity to consume, investment were increased by $1 billion.

First of all, incomes earned in the investment goods industries would go up by $1 billion, and consequently consumption expenditures would increase by $1 billion. Consequently, incomes in consumers' goods industries would also increase by $1 billion, and this would lead to a further rise in consumption by the same amount. This rise in consumption would generate an additional $1 billion increase in income earned in the consumers' goods industries, so that the total increase in income to this point would be $3 billion. Obviously the process would not end here. Rather, it would go on without limit, and each increase in income would be the same size as the ones before. After twenty steps the total increase in income would be $20 billion, and after a hundred steps, it would be $100 billion. Given time, incomes would increase by an infinite amount.

In the real world, such a reaction, even in its beginning stages, happens very rarely, because, when income rises, consumption is seldom raised by the same amount. And as we have seen, when the marginal propensity to consume takes a value between 50 and 75 per cent, the increase in income resulting from an increase in investment is finite. But during periods of extremely acute inflation, as, for instance, in 1923 in Germany, the marginal propensity to consume amounts to 100 per cent or even higher; for people try to spend money as soon as it gets into their possession. Obviously such a situation is explosive. A small increase in income will lead, if it is not checked, to a catastrophic increase in spending and in money income, though of course not in real income.

The Multiplier and the Marginal Propensity to Consume

There is a precise relation between the way income responds to a change in investment and the marginal propensity to consume. Let us now examine the nature of the connection. When the marginal propensity to consume is 65 per cent, a $1 billion increase in investment brings about a $2.8 billion increase in income. This ratio of the increase in income to the increase in investment, or more accurately, the ratio of the *changes* in income and investment, is known as the *multiplier*, and is symbolically defined as follows:

M = the multiplier
ΔY = as before, the change in income
ΔI = the change in investment

then $$M = \frac{\Delta Y}{\Delta I}.$$

Hence, when the national income increases by $2.8 billion as investment rises by $1 billion, the multiplier is 2.8. When the national income rises by $2 billion as investment rises by $1 billion, the multiplier is 2. And if the national income should rise by $11 billion when investment increased by $4 billion, the multiplier would be 11/4, or 2.75. The multiplier simply measures the responsiveness of the national income to investment changes. The relation between M (the multiplier) and R (the marginal propensity to consume) can be expressed in the formula $M = \frac{1}{1 - R}$.

Before seeing how this formula is developed, let us see whether it applies to the cases discussed above. In the first one, R was equal to 65 per cent, and the total increase in income was 2.857 times the increase in investment. The multiplier was therefore $\frac{2.857}{1}$, which equals $\frac{1}{1 - .65}$. In the second case, R was equal to 50 per cent and $M = 2$. Thus, M was clearly equal to $\frac{1}{1-\frac{1}{2}}$. The formula appears to hold for the two cases which have been worked out in detail.

Derivation of the General Rule

The total increase in income resulting from an increase of 1.0 in investment is calculated as follows for the general case. First, we

must take into account the increase in income earned in the investment goods industries. Since investment is increased by 1.0, income earned in the investment goods industries also increases by 1.0. With higher incomes, employees and employers of firms in investment goods industries will increase their consumption by $1 \times R$ (R generally being less than 1). Because of this increase in consumption, incomes in the consumers' goods industries go up by R. The total increase in income through two steps is, therefore, $1 + R$. When incomes in the consumers' goods industries increase by R, consumption itself will expand by $R \times R$, or R^2, and hence incomes in the consumers' goods industries will rise by an additional R^2, which will lead consumers to expand their purchases of consumers' goods by $R^2 \times R$, or R^3. With each increase in consumers' spending, incomes earned in the consumers' goods industries rise by an equivalent amount, or at this stage by R^3, and with this rise, consumption is itself boosted further, at this stage by R^4. The total increase in income, considering only these four steps, is therefore $1 + R + R^2 + R^3$. You should now be able to see that if we allow for the effects that will be felt after a large number of steps, the total increase in income comes to: $1 + R + R^2 + R^3 + R^4 + R^5 + \ldots R^{n-1} + R^n + \ldots$, etc. This, as you will recall, is an infinite geometrical progression, with R the common ratio. Since R is normally less than 1.0, the series is convergent. The formula for the sum of an infinite series of this kind is $\frac{A}{1 - R}$, where A is the first term (in this case, 1).

Now, what is the multiplier? It is defined, as we have noted, as the ratio of the change in income to the change in investment. The change in investment is 1. The change in income is $\frac{1}{1 - R}$. Therefore, the multiplier is $\frac{\frac{1}{1 - R}}{1}$, or $\frac{1}{1 - R}$.

From this formula it is apparent that if the marginal propensity to consume is high, the multiplier will also be high. This is what we should expect, for the total increase in income is simply the sum of all the reverberations that follow the initial increase. When the marginal propensity to consume is high, the reaction is considerable, and the increase in income is, therefore, relatively large. But if the marginal

propensity to consume is zero, the multiplier equals 1, for in that case there are no repercussions to add their effects on income to the initial change. The marginal propensity to consume in our economy appears, in normal conditions, to be roughly two-thirds, or, to use the average figure, 65 per cent. Thus, applying the formula $M = \frac{1}{1 - R} = \frac{1}{1 - .65}$, the multiplier would be nearly 3. In our economy, therefore, the effect of a change in investment is considerable, since it brings about an increase in income almost three times as large.

The Multiplier in the United States

More direct evidence on the value of the multiplier can be secured by comparing changes in income, or gross national product, between any two dates, and changes in investment between the same dates. Some comparisons of this kind are given in Table 46.

To calculate the multiplier in this way is to assume that all the changes in income were brought about only by changes in investment, which implies that the propensity to consume did not vary.[1] While this assumption seems reasonably accurate for the period under review, it is obviously not exact. On this account we should not expect the measure of the multiplier so obtained to be stable even if the multiplier actually is so. The computed results show a fair degree of stability — they suggest that a value between 2.5 and 2.75 is a good working figure — a rough approximation certainly, but nonetheless better than an uninformed guess.

The Multiplier Effect over a Period of Time

This snowball process which we have been describing obviously takes time, and it does not occur in the simple, clear-cut way our analysis may have implied. To illustrate the process in concrete but not necessarily in realistic terms, it is as though we started with a level of investment which had been stable from January to May. On June 1, investment is raised, and it is maintained at this new, higher level, let us assume, for a year. Income, after being stable from January to May, will begin to rise after June 1. It may take two or three months

[1] It also assumes that the effects of the time lag between the change in investment, and the consequent change in income, were small.

TABLE 46

Comparison of Changes in Income and Investment Between Given Dates

The Multiplier in the United States
(in billions of dollars)

Year	Gross National Product	Investment	ΔY*	ΔI*	$M = \frac{\Delta Y}{\Delta I}$
1921	70.7	17.6			
1925	90.4	24.9			
1921–1925			19.7	7.3	2.7
1926	95.6	26.6			
1930	88.2	23.3			
1926–1930			—7.4	—3.3	2.2
1931	72.1	17.9			
1933	54.8	12.4			
1931–1933			—17.3	—5.5	3.1
1934	63.8	16.1			
1937	87.7	25.2			
1934–1937			23.9	9.1	2.6
1938	80.6	22.1			
1941	120.2	45.6			
1938–1941			39.6	23.5	1.7

*ΔY represents the change in gross national product: ΔI the change in investment. The reader is reminded that investment includes private investment (before subtracting depreciation), and total government expenditures on goods and services (which we call government investment).

before it reaches the level at which it is adjusted to the higher investment figure.[1] Then it too remains stable until investment changes again. We should not, of course, identify the actual course of the increase in income with the step-by-step analysis by which we demonstrated the multiplier effect. For there is no reason to suppose that each round in our analysis takes a set number of days or weeks. The step-by-step analysis is not a method of tracing out the actual changes in income as it rises from the original to the final higher level. It is simply a logical device which enables us to determine the size of the increase in income once investment has increased.

[1] During the period in which it is rising, either investment or the propensity to consume, or both, will be lower than in the final picture.

The Multiplier Effect of a Reduction in Investment

Unfortunately, a multiplier which is large when investment is increasing is likely to be large when investment declines. For there is no reason to suppose that the marginal propensity to consume is smaller for reductions in income than it is for increases. Thus in an economy in which the multiplier is three, a small drop in investment brings about a large decrease in the national income and, of course, a large increase in unemployment. It is perhaps desirable to trace out, in some detail, the consequences of a decline in investment.

Let us consider a fairly simple example, in which the marginal propensity to consume is equal to 50 per cent, and investment drops from $10 billion to $5 billion a year. When investment falls by $5 billion, factories once busy producing steel, machinery, and construction materials become idle. Workers employed in those factories are thrown out of work and their incomes disappear; and the owners of the factories receive much smaller profits and perhaps even suffer losses. The total decline in the income of these employers and employees amounts to $5 billion. This is the primary reduction in income. Since the marginal propensity to consume is 50 per cent, consumers' expenditures will fall off by $2.5 billion because of this reduction in income. The depression now begins to spread. Shoe factories, retailers, barber shops, drugstores, all do less business. Some of their employees are thrown out of work, and most of the owners receive lower incomes. Thus, to the original $5 billion decline in income earned in producing investment goods, we must now add the $2.5 billion drop in income from consumers' goods industries. From this there follows a secondary drop in consumption of $1.25 billion, and of course an equal decline in income from the production of consumers' goods. Hence the total drop in income is already $8.75 billion. But obviously, the series does not end here, for each drop in expenditure generates a further reduction in income, and so on *ad infinitum* — or at least for a very long time. To the original $5 billion we should have to add $2\frac{1}{2} + 1\frac{1}{4} + \frac{5}{8} + \frac{5}{16} \ldots$, and so on, billion dollars. The sum of this series, employing the formula for the sum of an infinite, convergent, geometrical progression, is $10 billion. The total decline in income is therefore double the amount of the original decline in investment which started all the trouble. Obviously, the situation would be even more unfortunate if the multiplier were larger than two. Thus, if it were four, a $5 billion decline in investment would

ultimately reduce income by $20 billion, and incidentally, the multiplier would have that value if the marginal propensity to consume were three-fourths.

The Uses of the Multiplier

The multiplier concept and a knowledge of its value are helpful tools in developing controls for the business cycle and in forecasting business conditions. If the multiplier is three, the national income changes by three times as much as the change in investment. Hence it follows that if we want the national income to increase by $30 billion, we have to find ways of increasing investment by $10 billion. Or if the multiplier is 2.5, we must increase investment by $12 billion in order to increase the national income by $30 billion. If we also know the relation between changes in employment and the national income, we can determine how great an increase in investment will be necessary in order to rid the economy of unemployment. Suppose, for instance, that there are 15 million unemployed who should be given jobs, and that the national income would be $115 billion higher if all these men were employed for a year. Then, if the multiplier is 2.75, it follows that investment must be raised by $\frac{115}{2.75}$, or about $41.8 billion, to achieve this result. If investment were running originally at a rate of, let us say, $23 billion a year, it would be necessary to raise it to approximately $65 billion in order to reduce unemployment to a satisfactory level. Thus, a knowledge of the multiplier enables us to see in numerical terms what has to be done.[1] It does not, of course, tell us how to cure unemployment; for we would still have to determine how to raise the investment level. But at least the analysis sets out some essential guideposts.

A knowledge of the multiplier is also essential if our forecasts of business conditions are to be reasonably accurate. Suppose, as in early 1945, that the gross national product is about $200 billion on an annual basis, and that investment is expected to decline, after reconversion is completed, by about $15 billion.[2] If the multiplier is 2.7, we must expect a decline in the gross national product of about $40

[1] Naturally, if we cannot estimate the multiplier with any precision, we cannot specify the required dosage of investment very accurately.

[2] These figures are chosen purely to illustrate a way in which the multiplier can be used. They are obviously not predictions.

billion and a related increase in unemployment. Obviously, in order to make a good forecast, not only do we need an accurate estimate for the multiplier; we must also predict future changes in investment accurately. At any rate we should, in preparing our forecast, know what to look for. To follow this procedure is better than to use astrology or to read tea leaves, though how much better depends upon how accurately we can estimate the value of the multiplier and predict the future course of investment. Since the multiplier simply relates changes in income to changes in investment, and since investment is liable to large fluctuations and is to some degree subject to social control, the multiplier concept is a useful analytical tool.

Consequent Changes in Investment: The Acceleration Factor

One of the editors to whom these pages were submitted wrote the following on the manuscript: "Query — doesn't the continued rise in consumers' goods production cause some increase in investment goods? More books need more presses; more cloth takes more looms, and so on." He was right, of course. It is highly probable, as we pointed out in the previous chapter, that once consumers' spending begins to increase, business firms will be induced to expand their investment still further. When more books are sold, printers are encouraged to buy more presses. When sales of men's clothing increase, textile mills and clothing establishments are induced to put in more looms and more display cases. This will mean a still higher rate of investment. We assumed in a previous example that investment rose initially from $20 to $21 billion, and that as a consequence consumers' spending would also increase. We now see that this would probably induce business firms to invest at an even higher rate — perhaps $21.5 billion per annum — which would raise consumers' spending still further, so that investment might also be raised again, perhaps this time to $22 billion. The total effect on income would thus be considerably greater than the product of the initial rise in investment and the multiplier. If the *total* rise in investment were $3 billion, instead of the original $1 billion, the increase in income with a multiplier of 2 would be $6 billion, not $2 billion, as we figured earlier before counting in the effects of the additional rises in investment.

The fact that the level of investment depends (among other things) on the rate of change in consumption, should now be clear. This phenomenon is known as the *acceleration effect*, and the acceleration

factor, or coefficient, which measures this effect, is the ratio of the secondary change in investment induced by a change in consumers' expenditures, to the change in consumers' spending. If, for instance, a rise in consumption of 2 causes investment to be higher by three-fourths than it would have been with no rise in consumption, the acceleration factor or coefficient is three-eighths.

Since the acceleration factor is unlikely to be zero — that is, since changes in consumption will almost certainly affect investment — we should not apply the multiplier against the initial change in investment to determine the effect on income, but rather against the total change. For example, suppose the multiplier is 2 and investment initially falls by $4 billion. If there were no further variations in investment, the total decline in income would be the product of the change in investment and the multiplier, or $4 \times 2 = \$8$ billion. But if, as consumers' spending fell, investment were further reduced, income would fall by more than $8 billion. For we should have to consider the multiplying effects of the secondary fall in investment which was caused by the decline in consumption. Thus, if investment finally settled at a figure $8 billion below the original level, the total reduction in income would be 8×2, or $16 billion.

Summary

Changes in investment are the most important factor in bringing about changes in the national income and employment. Hence it is vitally important that we understand how variations in investment produce their effects. To determine how these variations affect the national income is the purpose of the present chapter. A change in investment affects the national income directly by changing the income of those who are engaged in the production of investment goods. It affects the national income indirectly through its effect upon the consumption of those whose incomes are directly altered, and hence upon the incomes, consumption, and back again to the incomes of those who produce consumers' goods. And the sum of these indirect effects upon income may be larger than the direct or primary effect. When it is larger, the multiplier exceeds two. Finally, a change in investment affects income still more remotely through the secondary change in investment that is induced by the change in consumption.

To take a very simple example, suppose that the multiplier is 3 and the initial change in investment is 2. Suppose that as a consequence

of whatever changes take place in consumption, investment is increased by an additional 3. Then the *total* change in investment is 5, and that in income is 5 × 3, or 15. The direct effect is to raise income by 2 (through direct change in investment). The indirect effect is to raise income by 4 (through resulting change in consumption). The remote effect is to raise income by 9 (through induced change in investment and thus in consumption). The total effect, therefore, is to raise income by 15.

An analysis of the effects of changes in investment upon income is of especial importance to us because our economy is subject to enormous fluctuations, as we have already seen, in investment. In some years during the nineteen-thirties, investment was as low as $13.5 billion. In 1944 it reached a peak of almost $100 billion. When it is realized that the effect upon the national income of a change in investment is perhaps two or three times the size of that change, its importance can be readily judged. If, for instance, there had been no change in the propensity to consume, the increase in investment between 1933 and 1944 would have brought about an increase in the gross national product of about $220 billion.[1] The propensity to consume did change in this period, but this is unusual. Normally, changes in investment are of critical importance in plunging the economy into deep depression or raising it to high prosperity. For the propensity to consume is, in contrast, relatively stable. Investment is to our economy much the same as an accelerator to an automobile. Most of the changes in the rate at which our economy operates result from changes in the amount of investment. For that reason, an understanding of the way in which changes in investment affect the national income is of critical importance.

[1] Assuming a multiplier of about 2.6.

30

The Determinants of the Propensity to Consume

CONSUMERS' SPENDING is done to acquire such things as food, clothing, amusements, reading, education, and house furnishings. What distinguishes consumers' spending from investment spending is that the buyer makes his purchase to satisfy directly either his own wants or those of his family. As a consumer, he does not buy things for resale. He buys food to eat and clothes to wear, books to read and radios to listen to. Moreover, since in dealing with our own economy, we regard *all* sales to foreigners as sales of investment goods, consumers' spending in our terminology is confined to the purchases of persons who reside in the United States.

We have already seen that income is one of the most important factors in determining how much consumers spend. When their income is high, they spend freely; when it is low, they spend much less so. The influence of all the other factors that determine consumers' spending we have grouped together as the propensity to consume. The propensity to consume, then, indicates the level of consumption forthcoming at each level of the national income. Since the propensity to consume, together with investment, determines the national income, we must find out what elements go to make it up if we are to understand the workings of our economy. How can we account for the present level of the propensity to consume? What could be done to raise or lower it, if that seemed desirable? These questions the present chapter will attempt to answer.

Changes in the Propensity to Consume

First of all we must be clear about what is implied by the phrase, "a change in the propensity to consume." The analogy with such terms as "a change in demand" or "an increase in supply" will prove helpful here. Briefly, it means that consumers' spending *at a given level of income* is different from what it was before. Thus, if, when the national income is 100, consumers' spending rises from 80 to 85, the propensity to consume has increased. Or if, when the national income is 150, consumers' spending declines from 100 to 95, the propensity to consume has declined.

Changes in the propensity to consume can be illustrated graphically. In Figure 61, the symbol ϕ_1 refers to the original propensity to consume. At income 100, consumers' spending, as we can read from ϕ_1, is 80. Now, if consumers should change their spending habits so that with the same income they would spend 85, ϕ_1 would no longer describe the situation. A new curve denoted by ϕ_2, which passes through the point B (100, 85), would have to be drawn to picture the

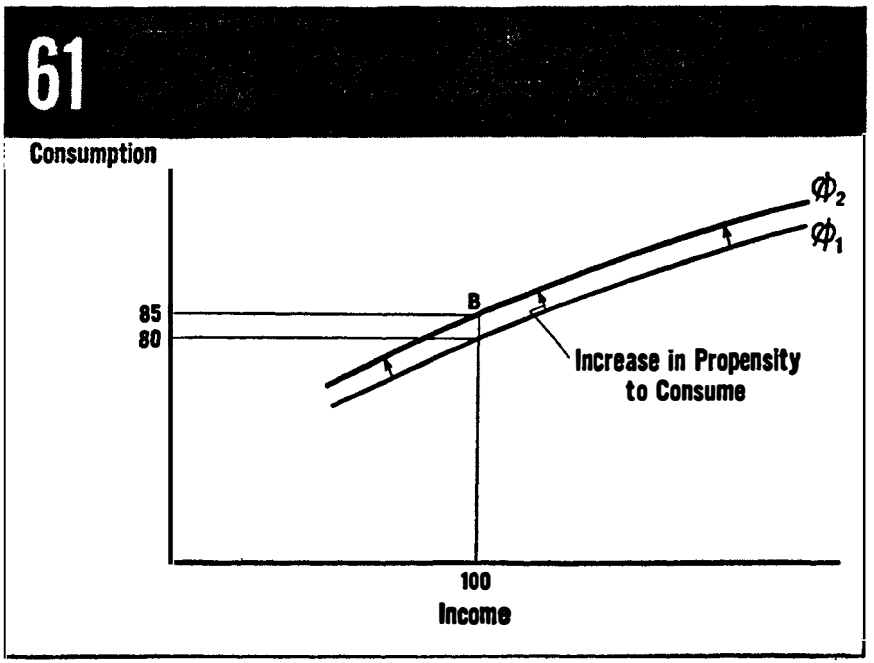

Figure 61. *Change in Propensity to Consume*

increased propensity to consume. Conversely, if the propensity to consume had fallen, the new curve would, of course, have to be drawn below ϕ_1, instead of above it.

We must distinguish carefully between a change in the propensity to consume and a change in consumption, just as earlier we had to distinguish between a change in demand and a change in the amount demanded. A change in consumption means only that consumers are spending a different amount than before. This may happen because of a change in income, or because of a change in the propensity to consume, or because of a combination of these changes. But a change in the propensity to consume is something quite different from this. It implies a change in the spending habits of the economy, and it is recognized by the fact that *even when income is unaltered*, spending on consumers' goods is at a different level than before. A change in income does not bring about a change in the propensity to consume, but it would certainly bring about a change in consumption. This exactly parallels a point made earlier, that a change in price does not bring about a change in demand, but certainly brings about a change in the amount demanded. The level of consumption is measured by the amount spent on consumers' goods. For any period it is a definite sum, say $100 billion. The propensity to consume cannot be expressed as a sum or a fraction: rather it is, as we have already seen, the *relation* between income and consumption. It has to be formulated either graphically, as above, or else in the form of a table or schedule. The difference will perhaps be clear if it is noted that the propensity to consume could change in one direction, while at the same time consumption changed in the other. Thus, for instance, if there were an increase in the propensity to consume and at the same time a sharp fall in investment [1] which caused a reduction in income, consumption might fall. Thus, consumption and the propensity to consume are quite different concepts.

Propensity to Consume and the Distribution of Income

The amount spent on consumers' goods at any level of the national income depends, for one thing, upon who gets that income. That is, the propensity to consume depends upon the distribution of income, and any change in income distribution produces a change in the

[1] It is not suggested that the increase in the propensity to consume would discourage investment; but a decline in investment could of course occur for other reasons.

propensity to consume. Let us see why this is so. Let us assume first a very simple case in which we have only two income recipients, who have between them a total income of $7000 a year. If this income is divided equally between them, each man will spend about $2700 of his $3500 on consumers' goods, according to the average figure for 1942. Total consumer spending out of this $7000 "national" income is therefore $5400, when income is distributed equally. But let us see what happens when income is distributed unequally. If one man gets $6200 and the other $800, the wealthier spends $4150 on consumers' goods and the poorer spends $800 — again according to the average figures for 1942. Total spending on consumers' goods with unequal distribution of incomes is therefore $4950, or $450 less than when income is distributed equally. It appears from this simple example that the propensity to consume rises as the distribution of income becomes more nearly equal, and falls as distribution becomes less equal. This result is based upon differences in the consuming habits of, and the taxes borne by, those in the upper and lower income brackets.

Some interesting estimates of how a change in the distribution of the national income would affect the propensity to consume have been made by the National Resources Committee. This agency determined that for the period from July 1, 1935, to June 30, 1936, when the national income of the United States was about $59 billion, consumption came to approximately $50 billion.[1] Basing its computations on a study of the spending habits of families in various income brackets, it then computed how much would have been spent on consumers' goods if the national income had been distributed equally among all the families in the country. The calculation showed that, instead of spending only $50 billion of the $59 billion on consumers' goods, we would have spent $55.7 billion. This, of course, implies a markedly higher propensity to consume, and is consistent with the results of the simplified case we considered above. We may conclude, therefore, that the more nearly equal is the distribution of income, the higher is the propensity to consume; and the more unequal the distribution, the lower is the propensity to consume.

[1] While these estimates of income and consumption are somewhat different from those now accepted, the difference in figures need not concern us in the present argument.

The Propensity to Consume and Attitudes Toward Thrift

The miser spends only a small part of his income; the spendthrift may spend all of his, or even more. Obviously, psychological attitudes toward spending and thrift affect the propensity to consume. Attitudes toward thrift are, of course, very complex. The miser wants to accumulate and save simply because he is gratified by hoards of wealth. Most of us feel that we should put something aside for a rainy day; that we should have savings on which we can draw in case of sickness or unemployment. Many of us have been brought up to believe that saving is a virtue, and that society prospers when we all save. Obviously a complex of motives conditions our spending, and any change in these attitudes will alter our spending habits. A clear instance of this occurred during the last war. Anxious to reduce purchases of consumers' goods, the government employed a variety of techniques, such as taxes, bond sales, and the encouragement of thrift. As a result the propensity to consume was reduced far below its normal level. We cannot, of course, determine the effectiveness of these measures individually, but there is no reason to doubt that the publicity and propaganda directed toward stimulating thrift was to some extent successful.

Since a good deal of family saving is done to provide against emergencies, any measures taken to reduce the need for this kind of saving would probably affect the propensity to consume.[1] Thus, when the government provides unemployment insurance, either by seeing to it that jobs are always available or by payments of insurance benefits for the unemployed, less must be set aside for this contingency. Likewise, if medical insurance were generally available, families would require less to meet the costs of illness, and consequently they would tend to save less than if they had to pay all such expenses themselves. Such measures would therefore bring about an increase in the propensity to consume.

The Propensity to Consume and Expected Changes in Price or Supply

Another determinant of the propensity to consume is any expectation of changes in price. When consumers expect prices to rise, they spend on consumers' goods a larger sum out of a given income than

[1] While most of us save for these reasons, most saving is done by families who do not have any real fear of the economic future: by families whose incomes are very high. Creating greater security would probably not affect their saving markedly.

when they expect prices to remain the same or to fall. We all like to buy at the lowest possible price. If we have any reason to believe that the price of something we want is going to be raised, we normally try to buy it before this happens. If general price increases are expected, we stock up as well as we can on durable goods, clothing, canned foods, and so on. Contrariwise, if we expect prices to be lower a month from now than they are today, we may postpone our purchases a while. Therefore the propensity to consume is higher when consumers expect prices to rise than it is at other times.

Occasionally, the belief that prices are due for a sharp change brings about an enormous shift in the propensity to consume; this is most likely to happen in periods of extreme inflation, such as the great German inflation of 1920–24. In those years prices rose rapidly, and toward the end at breakneck speed. No one had any reason to hold marks, for if he delayed his spending until the next day, he would be unable to buy nearly as much with them. Hence, as soon as anyone received money income, he tried to spend all of it. You have probably heard stories like the account of the village school teacher who received her pay on Friday evenings, and then immediately had to bicycle thirty miles to the nearest market town to buy provisions. She had to hurry because if she had waited until Saturday morning, prices would have been out of reach. The rush to acquire consumers' goods before prices rose still further, or, as it is often put, to get rid of marks before they lost their value, certainly affected the relation between income and consumption. To say that at such times the propensity to consume is abnormally high deserves a place in the *New Yorker Magazine's* Department of Understatement. In a small way we had much the same experience in the United States in the first weeks of July, 1946. Price-control legislation had been allowed to lapse, prices rose sharply, and consumers' buying increased considerably for a short time.

Consumers react in much the same way when they expect that goods are going to become scarce as when they expect prices to rise. During the war a rumor that a certain commodity was going to be rationed was enough to send up its sales enormously. Housewives — or perhaps in fairness we should say some housewives — bought up all they could of whatever they thought would soon be rationed or scarce. Thus, the expectation of shortages has the same effect on the propensity to consume as the expectation of a rise in prices. It is

worth noting, incidentally, that when consumers react in this way, they intensify the very trouble they are trying to avoid. When they rush to the stores to buy everything in sight before prices rise, they give the seller an even stronger motive for raising his price than he had before. The fear of price increases tends to promote these increases, and the fear of shortages tends to make the shortages real.

The Propensity to Consume and Taxes

Taxes also play a part in determining consumption and the propensity to consume. By subtracting something from the income available to the consumer, a tax generally forces a certain amount of retrenchment, and thus reduces consumption. The amount by which consumption is reduced depends, of course, upon the kind of tax and consequently upon who has to pay it. A tax which falls heavily upon the low income group will cut consumption by approximately the amount of the tax, since all, or nearly all, the income of this group is spent on consumers' goods anyway. But a tax which falls on very high incomes will probably force a reduction in *saving* rather than in consumption, since, as we have seen, there is usually a considerable gap between the consumption and income of the wealthy, and a change in their income does not greatly change their consumption. Thus certain levies, like sales or property taxes, lower the propensity to consume, since they tend to take a larger percentage of low than of high incomes; while others, such as the income tax, affect it much less strongly.

The Rate of Interest and the Propensity to Consume

A change in the rate of interest may also alter the propensity to consume, though the direction of the change is not certain. The effect is probably greatest with durable consumers' goods such as automobiles, refrigerators, furniture, and other commodities often bought on the installment plan, for a rise in the interest rate makes installment buying more expensive and is likely to discourage such purchases. This effect is reinforced by another: that a high interest rate encourages potential lenders to reduce their own consumption in order to increase their ability to lend. When the interest rate is raised, it becomes more advantageous to lend. To lend is not of course the same thing as to save, for the lender simply makes money available to borrowers which has been accumulated somehow, perhaps years

before. But naturally, the more one saves, the more he has to lend. Hence we may suppose that anything which makes lending more attractive will also encourage saving and thus lead to a reduced propensity to consume. Therefore, on this account too, a rise in the rate of interest would reduce the propensity to consume.

But opposing tendencies also operate. A rise in the interest rate may lead to a rise in consumption by those who save in order to build up a certain amount of wealth. When the interest rate is only 3 per cent, an annual saving of approximately $372 is enough to accumulate $10,000 in twenty years, but when the rate is 6 per cent, an annual saving of only $272 is needed to accumulate the same amount in the same time. The most important application of this for the ordinary consumer is in connection with life insurance. The cost of a policy varies inversely with the rate of interest; the higher the rate, the less it costs to purchase a certain amount of insurance. Hence, at the higher interest rate the head of a family who does most of his saving in order to provide protection for his dependents may build up the same reserves with less saving, and may therefore have more to spend on consumers' goods. Thus, whether a rise in the interest rate raises or lowers the propensity to consume depends on the relative strength of these opposing forces, and there is no reason to suppose that the answer will always be the same.

It is well, in this connection, to emphasize again the distinction between consumption, saving, and the propensity to consume. We have seen that a rise in the interest rate may lower the propensity to consume, since less may be spent on consumers' goods and consequently more may be saved, out of a given income, than at the original interest rate. But this does not mean that saving would increase, for — and this is the important point — the increase in the interest rate may discourage investment and thus reduce the national income. And even though the propensity to consume should decline, saving would fall if there were a large enough decrease in the national income. Thus, to say that when interest rates rise people try to save more and the propensity to consume declines, is not to say that consumption falls and saving increases. In the situation we have postulated, consumption would fall on two accounts: because of the decline in the propensity to consume, and because of the fall in income. And saving would decline in spite of the fact that people wanted to save more, because their incomes were lower.

In short then, a change in the rate of interest may either raise or lower the propensity to consume. However, because of the opposing forces set into motion by such a change, we may generally conclude that the propensity to consume will not vary by a large amount. The effect on saving may of course be quite different.

Stocks on Hand and the Propensity to Consume

In determining the propensity to consume for durable consumers' goods, an additional and very important factor has to be taken into account: the supply already in the hands of consumers. Very few families want more than one refrigerator or more than a few bathtubs or radios. Hence, after a period of heavy buying of automobiles, refrigerators, pianos, and so on, it becomes more and more difficult to encourage consumers to go on buying at the same rate. This tendency to slacken up is often countered by advertising which persuades buyers that only the newest model is worth having, as with automobiles; or by improvements which make the older goods partially obsolete, as with tone control, short wave, and frequency modulation in radios. But in the absence of such counteracting forces, we may expect a gradual decline in the propensity to consume durable goods, provided that the rate of accumulation of such goods is high enough. Conversely, when these goods are very scarce in relation to demand, as happened during and for some time after the war, there is a gradual growth in demand, for at such a time the rate of accumulation is much below normal. Once it becomes possible to buy such things again, the propensity to consume rises. Families replace their worn-out automobiles, radios, and washing machines, and for a time the level of consumption is abnormally high compared with income. Likewise, after periods of depression when purchases of such goods have been abnormally low, we may expect a gradual increase in the propensity to consume as soon as incomes cease to fall. For as existing stocks wear out and as new products and improved models are developed, the demand gradually rises again even without a rise in national income. Thus just as the wearing-out of capital equipment eventually stimulates the demand for investment goods, so the wearing-out of durable consumers' goods raises the propensity to consume. This of course is an additional factor that helps to bring depression (or boom, if the goods are accumulating) to an end.

Imports and the Propensity to Consume

We have been careful to define spending on consumers' goods as spending for goods produced in this country. Buying goods produced elsewhere does not directly add to the amount of spending on the products of the American economy and creates no income in this country. It is on this account that the propensity to consume is limited by definition to the relation between income and the amounts spent upon goods produced in this economy.

Any change in the attractiveness of goods produced elsewhere is likely to affect the propensity to consume. A reduction in the price of English woolens would increase their sales and probably reduce the amount spent on American woolens. And when a war cuts off the supply of French wines and natural rubber from the East Indies, consumption of the closest substitutes our own economy can offer will be very likely to rise. Generally, if foreign goods become more attractive, so that the amounts purchased increase at each level of income, spending on consumers' goods produced in the United States can be expected to decline.[1] The reverse, of course, holds when foreign goods become less attractive. The attractiveness of foreign goods may vary because of a change in their price, their quality, or their availability. A change in their price in the United States could occur for any of the following reasons: a change in the price in the country of origin, a shift in the rate of exchange, a change in the cost of transportation, or an alteration in the tariff levied against them by this country. A complete analysis of the factors that influence imports will have to wait for the analysis of International Trade in Part Five of this book.

Corporation Financial Policy and the Propensity to Consume

The owners of a corporation share its profits. But they do not always share exactly what the corporation earns, for it may not distribute all of its profits as dividends, or it may pay out more than it has actually earned. The difference between dividends received by shareholders and the profits earned by corporations is often important, as can be seen from the following data. In 1943 corporations earned about $9.8 billion after taxes, but they paid only $4.3 billion in dividends and withheld $5.5 billion as undistributed profits. In 1931 the situation was just the opposite. In that year net corporation profits

[1] Though, as we shall see later, our exports may, as a result, rise by more than enough to make up the difference.

were minus $1.6 billion, but corporations paid $4.3 billion in dividends to shareholders, as in 1943. Thus they paid out $5.9 billion more than they earned. Hence, in the economy as a whole, there is often a large difference between income payments — including dividends received — and incomes earned — including the net profit of corporations. Whether the difference is large or small depends upon how much of their profit is distributed by corporations.

All this has a bearing upon the propensity to consume in this way: The owner of a corporation thinks of his income not so much as his share of the net profits of the corporation, but rather as his dividend receipts. When he receives dividends of, say $10,000 a year, he bases his consumption expenditures upon this figure rather than upon his share in what the corporation makes, which may be −$2000 or +$20,000. His consumption is geared to his receipts of dividends more closely than to the profits of the corporation of which he is a part owner. He may spend approximately the same amount on consumers' goods whether his share in corporation profits is −$2000 or +$20,000, so long as in each case his dividends are the same. Hence the consumption expenditures of corporation owners depend more upon their dividends than upon the profits of the corporation. Now, the national income, or the gross national product, is measured to include corporation profits, not shareholders' dividend receipts. Therefore the propensity to consume — that is, the relation between consumers' expenditures and the national income — is conditioned by the relation between dividends and corporation profits.

More concretely, suppose that the national income stood at $150 billion, corporation profits at $10 billion, and dividends at $6 billion. Among other things, consumers' expenditures would depend upon the $6 billion paid to shareholders. If there were no change in the national income, and corporation profits remained the same, but if dividends were increased to $8 billion, consumers' expenditures would certainly rise; for shareholders receiving higher dividends would increase their spending on consumers' goods. Hence the relation between national income or gross national product and consumption would change. In short, the propensity to consume would increase.

The policy of corporations in other financial matters, such as determining the size of the allowance for depreciation, also influences the propensity to consume through its effect on the dividends received by shareholders and thus on their consumption.

Summary

Consumers' spending depends most directly upon their incomes: the higher the national income, the more they spend. But many other factors, as we have seen, also affect the level of consumption, and the combined influence of these factors constitutes the propensity to consume. A change in any one of them — for example, in the distribution of income, or in corporation policy in respect to dividends — affects that propensity. The list of factors which we have discussed is not meant to be complete or exhaustive. Rather it is suggestive. Still others can exert an influence upon consumption, such as the prices of consumers' goods, consumers' holdings of cash and other liquid assets, the stock market, expected changes in income, and the size of government pensions. A complete list would require no less than a catalogue of the whole economy, for its parts are closely interwoven.

Although the propensity to consume can change for any of a great variety of reasons, it has generally been a relatively stable and constant function in the United States. Evidence for this can be seen in the following data. In 1922 the gross national product was $72.7 billion, and consumers' expenditures were $53.9 billion. In 1931 the gross national product was $72.1 billion, very slightly less than in 1922, and consumers' expenditures were $54.2 billion, only a trifle higher than in 1922. Again, in 1920 the gross national product was $86.6 billion and consumers' expenditures were $60.1 billion, while in 1939, the gross national product was $88.6 billion, and consumers' expenditures stood at $61.7 billion. Thus the relation between consumption and income was about the same in 1922 and 1931, as well as in 1920 and 1939. Indeed, the data for most of the years between 1920 and 1941 show a relatively stable relation which changed only slightly during the period. After 1941, however, the propensity to consume fell sharply. Although the gross national product of 1943 was about $90.3 billion higher than that of 1940 — that is to say, about twice the amount of the earlier year — consumers' expenditures increased by only $25.8 billion, or by 39 per cent. This suggests, although it does not prove, that the propensity to consume was lower in 1943 than in 1940. The only other explanation for so small an increase in consumption is that the marginal propensity to consume was only 29 per cent, the ratio of the change in consumption to the change in income. But this is an unbelievably low figure. We may therefore conclude

that the propensity to consume did fall. With the end of the war it rose again. The gross national product was about the same in 1944 and in 1945, but consumers' expenditures in 1944 were only $98.5 billion, and in 1945 they reached $104.9 billion. Between the first and the last quarter of 1945, the gross national product declined from an annual rate of $204.5 billion to $182.8 billion, but for those same two periods, consumers' expenditures rose from $105.0 billion (annual rate) to $110.9 billion. Since it moved in the opposite direction, such a change in consumption could not have been the result of the change in income; it must therefore have resulted from an increase in the propensity to consume. Evidence from other countries, Great Britain, for example, confirms the statement that in normal times the propensity to consume has been relatively stable, but that under exceptional circumstances, as for instance during a war, it is liable to change.

The Marginal Propensity to Consume

The national income depends upon investment and the propensity to consume. If investment alters, the size of the *change* in the national income depends upon the *change* in investment and the *marginal* propensity to consume. Thus, the marginal propensity to consume plays the same role for changes in income as the propensity to consume does for the level of income. We saw that if the marginal propensity to consume were one-half, the multiplier would be 2, and that with such a multiplier any change in investment would produce twice as large a change in the national income. Or if the marginal propensity to consume were two-thirds, and the multiplier were accordingly 3, any change in investment would produce three times as large a change in the national income. The marginal propensity to consume thus determines how widely the effects of any change in investment will spread. It is low when consumers respond only moderately to changes in their income — when, therefore, only a small part of the effects of any change in their income is passed on. It is high when consumers respond vigorously to changes in their income — when, therefore, most of the effects are passed on as changes in the income of other consumers. The marginal propensity to consume thus measures an important characteristic of the economy — a characteristic which has to do especially with the economy's stability. In short, it determines whether a change in investment affects income and employment strongly or slightly. It is therefore important to understand the fac-

tors upon which the marginal propensity to consume depends. Let us see what these factors are.

The Distribution of Changes in Income

Whether consumers will spend a small or a large fraction of a change in their income depends upon whose incomes are altered. When those in the upper brackets receive most of an increase in the national income, the change in consumers' expenditures is less than when other income groups enjoy most of the increase. Hence the marginal propensity to consume depends upon how the change in income is distributed among the various income classes.

Studies of consumers' expenditures by income groups throw light upon this problem. In 1943, according to data provided by the Office of Price Administration, consumers in the $500–$1000 income group on an average received $758 and spent $762 on consumers' goods. Those in the $1000–$1500 bracket on an average received $1249 and spent $1163 on consumers' goods. The second of these groups averaged $491 more than the first in income, and spent $401 more. If, when a family's income rises from $758 to $1249, that family would spend as much as others in its new income group, the marginal propensity to consume in the lower income brackets would be 401/491, or about 82 per cent. Similar comparison of the $2000–$2500 group with the $2500–$3000 group suggests a marginal propensity to consume of only 62 per cent in the $2000–$3000 bracket. Finally, for the $5000–$7500, and the $7500–$10,000 group, the marginal propensity to consume appears to be as low as 30 per cent. From this we may fairly conclude that a change in income experienced mainly by those in the upper income brackets will produce only a relatively small change in consumption, but that such a change affecting mainly those in the lower income brackets will produce a change in consumption nearly as great as the change in income. Hence the way in which a change in income is distributed among different income classes is an important factor in determining the marginal propensity to consume.

It is often suggested, in accordance with this observation, that adding to the income of those in the lower income groups is a more helpful recovery measure than adding to the income of those in the higher income brackets. People in the low income groups, it is argued, can be trusted to increase their expenditures by a large fraction of the increase, and thus to spread the recovery more effectively

throughout the economy. Other considerations, of course, must also be taken into account in judging this policy — as, for instance, its direct consequences on investment. These we shall discuss later.

The Marginal Propensity to Consume in Prosperity or Depression

The marginal propensity to consume is likely to be different after a long depression than after a considerable period of prosperity. Here two forces are at work which to some degree offset each other. The first of these is that during depression, many persons are likely to go heavily into debt. Then, when a revival begins and their incomes increase, they are likely to use a large part of the gain to pay off their debts instead of using it to buy consumers' goods. When this happens, the marginal propensity to consume will be relatively low during the early stages of the revival. The force opposed to this is the desire, after a long depression, to replace worn-out automobiles, household furnishings, clothing, and so on, as soon as incomes allow, which means a relatively high marginal propensity to consume during the revival. Thus these two forces tend to cancel each other out. At present there is no convincing statistical evidence to show which is the stronger.

Imports and the Marginal Propensity to Consume

The nature of a country's economic dealings with other countries also helps to determine the marginal propensity to consume. When incomes increase, consumers' spending also increases, as we know. If much of the increase in spending goes for goods produced abroad, less of it will be spent on home-produced goods. Generally, therefore, if imports respond strongly to changes in income, the marginal propensity to consume home-produced goods is low; while if imports are only slightly affected by changes in income, the marginal propensity to consume is higher.

A concrete example will illustrate: Suppose that when income increases, consumers' expenditures on all goods, foreign and domestic, increase by 75 per cent of the increase. Specifically, let us say that when income rises by $1 billion, consumers increase their spending by $750 million. If their imports rise sharply, say by $300 million, the marginal propensity to consume will be only 450/1000, or 45 per cent; but if their imports rise by only $50 million, the marginal propensity to consume will be 700/1000, or 70 per cent. Thus a country

which depends on others for much of its supplies, and which therefore may be expected to import a good deal more when its national income rises, will have a relatively low marginal propensity to consume. But a country which is relatively self-sufficient, and therefore does not greatly increase its imports when its national income rises, will tend to have a high marginal propensity to consume.

Hence, as might be expected, a change in income in the United States will have a much smaller effect on imports than such a change in Canada, Great Britain, or most other countries. Consequently, the marginal propensity to consume and the multiplier are higher for the United States than for most other countries. When investment increases in this country, most of the resulting prosperity is retained at home; consequently, the increase in domestic income is large, and relatively little of it leaks away to foreign countries.[1] When investment in the United States declines, the reduction in our income is sharp, and relatively little of the depression leaks away to foreign countries. Since most of the multiplier effect is confined generally within our own border, our multiplier is large.

In countries where the marginal propensity to consume is low because they depend heavily on trade with the rest of the world, much of the effect of a change in investment is dissipated abroad. Foreign trade acts in a sense like a safety valve for such countries. If investment rises, the increase in income is dampened because imports go up rapidly. Such a rise in imports, however, tends to create more prosperous conditions in the supplying countries. If investment falls, the decline in income is also dampened. Imports fall sharply, and thus some of the depression originating in such countries is exported to other countries.

Summary: The Marginal Propensity to Consume

The marginal propensity to consume is of decisive importance in determining the stability of the economy. The higher it is, the greater is the shock to the economy from any change in investment, for with

[1] The leakage measured against the national income of the United States is very small, but because the incomes of foreign countries are so much lower than our own, the effect of this leakage upon foreign countries is relatively large. Our imports may increase by only $200 million when our income increases by $5 billion; but a $200 million increase in foreign exports, and hence in the investment of foreign countries, is a relatively large increase for them. Hence a leakage of prosperity or depression which may be of very slight importance to us may be of great importance abroad. This matter will be discussed in detail in Chapter 42.

a high marginal propensity to consume, the multiplier also is high. The marginal propensity to consume depends, among other things, on the following factors: the way in which changes in income are distributed among different income classes, the nature of the country's economic relations with other countries, and the stock of goods in the hands of consumers. Other factors, such as the nature of the tax program, and the effect of a change in income on wages and prices, should also be considered.

31

Changes in the Propensity to Consume

SINCE THE PROPENSITY TO CONSUME has been a relatively stable function in the United States, it may well be asked why it is worth while to consider a variable which has shown almost no tendency to vary. There is some justification in this question, since as we have just seen, there were almost no changes in this factor between 1921 and 1941. Nevertheless, it definitely is worth our while to consider what happens to the economy when the propensity to consume changes, and for several reasons. In the first place, the propensity to consume does change sometimes, as in the period from 1941 to 1946, when it fell sharply and then quickly rose again. And since, together with investment, it determines the level of income and employment, these variations have a very important effect on the economy, for a change in this factor, like a change in investment, sets in motion a multiplier process and produces a change in income which is several times as large as the initiating change in consumption. Moreover, many economists are convinced that to raise the propensity to consume would be a most helpful stimulus to the whole American economy. Finally, an examination of this factor will round out our analysis and fill a gap in our understanding of fluctuations in the national income and employment. For all these reasons it is important to determine the consequences of a change in the propensity to consume.

A change in the propensity to consume does not necessarily mean a change in consumption or in saving. It simply means that spending on consumers' goods and saving at each income level are different from what they were before; and to say this is not to say that either

consumption or saving changes, for income itself may be altered. A numerical example will clarify this distinction between a change in consumption and a change in the propensity to consume. Suppose that originally the propensity to consume is such that:

TABLE 47

Propensity to Consume

When income is:	60	70	80	90	100
consumers' spending is:	50	56	62	67	72
and saving is:	10	14	18	23	28

Then when the propensity to consume is lowered, for example, the table becomes:

TABLE 48

Changed Propensity to Consume

When income is:	60	70	80	90	100
consumers' spending is:	46	52	57	62	67
saving is:	14	18	23	28	33

It will be seen that at any income level, say 90, consumption is lower than before and saving is greater. But if income should fall from 90 to 80, then in spite of the lower propensity to consume, saving would remain at 23; while if income should rise from 90 to 100 (a most improbable development), consumption would remain at 67. Hence consumption and saving are affected not only directly by the change in the propensity to consume, but indirectly through the effect of that change on the national income. Thus it is clear that if a fall in the propensity to consume, implying a greater desire to save, brings about a sufficient reduction in income, saving will actually fall. In other words, an economy which tried to do more saving might find that it actually did less; though obviously this would depend upon what happened to the national income. What then happens to income,

employment, and saving when the propensity to consume alters? Let us analyze this problem in detail.

Effect of Decrease in Propensity to Consume; Investment Constant

The analysis of the effects of a change in the propensity to consume can be made more conveniently if we assume at first that investment remains constant throughout; that is, it is not altered because of the change in the propensity to consume. While this assumption is not realistic, as we shall see below, we can correct our results to allow for induced changes in investment at a later stage. We shall suppose that investment is maintained at a rate of $20 billion a year, and that in accordance with the original propensity to consume, consumption comes to $80 billion when the national income is $100 billion. Now let us suppose that the propensity to consume falls to such an extent that consumers now desire to spend only $78 billion out of a $100 billion income. Let us also assume that the marginal propensity to consume is equal to one-half; that is to say, that when income changes by $1 billion, consumption changes in the same direction by $500 million.

If investment remains constant, the incomes earned in the production of investment goods will not change. But as a result of the decline in consumption expenditures, the incomes of those engaged in consumers' goods industries will fall by $2 billion at the first stage. Since the marginal propensity to consume is equal to one-half, consumption will then decrease by an additional $1 billion. From this point on, the analysis goes along in exactly the same way as that of the multiplier given in Chapter 28. This $1 billion decrease in consumption will lead to a further decline in income of the same amount, and this in turn will lead to an additional decline in consumption of $500 million. Incomes earned in consumers' goods industries will fall by an additional $500 million, and the total reduction in income up to this point will be $3.5 billion. Further decreases in consumption will generate further decreases in income, until finally both consumption and income will decrease by $4 billion. At the end of the process, the national income will thus have fallen from $100 billion to $96 billion, while consumption will have declined from $80 billion to $76 billion.

We have assumed, of course, that investment has not altered; that it remains at $20 billion. If so, then both before and after the change

in the propensity to consume, saving will also equal $20 billion, the difference between income and consumption. Thus while the decrease in the propensity to consume implies an increased desire to save, no more saving is done than was done before. What maintains saving at the original level, in spite of the increased desire to save, is the substantial decrease in incomes. The following table illustrates the situation both before and after the fall in the propensity to consume.

TABLE 49

Effect on Saving of Decrease in Propensity to Consume

Original Situation		Situation after Decline in Propensity to Consume
Investment	20	20
Consumption	80	76
National income	100	96
Saving	20	20

An Increase in the Propensity to Consume

The analysis of the effect of an increase in the propensity to consume follows exactly the same pattern. Let us trace it through briefly. We shall assume that the marginal propensity to consume is two-thirds, and that, as before, investment does not change. We may further assume an increase in the propensity to consume such that expenditures on consumers' goods rise from $80 billion to $82 billion out of a national income of $100 billion. The initial $2 billion increase in consumption will generate a $2 billion increase in the incomes earned in the consumers' goods industries. This will produce a further rise of $1⅓ billion in consumption, which will lead to an equivalent increase in income. Again there will be an increase in consumption, this time by $.89 billion (= 2/3 × 4/3 billion), and yet another increase in income, and so on. The total increase in income would be $6 billion (= 2 + 4/3 + 8/9 + 16/27, etc.). Consumption, of course, would increase by the same amount. The national income would thus rise from $100[1] to $106 billion, while consumption would increase from $80 to $86 billion. Saving, which originally stood at $20 billion, would be maintained at this same level, paradoxical though it may seem.

[1] If investment were $20 billion.

In short, when investment does not vary, a change in the propensity to consume leads the national income to change in the same direction, whether upward or downward. The change in consumption initiated by the change in the propensity to consume brings about a change in income. The change in income in its turn produces another change in consumption, the size of which depends upon the marginal propensity to consume. A multiplier process similar to that initiated by a change in investment is set in motion. Hence the total change in income is considerably larger than the initial change in consumption. With a marginal propensity to consume of one-half, the total change in income is twice as large as the initial change in consumption. Generally, if the marginal propensity to consume is R, the total change in income equals the initial change in consumption $\times \left(\frac{1}{1-R}\right)$.

The level of saving, however, does not change so long as investment remains the same, even though the desire to save should alter. When the propensity to consume falls, saving remains constant as long as investment does not change. This is hardly what we should expect if we were to notice only that a decline in the propensity to consume meant an increased desire to save. But because of the consequent decrease in income, saving also would be reduced. It is easy to understand how this result is obtained if we regard saving for what it really is — simply not spending. We must be careful not to regard it as something active or positive, as for example, buying a new capital good. It is no such thing; it simply means that income recipients have refrained from spending all of their income on consumers' goods. Thus when the community tries to save more, it merely refrains from spending as much as before on consumption. Unless business firms or government bodies are led, as a consequence, to buy more investment goods (and services), income declines and saving itself does not increase.

Investment and the Changed Propensity to Consume

We must now re-examine the assumption that investment remains fixed, for the question arises, Can we reasonably expect investment to be unaffected when there is a change in the propensity to consume? The answer is that private investment surely would alter. When the propensity to consume changes, the first effect is a change in the

sales of consumers' goods in the same direction. When sales fall, as they would if there were a fall in the propensity to consume, do businessmen still have the same incentive to order investment goods? When the demand for their product declines, are they led to expand inventories, to put in more equipment, and to add to plant at the original rate? Certainly not, if they base their investment decisions upon profit considerations. When sales are falling, rather than add to inventories at the old rate, they are tempted to reduce them; rather than continue installing new machinery as before, they cut down their orders. The equipment already on hand is found to be adequate for the reduced demand, hence the marginal efficiency of capital of these types declines. Therefore when the propensity to consume declines, private investment would ordinarily be expected to fall too. And there is no reason why either foreign or public investment should increase when the propensity to consume falls. Therefore we may normally expect investment to fall when the propensity to consume falls. The same considerations suggest that when the propensity to consume rises, investment would normally rise too. Indeed, this conclusion simply embodies the acceleration principle, which we have already discussed (see Chapter 29). Hence we shall have to take into account not only the effects of the changed propensity to consume on income, employment, and saving, but also the effects of the induced change in investment.

What, then, are the effects of the combined change in investment and the consumption function? We have seen that when the propensity to consume increases and investment is steady, income and consumption rise and saving remains constant. And when investment increases and the propensity to consume is constant, income and consumption rise and saving also increases. Therefore when both investment and the propensity to consume increase, we must add the two results noted above. This would give us an increase in income, consumption, and saving.

This result can be best illustrated arithmetically. Suppose we assume a marginal propensity to consume of one-half, and accordingly, a multiplier of 2. If the propensity to consume increases in such a way that consumers' spending rises by $2 billion at any income level, and if investment is assumed to remain the same, income and consumption would be raised by $4 billion and saving would not change. Now let us assume that because of the increase in consumers' spend-

ing, investment rises by $3 billion. There would then be a further rise in income of $6 billion, since the multiplier is two, and consumption and saving would each increase by $3 billion. Hence the total effect of the rise in the propensity to consume and the induced rise in investment would be: a rise in income of $10 billion ($4 billion + $6 billion), a rise in consumption of $7 billion ($4 billion + $3 billion), and a rise in investment of $3 billion. Since income goes up by $10 billion, and consumption by $7 billion, it is evident that saving increases by $3 billion. It appears, therefore, that when the propensity to consume increases, the economy will consume, invest, and save more — and of course enjoy a higher national income. And yet, an increase in the propensity to consume suggests a weakening in the desire to save. But saving, as we have already pointed out, is a resultant, and like consumption, depends upon the national income. And if the national income actually increases, because of a reduced desire to save, saving may itself increase, and indeed will do so if investment rises.

When the propensity to consume declines, consumers reduce their spending. The effect of this, as we have seen, is that income and consumption fall even if investment does not. But surely investment will fall too, because of the reduced demand for consumers' goods. And the fall in investment will cause a further fall in income and consumption. Assume the following values:

TABLE 50

Effects of Reduced Propensity to Consume

Marginal propensity to consume		2/3
Hence multiplier		$\left(\frac{1}{1-2/3}\right) = 3$
Decline in propensity to consume such that at any income, consumption is reduced by		4
Consequent decline in investment		2
Then change in	**Income**	**Consumption**
Because of reduced propensity to consume	−12	−12
Because of reduced investment	− 6	− 4
Total change	−18	−16
Consequent decline in saving:	2	

In this case the initial decline in the propensity to consume brings about directly (and indirectly, because it causes a fall in investment) a total decline in income of 18. Since consumers' spending falls by only 16, saving declines by 2. We have once again the paradoxical result that although the economy desires to increase its saving — witness the decline in the propensity to consume — saving actually falls. Saving is thus rather like the lake of Tantalus: the more avidly it is wanted, the less it is supplied.

This conclusion is relevant to the course of the business cycle. If, after a long period of prosperity, consumers become pessimistic and begin to expect a depression, they cut their spending in order to protect themselves against a decline in income. Any tendency for investment to decline is thereby reinforced. Income and employment fall because of the reduction in consumers' spending. With the consequent decline in investment, income falls off by so much that saving also declines, in spite of the desire to save more than formerly. Therefore pessimism inevitably produces the very condition it fears. The effort to increase saving only succeeds in driving the national income lower. And the consequent decisions of business firms to retrench and to guard against unwanted expansion bring about the very condition that makes all expansion unwanted.

It is perhaps difficult to reconcile the fact that in such circumstances saving is bound to decline, with what appears to be the freedom of each individual to save as much or as little as he pleases. The reason it is hard to reconcile this individual freedom with what we might call a collective compulsion, is that we perhaps fail to notice how our actions affect others and compel them to adopt certain courses. Thus it may at first glance seem unreasonable to conclude that total saving should decline if everybody tries to save more. For we each feel perfectly free, within limits, to save what we like out of our incomes. But if we remember that to save more simply means to spend less on consumers' goods, it is easier to see how the above result can come about.

It is of particular importance to remember that one person's spending *is* somebody else's income; that, in our economy as in any other, we *do* live in one sense by taking in one another's washing. If we spend less, we of course increase our own saving. But as soon as we do so, we reduce the incomes of individuals who otherwise would have produced goods and services to meet our demands. When the

incomes of these people are reduced — to zero, if they become unemployed — it is obvious that their saving declines, perhaps even to a negative figure. Thus, increased saving in some parts of the economy is offset by decreased, and perhaps even negative, saving in other parts of the economy. The net effect upon saving, of course, depends upon what happens to investment. If investment is maintained at the same level, total saving does not change, although some parts of the community save more than formerly and others have to scale down their saving. On the other hand, if investment declines as a result of the decision to spend less on consumers' goods, total saving also declines at the same rate.

At this stage, the point must again be stressed that the propensity to consume is normally rather invariable. Except when the government takes strong measures to change it, as in the years from 1942 to 1945, it is relatively steady. Most of the income changes in our economy seem to result not from variations in the propensity to consume, but rather from fluctuations in investment. Hence the processes analyzed in this chapter, processes which are initiated by shifts in the propensity to consume, rarely take place in our economy. But they do happen now and then, as during 1945. Furthermore, if active steps are taken to raise the propensity to consume, as many economists now suggest, such a process would again be set in motion.

The Propensity to Consume, Investment, and the National Income

It is worth while at this point in our analysis to indicate explicitly what must be done in order to reach and maintain a high level of employment. Income and employment depend on investment and the propensity to consume. To reach full employment in our economy with the present propensity to consume requires something like $60 to $70 billion of investment a year. Thus, if nothing is done to alter the propensity to consume, investment must be maintained at this enormously high figure if we are to enjoy full employment. But there is an alternative, or more accurately, a complementary, route to full employment: that is to increase substantially the propensity to consume. If, when the gross national product is $200 billion,[1] consumers spend not $130 to $140 billion, but, let us say, $170 billion, much less investment would be needed in order to achieve full employment. Whereas with the lower propensity to consume it would

[1] We shall suppose that this corresponds to full employment.

take $60 to $70 billion of investment to produce full employment, with the higher propensity it would require only $30 billion. In short, a high level of employment can be secured by maintaining a high level of investment; but it can also be secured by increasing the propensity to consume. Thus if it should prove difficult to raise investment sufficiently, it is still possible to prevent serious unemployment if the propensity to consume can be sufficiently increased. Methods for raising the propensity to consume are to be discussed in Chapter 35.

Summary

As long as investment is constant, a change in the propensity to consume changes income and employment in the same direction. If the propensity to consume increases, income increases, and vice versa. Generally, however, we may expect that investment will not remain constant, but will also change in the same direction as the propensity to consume, thus reinforcing the direct effect of that original change. Paradoxical as it may seem, saving is likely to vary inversely with changes in the desire to save. When the propensity to consume goes up, and accordingly the desire to save declines, saving is likely to increase, and will do so if investment reacts to the changes in propensity to consume in the way suggested above. For an effort to save less increases the incomes of some people in the economy so that their saving inevitably rises. Thus, changes in the desire to save and in saving can usually be expected to be opposite in direction.

Outline of the Theory of Employment

In these chapters we have sketched only an outline of the theory of employment. Since this analysis has been somewhat complex, the chart in Figure 62, which restates the structure of the theory in graphic form, may be helpful. Briefly, the theory can be summarized as follows: Income and employment depend upon investment and the propensity to consume, and the total of investment equals the sum of private, public, and foreign investment. Private investment depends upon the rate of interest and the marginal efficiency of capital (or, in more concrete language, the schedule of anticipated yields on new investment projects) which in turn depends upon such things as the rate of development of new techniques of production and the stock of capital goods on hand.

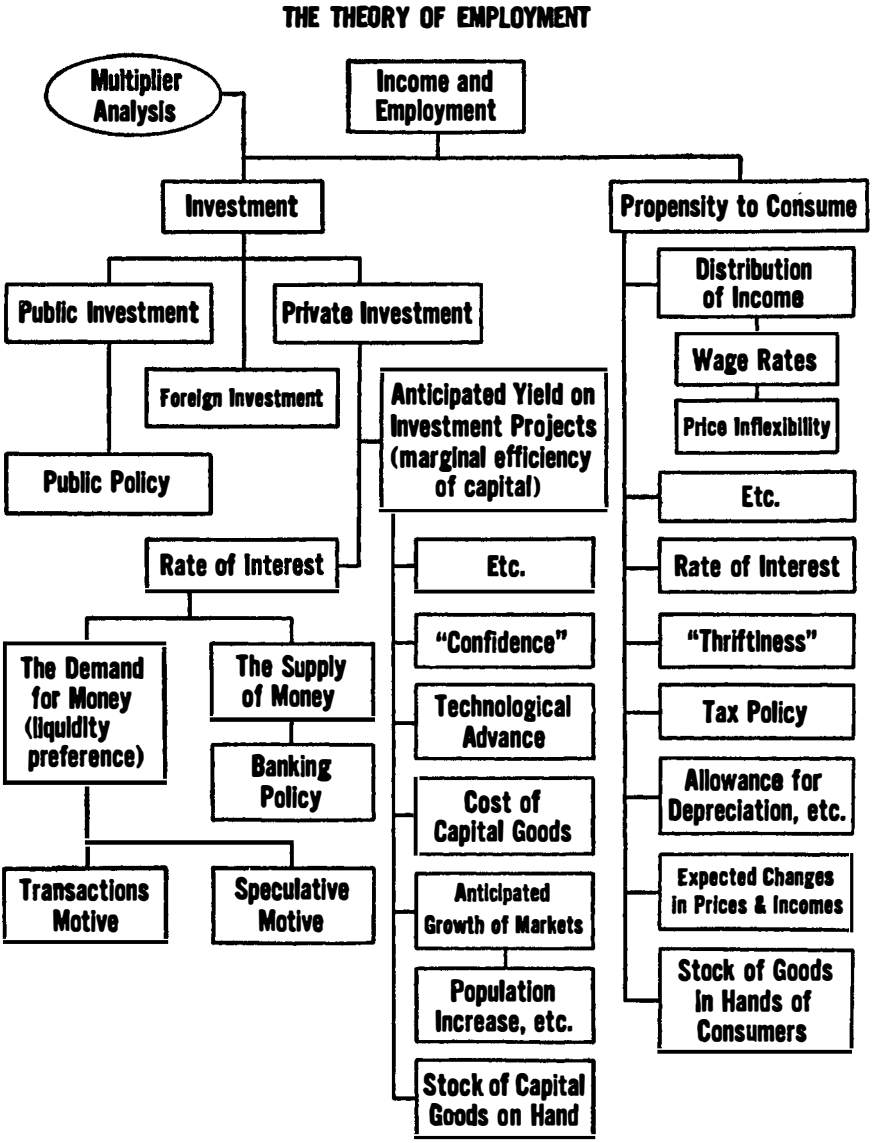

Figure 62. *The Theory of Employment*

But certain points must be kept in mind in going over such an outline. We must be particularly careful not to think of this brief outline as giving a *complete* picture of the forces that determine employment and the national income. For there are many other factors that influence the propensity to consume and the marginal efficiency of capital besides those listed in Figure 62 or discussed explicitly in the appropriate chapters. It is even more important to see that the individual determinants, far from being independent of one another as the outline suggests, are in reality interdependent. For instance, the diagram may give the impression that the rate of interest is not in any way related to the rate of technological advance. But we have already seen that the level of income and the volume of transactions play an important part in determining the interest rate, and the level of income partly depends in its turn upon the rate at which new methods of production are introduced. Thus the rate of interest does to some extent depend on the rate of invention. And this relation is also reversible, for the interest rate is a factor in determining how quickly inventions are put to use and how eagerly new processes are sought. Indeed, it is difficult to discover any two determinants that are not in some way connected with one another. The stock of capital goods on hand is certainly related to the rate of growth of demand. Expected changes in price are not independent of tax policy, and so on. Instead of drawing the outline in the form of a tree upside down, it would be more accurate to draw it as a spider web, with many horizontal linkages in addition to the vertical ones that we have sketched. Unfortunately, any diagram which tried to indicate all the possible interrelations would have to be so crisscrossed with lines that it would be almost impossible to follow. The simple tree diagram, though somewhat inaccurate, is obviously much more convenient to use. But the reader who uses it should be careful to treat each variable as though it were connected with many others. Whether the connection is important or not can only be judged by reference to the problem in hand.

32

Prosperity and Depression

OUR OUTLINE OF THE THEORY OF EMPLOYMENT is now complete. The skeleton of the analysis should be clear: that income depends upon spending; that spending is determined by investment and the propensity to consume; that investment depends upon the marginal efficiency of capital and the rate of interest; and so on. But much has yet to be done to give body and substance to this framework. We shall have to see more clearly the relative importance of the various factors; we shall have to determine which variables matter most and which least; and we shall have to know more about their interrelations. Yet it is not enough to seek a fuller understanding of the theory itself: we should also see clearly its implications for our own economy. What sort of policies should we pursue to prevent serious unemployment? Should we encourage or discourage invention? Should wages be raised or not? Are the economic consequences of government control over prices harmful or helpful? On these and other problems considerable light is thrown by the theory of employment, and the discussion of them in its turn throws considerable light upon the theory.

In the next few chapters we shall attempt to add the necessary flesh to the skeleton of the analysis which has so far been presented. This can best be done by applying the theory to some important economic problems. In the course of this exposition it will be necessary to present a good deal of statistical material bearing on our economy and, occasionally, on others also. While statistics by themselves interest almost no one, it is hoped that the reader will gain a

the cycle does not closely fit any actual cycle. For one characteristic common to all cycles is that each one is a little different from all the others. Another feature which they all share is that investment is high during prosperity and low during depression. This accords with our observation that the propensity to consume [1] is relatively constant, and for this reason we shall not devote much attention to it in the present discussion. Changes in investment, however, are of decisive importance in bringing about depression and prosperity, and consequently we shall direct most of the present analysis to an explanation of such changes.

The Economy in Depression

Let us begin with a description of the economy during a depression. At such a time investment is very low. Business firms have reduced their orders for new equipment, buildings, and goods for inventory almost to zero. Foreign demand for goods is likely to be low, too. And the government, although it may be purchasing more than usual, has evidently not increased its purchases by enough to offset the decline in private and foreign investment; for otherwise there would be no depression. Because orders are low, firms do not find it worth while to maintain equipment, machinery, and factories, and the productive capacity of the country is slowly falling. Inventories in the hands of retail, wholesale, and manufacturing firms are being depleted, perhaps at a rapid rate. Interest rates are likely to be low, partly because banks find few who wish to borrow, and partly because the Federal Reserve Banks have probably taken positive steps to reduce them still further. If the depression has been a long one, the cost of investment goods is probably low in consequence of the reduced demand for such goods. These are the favorable factors in the situation: the reduced stock of capital goods on hand, the low rates of interest, and the low cost of new capital goods.

But there are many unfavorable factors. Prevailing attitudes are likely to be very pessimistic. While newspaper editors will probably have managed to persuade themselves that all is really for the best in this best of all possible worlds, they may have been less successful in convincing the ordinary businessmen whose factories are operating far below capacity. At the same time, prices on the stock market are probably very low. Moreover, social disorders are most likely to

[1] In the sense of the schedule that shows the relation between income and consumption.

occur when unemployment is very high. The government may be compelled in deep depression to adopt policies that seem harmful or ruinous to the fortunately situated members of society. On the whole, it is hardly a time at which we should expect investment to revive.

The Revival

Eventually, however, revival begins. When the decline in sales comes to an end or even slows down, it becomes profitable to put a stop to the process of depleting inventories and to the running-down of other kinds of capital goods. Orders will have to be increased in order to get the goods needed to maintain inventories, and to keep machinery and equipment from falling further into disrepair. Once the rate of decline in sales begins to slow down, there is a rise in anticipated profit yields from new investment projects. Hence orders for investment goods go up, and activity in the investment goods industries expands. This itself brings about an increase in consumption, so that the situation begins to look more hopeful. Further expansions in investment are now profitable, and the effect becomes, in a sense, cumulative. Naturally the process could also be initiated or bolstered by any external change that brought about an improvement in markets — such as a government demand for munitions, an important improvement in methods of production, or the development of a new product.

It is important to note that no matter how the process is initiated it is to a certain extent cumulative. The initial increase in investment brings about an increase in consumption which encourages a further increase in investment. And so the recovery continues, its course conditioned by a combination of the multiplier and acceleration effects. In the early stages of the revival, interest rates continue to be low. Although there is an increasing demand for loans, the banks probably have large excess reserves and thus are not forced to restrict their lending or to raise interest rates, nor do they find it profitable to do so. Furthermore, as people become more optimistic, and the risks of default on loans appear to grow less, the price of industrial bonds may actually rise. Such a rise in price amounts to a reduction in the rate of interest at which corporations are able to borrow. The cost of investment goods also continues for some time at a relatively low level. During this period, therefore, there are few obstacles to continued expansion.

Prosperity and the Conditions That Bring It to an End

But as the expansion continues and revival grows into prosperity, the difficulties increase in number. Interest rates gradually rise. With the growing demand for investment goods, their price also begins to increase. Moreover, as a result of the high level of investment, the stock of capital equipment on hand gradually accumulates. This most important factor, as we have seen earlier, eventually acts as an effective restraint upon further investment. For some time, however, even though capital goods are accumulating and industrial capacity is growing, there will be no sign of saturation in the investment market, since the demand for all kinds of commodities is high enough to support profitably a great volume of equipment. But eventually the growth in capital equipment catches up with the demand for the products of that equipment. When that time comes, when the market for investment goods becomes saturated, the continuance of investment at a high level becomes more and more doubtful. The mathematically inclined reader may even demonstrate that, once the rate of increase of consumption begins to decline — that is, once consumption begins to level off — investment is bound to fall, unless new, favorable developments intervene.

As investment markets become saturated, and as interest rates and the cost of investment goods rise, it is no wonder that investment begins to decline. Indeed, there is no reason to suppose that investment has to reach an extremely high level before such a decline begins. As we shall find in the third part of this chapter, investment in 1937 was checked before it had reached a level high enough to provide us with anything like full employment. But generally it takes several years of high investment before the saturation point is reached. For instance, investment in housing continued at a high level from 1923 to 1928, though this was rather a long investment boom. With inventory accumulation, a high level of investment rarely lasts longer than two or three years. In fact, we find that if inventories are accumulated at a rate that exceeds $2 or $3 billion a year, they grow within a relatively short time to a point at which it no longer appears profitable to add to them further. A high level of investment in other kinds of capital goods can last two to five or even ten years, depending upon the stock on hand at the beginning of the boom, on how demand is changing, and on other factors. But in a modern economy, if prosperity based on private investment is to continue much longer

than a decade, it will probably have to depend on the rapid development of investment outlets which exploit new products and new methods of production. For in a modern capitalist economy, so much investment is needed each year to support prosperity that when prosperity exists, the stock of capital goods accumulates with exceptional rapidity. And when capital goods pile up very quickly, a point is soon reached at which further rapid accumulation or expansion becomes unprofitable in the absence of frequent important inventions.

The Beginning of Depression

When it no longer appears profitable to add to the stock of capital goods and inventories at so rapid a rate, investment declines. With this decline in investment, unemployment becomes worse and consumption falls off, so that investment, which had been geared to the high and rising level of consumption characteristic of prosperity, thus loses another support and declines further. Moreover, at this point interest rates probably rise. The upshot of all this is that the bubble of prosperity has been pricked, and optimism gives way quickly to uncertainty and pessimism. Lenders more carefully scrutinize the security of borrowers. There is probably a good deal of panic selling of bonds and stocks, their prices fall, and as a result interest rates rise further. Banks may lose confidence and begin to call in their loans. Businessmen, seeing sales falling, interest rates rising, and prices on the stock market dropping swiftly, are likely to curtail their investment expenditures very quickly. Hence all the forces of nature seem to conspire to discourage investment. And the depression deepens.

The Course of the Depression

The forces of depression, like those of revival, are cumulative. Lowered investment reduces consumption, and this leads to a further drop in investment. The amount of capital equipment on hand, which at the top of the boom seemed to be reasonably well adjusted to the size of the market, now seems entirely too high. Consequently, investment falls to a very low figure as business firms hasten to reduce excessive inventories and cancel plans for expansion of plant and equipment, unless the decline in private investment is offset by a large increase in foreign or government investment. The economy now experiences a deep depression. But as it deepens, interest rates and the cost of investment goods begin to decrease; and gradually, as

the decline in sales slows down, and as surplus plant, equipment, and inventories are disposed of, conditions become more favorable for revival. However, unless something happens from outside, like the development of an important new product, or active intervention by the government, it is generally a long time before the conditions arise that bring about a new revival.

Thus there is something natural and almost inescapable about the cycle of depression and prosperity. The movement from one to the other is cumulative, though each germinates the very forces that later create the other. But the fact that in this sense the cycle is "natural" should not lead us to conclude that nothing should be done to check it. It is "natural" for human beings to lose their teeth, but that has not kept dentistry from developing. It is "natural" for them to remain glued to the earth, but that has not kept engineers from building aeroplanes. And it is "natural" for perhaps 90 per cent of the members of an unvaccinated community to be susceptible to smallpox, but that does not imply that nothing should be done to prevent such a disaster. Our very control over nature, of which we are so proud, implies that we do not surrender to "natural" forces. Yet in economic and social affairs, there are many who still feel that "tinkering or meddling with natural laws," as we phrase it, is undesirable, and that trying to prevent recurrent depressions is little better than sacrilege.

PRICES DURING THE BUSINESS CYCLE

The Supply Curve: Output and Price

Before we turn our attention to the cycles of prosperity and depression experienced by the American economy in the two decades before World War II, let us examine the price changes to be expected during the model cycle. Prices are set by firms, and change when firms change them. We saw in Part Two that the price charged for any commodity depends upon the elasticity of the average revenue function and the marginal cost of producing the equilibrium output of that product. Generalizing from this, we constructed a supply curve for the industry, and an aggregate supply curve for the whole economy. The industry supply curve, illustrated in Figure 63, shows how much the industry will produce at each price level. Reflecting the behavior of marginal costs described in Chapter 7, the industry supply

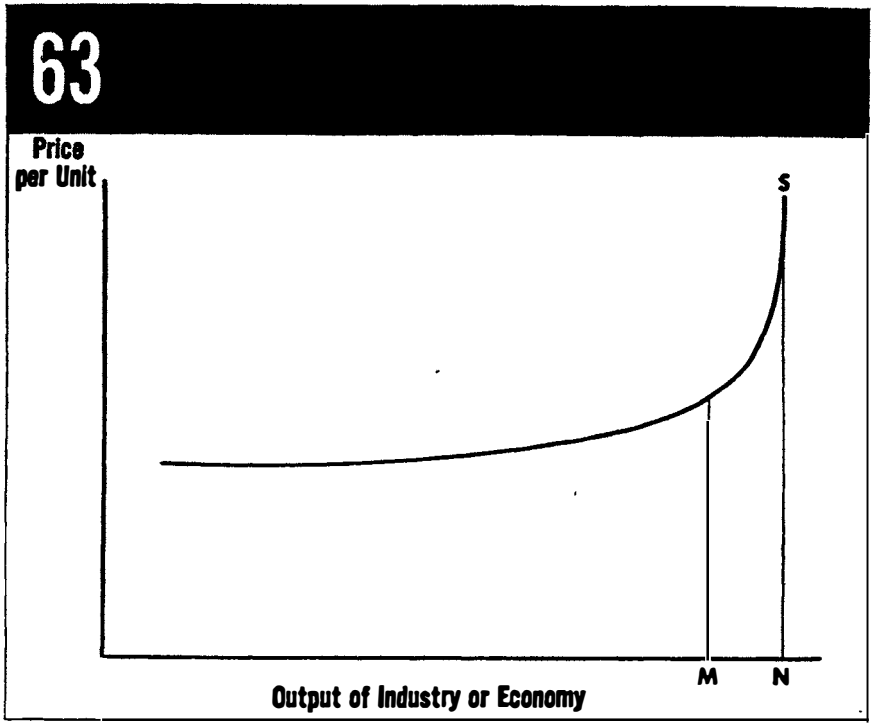

Figure 63. *The Aggregate Supply Function*

curve indicates that only a small increase in price is necessary as output increases from a very low level, but that a greater price rise will be associated with an increase in output when it is already high. This is shown by the fact that the slope of the supply curve is steeper for outputs in excess of M than for smaller outputs. Finally, once output has reached its maximum, no further expansion can be secured in the short period no matter how high a price is offered. At this point, where output is N, the supply curve becomes vertical.

A supply function, developed from the supply curve for a single industry, can be plotted to illustrate the relation between the general price level and the output of the whole economy. This we called the *supply function for the economy.* The curve of this function is shaped like the industry supply curve, since exactly the same forces determine its form. It shows that when total production or the national income is relatively low, very slight changes in price are associated with changes

in output. But when total output or the national income is very high, much larger changes in price will accompany changes in output. Obviously, therefore, it is important to bear in mind the nature of the supply function in accounting for price changes during the business cycle.

Changes in the Supply Curve: Wage Changes and Price

Prices do not depend only upon the level of output. Since changes in wage rate or in the prices of other factors of production also affect marginal costs, prices can be expected to vary with such changes. We saw earlier that prices would rise or fall if wage rates did, even when total output remained constant. We also saw that improvements in the technique of production or expansion in the size of firms would reduce the cost of production and thus lead to a reduction in prices. Whereas the effect on price of a change in output can be determined by referring to the appropriate supply curve, the effect of a change in wage rates or an improvement in methods of production

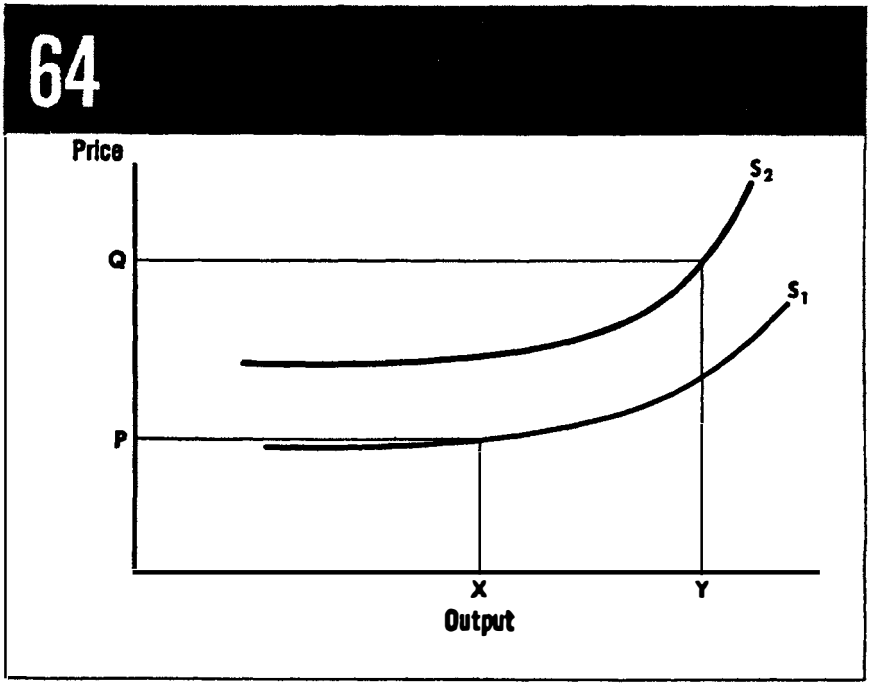

Figure 64. *The General Price Level and Changes in Supply and Output*

can only be seen by drawing a new supply curve. For as we pointed out in Chapter 19, a supply curve is appropriate to only one set of circumstances, in technology, size of firms, and wage rates, so that a change in any of these cost determinants necessitates a new supply function, as illustrated in Figure 64. Here the appropriate supply curve is S_1 when the wage rate is W_1, and S_2 when the wage rate is W_2. If the wage rate goes up from W_1 to W_2 as total output rises from X to Y, the general price level will increase from P to Q.

In any real situation, what happens to prices depends upon a complex of forces of which the change in total output is only one, although sometimes the most important one. If wage rates increase during the upswing, and at the same time there are only moderately important improvements in the technique of production, we should expect prices to rise. For the effect on price of the increases in output and wages should more than offset the effects of the moderate improvements in methods of production, though these effects would of course prevent the total increase in price from being as great as it would be otherwise. If, during the downswing, wage rates are reduced and technical improvements continue to be made, these factors combine with the decline in production to reduce prices.

Prices During the Revival

In the early stages of revival, none of the forces making for an increase in price operate with any great strength. In the first place, since production is quite low, the rise in output scarcely affects prices. Secondly, wages generally rise rather slowly at this stage.[1] Furthermore, improvements in technique continue to be made and exert a downward pressure on prices. The total effect to be expected is, therefore, that prices rise only very gradually in the early phase of the revival.

But as the revival advances into prosperity and eventually into boom, the forces making for price rises grow in strength. Increases in output are associated with larger and larger rises in price as output approaches capacity. Wages begin to increase much more rapidly once unemployment has been reduced to a low figure. The adoption

[1] Here are a few figures to illustrate the complexity of the situation. Between June, 1933, and June, 1934, wage rates rose from about 45 cents an hour to 56 cents an hour, or by almost 25 per cent. But this was very early in the revival. And yet between 1922 and 1929, years of revival leading to very high prosperity, wages in manufacturing increased only from 49 cents an hour to 59 cents.

of improved methods of production, of course, continues, but can scarcely be expected to offset the forces making for higher prices. Thus, as the revival continues, prices rise with increasing speed. If a very high level of output is reached before depression occurs again, prices will almost certainly be far higher than at the bottom of the preceding depression. This would occur because with a very high output, and higher wage rates, firms would find it profitable to raise their prices. It would only be prevented if firms were not allowed to do what seemed most profitable to them, or if marginal costs at the peak of prosperity were not much higher than in depression — perhaps because wage rises had been checked, or important improvements in production methods had been devised.

Prices During the Depression

In the early phase of the downswing, prices will probably fall. Wages will not be reduced rapidly, though sizable wage cuts can be expected later as the depression deepens. But the cut in production will be associated with fairly large reductions in marginal costs and thus in price. As the depression is intensified, wage reductions become more frequent, and the decline in price therefore continues. Prices should reach their lowest point at the bottom of the depression, since output is then lowest, and wages are likely to be near the minimum. In conclusion, it is reasonable to suppose that prices would rise very slowly during the early stages of recovery, and faster as recovery advanced to full prosperity. After the collapse of the boom, prices would begin to fall again, and would continue to do so until depression once more gave way to revival.

Note that in this account of price changes we made no use of certain well-known concepts, such as the amount of money in circulation, its velocity of circulation, the amount of backing the currency had, or whether the government budget was balanced. This does not mean that these factors play no part in determining prices. Some of them undoubtedly affect prices, but they must do so by producing a change either in the marginal cost of the equilibrium output [1] or in the elasticity of demand. For otherwise prices would not vary. Hence when we consider the influence of any of these factors, we should do so either as they affect output (by changing investment or the propensity to consume), or as they bring about a change in wage

[1] By changing the marginal cost curve or by changing the equilibrium output.

rates or other factor costs. For example, a rise in the amount of money in circulation is generally regarded as a cause of price increases. We can see that this may be so, because (see Chapter 25) such an increase generally reduces the rate of interest; and the lower interest rate encourages investment and may raise the propensity to consume. Hence output will be increased. The marginal cost at the higher output is likely to be higher than at the original output, especially if that output is reasonably close to capacity. And wage rates may go up if the rise in employment adds enough strength to labor's demands. Hence it can safely be said that prices may increase when the amount of money in circulation is raised. But notice that there are many links in the chain of events, any one of which may not apply in an actual situation. If, for instance, the interest rate does not fall because of the increased supply of money, the process ceases. If neither investment nor the propensity to consume rises because of the decline in interest rates, the process stops at that point. Or if output is low and unemployment is high to begin with, and if consequently there is no increase in the marginal cost as output rises, prices will not rise. Finally, even though marginal costs increase, if the government prevents firms from raising their prices, in a sense imposing on them a more elastic demand function, prices will remain steady. In short, we can adequately account for changes in the general price level by the use of the theory of the firm and the theory of employment which we have already worked out. There are no short cuts that are likely to be helpful or accurate.

BUSINESS CYCLES IN THE UNITED STATES, 1920–1941

Between 1920 and 1941 there occurred in the United States two periods of prosperity when output was almost at capacity, and one period of depression, which happened to be the deepest in our history. Moreover, several minor depressions interrupted the revivals between 1920 and 1941. The investigator of business cycles can find in these twenty-one years examples of almost every kind of economic condition. When the period opened in 1920, the economy was enjoying very high prosperity. In 1921 there was a sudden and sharp decline in activity, but this depression was short-lived. From 1922 to 1929, there was a prosperity plateau from which, according to the happy beliefs of the period, we were never to fall. But we did. The turning

point came in late 1929, and for the next three or four years the economy experienced a depression that swiftly grew in strength until by late 1932 and early 1933 we were employing no more men than had been at work in the depression year of 1921, although our labor force had grown by about 8 million men. Then came a very sharp revival, which lasted until the final quarter of 1933, followed by a sharp, short relapse, and then a rather steady rise until 1936. For the next eighteen months, that is until the middle of 1937, the pace of revival quickened. By that date employment was almost as high as in 1929, but the labor force had grown by about 5 million. Then came a very sharp depression, but by late 1938 another revival was under way which continued without interruption through 1941. Figure 65 records the course of the gross national product and of non-agricultural employment in the United States from 1928 to 1941, and illustrates the historical outline presented below.

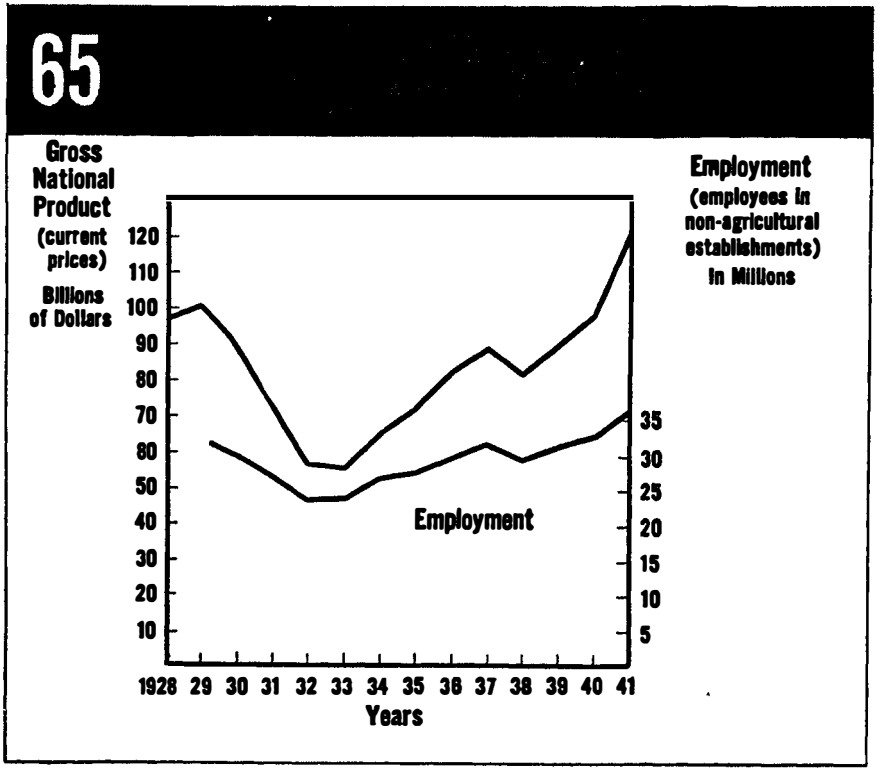

Figure 65. *Output and Employment: 1929–1941*

Since, as we have seen, changes in investment are of decisive importance in accounting for these variations in income and employment, the following pages are devoted mainly to an analysis of the variations in investment.

The Period from 1920 to 1921

The goose hung high in 1920. Business had picked up after a slump following the armistice of 1918, and production had advanced substantially. The high level of income and employment depended of course upon high investment. The data shown in Table 51 throw a good deal of light on developments in these years.

TABLE 51

Investment Level, 1919–1921

	(in billions of dollars)		
	1919	1920	1921
Gross national product	77.5	86.6	70.7
Private investment	14.1	16.3	7.8
in plant and equipment	8.3	9.2	5.6
in housing	2.3	2.2	2.2
in inventories	3.6	4.9	0.1
Public investment *	8.6	8.0	8.4
Foreign investment †	3.1	2.2	1.4

* It is desirable to state again that public investment is here measured as total government expenditures on goods and services.

† The data for foreign investment do not measure the concept we found most convenient to use for theoretical analysis. Instead they measure net exports (that is, exports minus imports) and monetary use of gold and silver. However, if we had used data for exports, as demanded by our analysis, we should have had to revise the measure for consumption too. We thought, therefore, that it was preferable to use the government statistics and to warn the reader that foreign investment as here measured differs from foreign investment in our sense. Since the absolute figure for foreign investment is small no matter how it is measured, it makes no great difference in any case.

In 1920, investment in plant and equipment by manufacturing firms stood at over $3 billion, higher even than in 1929 in dollar value, though not in physical terms. Much of this investment consisted of a very rapid accumulation of goods for inventory. For the whole of the economy, manufacturing, trade, and so on, investment in inventory was estimated at almost $5 billion, the highest in our history

before 1946.[1] This latter type of investment, for reasons that should be obvious, does not long continue at a very high level. A year or two generally seems to be time enough in which to pile up all the inventory that is wanted when the accumulating is done so rapidly. After this, as we noted earlier, investment in inventory falls off, and unless investment of some other kind rises to offset this decline, the economy is plunged into sharp depression. This is what happened after 1920. In 1921 investment in inventory fell to less than $100 million, while investment in plant and equipment by manufacturers fell by over 55 per cent to about $1.4 billion. The decline in inventory investment is scarcely surprising in view of the fact that inventories had grown by $8.5 billion in 1919 and 1920 alone. After so rapid a growth business firms did not find it profitable to expand inventories further, and accordingly investment in 1921 was negligible. With a decline in private investment of about $8 billion, the gross national product was reduced from about $86.6 billion in 1920 to about $70 billion in 1921. Within a year, unemployment increased from about 500,000 to roughly 5 million men. The crash in 1921 can obviously be regarded as an inventory depression — though, unfortunately, suitable statistics are not available for a more detailed analysis.

The Period from 1922 to 1924

The data in Table 52 offer a broad summary of the events of the period, 1922 to 1924:

TABLE 52

Investment Level, 1922–1924

	(all in billions of dollars) 1922	1923	1924
Gross national product	72.7	84.5	83.7
Private investment	9.8	15.3	12.6
In plant * and equipment	5.7	8.4	8.0
In housing	3.2	4.2	4.8
In inventories	0.8	2.8	−0.2
Public investment	8.5	8.7	9.1
Foreign investment	0.5	0.3	0.7

* Including other business construction: for example, office buildings, rail lines, and so on.

[1] In one month of 1946, July, investment in inventory was made at the annual rate of $16.8 billion. This elephantine figure, which represents hoarding on a scale almost beyond belief, suggested trouble ahead.

Revival began in 1922. The most favorable factors were increased investment in housing and inventories; though at the same time foreign investment declined rather strongly and most other items showed little change. In 1923, the pace of the revival quickened. Investment in housing and inventories was much higher than in the previous year, and investment by business firms in plant and equipment rose to a level 50 per cent higher than that of 1921. A mild inventory depression set in in 1924, with investment in inventories in that year about $3 billion below the figure of the previous year. The decline in inventory investment is not surprising in view of the fact that inventories were $3.6 billion higher by the end of 1923 than at the beginning of the previous year. Business firms apparently anticipated no profit from adding another $2.8 billion worth of goods to their stocks in 1924; instead they reduced their inventories by a small figure. Incidentally, although sales in 1923 and 1924 were slightly lower than in 1920, inventories were $3.7 billion, or about 12 per cent higher. To repeat: it is not surprising that investment in inventories declined in 1924. It had done so after 1920, and in one sense the situation in 1924 was even less favorable to continued expansion in this category. Most other types of investment changed relatively little in 1924 as compared with 1923, except that in housing, which continued to climb and in that year accounted for 38 per cent of all private investment.

The Prosperity Plateau: 1925 to 1929

Inspection of the figures shown in Table 53 for the years from 1925 to 1929 shows that this was a period of relatively stable and high income. Employment was good, and it must have looked at the time as though depression had been banished forever. The gross national product was higher throughout the period than it had ever been before. But we, who are equipped with the powerful faculty of hindsight, can detect elements of instability in the situation — things which became increasingly important as the end of the period approached. Table 53 will picture the essentials of this period and suggest what these weaknesses were.

The gross national product of 1925 was well above that of 1924. Disinvestment in inventory had come to an end. Except for a small decline in the foreign category, all types of investment increased, and that in housing reached the highest level it has ever attained — over

TABLE 53

Investment Level, 1925–1929

	(in billions of dollars)				
	1925	1926	1927	1928	1929
Gross national product	90.4	95.6	94.2	95.9	99.4
Private investment	15.3	16.8	15.2	14.9	17.6
in plant and equipment	8.9	10.0	9.7	10.1	11.5
in housing	5.1	5.1	4.8	4.4	3.7
in inventories	1.0	1.6	0.2	−0.4	1.6
Public investment	9.6	9.8	10.6	10.6	11.0
Foreign investment	0.3	0.1	0.5	0.7	0.4

$5 billion — or the equivalent of about 1 million new houses in a single year. The situation was even more favorable in 1926. Investment in housing remained at the peak figure; investment in inventories climbed, though again too rapidly, as could be seen in the following year; and business firms engaged in even more projects for expanding plant and installing equipment. Housing still accounted for about 30 per cent of all private investment, and the construction of plant and equipment accounted for about 60 per cent. It is notable, too, that government investment had been well maintained throughout the period, and that by 1926 it accounted for about 37 per cent of all investment. One feature of the slight decline of 1927, the drop in inventory investment, is not surprising, for inventories had again piled up quickly in 1925 and 1926. It is worth noting that, except in 1941, whenever investment in inventories in successive years exceeded $2.5 billion, it fell substantially in the next year. The decline in total investment in 1927 as compared with 1926 was $1.6 billion; the decline in investment in inventories was $1.4 billion. In the same year, investment in housing fell, and for the first time in the whole period it was lower than in the previous year.

The situation was not very different in 1928 from that in 1927. There was a further slight decline in inventory investment. Both government investment and expenditure on plant and equipment increased. More ominous was the decline in housing activity. Although the gross national product was about $16 billion higher than in 1924, and the average family therefore had more money, housing investment in 1928 was only slightly higher than in 1924. But in other ways the figures did not yet reveal anything dangerous.

In 1929 the gross national product reached its peak, almost $100 billion. This was sustained by the highest level of investment ever reached, and in almost every category there was an increase over the previous year. The rise was especially marked with inventory investment, which stood $2 billion higher than in 1928. Government investment was higher in 1929 than the year before, and so was the expenditure of almost every industry on plant and equipment. For instance, total capital expenditures for plant and equipment in the automobile and automobile equipment industry stood at $149 million in 1928, and at $186 million in 1929. Plant and equipment investment by public utilities was $1.6 billion in 1928, and $1.9 billion in 1929. Only investment in housing stood out against this tendency, and by 1929 it had fallen by over 25 per cent from its peak. In that year only 509,000 new non-farm dwellings were started, whereas, in 1925 the figure, as noted above, was just about 1 million. The decline in housing investment, which occurred in spite of the fact that the national income was higher than ever, and the rather high level of investment in inventories, were perhaps the two most ominous developments. Unless either the decline in housing investment could be halted, or investment in some other categories could be substantially increased, income was bound to fall. And with the fall in income, investment in plant and equipment in most industries could be expected to decline. The long-term situation was thus becoming unstable. Moreover, the short-term situation also was unstable, because by that date inventories had increased rather quickly. In fact, they had gone up $1.6 billion in 1929 alone. By the end of 1929, inventories were roughly $3.0 billion higher, or about 15 per cent above the level of 1920, while the gross national product was $13 billion higher, which was also a 15 per cent increase. But as the events of 1921 demonstrated, the inventory situation in 1920 was unstable; and we may suppose that it was no less so in 1929. We shall later examine some of these series in more detail.

Depression: 1930 to 1933

The prosperity of the nineteen-twenties came to a sudden end. To take just one example, quarterly data on investment in plant and equipment by manufacturing concerns show that this item reached its peak in the second quarter of 1929 and remained very close to that level until the end of the year. But it was 16 per cent below the

peak in the first quarter of 1930, 30 per cent below in the second quarter, 39 per cent below in the third, and 48 per cent below in the fourth. And the decline continued steadily throughout 1931 and 1932. By the fourth quarter of the latter year, investment of this type had fallen to less than 17 per cent of its 1929 peak level. In fact, it was so low that only about 30 per cent of the depreciation of manufacturing plant and equipment was being made good, so that plant and equipment were wearing out faster than they were being replaced. More complete data for the whole period are shown below:

TABLE 54

Investment Level, 1930–1933

	(in billions of dollars)			
	1930	1931	1932	1933
Gross national product	88.2	72.1	55.4	54.8
Private investment	11.3	6.0	1.9	2.7
In plant and equipment	9.5	6.3	3.4	2.8
In housing	1.9	1.6	0.7	0.5
In inventories	−0.3	−2.0	−2.3	−0.7
Public investment	11.2	11.5	10.2	9.1
Foreign investment	0.8	0.4	0.3	0.6

The gross national product was valued at almost $100 billion in 1929; its value was just over half that figure in 1932 and 1933. The decline in investment was enormous. It went from about $28.6 billion in 1929 to $12.4 billion at the low point of the depression. In every single category, investment was far below the 1929 level. For instance, investment in steel mills, blast furnaces, and rolling mills fell from $150 million in 1929 to $40 million in 1932. Investment in plant and equipment in the automobile and automobile equipment industries declined from $186 million in 1929 to only $48 million in 1933. Public utilities reduced their investment from $1.9 billion in 1929 to about $405 million in 1933. The railroads reduced theirs even further, from $840 million to $101 million. Even government investment declined from $11 billion in 1929 to $9.1 billion in 1933, presumably because of the effort to balance the budget.

Between 1930 and 1932, investment in plant and equipment by manufacturing concerns fell short of depreciation charges by approximately $2 billion. Plant and equipment simply wore out without

being replaced. Inventories of business firms likewise declined — in physical terms, standing in 1932 at roughly 80 per cent of the 1929 figure. The stock of housing also fell. In short, we actually had less capital goods in existence by the end of 1932 than at the end of 1929, and our capacity to produce in the latter year was in many industries below the capacity at the beginning of the depression.

Between 1929 and 1932, income fell steadily by about 15 to 20 per cent a year, but between 1932 and 1933 the decline was practically halted. One cause of this was increased investment — or perhaps more accurately, reduced disinvestment — in inventory. During 1932 inventories were cut by about $2.3 billion. During 1933, they were reduced by only $0.7 billion more, which was equivalent in its effect upon income to a rise in investment of $1.6 billion ($2.3 billion — $0.7 billion). But all other categories of investment were slightly lower in 1933 than in 1932. The reduction in inventory disinvestment was to be expected, for just as a few years of high investment bring about an accumulation of inventories so great that further rapid additions are not considered profitable, so several years of rapid disinvestment deplete inventories to the point where further rapid liquidation becomes unprofitable. When business firms finally decide that inventories are low enough and should not shrink further, inventory investment rises from perhaps −$2 billion to zero. But, as we have seen, this is equivalent to a $2 billion increase in investment. Hence by 1933 the stage was set for a revival in at least a few types of investment.

Revival: 1934 to 1937

The revival of 1934 to 1937 was nearly as spectacular as the decline of 1929 to 1933. By the end of the later period the gross national product was not far below what it had been in 1929; in physical terms it was actually higher. Investment rose from $12.4 billion at the lowest point to $25.2 billion. But the pattern of the revival was quite different from that of the nineteen-twenties. Let us examine some of the basic data (Table 55).

The general increase in investment is noteworthy. In 1937 expenditures in certain industries for new plant and equipment, particularly for the latter, were actually higher than in 1929. For example, they were $316 million in 1937 compared with $150 million in 1929 in the steel industry, and $144 million in 1937 compared

TABLE 55

Investment Level, 1934–1937

	(in billions of dollars)			
	1934	1935	1936	1937
Gross national product	63.8	70.8	81.7	87.7
Private investment	4.6	6.3	10.3	11.1
In plant and equipment	3.9	4.9	6.5	7.9
In housing	0.7	1.0	1.5	1.9
In inventories	−0.1	0.2	2.2	1.1
Public investment	10.8	11.9	12.6	13.6
Foreign investment	0.7	0.4	−0.3	0.5

with $125 million in 1929 in the petroleum refining industry. In other categories, investment in 1937 was far lower than in 1929. Investment in housing reached only about 50 per cent, in public utilities about 60 per cent, and in the stone, clay, and glass products industries only 41 per cent of the 1929 level. Undoubtedly investment in the two industries last named was low because of the much smaller amount of investment in housing and in new plants. Private investment in the construction of both industrial plant and private housing made up 23 per cent of the total in 1929, but in 1936 it accounted for only 12 per cent, and in 1937 only 14 per cent. In contrast, inventory investment in 1936 was 10 per cent of the total, while in 1929 it accounted for only 6 per cent of all investment. In 1937 though most kinds of investment were higher than in 1936, inventory investment was lower. However, the annual data obscure a most interesting development during that year. For the first part of the year, investment in inventories was very high, but it fell sharply, and undoubtedly was negative in the last months of the year.[1] By the end of 1937, the gross national product and employment were much below the 1936 level. Evidently, a high rate of investment in inventories in 1936 and early 1937 had again built up stocks to the point where further expenditures were not considered profitable; and when this point was reached, total investment once more declined, and with it, the national income.

There is also some evidence, although it is not striking, that the

[1] Inventories of department stores rose by 10 per cent between January and June, 1936, and by 6 per cent between July and December, 1936. They rose by 15 per cent between January and June, 1937, and fell by 9 per cent between July and December, 1937.

propensity to consume shifted downwards in 1937. The gross national product was $6 billion higher than in 1936, but consumption increased by only $3.4 billion. An increase of about $4 billion would have been expected with a marginal propensity to consume of two-thirds, the average value for the whole period. This deficiency in consumption may have been caused by the introduction of the Social Security Program, under which taxes yielding about $1.5 billion a year were first collected in January, 1937.

In short, the revival of 1934 to 1937 differed from that of the nineteen-twenties principally in that during the later upswing, housing and construction of plant accounted for a relatively smaller part of the total investment, whereas inventory and government investment played a more important role.

Recovery: 1938 to 1941

The "recession" began in 1937 and carried through 1938. Then the economy experienced a recovery more spectacular than the prosperity of the twenties. The basic statistics are set out in Table 56.

TABLE 56

Investment Level, 1938–1941

	(in billions of dollars)			
	1938	1939	1940	1941
Gross national product	80.6	88.6	97.1	120.5
Private investment	6.5	9.9	13.0	17.7
in plant and equipment	5.8	7.1	8.9	11.4
in housing	1.9	2.0	2.4	2.8
in inventories	−1.3	0.9	1.8	3.5
Public investment	14.4	16.0	16.7	26.5
Foreign investment	1.2	1.0	1.8	1.7

Again in 1938 the chief difficulty was the necessity for reducing inventories, with the result that investment in this category once more became negative. Other declines in investment were more moderate. Recovery from the 1938 recession began in 1939, when the gross national product reached the 1937 level. In 1939, business firms again found it profitable to expand inventories, as well as to build more plant and install more equipment. With the outbreak of war in Europe in September, 1939, the revival was encouraged further. In 1940, the gross national product reached 1929 levels in dollar terms, and since prices were lower than in 1929, the physical volume

of output in the latter year was well above that of 1929. Investment of all types rose steadily to a new peak in 1941, our last year of peace — if not of normality. The 1941 gross national product was 20 per cent above the 1940 figure, certainly the most rapid rise since 1918. Again, all types of investment rose, but the most spectacular increase was in government investment, which alone *increased* by almost as much as the total investment of 1932 or 1933. From 1941 to 1945 ours was a wartime economy. Since the problems of such a period are vastly different, we shall discuss them in another chapter.

Summary by Decades

Taking the period from 1921 to 1929 as a unit, let us analyze the importance of various types of investment. Housing accounted for almost 18 per cent of all the investment done in the period, though by 1929 the figure had fallen very far below what it had been at the peak. Investment in inventories amounted to less than 4 per cent of the total, but in some years it accounted for more than 6 per cent — for example in 1923, 1926, and 1929 — and in 1920 it accounted for more than 11 per cent. Government investment made up 40 per cent of the total. More than half of the government investment in this decade was done by local units; federal government investment made up 35 per cent of the total, and state government investment came to about 15 per cent. Other categories of investment contributed to the total as shown in Table 57.

TABLE 57

Categories of Investment as Percentages of the Total: 1921–1929

Housing	18	
Plant and equipment	21.5	
Manufacturing industries		8.5
Public utilities		6.0
Railroads		3.0
Others		4.0
Inventory	4	
Commercial building	10	
Agriculture	3	
Foreign (see footnote, p. 455)	2.5	
Government	40.0	
Federal		14
State		6
Local		20
Total	100	

Investment in housing, in plant and equipment for manufacturing industries and public utilities, in commercial building, and in federal and local projects thus accounted for more than 75 per cent of the investment expenditures of the period.[1]

The pattern was markedly different in the following decade. The decline in the importance of housing and the increased importance of government investment have already been noted. The information can be conveniently summarized as in Table 58.

TABLE 58

Categories of Investment as Percentages of the Total: 1930–1940

Housing		7
Plant and equipment		20
Manufacturing	5	
Other	15	
Inventory		—*
Commercial building		6
Agriculture		4
Foreign		3
Government		60
Total		100

* This was a small negative figure.

Summary by Type of Investment

Investment in inventory was extremely variable. Considerable expansion of inventory, and thus positive investment, occurred when sales or gross national product was increasing rapidly, as in 1922–23, 1924–26, 1935–37, and 1939–41. At other times, and especially after a period of rapid inventory accumulation, inventory investment was low or even negative, as in 1924, 1927, 1930–33, and 1938.

Examination of data which are too detailed to be presented here indicates that a considerable part of the investment in plant and equipment for the period from 1920 to 1929 was based upon the development of the automobile and electrical equipment industries. Investment in plant and equipment for the automobile and automobile equipment industries, the rubber industry, and petroleum refining amounted to over $2.5 billion between 1920 and 1929. In addi-

[1] It must be pointed out that this breakdown is very crude. Figures have been collected from a variety of sources and it was not always possible to reconcile them.

tion, a substantial part of the investment in the steel industry, which came to $1.6 billion, could be traced back to the development of the automobile. The textile industry, crude petroleum production, and the glass industry were also given an incentive to expand because of this development. Part of the indirect boost to investment in plant and equipment in the cement, lumber, plumbing supplies, and certain other industries, can also be credited to the automobile, which greatly stimulated house building and highway construction during the period. In fact, neither suburban developments nor highways would have been possible or necessary without it. Other newly developed products to which we have already alluded were of considerable importance in stimulating investment in plant and equipment in that decade. In the next decade, the development of synthetic textiles, of new processes for rolling sheet steel, and of automatic and semi-automatic controls for machinery were of great importance.

Certain other factors affected investment in plant and equipment during part of this period. Because of the requirements of war in 1917–18, capacity in a number of civilian goods industries was low in 1920. Thus the automobile industry could produce no more than 2 million automobiles in that year. At the beginning of 1919, the rated capacity for the production of steel ingot was 61 million tons—whereas by 1942 it had increased to 89 million and by 1945 to 96 million tons. The inadequacy of existing capacity was made the more emphatic because markets had been expanding in the previous decade. A crude measure of the growth can be seen in the change in the national income. It stood at less than $30 billion in 1910; at more than twice that figure in 1920. Increase in price, of course, accounted for the greater part of this change, but even allowing for this factor, there was a growth of about 15 per cent. Population had increased by about 15 million between 1910 and 1920, and by another 16 million between 1920 and 1930. Furthermore, in the first few years following the armistice of 1918, our foreign markets were larger than they had ever been. All this is significant to an understanding of the nineteen-twenties.

But it is just as important to see why investment in plant and equipment was not maintained at the 1929 level during the nineteen-thirties. The chief explanation seems to be that we built up our capacity rapidly; we built ourselves out of a market. By 1929 we could produce with existing plant and equipment more than 5.5 million automobiles a

year. By the same year our capacity in the electric power industry was more than double what it had been in 1920. And much the same thing happened in most other industries. It was not until markets again grew larger and the development of new processes made replacements profitable, that investment in plant and equipment once more achieved a high level.

About 7 million new houses were built between 1920 and 1930, exclusive of farm dwellings. Four factors appear to have been decisive in creating this demand: the small amount of building done in the previous decade; the rapid growth of population; shifts in population from the farm to the city, from south to north and from east to west; and the development of the automobile, which made possible the development of suburban areas.[1] Any short summary of the factors that affect so complex an activity will invariably be an oversimplification, but many of the forces already mentioned as instrumental in the great housing boom of the nineteen-twenties had, by inversion, some effect on the decline in house building during the nineteen-thirties. Perhaps, however, the chief factors in the decline were two. One of these was that the rate of population growth was declining, for there was during the period an increase in population of only 9 million compared to one of 16 million in the previous decade. The other was that the high rate of investment in housing during the nineteen-twenties had nearly saturated the market. In that decade a new house had been built for about one out of every four families in the country, a rate far in excess of that at which houses were falling into disrepair. The decline of investment in housing after 1926, despite the high prosperity of the next three years, is a most significant factor.

Interestingly enough, government investment was remarkably steady throughout the twenty-year period. As we have already noted, in the nineteen-twenties, local government bodies did most of the investment for roads, highways, school buildings, and municipal buildings, while in the nineteen-thirties, especially after 1933, the federal government did the bulk of the investment. The importance of government investment was much higher in the second decade, chiefly because most other kinds of investment were relatively low in that period. Until 1940, at least, the absolute growth in government investment was moderate.

[1] Interest rates, building costs, and other factors were also favorable.

General Summary

A thorough study of the reasons for changes in the individual types of investment would require the space of this entire book, at the very least. The account which has been given here must be regarded as no more than suggestive. But two points deserve particular emphasis. One of these is the interrelation between these various categories of investment and the fact that a change in one may affect another, as when the development of the automobile stimulated investment in the steel industry and encouraged local governments to pave streets, thereby increasing investment in the cement industry. The second point to be emphasized is that the reader should seek his own statistics to supplement those which have been presented here. An enormous mass of data are available, and they will all throw further light on the questions of where, as an economy, we have been, and where we are going. Only by becoming familiar with the essential facts about our economy can we hope to determine how to make it operate efficiently.

33

Full Employment and Beyond

IN THE UNITED STATES we have full employment when investment reaches $65 billion a year, with the propensity to consume generally characteristic of this economy. With full employment, our national income and gross national product are at their ceiling values; no further increase in output or employment is practicable. Consideration of this feature of full employment should raise a question of very great importance. We showed in Chapter 29 that when investment increases, the economy's output of all kinds of goods and services also increases. What happens when investment increases beyond the level at which full employment occurs? Obviously output and employment cannot expand beyond this point. What, then, are the effects of such an increase in investment?

The problem is a real one, though perhaps unfortunately not a common one. Its reality can be seen in our experience between 1943 and 1945, when investment soared far above the $65 billion figure required for full employment. In 1943 it stood at $96 billion, in 1944 at $98.1 billion, and in 1945 at $92.4 billion. But this is a rare phenomenon in our economy. In the last three or four decades we have usually had a considerable amount of unemployment. And since we have failed to make full use of the resources available, our output has been well below the peak, and therefore we have generally been able to expand it. This, of course, implies that investment, instead of being above the full-employment level, has usually been substantially below it. Our typical problem has been unemployment, indicating too little investment. And other economies have been faced with the

same difficulties. Such modern capitalist economies as Great Britain, Germany before 1945, France, Canada, and Australia have had to deal with heavy unemployment through most of the last few decades. But occasionally these countries too have been confronted with the problems that arise when investment is too high. Just as with us, this situation is especially likely to occur with them during a war.

In less advanced economies the problem of unemployment is usually much less acute. They have their problems, such as crop failures, breakdowns in transportation, and so on, and the consequences are terribly serious. But they rarely have to worry about unemployment arising from deficient investment. Instead, they are often forced to cope for long periods of time with difficulties of the kind we encounter during a war.

The economic problems that arise when investment goes beyond the level needed to secure full employment are in essence the same whether the economy is at war or not, though the methods of control may be somewhat different. During a war, every effort is made to facilitate the production of investment goods; in times of peace, an attempt is usually made to limit the increase in investment. Our own recent history gives us no examples of such problems except during the two great wars, but in the nineteenth century they did arise during periods of peace. In this chapter we shall examine both these situations.

THE ECONOMY AT WAR

Public Investment in War

It is not difficult to understand why unemployment disappears and shortages develop in time of war. War requires an immense output of all kinds of munitions — tanks, airplanes, guns, ammunition, naval and merchant vessels, and smaller items of almost infinite variety. The government must purchase vast amounts of munitions. It must feed, clothe, house, and protect the health of its armed forces. Finally, it must finance the expansion of plant and equipment for the production of the myriad goods and services required. Between 1942 and 1945, our government spent an average of $84 billion a year on new goods and services. Such spending of course constitutes government investment, and in these years it went far beyond the level normally required for full employment. If to this figure we add the amount

spent by business firms and foreigners — a comparatively small sum as we shall see — it is clear that investment is boosted well above the figure needed for full employment.

Increasing Investment with Full Employment

When investment increases beyond the full-employment level, output cannot rise, but prices can. In fact, the full effect of the increased investment is focused on prices, and unless active steps are taken to prevent it, inflation is inevitable. Let us see why. We shall assume that, to begin with, the economy is enjoying full employment, which implies that investment is at the level which insures it. The labor force is distributed between the investment and consumers' goods industries in accordance with the propensity to consume. If the propensity to consume is such that two-thirds of the full-employment income is spent upon consumers' goods, approximately two-thirds of the labor force will be at work in consumers' goods industries and the remaining third in investment goods industries. Under these circumstances, unless wages rise or business firms adopt improved methods of production, there will be no tendency for prices to change.

But if the government finds the production of investment goods to be inadequate, the situation will have to be altered. If it needs more aircraft, tanks, merchant vessels, and other commodities than it is already getting, it expands its orders for these munitions. Government investment rises with the expansion in output to meet these orders. But how can output be increased?

At this point we encounter the critical difference between an economy with unemployment and one with full employment. So long as there is some unemployment to begin with, output can be expanded by calling upon the unemployed to fill the new jobs opened by the increase in investment. But when there is no unemployment, when all those who want to work already have jobs, then an increase in the output of goods of one type can be obtained only by reducing the output of goods of other types. For the added labor force needed to build more tanks and guns can be found only by reducing the numbers employed in other industries. If we make more munitions, we must make less butter — or fewer suits, tablecloths, and other non-critical items. This difference between an under-employment and a full-employment economy cannot be too strongly emphasized. Increased government investment in an economy having unemploy-

ment increases not only the output of investment goods but also that of consumers' goods.[1] But increased government investment in an economy already enjoying full employment reduces the output of consumers' goods or possibly of other kinds of investment goods. In short, when there is unemployment, we can have both more guns and more butter; when there is full employment, we must choose between the two.

It is a palpable, physical fact that when the output of goods for the government expands beyond a certain point, the output of other goods declines. In practice, it is ordinarily the output of consumers' goods that falls by the largest amount. We must not lose sight of this fact, that consumers' goods production falls, no matter how much it may be obscured by the financial overlay. And there is no way of avoiding the difficulty; some members of the economy must reduce their physical consumption. No matter how hard people try to avoid it by paying premium prices, by thronging the stores, and by buying more than they need, some of them are bound to get less than they want. It is like a game of musical chairs. With twenty people circulating round nineteen chairs, someone will be left standing when the music stops, no matter how fast they rush and no matter how hard they push. In the same way, when the output of munitions is expanded beyond a certain figure, fewer consumers' goods become a ailable, and some consumers must do without.

Shortages do not, of course, spring into existence suddenly, at the critical moment when investment first exceeds the full-employment level. Trouble begins once investment climbs above the point where unemployment is moderate. One of the first symptoms is the increasing difficulty of finding workers for the new jobs that continually arise. There may be shortages of workers with special skills, and it may become impossible to expand the output of certain goods to meet the increasing demand. Thus local shortages may become manifest long before all unemployment has disappeared. Generally, however, so long as there is idle capacity, it is possible to expand the output of munitions and at the same time to increase the production of most kinds of consumers' goods to meet the increasing consumers' demand. No real choice has to be made at that point between guns and butter; we can still have more of both. So long as appreciable numbers of unemployed men and much idle equipment remain, there will be

[1] Unless other investment declines. See Chapter 35.

some increase in the output of consumers' goods with each increase in government investment.

Let us see in quantitative terms how consumption moves when investment increases, but there is still idle capacity. The data given in Table 59 show what happened in the United States in the years from 1940 to 1942:

TABLE 59

Relation Between Investment Increase and Consumption Increase, 1940–1942

	(in billions of dollars)		
	1940	1941	1942
Gross national product	97.1	120.2	152.3
Total investment	31.5	45.6	70.3
Government investment	16.7	26.5	62.7
Consumption	65.7	74.6	82 0

It will be seen that the total output of goods and services increased with the increase in investment. Between 1940 and 1941, investment grew by $14 billion, and the value of output increased by $23 billion. The ratio of the increase in income to that in investment is 1.6, which is rather lower than normal; and hence it suggests either that the propensity to consume was slightly reduced, or that the marginal propensity to consume is lower when income is higher than when it is less. But consumption in terms of dollars also increased as investment increased. While prices rose somewhat, it is clear that the quantity of consumers' goods produced did so too. Between 1941 and 1942, investment grew by $24.7 billion, and the gross national product by only $32.1 billion. Thus the ratio of the increase in income to the increase in investment was even lower than it was between 1940 and 1941: only 1.2, an exceptionally low figure. Consistent with this is the relatively small increase in consumption as measured in money, an increase of only $7.4 billion. But prices were somewhat higher in 1942 than in 1941, so even this figure overstates the increase in the physical output of consumers' goods. In fact, it is estimated that the production of consumers' goods fell by about 2 per cent, though incidentally this decline was more than made up by a 3 per cent rise between 1942 and 1943.

While the total output of consumers' goods increased from 1940 to

1942, the output of particular commodities declined, or at any rate did not rise, even though it would ordinarily have been expected to do so with an increase in demand. In 1941 consumers spent over $3.2 billion for automobiles and parts; in 1942 they spent only $0.5 billion. In 1941 they purchased $4.3 billion worth of furniture, furnishings, and equipment; in 1942 they bought $4.1 billion worth of such goods. By 1942 and 1943 the output of many consumers' goods, particularly durable items, was very substantially below the 1941 figure.

Both the United States and Great Britain reacted markedly to the increase of government wartime investment. But even at the peak of our wartime production effort, our output of consumers' goods, measured in constant prices, scarcely had to be reduced, although as we have seen, the production of many individual commodities had to be sharply curtailed.[1] In 1944 the output of consumers' goods expressed in physical terms was about three-tenths of 1 per cent below the output of 1941 — and incidentally our civilian population fell by somewhat more than that figure. But even this is striking, since it shows that despite the increasing investment, and thus higher incomes and higher demand, there was no *increase* in the output of consumers' goods. Even in the United States consumers did not get all they wanted; though they did consume more than they did in any previous year except 1941. In Great Britain, however, the reaction was more violent. By 1941 the output of consumers' goods had declined from its 1938 pre-war level by about 20 per cent, and between 1941 and 1944 there was a further decline of about 3 per cent. Thus the increased production of investment goods in Great Britain forced a reduction in the output of consumers' goods of about 23 per cent. When investment is pushed beyond a certain point, consumers cannot get all the goods they want.

The Problem Restated

The economic problem imposed by war is, therefore, created by the necessity of carrying on investment at a rate higher than is needed to provide full employment, or alternatively, by the need of cutting consumers' goods production in order to permit so large an output of investment goods. The result is that at the very time when the pro-

[1] This indicates that even when we were producing munitions at the highest level we reached, we were making nothing like a *maximum* production effort.

duction of consumers' goods must be restricted, consumers have more money than ever before. Everyone is employed, wages and profits are high, and naturally the demand for consumers' goods is abnormally large. Hence there are fewer goods available to consumers than they want to buy. The question then is, which consumers will do without? Who will be left standing when the music stops in the game of musical chairs?

We shall suppose that the government takes whatever steps are necessary to make possible a sufficient production of the goods it needs. It may control manpower, forcing workers to leave factories producing consumers' goods and to work in munitions plants. Or it may reduce the output of consumers' goods by controlling the flow of raw materials, allocating less copper, cloth, or leather than is needed for capacity output. Since their effect may generally be clearly seen, we shall not investigate these controls over production any more closely.

It will, however, be useful to examine carefully the consequences of certain fiscal policies of the government as distinguished from production policies. The production policy determines the level of consumer's goods production, but the fiscal policy determines which classes of consumers must make the sacrifices. Someone has to walk home with a shopping bag that is not as full as he would like. The tax policy, the nature of the rationing scheme, the controls over prices and wages, or even the absence of any fiscal policy whatsoever, determine in essence who the disappointed consumer is to be. No matter what the fiscal policy, the nation's shopping bag is going to be filled to a level determined not by these fiscal policies, but rather by the availability of men and machines for the production of consumers' goods, or in other words, by the production policies. The output of consumers' goods does not depend (except indirectly) upon whether the government raises taxes or not; nor does it even depend upon whether the government adopts any fiscal controls. Instead the fiscal controls determine how consumers will divide up the output of consumers' goods.

The Absence of Fiscal Control — Inflation

If the government is unwilling to adopt fiscal controls of any kind, it is choosing inflation. And inflation, as we shall see, has its own special methods of distributing consumers' goods. If the government

refuses to face the problem of distributing consumers' goods, natural economic laws do, and the result is not always very attractive. Let us suppose that the government determines to take only those measures needed to increase the output of munitions. Since government investment is very high, the national income is at the ceiling level. With high wages, high profits, and so on, the demand for all kinds of consumers' goods is unusually high. But the ability of firms to meet that demand is reduced because of shortages of labor and raw materials.

Let us consider one firm in this situation. Not only is there a heavy demand for its product, but marginal costs at any output are higher than before, and capacity output is much lower. Figure 66 illustrates this situation. The original demand is denoted by AR_1, the increased demand by AR_2; marginal costs originally are shown by MC_1, and after production controls are imposed and wages raised, by MC_2. If the firm is free to act as it wishes, it will decide to raise the price from P_1 to P_2. It could of course raise it even further, or not at all, since there is no civil law to compel it to increase price by exactly the amount shown. But if it seeks to maximize profits, it will adopt such an increase. Now what is true of this firm applies to the remaining firms in the economy. Prices generally will increase rapidly. Figure 66 shows that the amount of goods purchased is now greatly reduced, from O_1 to O_2. The rise in price has in effect reduced the amount demanded, for the price rise has persuaded some buyers to withdraw from the market, so that the customers who are to do without are "automatically" determined. As the firm finds that at each price there are many more customers who want the product than can be supplied, it raises the price higher and higher. And each time it does so, some of its customers leave the store with shopping bags only half filled.

Who, then, get the goods? And who make the sacrifices? This is not hard to guess. If the price of sirloin steak rose to $3 a pound, most of us would have to eat our memories of steak. The only people who could buy it at that price would be those whose incomes were very high. Generally a rise in prices brings about a reduction in consumption, especially among the lower income groups. Those in the higher income brackets, whose incomes provide a margin or surplus above normal consumption expenditures, can buy consumers' goods in approximately unchanged quantities if they choose to do so.

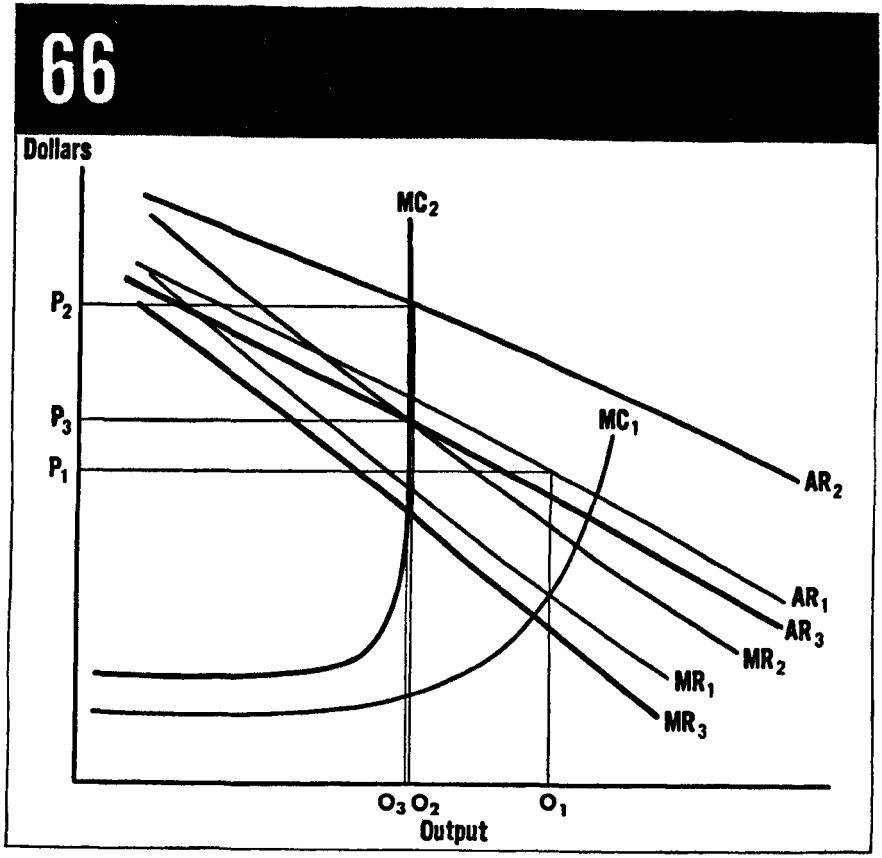

Figure 66. *Demand Change; Output Near Capacity*

Total consumption declines as the result of a do-nothing fiscal policy because large numbers of people can no longer afford to live on the same scale as before. That this is an advisable method of reducing consumption is questionable. From the point of view of equity, it commends itself to almost nobody. Obviously, a smaller total sacrifice would be required if those who have been living most comfortably were forced to reduce their consumption to a more reasonable level. To compel those who have been eating the least to eat even less seems hardly just.[1]

[1] Inflation not only forces those whose incomes are low to consume less; it works to the advantage of the debtor as against the creditor. The man with a savings deposit, a life insurance policy, or a government bond, loses out. The individual who owes money, and the owner of stocks, gains.

An additional objection to inflation is that it lowers efficiency at the very time when the maximum of production efficiency is required. If the output of munitions and consumers' goods is to be high, labor must be properly fed and housed. No man can work properly if his diet is deficient and his living conditions are very poor. Nor will he work hard if he sees that the richer members of society are still living as usual, regardless of his own sacrifices. The only way by which workers can rectify the situation is to press for higher wages, hoping that prices will lag. But this means constant "labor troubles" and strikes, and striking workers produce nothing. Thus to permit inflation is to sponsor a method of distributing scarce consumers' goods which is not only unjust but also inefficient, because it reduces productivity when the highest possible output is badly needed.

Furthermore, inflation tends to become cumulative, so that these difficulties tend to grow more and more serious. When prices are rising, consumers will try to buy more than they need for their current requirements in order to beat anticipated price rises. Thus, while high prices tend to discourage buyers, the general expectation of still higher prices tends to pack stores. For some time, anyway, consumers' demand will be likely to grow, thus adding still more to the upward pressure on prices. And the faster prices rise, the greater is the incentive to buy, even though the buyers know the price is high, and even though they go into debt to make their purchases. But obviously the winners — and of course with several chairs removed, not all can be winners — will be those whose incomes are high enough to finance purchases almost regardless of price.

The actions of sellers as well as of buyers tend to make inflation cumulative. Just as buying is stimulated by rising prices, so selling falls off in anticipation of still greater profits yet to come. It is obviously profitable to accumulate stocks of goods if you can hope to sell them shortly at an increased price. When sellers restrict the amounts they put on the market, the pressure on price becomes even greater than before. Hence, because of the reactions to rising prices of both buyers and sellers, the inflation tends to grow cumulatively. And if a 50 per cent inflation makes for a certain amount of injustice, a 500 per cent inflation is very much worse.

Taxation and Consumption

What can a government do to help prevent all this? For one

thing, it can alter its tax program, and if it does so successfully, it will have reduced consumption and curbed price rises as a result. Since there are no more consumers' goods to be purchased whether prices are kept down or not, it is evident that if they are kept down the altered tax program will have compelled a reduction in consumption, perhaps different in nature from that occasioned by inflation, but certainly not different in total amount.

When taxes are used as a method of checking inflation, they exert their chief influence by reducing consumers' spending. An increase in tax rates generally reduces the propensity to consume by leaving less for consumers to spend out of any given level of earnings. And a reduction in consumers' spending will to some extent avert price rises. Reference to Figure 66 will show why this is so. If instead of increasing from AR_1 to AR_2, demand falls to AR_3, the business firm will not find it profitable to raise prices by nearly so much. But notice that it will produce about the same output, with marginal costs at MC_2, whether the demand is AR_1, AR_2, or AR_3. With the same output, the level of consumption is the same, though of course lower than it would have been, if labor and raw materials had continued to be available to the firm in the original amounts. Thus consumption is reduced by the same quantity whether taxes are used to curb prices or whether inflation is allowed to go completely unchecked.

When inflation is permitted, the storekeeper, or the producer generally, continues to raise prices until enough people leave the store without buying, for the remaining goods to satisfy the wants of those who remain. When increased taxes are imposed, the heavy hand of the tax collector keeps the store from being overcrowded. Those who have so little left after paying taxes that they cannot afford to enter the stores do not have to be driven away by higher prices. Thus it is not necessary to raise prices to keep the number of customers small enough for the supply of goods. Increased taxes force reductions in consumption on those whose incomes, after they have paid the taxes, are not high enough to permit the purchasing of consumers' goods at the rate they would otherwise have intended.

Whether taxation as a means of reducing consumption is preferable to inflation depends, of course, upon what kinds of taxes are levied. To take an extreme case, suppose that the taxes are imposed only upon those in the low income groups. Their demand for con-

sumers' goods would necessarily decline, and the pressure for price increases would accordingly be reduced. Actually, however, the lower income groups would be no better off under such a policy than with inflation, for in either case their consumption would be reduced. Hence such a tax program would be open to the same objection as inflation. But the taxes may be levied so as to bring about a fairer reduction in consumption. Suppose, for example, that very heavy taxes were levied on high incomes. Those whose incomes were high would thus be compelled to reduce their consumption, and the lower income groups would accordingly be compelled to reduce theirs by less. But it is obviously not enough simply to tax the very wealthy at a high rate; the rate must be *very* high. For unless they are actually persuaded to reduce their consumption, the situation is not eased at all, since other income classes in that case would be called upon to reduce theirs. And it is to avoid this very thing that a program of high taxes is adopted. But it is difficult, by using taxes, to force a substantial reduction in the consumption of those whose incomes are high. To do so may require rates of taxation far above normal levels. Yet, to repeat: if such rates are not put into effect, only those in the lower income brackets are forced to curtail their consumption, which would be as bad as inflation and would also involve the working out and administration of an ambitious tax program. If a tax program is to be better than inflation, it obviously must secure better results. If it forces the largest cuts in the consumption of those with the lowest incomes, and permits those with high incomes to maintain their consumption, it is obviously no better than very high prices. But it will certainly be a great deal more trouble.

We are often tempted to adopt an ostrich-like policy when faced with the problem of reducing consumption. The opinion is frequently ventured that taking large amounts in taxes during a war imposes too great a sacrifice upon the taxpayer. This view neglects the real facts of the case and emphasizes the financial overlay. The total amount by which consumption must be reduced is determined, as we have seen, by the resources of the economy and the extent to which they are required to produce an adequate flow of munitions. A government that adopted a policy which appeared to be easier and less austere with respect to taxes would, in fact, simply be forcing a reduction in consumption in some other way. Less butter, less clothing, fewer refrigerators, and fewer automobiles are provided whether taxes

are high or low, whether there is inflation or not. But *who* will eat less and secure less clothing, and *who* will be unable to purchase the refrigerator or the automobile, depends upon the fiscal policy adopted. Inflation is one way, and a poor one. Taxes as a check on inflation are better only if they compel classes that would escape real sacrifices during inflation to reduce their consumption. And this, as we have seen, occurs only if the tax rates on high incomes are exceptionally high.

In times of war all governments increase their tax rates. Generally, income tax rates have been sharply stepped up and sales taxes on many kinds of commodities have been substantially increased. But the tax weapon is never used alone. To reduce consumption satisfactorily requires other measures as well.

Price Control and Inflation

Another way to prevent inflation is to control prices, and this has been done extensively by countries at war. If the controls are strictly enforced, some consumers will find themselves unable to buy as much as they want simply because there is not enough available to satisfy all the demand. But those who do without will not necessarily be the poor. Rather, they will be the ones who didn't get there first, regardless of income.

Ordinarily, prices are increased when firms find it profitable to raise them, but firms are not compelled to increase their prices, although, of course, they find it profitable to do so when demand increases and labor and raw materials become scarce. The government may prevent price rises, however, by ruling that goods may not be sold at prices above a certain maximum, though allowing firms to produce as much or as little as they like (and can) in any manner they like. Since government determination of price is substituted for a determination which seeks to maximize profit, it is no wonder that most business firms are unfriendly to price controls.

But the enforcement of price controls is difficult. If demand is not reduced to keep pace with the dwindling supply of consumers' goods, sellers are under great pressure to charge more than the law permits, for they find that they can sell more than they have on hand at the maximum price allowed. nd buyers are anxious to secure, at that price, more than sellers can provide. In this situation, sellers are naturally eager to raise prices, and some buyers are willing to pay

more than the legal price in order to be sure of getting what they want. Under such circumstances black markets are likely to develop, and sales at prices above the legal maximum will be common. Unfortunately, price controls will almost certainly invite black markets. This is not, of course, a serious argument against the use of such controls. It simply means that the controls can be circumvented. When black markets become general, the results are not very different from those of inflation. For since there are too few goods to satisfy the wants of all buyers at legal prices, the goods go to those who are willing to pay illegal prices and able to pay high prices, to those with both a full pocketbook and an empty conscience. The rest of the community then finds that the goods are not available at the prices they are willing to pay. It is these people who are compelled to reduce their consumption.

However, when maximum prices are efficiently enforced, there is a demand which cannot be met out of the available output. Who then is able to get the goods? Obviously, those who arrive first, if the sellers play no favorites. There are likely to be runs on the stores, and those first in line will get what is to be had. The goods then go not to the wealthiest, as they do with inflation, but to those who are most agile in shopping and most patient in waiting for sales to open. But there is no reason to suppose that these will be the persons who most need the goods, any more than there is to suppose that those who can best afford to pay need the goods most. Thus, while rigidly enforced price controls check the injustices of inflation, they do not, without other measures, insure that the sacrifices are imposed where they can be most easily supported.

In summary, price controls, if they are enforced, prevent inflation. Hence there is a different pattern of sacrifice by consumers than with inflation. When the government sets maximum prices it is, in effect, saying, "No, you are not going to get butter just because you have more money than anyone else. You have to take your chance along with the rest. First come, first served. Ready? Go!" That solution is obviously not ideal either. There is the added difficulty to be considered that sellers, and some buyers, have an incentive to trade at illegal prices. When they do so, the way in which consumers' goods are distributed is no better, and it is probably worse, than with simple, legal inflation. Thus, price control legislation is scarcely a weapon to be used by itself.

Rationing and Inflation

Beyond taxation and price control, the next step in preventing maldistribution in time of shortages is rationing. Obviously, when goods are rationed, the amount each consumer can have is directly determined, so that by adjusting details of the ration scheme, it is possible to fix the amount by which each class of consumer is to reduce his consumption. In contrast to other fiscal policies, rationing is so obviously fair that almost all governments make use of it during a war. What rationing does, in essence, is to create a second kind of money — the ration coupon — and the buyer must use both types of money to purchase a rationed commodity. When the government decides how to distribute the ration money, it effectively determines who will share in the supply of goods. If everyone, from the infant to the aged, is given the same number of ration coupons in a year, then each has an equal claim upon the economy's output of the rationed commodity. If rations are adjusted to fit the needs of different kinds of consumers, the claim of each person depends upon the rationing authority's judgment of his need. In any case, a human agency, the government or the ration board, determines how much each person may claim. This will obviously give a very different result from that obtained when the size of one's claim depends only upon the fullness of his pocketbook.

There are various ways in which the principle of rationing can be applied. Rationing may cover individual commodities, like sugar or shoes. In the United States between 1942 and 1945, each holder of a ration book could purchase a certain number of pounds of sugar a year, but he could not, by purchasing less sugar, claim more of something else. It is also possible to ration groups of commodities together, as for example meat and meat products in the United States, and clothing in Great Britain during the last war. Under this scheme, the purchaser was not limited to, say, three ounces of bacon, four of beef, and two of pork. Instead, he could purchase various combinations of meat, the amounts depending upon the total number of points in his possession, and the point-value of each kind of meat. A still more general rationing scheme has been proposed, although it has not been used so far as is known, by which the total amount of money which could be spent by any one individual is limited to, let us say for illustrative purposes, $500 a year. Within that restriction, consumers are free to buy whatever they want. Thus various schemes of

rationing are possible. It can be confined to a single commodity or to a class of commodities, or it can be applied to consumers' goods of all kinds.

In general, rationing, and for that matter price controls and price increases, tend to extend over a wider and wider field as time goes on. The first symptoms of shortages in consumers' goods appear long before full employment is reached. Particular kinds of labor become unavailable for the production of consumers' goods even when there is still a large amount of general unemployment. The rationing of products in short supply will then avoid pressure to raise prices. But if there are substitutes for the rationed commodities, the demand for these will rise sharply once rationing is introduced. Because of the increase in demand, the substitutes will not be available in sufficient amounts to supply all buyers, so that if a price increase is to be avoided, the substitute commodities also will have to be brought under price control or will have to be rationed. And so, as more and more shortages develop, rationing is extended over more and more commodities.

As we have said, rationing reduces the inflationary pressure on prices. If the ration is properly chosen, the total amount demanded will not exceed the amount that producers are willing to put on the market at a suitable price. Therefore the firms that supply the commodity will have no reason to raise prices. Consequently when goods are rationed, it should not be necessary to invoke price controls. But rationing also is open to certain objections, though it is an equitable method of distributing scarce consumers' goods, and though it makes orderly marketing possible. The administrative difficulties of any rationing scheme are considerable. Books of coupons must be provided to every purchaser, and if the ration is to reflect differences in needs, the distribution of these books is likely to be complicated. Furthermore, retailers must keep accurate sales records, must arrange for the transfer of coupons to wholesalers, and so on. It is of course not surprising that a supplementary monetary system will require supplementary institutions to do the work normally done by banks and clearing houses. But the administrative difficulties have apparently not been great enough in any country to do more than *retard* the adoption of rationing during a war. Finally, rationing may mean a more serious interference with consumers' preference than is necessary. Some individuals may want more than a pound of sugar every other week, and others may desire much less, whereas most rationing

schemes do not allow for such differences. Obviously, however, if the ration is based not on individual commodities but on classes of commodities, such as food or clothing, this objection is not particularly serious, for the consumer is reasonably free to follow his preference within groups of rationed items. While he is not free to spend all his meat points for clothing, he is at least free to follow his preference for beef, pork, or lamb.

The "Deferred Pay" Scheme and Inflation

An interesting combination of taxing and borrowing which was employed in both England and Canada between 1941 and 1945 is known as the "deferred pay" scheme. Taxes are levied at high rates in order to reduce consumers' incomes and thus to reduce their expenditures. But a part of the tax thus paid is credited to the account of the taxpayer for his use after the war. According to the rates in effect in England in 1941, a married man with two children, if he had an income of $3200 [1] a year, had to pay a tax of approximately $844.45, of which about $156.67 was set aside as a post-war credit for him. This sum could be regarded as a forced loan to the government. From the point of view of cutting down consumption, a tax of this sort is almost as effective as any other; and since the taxpayer loses the use of the funds only temporarily, he is not so unwilling to pay it, even if it is very heavy. Finally — and this is perhaps one of the most important features of the plan — the refund is to be made at some future time when consumers' purchases would otherwise be deficient in amount. It is expected that this support to consumers' spending will prevent income and employment from falling at some date in the future when investment becomes deficient.

Other features of the scheme are also interesting. One of these is that the smaller the amount paid in taxes — and therefore, of course, the lower the income from which it is collected — the greater the proportion of the amount paid that is deposited to the account of the taxpayer. Thus, while a deposit credit of $156.67 was set aside for a tax of $844.45, the credit was only $193.33 for a tax of $1204.50, and it was as high as $65.33 for a tax of $97.50. Thus the smaller the tax, the larger the return, and vice versa. Since the amount credited to the taxpayer is strictly a part of the national debt, this scheme insures that the ownership of the public debt is widely dis-

[1] Pounds are here translated into dollars at the rate of £1 to $4.

tributed. Another feature of the plan is the timing of refunds. Obviously no purpose would be served in repaying at a time when output is at capacity or when consumption must be reduced. But if the credits are made available when the demand for consumers' goods is not sufficient to employ all consumers' goods resources to capacity, then the release of the deposited funds can supplement consumers' expenditures enough to induce firms to maintain output and employment at prosperity levels. In other words, this scheme shifts consumers' purchasing power from a time when it would otherwise run to waste in higher prices, to a time when it will actually increase employment and the supply of consumers' goods. Furthermore, because a large proportion of the refund goes to persons of low income, we may be certain that consumers' expenditures will increase greatly when the deposits are withdrawn.

It may be asked whether such a scheme might bring about inflation. It could, even in times of peace. But the government would have no difficulty in controlling the inflation. There is no reason why all of the deposits should become withdrawable at any one time. If only a relatively small addition to consumers' expenditures is needed to provide something like full employment, it might be wise to release only 10 per cent of the total on deposit. The remainder would be retained as stored-up demand to be used later when consumers' spending again becomes deficient. It should be clear that the effect of forced saving is to reduce the propensity to consume when increased consumers' expenditures are likely to increase price rather than output. Of course the propensity to consume is increased again as soon as consumers are allowed to draw on their deposits. Therefore if the refunds are made at the right time, the increased propensity to consume will bring about an expansion in the output of consumers' goods instead of a rise in price.

This method of war-financing was not tried in the United States, although the extensive sale of War Bonds brought somewhat similar results. Since in many firms employees were persuaded to set aside as much as 10 per cent of their incomes for the purchase of these bonds, buyers accumulated larger post-war credits in the form of bonds which could be cashed in after a brief time. There are two important differences, however, between the purchase of War Bonds under persuasion and the enforced deferred pay scheme used in Great Britain. For one thing, since our War Bonds can be turned into cash

at will, the government cannot control the amount added to consumers' spending at any one time. If consumers want to cash their bonds at the height of an inflationary boom, the government cannot make them wait until a time which would be more advantageous to themselves and to the economy as a whole. Secondly, the ownership of the War Bonds, or the post-war credits which they provide, is not widely and evenly distributed. While detailed information about Series E Bond holdings is not yet available, we do know that persons in the high income brackets hold the bulk of the liquid assets in the possession of individuals.[1] A survey made in the first quarter of 1946 shows that the 15 per cent of families whose incomes were highest (over $4000) held 46 per cent of all liquid assets. And such families are not likely to use their assets for consumers' goods. The 20 per cent of the families whose incomes were lowest (under $1000) held only 7 per cent of all liquid assets. These are the families which would be most likely to use their assets to finance the purchase of goods and services.[2]

In this section we have considered various wartime fiscal policies, especially with a view to their effects on consumption. This is not because no other problems of the wartime economy deserve consideration, but rather because the important question is how and how much consumption is reduced. If the problem of controlling consumption were solved differently, for instance by setting up communal kitchens and forcing everyone to eat in them, the question of how the war should be financed would be relatively unimportant. The government can always get money to buy ships, planes, and tanks, even if it has to print it. If consumption can be reduced fairly by other means than taxing, then there is very little to choose between having the government finance its purchases by printing money, by taxing, or by borrowing; although borrowing or printing money would at least promise higher consumers' expenditures in the post-war period and thus a higher level of employment.[3] For that reason either borrowing or printing money might be preferred. But if we depend upon our fiscal policy to reduce consumption, the choice of the

[1] These assets consist of savings deposits, demand deposits, and United States Government Bonds.

[2] For further information about holdings of liquid assets and how the holders expect to use them, see articles in *The Federal Reserve Bulletin* for June, July, and August, 1946.

[3] The chief danger would be that the level of consumers' expenditures would be too high in the post-war period if war expenditures were financed by printing money.

methods by which a war is financed is extremely critical. We should certainly not choose a method just because it provided the funds to the government efficiently and painlessly. Much more important is the success it promises in cutting down consumption in an equitable manner. This is the heart of the economic problem posed in wartime when investment goes far beyond the full-employment level.

THE ECONOMY DURING PEACE

During a war, investment, and especially government investment, is given top priority over all other kinds of production. For that reason, as we have seen, it becomes necessary to reduce consumption. There is less labor available for the production of consumers' goods simply because more is needed for the production of munitions. Consumers' goods production is in a sense a residual claimant for labor during war. But in times of peace, investment is not ordinarily regarded as more important than consumption. For that reason, efforts to control inflationary developments when there is peace are usually directed not only to limiting consumption but also to reducing investment.

The situation that requires this control arises when there is so much investment that the total demand is far above the level that would have induced employers to hire all those who want work. Thus, if spending at the rate of $200 billion a year will lead business and government to employ the total labor force, then control of inflation will be needed when spending rises to, let us say, $225 billion a year. Since with annual spending at $200 billion, firms would be producing at capacity and could find no unemployed labor for further expansion, increasing the rate of spending could not lead to increased output but only to higher prices. For as we have seen, when increasing investment raises demand beyond the full-employment level, prices rise steadily. And rapidly rising prices of course constitute inflation.

We have already seen at least some of the reasons why inflation is unwelcome. Inflationary price rises do not evoke further increases in output; rather, they come about because output has not been increased to meet increasing demand. They are to some extent to be regarded as symptoms that demand has grown beyond the level which can be satisfied by a fully functioning economy. But worse than that, inflation provides evidence that relatively scarce consumers' goods are

going to those who can most easily be without them — those whose incomes are highest. Since more was wanted at the original price than could be made available, it is clear that some consumers have to do without. The necessary reductions in consumption must therefore be made by the lower income groups and those whose incomes are fixed. Thus, inflation means that those whose consumption is already low must reduce it further. So inflation may be condemned not only because it does not stimulate output, but also because it signifies a redistribution of real income away from the poor and toward the wealthy. While, like the common cold, it sometimes occurs, such a redistribution finds few who favor it — openly, at any rate.[1]

In wartime, as we have seen, the control of inflation is difficult because of the heavy pressure to expand the output of munitions, *i.e.*, to increase government investment. The only things to do are to reduce private investment, when this does not interfere with war production, and to reduce civilian consumption. In what we perhaps optimistically call normal times, however, it is not essential that investment be raised or even maintained. It is possible to reduce both investment and the propensity to consume.

Inflation Controls

Measures for lowering the propensity to consume have already been considered. Taxes, propaganda to stimulate saving, discouragement of installment selling by high interest rates, the adoption of "sound" financial practices by business firms, especially in connection with depreciation allowances and the withholding of profits — these are the conventional devices, and they do somewhat reduce the propensity to consume and thus curb inflationary pressures. Indeed, they are regarded so highly — for obviously they can be made to prevent inflation — that many of us are tempted to urge their use in all economic situations at all times, without even considering whether inflation is the problem.

We have also developed techniques for controlling investment, and especially for preventing it from increasing too greatly. Through

[1] There were some amusing illustrations of the prevalent attitude toward inflation during the brief period in which price controls lapsed during July and August, 1946. The newspapers were filled with advertisements in which the advertisers pledged that they would "hold the line" against inflation and would not raise their prices except to the extent that costs had increased. Everyone opposes inflation — but . .

constant practice we have even learned to apply these controls with great efficiency. Raising the interest rate and in other ways making it difficult to borrow is one such step. As we saw when discussing banking operations, the controls of the Federal Reserve System are extremely effective in forcing commercial banks to restrict their lending operations and thus to raise the interest rate. These controls have been employed very frequently. The efforts to keep government investment low have also been persistent and effective. Government expenditures are generally severely limited, and constant efforts are made to maintain a balanced budget. Such techniques are regarded as constituting a desirable fiscal policy, and any departure from these low-expenditure, balanced-budget precepts is thought to be unnatural and harmful. Certainly these judgments are sometimes correct. If investment is so high that, together with the propensity to consume, it produces inflationary pressures, government investment should be re-examined with a view to cutting it down. Thus the control of inflation during peacetime is not difficult. Provided that the government exerts adequate authority, it is generally possible to reduce both the propensity to consume and the amount of investment to the desired level. The methods employed are generally those described as sound finance, and they are logical and appropriate for checking a tendency toward inflation. But since, as we have seen, inflation threatens only when the economy is approaching full employment, these measures should be regarded as peculiarly applicable to an economy in which full employment is the rule.

During the greater part of the nineteenth century and the first two decades of the present one, economic controls in this country were directed not so much to reducing unemployment as to avoiding too much prosperity, in other words to checking inflation. We developed at this time not only our most important economic institutions but also our habits of thought about questions of economic policy. For this reason it is perhaps unfortunate that the mental approach and the economic controls required by an economy with full employment are generally the reverse of those needed to solve the problems of an economy which generally provides less than full employment. In an economy in which unemployment is likely to be present, efforts should be made to stimulate investment and to raise the propensity to consume. And as we have seen, in an economy which normally enjoys full employment and which, therefore, is subject to a persistent pres-

sure toward inflation, investment must be kept from going too high, and the propensity to consume must be kept down.

The lessons so patiently learned from the experiences of our economy before World War I may, of course, be applicable to our own economy now. Some of them obviously are. But we must take great care not to prescribe the medicine for the disease of too much income, to that quite different and characteristically modern malady, too little income.

Summary

When investment pushes beyond the level that provides full employment, either certain controls have to be adopted or inflation will develop. The chief objection to inflation is that it forces a most unfair distribution of the inadequate supply of consumers' goods. Alternatives to inflation — taxes, price controls, and rationing — at least divide the goods more fairly. When it is not essential to allow investment to go to a very high level, inflation can also be avoided by checking investment. In fact, control of investment by keeping the interest rate high, limiting government expenditures, and encouraging saving, has often been employed in this country. Naturally, measures suitable to control inflation are inappropriate when unemployment is severe.

34

Remedies for Unemployment: General Considerations

THE ECONOMIST is more like an engineer than like a physicist. He is interested in learning about the economy because he wants to help make the economy function properly. He studies unemployment, not for the intellectual joy of handling a knotty problem — there are much tougher ones in chess — but because he believes that unemployment is an evil that should be remedied. Thus all we have learned about the national income and employment leads to the question discussed in this chapter and the following — what can be done about unemployment?

The Conditions for Full Employment

We have seen that the national income, and hence the level of employment, depend upon the propensity to consume and the level of investment. With the propensity to consume given, there is full employment if investment is high enough. How high it has to be depends upon the amount of consumers' goods the community is willing to take when the national income reaches the full-employment level; more exactly, for full employment, investment must equal the difference between the full-employment level of the national income and the level of consumption with that income.

It has been estimated that when there is no unemployment in our economy, we produce approximately $200 billion worth of goods and services a year. At that level of output, and therefore of income,[1] we purchase about $135 billion worth of consumers' goods a year.

[1] *Gross*, not *net*.

Therefore there will be no unemployment if investment is done at the rate of $65 billion a year — $200 billion minus $135 billion.

The condition as to investment can also be expressed in the following way: With full employment, about 60 million men are working. They and their families, together with all the other income recipients in the community, purchase the output of about 39 million men who are engaged in the production of consumers' goods. Therefore, so that all 60 million men may find work, it is necessary that the remaining 21 million should be employed in producing goods and services *not* demanded by consumers — that is to say, investment goods. If fewer than 21 million men are actually employed in producing these investment goods, there is bound to be unemployment. And as we have seen in our analysis of the multiplier, the amount of unemployment would exceed the difference between 21 million men and the number actually producing investment goods. For in addition to those who might have been working in the investment goods industries, some who would have been working in consumers' goods industries will also now be unemployed, since the reduction of some incomes will lower the demand for consumers' goods. If investment is deficient, we suffer from unemployment. With our propensity to consume, we must invest $65 billion a year — that is, we must have jobs for 21 million men in investment goods industries — or take the consequences in unemployment. Can this level of investment be maintained? In other words, can we avoid unemployment? These are questions which this chapter and the one following will attempt to answer.

Enormous Investment Now Required in the United States

We require much more investment for full employment than other countries, and much more than we formerly needed ourselves. In 1929 investment came to $28.6 billion, and unemployment was very low. Perhaps $30 to $32 billion worth of investment would have been enough to assure full employment that year. Twenty years earlier, at a rough guess, we would have needed an annual investment of something like $5 billion for full employment.[1] But however rough the guess, there is no question that the amount of investment needed

[1] Of course, allowance must be made for the fact that prices were higher in 1945 than in 1910 or 1929. The cost of living in 1945 was about 80 per cent above the 1913 level and 6 per cent above the 1929 level.

for full employment in this country is enormously greater now than it used to be. In Great Britain, which after the United States probably requires more investment to produce full employment than any other country, something like $8 to $10 billion worth of investment a year appears to be enough. Even if this figure were multiplied by three to allow for the difference in population, it would be far below the amount needed in this country.

Why did this country need so much more investment in 1946 than in 1910 in order to achieve prosperity? The basic reason is that, *with full employment*, in 1946 we could produce about four times as much as in 1910 and about twice as much as in 1930. But while there has been an enormous increase in our capacity to produce, and hence in the national income with full employment, there has been no such increase in the amount that we are willing to spend on consumers' goods out of a full-employment income. For example, in 1929 our gross national product with full employment would have been about $105 billion, and of this amount we should have been willing to spend roughly $75 billion on consumers' goods. By 1946, our capacity to produce had increased so rapidly that our output with full employment would have been worth about $200 billion, of which we should have been willing to spend about $135 billion on consumers' goods. Thus, in twenty years our capacity to produce had increased by about $95 billion, and our consumption purchases with a peak income would have increased by about $60 billion. The gap between the national income with full employment and consumption expenditures at the same level is therefore larger by about $35 billion. This gap must be completely filled by expenditures upon investment goods if there is to be no unemployment.

The reason why this country requires more investment to secure full employment than either Great Britain or Canada requires is that we consume less in relation to our income. While with full employment, national income is considerably higher in the United States than in either of these other countries, average consumption is higher by a smaller absolute figure. To illustrate: with full employment, our output, and therefore our income per head, is about $670 more than that of Great Britain, but our consumption per head is only about $300 more.[1] Generally, therefore, as a country grows in wealth, and

[1] These figures are intended merely as illustrations. They do not pretend to give even crudely the actual differences involved. There are many difficulties in comparing the

as its capaci y to produce is thereby increased, more and more investment is required in order to provide it with full employment. The level of investment required in the United States was, to repeat, about $5 billion in 1910, about $30 billion in 1930, and about $65 billion in 1945–50. If our productivity continues to increase in the future as it has done in the past, the amount of investment needed to provide full employment will be even higher in 1970 than it is today.[1]

The fact that so much more investment is required now than in the past to bring about full prosperity indicates a change in the structure of our economy which is of the greatest significance. For a century before 1910, the level of investment required for full employment was somewhere between $0.5 billion and $5 billion. It was undoubtedly rising slowly throughout this period, but the rate of increase in absolute terms was very modest. In the next two decades, from 1910 to 1930, the rise was much more rapid, and it has evidently been accelerating ever since that time. Unless our consumption habits change profoundly, the amount of investment required for full employment will undoubtedly continue to rise — ever faster and faster. The implications of this for our economy cannot be overemphasized. Like a drug addict who gradually increases his dose from half a grain to five grains before he gets the stimulus he craves, so the economy requires more and more of the stimulus of investment to maintain its vigor. The whole structure of the economy has already been profoundly altered by this development, and in our thinking about economic matters, we must keep this fact continually before us. The drug addict eventually dies, and that is the end of his story. But what about our economy?

Investment Required and Investment Available

The fact hat a certain amount of investment is needed for full employment does not, of course, mean that that amount is sure to be forthcoming. Our own experience of unemployment in he nineteen-thirties shows clearly that investment does not always reach the required level. Investment is often too low to provide full employment.

national income of the two countries: the rate of exchange to be used in expressing British currency in dollars, the reconciling of differences in the methods of measuring the national income in each country, and so on.

[1] This assumes no substantial change in our spending habits. If, however, our propensity to consume should increase sufficiently, the amount of investment needed might be no greater and might even be less.

But this is not surprising. When business firms make their decisions, they cannot be expected to consider he effect of these decisions upon the national income. It is sheer coincidence, then, if the amount of private investment which is planned equals the amount needed for full employment, given the level of foreign and public investment. Let us review our analysis of the factors upon which private investment depends in order to clarify this point, that private investment may be far below, or above, the amount required for full employment.

Private investment is carried on by firms that wish to expand their capital assets. When a firm undertakes an investment project, it orders new plant or more machinery, or it adds to its stock of finished goods on hand or to its work in progress. When it decides to expand in any of these ways, it does so because it sees a profit in such an expansion. Rapid expansion seems profitable to business firms when: (*a*) the market for their product is very large in relation to their capacity to produce, (*b*) the interest rate and the price of investment goods are low, or (*c*) it is possible to adopt technical improvements hat reduce the cost of production. There are other conditions, of course, but these are the main ones.[1] But firms see no profit in rapid expansion, and investment is consequently low, when (*a*) the existing capacity of the industry is large, (*b*) interest rates and the price of investment goods are high, (*c*) the market is not growing quickly, and (*d*) technical advance is slow. There is therefore no reason to suppose that private investment will reach just the level needed to provide full employment. It may be higher than is needed, and produce inflation; or it may be lower, and produce unemployment. A business firm cannot be expected to invest lavishly just because unemployment would otherwise be severe.

When the amount of investment needed for full employment is small, private investment is likely to be too high rather than too low, for it is more likely that business firms will discover at least $3 billion worth of profitable investment opportunities in a year than that they will find at least $40 billion worth. Moreover, the demand determinants may be exceptionally favorable when little investment is needed, and hence it is even more likely that private investment ill be adequate when the amount required is small. We have seen that very little investment is needed when the national income at full employment is small; that is, when productivity is low because the econ-

[1] See Chapter 28.

omy does not possess a large stock of capital goods. And we have seen that when the stock of capital goods is low, the marginal efficiency of capital tends to be high. Hence, when little investment is needed for full employment, the marginal efficiency of capital is likely to seem high and private investment will generally be adequate. In concrete terms, this means that when the economy has relatively few factories, its total output is low, and most of its labor force is busy producing goods for consumers, leaving only a small number to be employed in producing new machinery, new plant, and so on. But because its industries are not fully developed, its businessmen find many profitable opportunities for enlarging existing factories and building new ones. The amount of investment needed for full employment is very small; the amount that businessmen are ready to invest is relatively high; and so investment is not likely to be deficient. In fact, it is likely to exceed the amount needed for full employment by a large margin.

When the amount of investment needed for full employment is very high, the problem of unemployment becomes very much more serious. It is more difficult to find a large number of investment outlets than a small number when other things are equal; but unhappily they are not equal. The need for a high level of investment as a condition for full employment indicates that the stock of capital goods on hand is high. And this, as we have seen, makes the task of finding enough profitable investment outlets much harder. An economy which must invest heavily to secure prosperity is therefore one already rich in factories, equipment, and other capital goods. Its labor is very productive, and when all who want to work are employed, its output of goods and services is immense. Indeed, its output is so high that people want to consume only a relatively small part of it, say two-thirds. Hence about two-thirds of the men are able to find work producing goods for consumers, and the other third must be employed building more factories, producing more equipment, and so on, if unemployment is to be avoided. But because there are already so many splendidly equipped factories, businessmen are loath to build new ones. The result is likely to be inadequate rather than excessive investment, and unemployment rather than inflation. Yet many could be found who would recommend a dose of anti-inflation medicine for this economy, just as for the one in which investment tends to exceed the level needed for full employment.

All this implies that as an economy grows wealthier, the pressures toward full employment exerted by private investment grow less and less effective. In the nineteenth century, when a low level of investment was adequate to the need, full employment and inflationary pressures were looked upon as the normal and natural state of affairs by the man in the street and the economist alike. But now a very much larger investment is needed to bring about full prosperity, so that our country is much more likely to suffer heavily from unemployment than it was a generation ago. Hence we are perhaps no longer justified in viewing the economy through the rose-colored glasses our grandfathers habitually wore. This, of course, does not mean that a rich economy cannot be prosperous. For obviously it *can* avoid unemployment. The offsets, as we may call them, to a rapidly increasing and very large stock of capital goods — the expectation of a rapid growth in the market, important technological developments, and so on — may be very strong and effective. If so, the rich economy will enjoy great prosperity. But these offsets do not grow strong just because the amount of investment *needed* for full employment is very high. Their strength depends upon different factors, such as the strength of monopoly and of labor unions, government patent and banking policy, and rate of population growth. If these are not favorable, the offsets will not overcome the effects upon investment of the large accumulation of capital goods. Private investment will then be low, and unemployment will be severe.

Private investment is not always done at precisely the rate which provides full employment; often it is either too high or too low. But private investment is not the only kind. Indeed, it made up less than half of the total in the United States between 1920 and 1941. The question then arises whether, in a rich capitalist economy, public investment will make up for the expected inadequacy of private investment and thereby bring about full employment. And the answer, judging from our own recent past, is clearly that it will not. So far, at any rate, government units have not determined the volume of their investment [1] by reference to the amount needed for full employment. State and local governments in particular have tended to make their investment outlays when there is a *need* [2] for highways,

[1] In the sense used in this analysis — the total spending by the government on goods and services.

[2] Perhaps interpreted in a very narrow sense.

schools, police services, or public parks, and when their tax receipts are high. But, unfortunately, it usually happens that when private investment is lowest, the need for such expenditures seems to be least pressing, and tax receipts for financing them are far below normal. Hence little is done, with the result that the difficulty is intensified. The federal government, however, has in recent years set its investment expenditures with some regard to the volume of unemployment. But while federal investment was relatively high during the nineteen-thirties when private investment was low, it was not enough higher than it had been in the nineteen-twenties to do much more than offset the reduction in state and local government investment. Since public investment is in the final analysis under our own control, and we can make it as high or as low as we please,[1] it is impossible to predict that it may not in the future be made to offset the fluctuations in private investment in order to give us full employment. Obviously, whether that is to be done or not depends upon our decision as citizens. But so far, no agency except the federal government — and that not extensively — has tried to set the size of public investment by reference to unemployment.

So long as we determine the amount of public investment on the bases used before and during the nineteen-thirties we cannot expect total investment always to provide full employment. If certain objective factors are favorable — such as the state of confidence, the rate of technological advance, and the existing stock of capital goods — total investment is likely to be adequate or even too high. If these factors are less favorable, total investment is likely to be deficient, and in consequence we are likely to suffer chronic and heavy unemployment. If an economy does not naturally tend toward a condition of full employment, controls become important. Before examining various plans for keeping unemployment down, however, it is necessary to give further attention to certain basic characteristics of our economy.

Prosperity and Debt

Investment, to be fully effective, must be financed in some way that does not simply take from consumers the money that they would otherwise have spent on consumers' goods. And this is true not only of private, but also of government investment. If an investing agency,

[1] Though there may of course be difficulties in changing it *quickly*.

whether the government or a business firm, orders an investment good worth $50 million and secures this sum from individuals who otherwise would have spent it all upon consumption, then it is clear that there will be no over-all rise in employment. For the increase in employment in the investment goods industries will be exactly offset by a reduction in employment in the consumers' goods industries. More investment goods and fewer consumers' goods are produced, and the increase in investment is exactly offset by the decline in the propensity to consume. If, however, the investing unit gets its funds without forcing a reduction in consumption, the full multiplier effect of the investment will be felt, for there will be no decrease in consumption to cancel the rise in consumption created by the investment expenditures. When investment increases and there is no reduction in the propensity to consume, employment and income rise by a relatively high multiple of the increase in investment.

If funds for investment are borrowed, the propensity to consume is usually not affected adversely. If, however, the funds are raised for government investment by increasing taxes, or for private investment by increasing the sums withheld from dividends, the propensity to consume is reduced. Thus, when the government finances its investment expenditures by borrowing, income and employment are higher than when it does so by raising the tax rate. Likewise, when business firms get funds for their investment projects by borrowing or selling securities, rather than by withholding large sums in the form of undistributed profits,[1] income and employment benefit. This is true because a lender generally does not reduce his consumption in order to make loans. Banks, in particular, can make loans, as we have seen, simply by creating money. Except perhaps during a war, private lenders are not likely to cut down their consumption in order to buy government bonds or industrial securities. Hence the propensity to consume is not reduced just because the government and private industry borrow heavily.

If, however, firms increase the rate at which they withhold their profits, or if the government raises the tax rate, the effect upon the propensity to consume is adverse. This effect may be slight. If, for example, a business firm secures its funds for investment by withholding out of profits money which would otherwise have been paid as dividends to the wealthy, consumption will not be seriously reduced.

[1] Or making abnormally large allowances for depreciation.

The same thing will be true if the government secures funds for investment by taxing high incomes. It is possible, therefore, to finance investment in other ways than by borrowing, without facing a serious reduction in the propensity to consume. But the smallest reduction in the propensity to consume would occur when the funds used for financing investment were obtained through loans. For after all, a tax on high incomes generally forces some reduction in consumption and so does the withholding of profits from dividends. In summary, therefore, we see that investment financed by borrowing is more "high-powered" and more effective in supporting income and employment than investment financed in other ways.

Both business and government usually borrow for large investment projects. When the government wants to spend $100 billion a year, as it did in 1944, it does not collect all, or even most of it, in the form of taxes. When private firms spend almost $20 billion on investment goods, as they did in 1941, they do not get all the funds they need from undistributed profits and depreciation allowances. When investment is high, borrowing is generally high too.

It goes without saying that heavy borrowing increases debt. The American Telephone and Telegraph Company announced in August, 1946, that it was going to borrow $351 million "because of the unprecedented demand for telephone services, and in order to extend and improve existing services." The money was to be raised through a new issue of $351 million of convertible debentures. Obviously the debt of the company would increase with the sale of these bonds. Likewise, when the government borrows, as it has to do when its expenditures exceed its tax receipts, the public debt rises. When private investment is high, we should accordingly expect to find an increase in private debt; when public investment is high, we should expect public debt to rise. The statistics on debt will illustrate this point. Table 60 gives figures for private and public investment and for private and public debt.

Let us look first at the figures for the nineteen-twenties. Total private investment between 1922 and 1929 came to $115.5 billion, or an average of $14.9 billion a year. Private debt increased in those years by $53.7 billion, or by $6.7 billion a year. Thus almost half of the funds for the very heavy investment expenditures made in those years was borrowed by private firms. In contrast, between 1930 and 1939, private investment amounted to only $69 billion, or

TABLE 60

Investment and Debt, 1921–1944

(In billions of dollars)

	Investment		Debt (at end of year)	
Year	Private	Public	Private *	Public
1921			104.5	29.7
1922	9.8	8.5	107.7	30.4
1923	15.3	8.7	114.4	29.9
1924	12.6	9.1	121.0	29.8
1925	15.0	9.6	130.1	29.9
1926	16.7	9.8	136.7	29.3
1927	14.7	10.3	145.1	29.1
1928	14.2	10.6	153.6	29.1
1929	17.2	11.0	158.2	28.8
1930	11.3	11.2	160.9	29.5
1931	6.0	11.5	143.3	32.4
1932	1.9	10.2	133.3	35.0
1933	2.7	9.1	124.4	37.3
1934	4.6	10.8	122.1	38.9
1935	6.3	11.9	121.3	41.9
1936	9.7	12.6	123.4	45.6
1937	11.1	13.6	123.9	47.2
1938	5.5	14.4	120.2	48.4
1939	9.9	16.0	121.8	50.9
1940	13.0	16.7	125.9	53.1
1941	17.7	26.5	136.3	64.1
1942	7.4	62.0	138.5	109.3
1943	4.0	94.8	141.0	161.8
1944	3.9	99.4	142.6	219.1

* Some of this debt is owed by consumers.

$6.9 billion a year, and private debt fell by $36.4 billion, or by $3.6 billion a year. During those years business firms and private individuals either paid back debts or went into bankruptcy so that a part of the debt was canceled. Between 1940 and 1944, private investment amounted to $46 billion, or $9.2 billion a year, and private debt increased by $20.8 billion, or at the rate of $4.2 billion a year. Thus when private investment was high, private debt grew rapidly; when private investment was low, private debt fell.

This point can be brought out even more emphatically if we call

all the years in which private investment exceeded $10 billion, years of high private investment. We may then call all others, years of low private investment. The average change in private debt in the years of high private investment was $6.2 billion; this corresponds to private investment at the rate of $14.4 billion a year. The average change in private debt in the years of low private investment came to −$3.2 billion; in these years private investment averaged only $6 billion per annum.

Changes in public debt are related in a similar way to the level of public investment. Between 1922 and 1929, public investment amounted to $77.6 billion, an average of $9.7 billion a year, and public debt declined by $0.9 billion, or $0.1 billion a year. During the nineteen-thirties, public investment was somewhat higher — $111.3 billion, an average of $11 billion a year. In this period the public debt rose by $21.4 billion in total, or by $2.1 billion a year. From 1940 to 1944, public investment amounted to $299.4 billion, or to $49.9 billion a year, and the public debt of course increased enormously — by $178.2 billion, or $29.7 billion a year. Thus public debt also rises when public investment is high, and falls when public investment is low.[1]

These relations of course reflect the financial policies followed by government and business. As we have already pointed out, other practices could be adopted. Government could get all the money it needs by taxes, and business firms could meet requirements by selling equities and withholding profits from stockholders. But if such practices were followed, the propensity to consume would be reduced whenever investment was raised.[2] If we suppose that $65 billion of investment a year is needed to bring about full prosperity with existing financial practices, it is evident that an even higher figure, let us say $80 billion, would be required to secure full employment if investing units financed expansion by taxation or equivalent measures.

Since prosperity depends upon a high level of investment, and since when investment is high, debt increases rapidly, we may expect to find debt rising quickly during periods of prosperity. Conversely, we may expect to find that debt rises slowly, or actually declines, during periods of depression. Using data from Table 60, we can

[1] Changes in public debt also depend upon the level of private investment, though inversely; and changes in private debt vary inversely with the level of public investment.

[2] Financing expansion by selling stock would not affect the propensity to consume.

determine the relation between the level of income and changes in the total debt, both public and private. If we designate as years of prosperity all those between 1922 and 1944 in which the gross national product exceeded $90 billion, and all others as years of depression, the following relations come to light: in years of prosperity the total debt increased by $225.2 billion, an average of $22.5 billion a year. In years of depression the total debt increased by only $2.3 billion, or by $0.2 billion a year. If we leave out the war years and consider only the period from 1922 to 1941, the results are similar. When conditions were prosperous, and the gross national product exceeded $85 billion, the total debt increased by $7.35 billion a year; when conditions were depressed, the total debt fell by $0.4 billion a year. Thus it is evident that debt increases quickly during prosperity, and increases slowly or actually falls during depression.

Obviously when prosperity is based on a high level of private investment, as it was from 1925 to 1929, the greatest increase will be in private debt; whereas if it is based on a high level of public investment, as during the war years, the greatest increase will be in the public debt. But in either case, a rise in the total debt is inescapable. Rapidly increasing debt is a price [1] we pay for prosperity. Only a radical change in the methods of financing expansion, or a great increase in the propensity to consume, would create prosperity without rapidly increasing debt in a modern capitalist economy.

This fact has one most important implication. If, short of a radical change in financial practices or spending habits, debt always rises with prosperity, there is no reason to oppose any one method of stimulating prosperity merely on the grounds that it would increase debt. Of course, if an increase in private debt is to be preferred to an increase in public debt, or vice versa, that would be a reason for endorsing one proposal and rejecting another. But that is the only logical basis (considering debt alone) on which to reject a suggested method for achieving full employment — unless we instead urge the adoption of those radical measures which may bring prosperity without increasing debt.[2]

[1] Whether the price is high or low is a question that we shall attempt to answer in Chapter 36.

[2] Some of these measures will be discussed in the next chapter.

Prosperity and Prices

We have already seen why prosperity is likely to be accompanied by unusually high prices. When industry is producing far below capacity, there is little inducement for business firms to raise prices as demand increases, for greater increases in profits are secured by expanding output. But with high prosperity and therefore near-capacity production, most firms are under strong pressure to raise prices when demand rises. Inflation is therefore most likely to occur when the economy is producing at or near capacity level. Prosperity and high prices are likely to go together, but this does not mean that prices are high only during prosperity. Prices were high in most European countries in 1945, but conditions were far from prosperous. The fundamental relation is that between prices and capacity, for when firms are operating close to capacity, prices tend to be unusually high. Unfortunately, capacity production may be far below the level needed for prosperity. If an economy cannot import raw materials, if its factories and industrial equipment have been partly destroyed, or if the efficiency of its labor is very low, capacity output may be very low and yet prices would normally be high. We should expect prices to be high whenever we have high prosperity. They may, however, be high at other times too.

Since prosperity and high prices go together, we cannot hope to achieve prosperity without at least a measure of inflation. Yet it is commonly argued that certain proposals for reducing unemployment would bring on inflation. While we are not yet in a position to pass judgment on any of these proposals, we must remember that the risk of inflation is one of the necessary prices to be paid for prosperity; though the risk may be slight if the banking system and the government can exercise adequate controls. Therefore in weighing the merits of a particular policy to relieve unemployment, we should not object that it would be inflationary unless it would raise prices more than is normal to a period of revival.

Interest Rates, Wage Rates, and Prosperity

Proposals for the relief of unemployment are often condemned on the ground that they lead to higher interest or wage rates. But we must beware lest these objections, like those noted above, should lead us to throw out the baby with the bath water. It is foolish to reject a remedy for unemployment because it brings prosperity. It has

already been shown that the intere t rate is likely to rise in periods of revival. Unless the banking system is willing to accommodate the increased liquidity preference, interest rates are bound to be higher in prosperity than in depression. Therefore, the fact that a remedy for depression would raise the interest rate is not a sufficient justification for rejecting that remedy, unless it were shown that an equally effective one would not cause as large a rise in the interest rate.

Just as prosperity is generally accompanied by increasing debt and higher interest rates, so it is likely to be accompanied by higher wage rates. Any measure that raises employment may also raise wage rates, since labor is generally able to get higher wages when unemployment is low. It is thus illogical to condemn any particular method of curing depression because it makes for higher wages. In short, we must be careful that our objection to certain con equences of an anti-depression medicine do not imply objections to prosperity itself. If we do not want high debt, high interest rates, high wages, and high prices, then in effect we do not want high employment and prosperity. For the conventional means of achieving prosperity generally yield these other developments too.

Natural Prosperity Again

We have already seen that severe depression often cures itself in time. When depression is deep, private investment is very low; and when investment is low, capital goods are wearing out. It took expenditure on repairs and replacements amounting to between $1.5 and $2 billion a year to maintain plant and equipment in manufacturing industries in the nineteen-thirtie . Hence, when expenditures for plant and equipment amount to less than $0.6 billion, as they did in 1932, the existing stock of plant and equipment is deteriorating. The deterioration of factories, equipment, and housing, and the gradual depletion of inventories that goes on during depression, eventually bring about a condition favorable to private investment, so that once more it rises and the economy enjoys a revival. Some people derive comfort from the knowledge that prosperity is bound to come, if only the depression is long enough and bad enough. Waiting is a reliable remedy, and an acceptable one if we are patient. But even that remedy, since it postulates a high level of investment, once it has begun to operate, brings with it a rapid increase in private debt, a rapid accumulation of capital goods, and in addition, higher wages,

prices, and interest rates. In short, even the natural remedy may bring the same troubles as the so-called artificial ones.

Summary

Our economy has experienced a profound structural change in the last three or four decades with the result that the conditions to be met to give us prosperity have become much more stringent. As recently as forty years ago, a few billion dollars of investment a year were all that we needed for prosperity. And that figure was not very different from what it had been during the preceding hundred years. But recently it has been rising quickly — more and more quickly with each decade, until now the enormous figure of $65 billion a year must be spent on investment goods if we are to enjoy full prosperity. An economy in which it was very easy — often too easy — to find the amount of investment needed for full employment has *in a few short decades* become one in which inadequate investment has become a prime difficulty. In other words, the basic problem of the economy is no longer to control inflation, but to control depression. Unfortunately, there is no reason to suppose that the economy is a self-adjusting mechanism, that it has a governor which insures stability at full prosperity.

Prosperity, when we get it, is likely to be accompanied by a rapid increase in the total debt, which rises when investment is high. Thus when prosperity is supported by high private investment, private debt will increase rapidly; when it is based on public investment, the public debt will increase. Prosperity also generally implies high prices, high wages, and high interest rates. And all these things are inevitable.[1] Thus, when specific remedies for depression are opposed on the ground that they will raise the debt (public and private), or prices, or interest rates, the real opposition is to certain characteristics of prosperity itself rather than to a particular way of securing it. It is important to bear this in mind in appraising the desirability of measures designed to cure depression.

[1] Or nearly so.

35

Remedies for Unemployment: Details

"SAVE MORE: spend more; help bring prosperity!" was a slogan widely publicized in 1932 and 1933, years of deep depression. That it was self-contradictory evidently did not upset its sponsor. It is a useful slogan to remember, not because one could follow it without developing a split personality, but because it provides a clear example of the confused thinking then prevalent about depression — and perhaps still prevalent.

What should be done to rid the economy of unemployment? The following is just a sample of the answers that have been seriously suggested: "Make Them Work," "Raise Wages," "Lower Wages," "Raise Tariffs," "Adopt Free Trade," "Balance the Budget," "Soak the Rich," "Cut Hours of Work," "Do Away with the Unions," "Do Away with Taxes," "Have Uncle Sam Spend More." Everyone seems to have his own remedy for depression, just as for the common cold. This confusion of counsel is useless as a guide to action, but it should warn us how difficult it is to make a simple and clear analysis of the remedies for unemployment. The subject itself is complex, as we have seen in our discussion of the ways in which the determinants of income and employment are themselves interrelated (see Chapter 27). Unfortunately, it is not only complex; it is also seldom discussed objectively and without prejudice. It is hard for us to be as coldly analytical about methods of reducing unemployment as about problems in chemistry or physics. Otherwise there would not be such a chaos of remedies. Differences of opinion there might be, but not such violent differences. The basic issues are complicated, not only

by the complexity of the problem, but also by the difficulty of analyzing it dispassionately.

As a preparation for this analysis, it is wise to outline the course of the argument to be followed in this chapter. Unemployment exists when the number of people who want jobs exceeds the number who have them. We can accordingly do away with unemployment either by increasing the number of jobs offered, or by reducing the size of the labor force or the number seeking work. How can employment be increased? As we already know, by increasing investment or the propensity to consume. We shall therefore examine the implications of raising each of the three types of investment, private, public, and foreign, and the problem of raising the propensity to consume. Finally, we shall examine certain proposals for reducing the size of the labor force.

INCREASING EMPLOYMENT

(a) Raising Private Investment

There is no question that increasing private investment is a means of curing unemployment, for to do so will hardly depress public investment, and it is even likely to stimulate foreign investment, as we shall see in Chapter 42. When total investment rises, employment and income rise with it, unless the propensity to consume falls — and there is no reason why an increase in private investment, and thus in the total, would lead to a decline in this factor. Economists do not question the *advisability* of increasing private investment as a remedy for depression. But they are compelled to question how it can be done. Is it possible to maintain private investment at a level which, together with normal public and foreign investment, insures full employment? Can we take effective measures to increase private investment when there is severe depression?

Social control over private investment is obviously imperfect, since it is undertaken by private firms when they believe expansion will pay, and no government body in a capitalist economy can tell a private firm to "expand — or else." The only control over private investment is to create conditions in which expansion looks profitable. But to create these conditions may at times be very difficult — perhaps even impossible.

Private investment and the rate of interest. What are some of the things

that can be done? Obviously reducing the rate of interest would stimulate private investment. Projects on the margin of profitability at one rate of interest would clearly be worth undertaking at a lower rate. There are, however, two difficulties here. First, it is generally believed that the number of marginal investment projects is so small that a reduction in the interest rate would not stimulate any great amount of private investment. If this is so, such a measure would be helpful, but by itself would hardly be a cure for unemployment. The second difficulty is in reducing the rate of interest by any sizable amount. We saw in our discussion of the banking system that the Federal Reserve Banks can encourage member banks to lend more freely but cannot force them to do so. Their control at present must be applied by enlarging the excess reserves of commercial banks. However, banks do not have to lower the rate of interest when their excess reserves rise. Our existing institutions do not give the government the power to reduce the interest rate as far or as quickly as may be necessary to maintain private investment. Hence we cannot expect purely monetary measures that affect the economy through the rate of interest to secure prosperity by themselves. Indeed, many economists think that we cannot expect them to be any more than slightly effective; other measures also would have to be adopted.

Private investment and tax deterrents. If there are tax or other deterrents to business expansion, their removal would obviously encourage private investment. The difficulty in this connection is to identify the real deterrents. Do high taxes on corporation profits keep private investment low? Do high wages discourage private investment? Would the destruction of the trade unions boost private investment? It is difficult to answer these questions fairly. Even if we should agree that, for instance, government regulation of security sales created obstacles to expansion, there may still be ample grounds for retaining this regulation.[1] Obviously private investment would be helped if we did away with every social control that hampered it. But this is not sufficient reason to abolish all such controls. There are other economic and social objectives to be sought in addition to the maximum of profitable investment opportunities to business firms. Removal of government deterrents to private investment would, of course, help employment, but for a variety of reasons we may want to keep certain regulations on the books.

[1] We select this control only for illustrative purposes; this is not the place to pass judgment on it.

Private investment and incentive taxation. More positive steps are sometimes possible. In some economies, special tax concessions are granted to firms that undertake investment projects. Thus, while a firm that does not invest has to pay taxes at the full rate, one which builds a new plant or installs new equipment has its tax rate lowered. Firms are thus induced, or perhaps it is more accurate to say paid, to invest. Presumably this practice could be followed to an unlimited extent. Indeed, the tax rate could be made negative; that is, the government could make a positive payment to a firm that was willing to expand. If the payments were made large enough, of course, a firm could be persuaded to do almost anything, even to build a completely new plant when its old plant was idle. There might, however, be certain social objections to this procedure. Since the firm would own the plant on its completion, this means that the government would have arranged for its purchase on especially low terms or would have donated it to the firm — that is, to the stockholders. It reminds one of the policy followed during the depression of the nineteen-thirties when the government's food-stamp plan enabled low income families to buy food at lower than market prices. Incentive taxation for private investment would be much like a food-stamp plan for stockholders. But since we usually do not regard the stockholders of our corporations as requiring government relief, incentive taxation in an extreme form would look to many people like a dole for the wealthy.

Other positive steps might also be taken to stimulate private investment. Encouraging invention would help, and, perhaps even more important, encouraging the *adoption* of improved methods of production. So would finding larger markets for our products. Measures to reduce the cost of new capital goods — of newly built houses, for example — would also encourage private investment. The list could undoubtedly be extended considerably, but the range of possible remedies has at least been indicated.

Maintaining high private investment. Certain difficulties of maintaining private investment at a high level have already been mentioned. When private investment is high, privately owned capital goods accumulate rapidly; and when they do, the marginal efficiency of capital falls quickly unless the offsets are extremely favorable. And when the marginal efficiency of capital declines, private investment falls and depression sets in. Hence we can hardly expect that measures which boost private investment for a while will continue to be effec-

tive for long. To maintain private investment at a high level might require ever more rapid reductions in the rate of interest, ever more favorable tax considerations for firms that invest, and ever faster reductions in the cost of capital goods.

We saw in the last chapter that there are two obstacles to prosperity in a wealthy economy: such an economy needs an enormous volume of investment in order to reach full employment; and because its stock of capital goods is already great, it does not always provide a large number of favorable investment opportunities. We must now add a third difficulty: it cannot long maintain full employment through high private investment, because its stock of capital equipment will rise rapidly toward the danger point even if it is low enough at first to provide sufficient profitable outlets for investment. In our own economy, for instance, the stock of capital equipment is now so great that private investment would have to reach the rate of $40 to $45 billion a year in order to provide full prosperity, if we assume government and foreign investment of about $20 to $25 billion. If private investment were done on such a scale, the stock of privately owned plant and equipment would increase at a tremendous rate. We can see how big these figures are by recalling that in 1938 our total railroad wealth was estimated at about $24 billion, and the total value of our privately owned electric light and power plants at about $12.5 billion.[1] This means that if we are going to rely on private investment for prosperity, we shall add to our capital wealth each year not much less (after providing maintenance) than our total wealth in railroads and electric power plants in 1938. But with such an enormously high rate of accumulation, the unexploited opportunities for profitable investment would diminish rapidly as the new wealth piled up. Measures which stimulate private investment in one year would probably be less effective in the next. To maintain the investment of business firms at a high enough level may therefore require ever larger stimulants to private investment. While this does not mean that we cannot hope for continued prosperity based on private investment, it does mean that we are not likely to get it. It seems to be an almost impossible task to raise private investment to the astronomical figure that is now needed; and an even harder problem to keep it there. But this does not imply that we must have depression.

[1] In 1938 prices.

(b) Raising Public Investment

Because public investment,[1] unlike private investment, is directly under social control, it is easy to turn on the tap and to turn it off again. While control over private investment is likely to be spotty and uncertain, because it is difficult to raise and very difficult to maintain it at a high level, public investment can be maintained at any rate we please — at $10 billion a year or at $100 billion. The money has to be appropriated by legislative action, and spent; and there is the investment.[2] Thus we can do as much public investment as we please. But are there no limits to the amount we should do? Can we raise income and employment by increasing public investment? And even if we get a temporary rise in income, do we not store up added difficulties for ourselves in the future?

These problems are quite different from those discussed in connection with private investment. We accepted without question the advisability of raising private investment as an anti-depression measure; our problem was *how* to do it. In contrast, there is no economic problem about how to increase public investment; this is determined directly by government bodies. But there is a question about the *advisability* of increasing it. It looks as if we can't do what we should, and perhaps we shouldn't do what we can. Specifically, does increased public investment raise income and employment? Must the investment be in the form of self-liquidating public works? When the government spends us into prosperity, are we not also paying cash down for a future depression? These questions deserve consideration, and these are the ones we shall analyze below.

Increasing public investment and income. First, does an increase in government investment lead to a rise in income and employment? Does it make any difference whether the project is useful or not? An increase in government investment means that the government is spending more than before for goods and services. This may be done by hiring more school teachers or policemen, by buying more aircraft or naval vessels, by building more parks or highways, or by giving larger contracts to private firms for the construction of dams for flood control. The government may undertake the project itself, as with

[1] Public or government investment here refers to that made by all types of government units — federal, state, and local.

[2] This, of course, neglects certain technical difficulties in spending it, which may be very serious.

the WPA and the public school system, or it may simply finance it — buying commodities produced by private firms under contract. If the government pays wages and purchases raw materials directly, its expenditures immediately become added income for those who receive payment, so that if the government increases its investment expenditures by $1 billion, incomes earned in the production of investment goods increase by that amount. Likewise, if the government undertakes projects to be carried out by private firms, the firms have to hire more men, so that wage incomes increase and the owners of the supplying firms earn higher profits. Again the increase in income exactly equals the increase in the government expenditure. And here begins the cumulative process which we have traced through before. Increased incomes allow more spending upon consumers' goods. Naturally this is true whether the wage earners who benefit are hired by the government or by private firms. Then the increased spending on consumers' goods yields more income to those who produce them. This in turn leads to still more spending. In short, the multiplier effect is now well under way.

The total increase in income would depend, as we have seen, upon the marginal propensity to consume. With a marginal propensity to consume of two-thirds, and accordingly, a multiplier of three, a $1 billion increase in government investment would increase income by $3 billion, if there were no other change in investment to be taken into account. And generally the greater the marginal propensity to consume, the larger would be the increase in income, other things being equal, when the government increases its investment.

But there is another factor which must also be considered. When government investment changes, it may lead to changes in private and foreign investment. For example, when government investment is increased by $2 billion, private investment may rise by $1 billion as a result of the larger market. If so, the total change in investment against which the multiplier must be applied would be $3 billion, assuming no induced change in foreign investment. Or, alternatively, when government investment is raised by $2 billion, private investment may be discouraged and may fall by $0.5 billion. In that case the total increase in investment would be only $1.5 billion. Hence we may conclude that the increase in income brought about by a given increase in government investment is greater the higher the multiplier and the more favorably private and foreign investment respond to an increase in government investment.

These conclusions are illustrated in Table 61.

TABLE 61

Increase in Income Resulting from a $2 Billion Increase in Government Investment
(in billions of dollars)

When induced private investment is	2	1	0	−2
When multiplier is				
3	12	9	6	0
2.5	10	7.5	5	0
2	8	6	4	0

The multiplier for government investment. There is no reason to suppose that the marginal propensity to consume is abnormally low for government investment expenditures; indeed, if anything, it is likely to be unusually high. This is true for several reasons. Government employees and businessmen whose profits derive in part from government funds are just as likely to expand their consumption when their incomes rise as anybody else. When the government undertakes investment projects as an anti-depression measure, it generally provides that a relatively large proportion of the total expenditures should be paid for wages. On such projects, a minimum of expensive machinery is commonly used in order to channel as large a part as possible of the total expenditure to wage earners. And as we have already seen (Chapter 30), the marginal propensity to consume is likely to be unusually high when increased income goes mainly to the low income groups. In contrast, a private firm is forced to carry out its project at the minimum cost without regard to the employment-creating potentialities of its investment activities. Consequently, a relatively large part of its total expenditure is likely to go for expensive but efficient equipment. Hence a part of this sum is earned as profit by those who supply the equipment and a relatively smaller portion of the total goes directly to wage earners. Therefore, we should expect the marginal propensity to consume to be somewhat lower for private than for public investment projects, though it is not likely that the difference will be very great. If this is so, an increase in public investment would lead to a somewhat larger increase in the national income than an increase of the same size in private investment.

Government investment and the acceleration factor. Two forces, operating in opposition to one another, influence the reaction of private investment to changes in government investment. An increase in public investment implies increased orders for the products of many business firms — for the raw materials and equipment required for the project itself and for consumers' goods demanded by those whose incomes have increased. For instance, when government builds highways it must purchase concrete, tar, gravel, mixing machinery, rollers, and so on, so that the supplying firms enjoy an increased demand for their products and may in consequence be led to expand capacity. Moreover, men given jobs on the project or in the supplying firms have higher incomes than when they were unemployed, and they are led to purchase more food, clothing, and other consumers' goods. Increased sales by the suppliers of consumers' goods may lead them to expand also. Thus, when government investment rises, the force that we have previously identified as the acceleration factor (see Chapter 29) comes into play.

But there is an opposing force to be considered, the vague but potent quality of confidence. It is commonly alleged that an increase in government investment reduces the confidence of the business community, and that when confidence falls, private business firms are less inclined to expand. The force of this argument is difficult to appraise. There is no apparatus available for testing business confidence, and even if there were, we should probably not know how much a drop in confidence of, let us say 10 per cent, would influence investment. Nevertheless, it is reasonably clear that because of the reduction in confidence, the acceleration factor is lower for government than for private investment: which means that an increase of $2 billion in government investment will induce a smaller rise in private investment than an equal increase in private investment would induce. Mathematically, at least, the acceleration factor could even be negative, for it is conceivable that when government investment is raised by $2 billion, private investment would decline, perhaps by as much or even more because of the loss in confidence.

Unfortunately, analyses of existing statistical data do not point to a definite conclusion about the relation between changes in public and private investment. Public investment increased from $9.1 billion in 1933 to $13.6 billion in 1937, and private investment rose from $2.7 billion to $11.1 billion in the same period. But the rise in private

investment does not actually prove that the acceleration factor for public investment was positive. We certainly cannot conclude that the whole of the increase in private investment was brought about by the increase in public investment, for other factors, such as the development of new methods of production, were also operative in the period. We might even argue — at least, the statistics at present available would not contradict us — that if public investment had remained constant after 1933, private investment would have risen even more rapidly, let us say from $2.7 billion to $15.6 billion. But as with discussions in other fields where proof is as yet impossible, it is a waste of time to do much more than indicate the issues. The author's opinion, however, is that a rise in public investment normally encourages private investment. Business confidence is indeed affected adversely, partly because firms expect higher taxes; but this is much more than offset by the growth in their sales, so that when public investment rises, private investment is generally stimulated. On the other hand, some economists, and certainly many politicians, believe that a rise in public investment is self-defeating and leads to an equal reduction in private investment. To one who accepts that extreme view, one query may be directed: When government investment is raised from $15 to $50 billion what happens to private investment? If it was originally $2 billion, does it fall to −$48 billion? If not, and indeed it could not, such a rise in government investment would increase total investment.

Public investment and income: Summary. Unfortunately, we cannot precisely say how an increase in public investment affects the national income and employment. Much depends upon the value of the multiplier and the acceleration factor for government investment. If, as this writer believes, the acceleration factor is positive, an increase in government investment would lead to a rise in the national income and employment. Only if it is negative and very large would increased government investment cause a decline in income and employment; for so long as total investment goes up when government investment is increased, the national income would rise. Hence we may conclude that in the short run it is possible to increase employment and income by raising government investment. Indeed, our experience between 1940 and 1944 strikingly illustrates this conclusion. In 1940 public investment was $16.7 billion, and the gross national product was $97.1 billion. By 1944, public investment was $99.4

billion, or $82.7 billion higher; and the gross national product was $197.6 billion, or $100.5 billion higher than in 1940. It is impossible to account for the rise in national income and employment in any other way, particularly when we recall that private investment and consumers' expenditures were severely curtailed by government action. If we take enough of this public investment medicine, it appears that we can cure any depression, so long as we are willing to keep on taking it. Incidentally, if we cut government investment at such a time, we are likely to reduce income even further and to lower the government's receipts from taxes. Then we can boast:

> And now the budget's balanced,
> Retrenchment is the hero.
> On either side is entered
> A solitary Zero.

Useful and useless public projects. The nature of the investment project does not seem to make much difference to the result. To put it bluntly, employment and income, in money terms, can be expected to respond equally whether the government sponsors useful public works like highway construction, or completely useless ones like digging ditches and filling them up again. In either case, because the income of the newly employed would be higher than before, they would increase their spending, so that the output of consumers' goods would be expanded and the upward swing begun. Naturally we should prefer projects which directly add to our real wealth. Flood-control projects, highways, parks, school buildings, research projects, housing, and so on, are better than leaf-raking and useless excavations. But the latter are better than nothing, for even though the projects are useless, carrying them out leads to an increased output of consumers' goods. And even though the men responsible for the increased demand were idlers and good-for-nothings, their dollars, in our economy, are as powerful as any others in increasing consumption, income, and employment.[1]

The long-run effects of public investment. To conclude that an increase in public investment would bring about a rise in employment does not necessarily mean that we should endorse that remedy. Such a policy may have adverse consequences in the long run, for it may

[1] The reader may care to work out for himself how employment is affected if the more wasteful projects put all the sums spent into the pockets of wage earners, or if they lead to an unusually large (or small) decline in business confidence.

secure temporary prosperity only at the cost of future depression. Certainly this argument deserves consideration.

Two principal reasons are given to support the conclusion that a high level of public investment leads eventually to depression. It is argued, first, that when public investment is high, public debt increases quickly. We shall postpone until the next chapter a detailed analysis of the effects of an increase in the public debt. It may be well at this stage, however, to remind the reader that the alternative route to prosperity — an increase in private investment — also brings an increase in debt, though to be sure in the private rather than the public debt. But why should it be better to increase the private than the public debt? Unless a clear answer can be given to this question, we may either have to choose between putting up with the problems of increasing debt or putting up with depressions.

The second objection to a large volume of public investment as a remedy for unemployment is that this measure merely postpones depression. It is argued that so long as we are building highways, schools, and bridges, and hiring teachers, policemen, and economists, we can enjoy prosperity, but that a time will surely come when we shall have built all the schools and bridges we need — and when that happens, public investment will fall and we shall experience deep depression. There are several answers to this argument. In the first place, there is no real danger of running out of worth-while public projects. To supply proper housing, medical and educational facilities, highways, airports, harbors, and so on, would require a great number of years of high public investment. Moreover, even if we should complete all useful projects — in other words, even if we should satisfy all demands for socially owned investment goods — the government, unlike a private firm, can undertake employment-creating projects simply to provide jobs. Finally, if it is objected that prosperity based on public investment must come to an end because we eventually must reach a point where we want no more public investment goods, cannot the same point be raised about prosperity based upon private investment? Indeed, it is very likely that private investment will decline even more swiftly as privately owned capital goods accumulate. In fact, it may be said of the objection to government investment on the ground that it merely postpones depression that it really amounts to saying that prosperity itself is bad because it may be followed by depression.

Private versus public investment. To argue, as we have done, that public investment, if it is extensive enough, can cure depression does not imply that measures for raising private investment should be rejected. Both kinds of investment may be desirable, and normally we do not have to choose between them. The objective should be to raise total investment, and this ought to be easiest to accomplish if steps are taken to raise investment in every category. Measures to increase private investment do not compete with an increase in public investment. The two are complementary, and in depression it will usually be necessary and desirable to increase both. This does not mean that they should be given equal priority in all circumstances. Sometimes it is more important to increase private investment simply because society is more in need of privately owned capital goods than of public goods and services. At other times, the need is greater for public investment projects, as during a war when private business firms in non-war industries are usually even prevented from building new plants. And at any time, many kinds of public investment, such as schools and school teachers, highways, and police services, are at least as important to us as a new plant for the production of, let us say, bubble gum or luxury jewelry. On the other hand, there are times when one new steel plant may be more useful to us than a hundred new national parks. In our economy, of course, we can usually build bubble-gum factories, schools, steel plants, *and* national parks. But though there may be no competition between them, there may still be reasons for preferring one project to another, and thus for taking steps to encourage it, whether it is sponsored by a government unit or by a private firm.

(c) Raising Foreign Investment

The proposal to increase foreign investment is mentioned at this point simply to complete the analysis. We are not at this stage prepared to analyze how it may be accomplished or to discuss the difficulties involved. These questions will be considered at length in Chapter 42. There is no doubt, however, that an increase in foreign investment would lead to an increase in employment and the national income, unless it caused adverse changes in private or public investment or in the propensity to consume. For if firms are able to sell more to foreigners, they hire more men, and so generate more wage income, just as they would if they had increased their sales to domestic

firms or to the government. An increase in foreign investment helps to raise employment.

(d) Increasing the Propensity to Consume

With full employment, we stand ready to consume a little more than five-eighths of what we produce. Out of a $200 billion gross national product, our consumption expenditures amount to about $135 billion. In order to reach full employment, therefore, we must sell the other three-eighths [1] of our output, or $65 billion worth, to non-consumer buyers — that is, to purchasers of investment goods. The difficulties of finding markets year after year for $65 billion worth of investment goods have already been noted, and they are obviously very serious. But the only reason these difficulties exist is that we, as consumers, are not willing to consume more than five-eighths of our full-employment output, and hence it is necessary to find other buyers for the remainder. Since, as consumers, we are willing to purchase only five-eighths of our full-employment output, we have a low-consumption economy. And that, from one point of view at any rate, is one of the main sources of our difficulty.

Would the difficulty be as serious if ours were a high-consumption economy? Would it be as hard to satisfy the conditions for full employment if, as consumers, we bought as much as seven-eighths of our full-employment output? Obviously not. If consumers' purchases absorbed seven-eighths of our full-employment output, it would be necessary to find non-consumer buyers for only one-eighth of the total, or about $25 billion worth of goods a year. This should ordinarily be a comparatively easy task, although in passing we should note that in seven of the twenty years between 1921 and 1940, we failed to sell to non-consumers even as much as $20 billion worth of goods.[2] Certainly, however, if we had a high-consumption economy, the attainment of full prosperity would be very much easier.

Not only would it be easier to achieve full employment if our propensity to consume were high; it would also be easier to maintain it. As we have seen, when there is full employment in a low-consumption economy, capital goods accumulate very rapidly, and this rapid accumulation discourages private investment. Hence, in a low-consumption economy, prosperity is likely to give way to depression.

[1] Approximately.

[2] The score would not be so bad if values were expressed in 1945 prices.

But in a high-consumption economy, relatively little investment will provide full employment, as is shown by the figure of $25 billion a year with the high propensity to consume in contrast to the $65 billion we require with our present low propensity to consume. Therefore, investment goods would accumulate much less rapidly and the offsets to this accumulation would have more time to become effective. Hence we should expect the task of maintaining full prosperity to be relatively simple. Furthermore, debt would grow much less quickly with full employment in a high-consumption economy. We have already seen that the greater the investment, the greater is the increase in debt. If we needed to invest only $25 billion a year to attain full prosperity, our total debt, public and private, would grow at the rate of perhaps $8 to $10 billion a year; whereas if we have to invest $65 billion a year, the total debt might grow at the rate of $40 to $50 billion a year. Thus a high-consumption economy would avoid the burden of heavy debt — if there is such a burden — or at least would postpone it.

Suppose we compare total investment to the pressure put on the accelerator of an automobile and the propensity to consume to the design of the engine. With an engine of the low-consumption type, the accelerator must be pressed all the way to the floor to gain a speed of sixty miles an hour. With an engine of better design, the accelerator need go only part way down to gain the same speed. If we further suppose that when either engine is raced at full speed for a time, its efficiency falls, the analogy is reasonably close. So if we want to travel a long way at a high speed, perhaps it would be better to redesign the engine than to prescribe exercises to strengthen the leg that must put pressure on the accelerator.

Measures to raise the propensity to consume. How can we raise the propensity to consume? Can it be raised enough to make any real difference in the amount of investment needed for full employment? Finally, is there a danger that measures to raise it will discourage investment? In Chapter 30 we discussed the factors upon which the propensity to consume depends, and by implication, therefore, the steps that would be necessary in order to raise it. Briefly, they are: distributing income more equally, reducing incentives to thrift, doing away with taxes that reduce consumption, reducing imports, and encouraging corporations to pay out more of their profits in the form of dividends.

But while these measures would undoubtedly raise the propensity to consume, it is probable that some of them would also adversely affect investment and thereby nullify some of the favorable effects of the increased propensity to consume. For instance, there is no doubt, as we shall see in Chapter 42, that our exports or our foreign investment would fall if we were to reduce our imports. It is also possible that vigorous steps to equalize income distribution would adversely affect private investment. Whether it would do so or not depends upon several factors — such as the favorable effect on investment of a growth in markets, possible adverse effects of very high taxes on high income brackets, and so on. There is no reason to suppose that doing away with high taxes on consumption would discourage private investment; indeed, the reverse is more likely. Hence there is every reason, when we are faced with unemployment and depression, for doing away with general sales and excise taxes, general property taxes, and payroll taxes. For there is no doubt that all of these bear very heavily upon consumption.

It is impossible to estimate how strongly the propensity to consume would respond to these changes. Much would depend upon the extent to which habits of thrift, both of the consumer and of the corporation, could be altered. It is not likely that a more equal distribution of income and the abolition of taxes which discourage consumption would increase consumption by more than $10 billion a year with a full-employment income — that is, from about $135 billion a year to about $145 billion. While this would help, it obviously is not a complete answer. It is improbable that any practicable change in the propensity to consume would solve the problem completely, though it would certainly bring the solution much closer.

Expansion in a high-consumption economy. A high-consumption economy suffers a drawback that may at times prove serious. The very feature that gives it merit when it is faced with unemployment may in other circumstances be harmful. The fact that relatively little investment is needed to bring such an economy to full employment signifies that investment at a more rapid rate threatens it with inflation. Therefore, in a high-consumption economy the maximum rate at which investment can safely be done is relatively low. This is in sharp contrast to the situation that prevails in a low-consumption economy, where a great deal of investment is needed for full employment, and consequently a very rapid rate of expansion is possible

without danger of inflation. In sum, a high-consumption economy cannot expand quickly, and must be content with a slow accumulation of capital goods, while a low-consumption economy can expand much more quickly and may accumulate capital goods very rapidly. This does not mean that a low-consumption economy *will* necessarily expand more quickly than a high-consumption economy. For though the maximum rate of expansion is indeed higher, obviously an economy does not always expand at its maximum rate. We could have accumulated capital goods at a rate of $40 to $50 billion a year during the nineteen-thirties, but we did not do so. We must be careful, therefore, to distinguish between the rate of expansion which an economy can achieve and the rate which it actually does achieve. Indeed, there is good reason to suppose that a high-consumption economy would normally expand more rapidly than a low-consumption economy, since its relatively high and steady level of consumers' purchases would prove a greater incentive to private investment than the low and unstable level in a low-consumption economy.

When strong independent forces maintain investment, and especially when it is important to produce an immense volume of investment goods, the low-consumption economy has the advantage. Thus, when it is necessary to produce munitions very rapidly, it is easier in a low-consumption economy to do so. In a high-consumption economy, the danger of inflation soon becomes very great, and steps must be taken to reduce the propensity to consume. The investment opportunities may, of course, occur in other fields than the production of munitions, such as in exploiting an important technological development. Obviously the society which can most quickly take advantage of these developments will be better placed than the one in which investment has to be restricted to avoid inflation.

Summary: Increasing Employment

To increase employment and maintain it at a high level in our economy is not easy. It is therefore doubtful whether there is any single cure for depression. Severe depression would have to be treated by our whole armory of weapons: raising private investment as far as possible, raising the propensity to consume, and increasing public investment. Measures to achieve this are not competitive but supplement one another. Of these, raising private investment may prove the most difficult, and raising the propensity to consume by any con-

siderable amount may also be hard, while increasing government investment should be relatively simple.

There is no doubt that an increase in private investment would lead to an increase in income and employment in the short run. But difficulties may have to be faced in the future because of the accumulation of privately owned capital goods which this entails. To raise the propensity to consume will almost certainly be favorable in the short run, and will avoid many of the long-run problems created by high private investment. Raising government investment will very probably raise income and employment, though perhaps by less than an equivalent increase in private investment. The long-run effects of an increase in government investment do not appear serious. There is no reason for rejecting any of these methods of achieving full employment; what we should reject is any policy that implies a toleration of unemployment.

REDUCING THE SIZE OF THE LABOR FORCE

Unemployment is measured by the difference between the number of people who want jobs and the number who have them. When 60 million men are willing to work — that is, when our labor force is 60 million strong — and there are only 53 million at work, 7 million are unemployed. The arithmetician can see two ways of reducing unemployment: either we can increase the number of jobs available or we can reduce the size of the labor force. We have already discussed the ways of bringing about the first of these results. We shall now consider how the labor force may be made smaller, and whether making it so would be a good solution to the problems of unemployment.

It is not hard to reduce the size of the labor force. In periods of depression, suggestions for doing it are numerous, and range from such ingenious schemes as "Settle them in Alaska," through "Don't let married women have jobs if their husbands are working" and "Don't let aliens work," to "Reduce the hours of work in the workweek" and "Give longer vacations." There are strong social reasons for rejecting some of these proposals outright, though there may be good social reasons for accepting others. We may not want to resettle the unemployed even though it would reduce their numbers, and we may want to reduce weekly hours of work quite apart from any effect

this would have on unemployment. But in relation to the problem under discussion, all these remedies are alike in this respect: they attack the problem by reducing the size of the labor force. Cutting the work-week from 48 to 40 hours would effectively reduce the labor force by 16 per cent. Making annual vacations of two weeks compulsory would reduce it by 4 per cent. Shipping 5 million men and their families out of the country would reduce our labor force by about 8 per cent. These and similar measures all have the same goal.

But will they always cure unemployment, even in this purely arithmetical sense? Not necessarily. It will be obvious that unemployment may remain high if the number of jobs that are open also declines because of our policy. If, as the size of the labor force were reduced from 60 million to 55 million, the number of jobs fell from 50 million to 45 million, unemployment would not be reduced but would remain at 10 million. True enough, the result might not be so drastic. Nevertheless, it is likely that if some of these measures were followed, the number of jobs would decline, particularly if our policy not only removed people from the labor market but also from the ranks of consumers. Shipping the unemployed out of the United States would certainly reduce the number of hands that could work, but it would also reduce the number of mouths to feed and hence the amount of labor required to feed them. In short, certain methods of reducing the size of the labor force would at the same time reduce the volume of employment. Hence unemployment might remain nearly the same.

Of course unemployment would be reduced if we could reduce the size of the labor force without reducing the size of the market for our products. One way to do this would be to reduce the hours of work. Thus, if unemployment were very severe, we could bring it to an end by limiting the working hours to perhaps twenty a week.[1] Would a reduction in the size of the labor force achieved in this way be an

[1] This is not to say that the amount of employment — man-hours needed to produce the output — would not be affected at all; but the effects would certainly be much smaller, and possibly they would even be in the opposite direction from the changes in the size of the labor force. The redistribution of income within the ranks of labor, possible shifts of income from employer to employee or vice versa, and probable changes required in methods of production, would all affect the amount of employment in man-hours. If half the labor force were unemployed, it might be necessary to limit the hours of work to 40 per cent or to 60 per cent of the original figure. But by *some* reduction in hours, we could cure unemployment. We confine our analysis of the effect of changing hours, as noted above, to the problem of unemployment. There may be other good reasons for supporting or rejecting such a move.

acceptable cure for unemployment? Before we can answer this, it is necessary to recall our objections to unemployment. They were of two kinds.[1] First, heavy unemployment deprives society of a large part of the output of goods and services which it could otherwise enjoy. When employment is 15 million below the peak, society is deprived of goods and services worth perhaps $100 billion. Some of us might want our share of this output — the new home and furnishings, the automobile, and so on. And second, the man without a job loses morale, suffers extreme poverty, loses working skills, is deprived of fair opportunity for his children, and so on.

Would a reduction in the hours of work per week, by ridding us of unemployment, effectively answer these objections? It would not to any great degree make available the automobiles or houses we could have had if employment in the sense of man-hours of work were raised. We should produce perhaps the same output with 60 million men working 25 hours a week as with 45 million working 40 hours a week. But if we produce the same output, it is clear that society in the aggregate is no better off.[2] We do not come into possession of the $100 billion worth of goods of which we are deprived by unemployment. All that happens is a redistribution of the total output. The amount is unchanged, but different people get it. Hence the first objection to unemployment is not really answered. We are, in total, as impoverished as ever. We have merely redistributed the poverty.

And how have we redistributed the poverty? Workers who were employed before get the greatest share of it. Their hours of work are reduced from 40 to 25, and their weekly pay is lowered.[3] Since these men are actually sharing the work, what they lose the previously unemployed will gain, for their hours of work are raised from zéro to 25. What some workers gain, therefore, is equaled by what others lose. Society in the aggregate is no better off than before. The previously unemployed, however, are now freed from some of the disabilities that go with unemployment. They are raised from extreme

[1] We omit from this discussion the international consequences of depression. They follow from what we have called our first objection, and they would only be removed to the extent that this difficulty is eased.

[2] Except because of the increased leisure.

[3] Even if hourly rates are raised in compensation, workers would probably find that their real income — the actual food and clothing they could have purchased — was lowered, for prices would be increased. This point is discussed in Chapter 38.

to moderate poverty. Since they can now practice their skills, their work efficiency is maintained. They no longer need bear the stigma of being on relief. Thus, our second objection to unemployment is answered. Reducing weekly hours of work, while it does not solve one very important problem of depression, does solve the other. Spreading the effects of depression more widely among workers effectively does away with the unemployed. But because it permits the same total amount of depression — since total income remains at depression level — there is still needless suffering. We are still unable to enjoy capacity production of electric irons, new dresses, nylon stockings, and new houses. And this means that we continue to export depression to other countries.

Hence we may conclude that reducing hours of work, or in other ways reducing the size of the labor force, is not an effective remedy for unemployment. It is far better to increase the demand for labor than to decrease the supply. Greater employment means a larger output and a higher standard of living for all; freezing total employment (in the sense of man-hours of work) and reducing the labor force does not raise our standard of living; it merely spreads an inadequate output among more people.

There comes a time, of course, when society gains from reducing weekly or annual hours of work. This point is reached when, with full employment at full hours, output is so high that society prefers more leisure to a part of that output. Such a move, however, is not motivated by a desire to cure unemployment, but rather to increase leisure when we want it more than we want the added income that would otherwise be ours.

Summary

The upshot of the analysis in this chapter and the preceding one is that unemployment can be cured. It is not something like an eclipse, which can be predicted but not stopped. We can either raise employment, or, in the absence of a better remedy, we can reduce the size of the labor force. Raising employment is not easy, and the task is made more difficult because of the man-made taboos which stand in the way. But it was done after 1940 by immense public investment in munitions. Obviously, though, this is not the only way, and generally not the best way; for certainly the public investment, if we are unable to work out more attractive remedies, such as

raising the propensity to consume, does not have to be in the form of munitions production. If, for instance, the government were to clear the slums, employment would respond just as favorably.

It is often alleged that we cannot afford to relieve unemployment because it costs too much to do so. This charge is difficult to answer because it is hard to understand the sense in which the term *costs* is here used. But it is easy to understand that the costs of depression are immense and avoidable, and no economic sacrifice which incidentally makes us all better off is too great if it avoids these costs.

Our knowledge of how capitalism works shows us that we can prevent that suffering.[1] And we have certainly found no reason to conclude that we have to scrap the system to do it.

[1] There are, of course, very serious political difficulties involved. We may not be able to work out a scheme compatible with our political institutions by which government or private investment could be raised quickly enough to offset a sharp decline in inventory investment.

36

The National Debt: A Digression

PERHAPS the advice of old Polonius to his son that he should "neither a borrower nor a lender be" is not good advice for the government, though it is generally urged. But it is obviously a rule which we all profess to follow, for no one likes to owe money. The debtor has always been a figure of scorn, a person of questionable virtue, or a weakling. "Stay out of debt" is a maxim of conduct which we all accept, even though most of us fail to practice it. What we feel to be true about the individual and debt applies equally, we believe, to business firms and to governments. And yet, though most of us do not like being in debt, an examination of the real world shows that a great number of individuals are in debt, that almost all business firms are in debt, and — as no one today needs to be told — that the federal government is heavily in debt.

Prevalence of Debt

One would suppose that because of the prevailing attitude toward debt, most individuals would refrain from borrowing. But do they? Most families that own their homes have borrowed at least a part of the funds to finance the purchase. Most people buy washing machines, automobiles, vacuum cleaners, and radios on the installment plan. Most families have charge accounts with a department store or with the grocery store on the corner. In fact, the total personal debt of Americans amounted to about $43 billion in 1940, or a little more than $1000 a family. Apparently, even though we think of the debtor as

misguided, we are generally willing to borrow and to become debtors ourselves.

What is true of us as individuals is equally true, and even more striking, of business firms. Most firms in the country have large debts outstanding, and some of the biggest of them are very heavily in debt. The United States Steel Corporation owed its bondholders alone almost $250 million in 1937. The American Telephone and Telegraph Company owed its bondholder-creditors more than $1 billion in 1938. The Standard Oil Company (New Jersey) owed its bondholders about $170 million. Yet none of these firms is regarded as a financial weakling. In 1940 the total debt of all American corporations, not counting what they owed each other, amounted to about $60 billion. But does this figure, immense as it is, make us fear for the strength of American business? Do we really anticipate the widespread failure and bankruptcy of American business firms because they owe so much? Certainly not.

How, then, is it possible for a firm which is heavily in debt to escape bankruptcy? The answer is obvious. So long as its assets exceed its liabilities, it is perfectly safe. The American Telephone and Telegraph Company, with its debt of more than $1 billion, had assets of more than $3.5 billion to set against this figure. The United States Steel Corporation, with its debt to bondholders of about $250 million, had assets of nearly $2 billion. These firms are actually much better off, and much further from bankruptcy, than the corner drugstore which owes $14,500 and has assets of only $15,700. So long as a business firm possesses sufficient wealth to satisfy its creditors, its position is secure.[1] Indebtedness on the part of individuals or business firms is not necessarily an indication of either moral turpitude or a weak financial position. What is important in determining the financial strength of a debtor is the relation between his debts and his assets.

This relation is the generally accepted standard used in judging the financial position of a private firm or an individual, but the debt of the government [2] is generally regarded in quite a different light. Nearly everyone complains about the government or public debt, and

[1] This wealth must, of course, be in a form which creditors will accept, or it must be possible to turn it easily into such a form.

[2] In this chapter, we shall generally concern ourselves with the debt of the federal government alone. The debt of state and local government bodies is relatively unimportant in the total.

very few of us ever try to compare the assets of the government to its debt. The debt of the federal government — that is, the national debt — is believed to impose a heavy burden on the economy in all circumstances, a burden quite different from that imposed by private debt. And this is believed to be true whether assets are available to set against the debt or not. Yet some of our most fervently held beliefs sometimes merit re-examination. In this chapter we shall endeavor to determine the precise nature of the difficulties which our economy must meet because of its public debt.

The United States public debt has increased with great rapidity. In 1914 it stood at only $1.2 billion. By 1919 it had risen to $25.5 billion; by 1929 it had fallen to just below $17 billion. It increased again during the nineteen-thirties, and by the end of the decade it amounted to almost $50 billion. And because of the enormous expenditures necessary during the war, it had risen to about $260 billion by August, 1945. Thus, in a little over three decades, the public debt of this country has increased by more than two hundred times. And what is true of the public debt of this country is equally true of the government debt of other countries. For example, the government debt of the United Kingdom in 1913 was only £0.7 billion, and by 1945 had increased to almost £25 billion. The French public debt stood at approximately 30 billion francs in 1914, and by 1944 at over 1500 billion francs. All the important capitalist economies have experienced a steady and at times an extremely rapid rise in their public debts. Since the history of our own and other countries suggests that the public debt, like the poor, will always be with us, it is important to see its exact effect on the economy.

The Debt and Prosperity

At the very beginning of this inquiry, it is important to emphasize a point which has already been developed:[1] an advanced capitalist economy can enjoy a high level of prosperity only when its total debt is growing rapidly, unless a drastic change in spending habits or financial practices can be brought about. With the relatively low propensity to consume which is characteristic of modern economies, a high level of income is possible only when investment is high. Current methods of financing are such that high investment generally

[1] See Chapter 34.

implies heavy borrowing, and thus a rapid increase in debt. Hence prosperity is generally accompanied by rapidly rising debt.

If the bulk of the investment that supports prosperity is undertaken by private business firms, the most rapid increase in debt will be for these firms. This is what happened in the decade between 1920 and 1930. If, on the other hand, the basis for the high national income is provided by the high level of public investment, the public debt will increase very rapidly. And this is what happened between 1941 and 1945.

The fact that either private or public debt must increase rapidly during prosperity suggests not only that we must examine the precise nature of the burden imposed by a public debt, but also that we should compare it with whatever burden a high level of private debt imposes. For though it may be claimed that, other things being equal, a high public debt interferes with the smooth working of the economy, this does not necessarily mean that public investment or a high public debt should be avoided. A large private debt may impose just as great a burden, and if that should be true, we should in a sense be faced with three unwelcome possibilities: severe depression, which nobody wants; a large public debt, which we shall assume hampers the efficient functioning of the economy; or a high private debt, which perhaps does the very same thing. Hence it is not enough to consider the effects of a large public debt. We must also compare their effects with those of a large private debt.

To utter the words "public debt" is sometimes very much like pushing a button in a juke box. Automatically a record starts to play, and the air is filled with words like "inflation," "deflation," "exhausting the national credit," and "national bankruptcy." Obviously, some of these terms are contradictory. A high public debt cannot plunge us into deflation and at the same time reduce the value of the dollar to the vanishing point. Others may be no more than meaningless incantations intended to frighten us. But let us consider each of these possible results.

The Public Debt and National Bankruptcy

Can a high national debt force our government into bankruptcy? Bankruptcy comes about when creditors do not receive interest on their loans or when they are not repaid as their loans fall due. The creditors then have the right to seize the debtor's assets. Is it con-

ceivable that the owners of United States government bonds — for they are the creditors — should fail to receive their interest payments, or the full repayment when their bonds mature? Such a thing would happen only if the government willed it — or, as we could equally well say, if you and I willed it. For the federal government can always secure the funds to pay interest or to repay the principal of its bonds as they mature. If necessary, it can collect very large amounts in the form of taxes, so that its power to tax would save it from bankruptcy. And if it should not wish to impose heavy taxes on the economy at a time when servicing the debt required high expenditures, it can borrow. Only if it could not do so would it be forced into bankruptcy. The only question, then, is whether the government can always find a lender or someone who will accept government bonds. In the final analysis this is no problem, for the simple reason that the government controls the Federal Reserve Banks and can always compel them to buy government bonds. Anyone who controls a bank and is free to make the rules under which it operates will have no trouble in borrowing money. The government is in precisely this position, and therefore can always secure funds. Thus it has unlimited resources for meeting interest charges and for making repayments on bond issues. Since it may either impose taxes, or borrow through its control of the banking system, there can be no question of the federal government going bankrupt.

Let us consider this point in detail. Suppose an issue of government bonds falls due this year, and the government must find $10 billion with which to discharge this debt. If it can borrow $10 billion from individuals or banks, each holder of maturing bonds gets as much as he is entitled to receive. Even if the government borrows from the very ones to whom it must make payment, the individual bondholder is satisfied and the debt is fully repaid. It should be noted, however, that by this method of payment, the total amount of debt is not reduced even though each bond is paid off; that is, if the government borrows to meet maturing claims, each creditor is fully satisfied, but there is no reduction in the amount of the federal debt. This kind of borrowing goes on, incidentally, almost every day. For example, in June, 1946, almost $12 billion in government securities fell due, not including the Savings Bonds redeemed. The greater part of the funds needed to meet these maturities was obtained by selling new government securities. Thus, as long as the government

is able to borrow, it can avoid "bankruptcy," even though the total debt may not be reduced.

The Debt and Public Credit

All this suggests the answer to the common fear that a high national debt will exhaust the government's credit. The government may be said to exhaust its credit only if it is unable to borrow; and, as we have just seen, its ability to do so can never be seriously in question. The government can always find a buyer for its bonds — the Federal Reserve Banks if necessary. Indeed, our government has had no difficulty in finding private individuals and institutions who are most eager to have government bonds. In fact, the public has displayed a growing willingness to hold government bonds in spite of the increased debt, as is shown by the fact that the interest rate on government bonds has been declining steadily for a number of years. In 1919, federal government bonds [1] yielded 4.73 per cent. By 1929, the interest rate on these bonds had fallen to 3.60 per cent, and by April, 1946, it had dropped to 1.44 per cent.[2] Such a decline in the interest rate must mean, not that the government is finding it increasingly difficult to borrow, but rather that lenders are increasingly willing to hold government securities. It is possible that one day the government may be unable to sell its bonds to private individuals, commercial banks, and insurance companies. But to date its experience has been just the opposite. There is no sign that a high debt exhausts the credit of the government of the United States. And since as a last resource "it can borrow from itself," there need be no fear on this account.

The situation for state and local government debt is quite different. We may conceive of a local government in a modern economy which is unable to get funds, either from taxes or loans, with which to satisfy the claims of its creditors. State and local debt is in this respect like private debt. It can be reduced by repayment, repudiation, or bankruptcy. State and local governments are no better able to borrow at will than are business concerns, and this inability means that bankruptcy is possible. The federal government, however, is in a unique position among debtors. It can always borrow, and hence

[1] Partially tax-exempt, and due in not less than eight years.

[2] These maturities were taxable — a factor which would tend to make the rate abnormally high. The rate on partially taxable bonds of the same maturity would be even lower.

can always meet the claims of its creditors. Only if we fail to note the special relation between the government and the banking system will we reach the false conclusion that the national government can be driven into bankruptcy and that it can exhaust its credit.

Interest on the Debt and Deflation

Even though a high federal debt threatens neither bankruptcy nor an exhaustion of government credit, it does have certain other consequences with which we must be concerned. Ordinarily, the higher the debt, the greater, other things being equal, is the amount of interest which must be paid on it. Thus, in 1929 the interest on the federal debt was less than $700 million, but in 1945 it amounted to almost $5 billion, in spite of the decline in interest rates. There are two ways, as noted above, by which the funds required to pay this interest can be secured: taxation and further borrowing. If the government collects taxes to pay the interest on its debt, it transfers money from the taxpayer to the bondholder, so that while the taxpayer has less, this is offset, quantitatively at any rate, by the fact that the owner of government bonds has more. Unfortunately, we do not know enough about the identity of the typical taxpayer and the typical bondholder to assess the consequences of this transfer. To the extent that the bondholder is himself a taxpayer, no transfer of income takes place. But most people are not bondholders and taxpayers to an equal degree; it is generally supposed that the typical bondholder has a somewhat higher income than the typical taxpayer. If so, the transfer is in the direction of those in the higher income brackets. And naturally, the greater the debt and the higher the interest payments to be made, the larger will be this transfer of income from the poorer taxpayer to the richer bondholder. As we have seen, a shift of spending power from lower to higher income groups generally reduces the propensity to consume, and such a reduction, unless it is offset by increased investment, lowers income, employment, and probably prices. Thus, in this way, at any rate, a high national debt is deflationary. And the higher the national debt, the greater the deflation will be, as long as interest payments are financed by taxation.

If this deflationary effect is thought to be serious, there are three ways of avoiding it. The first of these is to reduce interest rates on government debt still further. This is usually possible because the government controls a most important potential lender, the Federal

Reserve Banks, to whom it can sell its bonds at almost any price it wants and in that way set the interest rate.[1] A second way to avoid deflation through a shift in income from taxpayer to bondholders is to change the nature of the tax structure in such a way that a larger part of the tax burden falls on those in the higher income brackets and a smaller part on those in the lower brackets. If this is done, the transfer of income from taxpayers to bondholders reduces the propensity to consume less markedly because the typical taxpayer is closer in the income scale to the typical bondholder. Social security and excise taxes would have to be reduced, while income and inheritance taxes would have to be raised, to gain this objective. A third way to avoid the deflationary effects of financing interest payments on the national debt is to get the necessary funds, not by taxing, but by borrowing. Obviously, if this were done, the difficulties associated with the transfer in income would not arise. The money could be borrowed from private individuals, insurance companies, commercial banks, or the Federal Reserve Banks. Since the lender's income is not reduced when he lends and the creditor's income is increased by the interest he receives on the bonds he holds, the propensity to consume is not reduced, but in fact is likely to rise. However, borrowing to finance interest payments means, of course, a relatively rapid rise in the national debt.

When money to pay interest on the public debt is secured by taxation, there may also be a reduction in private investment. No matter who pays them, high taxes are likely to reduce private investment to some extent, for they reduce consumption and may raise business costs enough to make expansion unprofitable.

Another consequence of high public debt is that it is likely to affect the confidence of business firms adversely. Business firms generally disapprove of a high public debt, either because it is likely to mean high taxes or because it suggests an active federal government, the possibility of government regulation, or even direct competition by the government. No matter what the reason, business firms generally are somewhat less optimistic about the profit prospects of any investment project when the public debt is high. The unfavorable psychological atmosphere tends to discourage private investment. Thus a

[1] It is worth noting that when interest is paid on government bonds held by the Federal Reserve Banks, it is in a sense paid by the government to itself. For although member banks own the Federal Reserve Banks, their dividends are limited in amount. Any profits beyond that go to surplus, and the United States government "owns" the surplus.

very high public debt has a variety of deflationary effects. It tends to reduce the propensity to consume by transferring income from taxpayers to bondholders, and it tends to depress private investment. The reduction in the propensity to consume and in private investment brings about a reduction in the level of income, employment, and prices. To this extent, a high public debt imposes a burden upon the economy.

But does not a high private debt have exactly the same effect? When private debt is high, large interest payments must be made by the debtor firms, and, to make matters worse, the interest rate on private debt is generally higher than on public debt.[1] Private firms do not, of course, levy taxes in order to collect funds to pay these interest charges; instead they get them from the sale of goods. The existence of a high private debt,[2] therefore, means a relatively large transfer of income from the purchasers of goods to the bondholders. Such a transfer is to some extent a transfer of income from the lower to the higher income groups. On this account, we may expect the propensity to consume to be lower, the higher is the private debt. And this is precisely what we expect when public debt is high, though the effect may be even more serious with a large private debt.

The effect of a large private debt on private investment is also likely to be unfavorable. Firms that must pay more for their products because their suppliers are heavily in debt are not so willing to invest as they would be if they could get what they need at lower cost. Furthermore, when private debt is very high, consumption is relatively low, and this, too, discourages private investment. The unfavorable effects of a high private debt on private investment appear to be analogous to those of a high public debt. But one discouraging consequence of a large public debt does not result from private debt: lack of confidence. Hence at least this deflationary factor is weaker when private debt is high than it is when the public debt is high. Otherwise the deflationary effects of the two types of debt are not very different.

It may be well here to warn against a possible confusion. In an earlier discussion of private and public investment,[3] it was pointed out that expansion could be expected only when the debt was rising.

[1] In April, 1946, the federal government had to pay only 2.1 per cent on its long-term bonds. Corporate bonds yielded from 2.5 per cent to 3 per cent.

[2] Including common and preferred stock.

[3] See Chapter 34.

Yet it has just been said, tentatively at any rate, that a high debt has deflationary effects. This does not involve a contradiction. For a *high* debt is not the same thing as a *rising* debt. A debt can be high and falling; and hence probably deflationary on both counts. Or it can be high and rising, and hence expansionary on at least the second count. A rise in debt goes with high income and employment; a high debt, we conclude at this stage, tends to reduce income and employment.

The Debt Burden and the National Income

The amount to be paid by the government as interest on its bonds depends only on the size of the debt and the interest rate. In depression it must pay just as much as in prosperity, except as it may be able to scale down its interest charges by refinancing. Now if it must transfer $5 billion a year as interest on its debt, the effect on the propensity to consume and on investment is likely to be much more serious when the national income is $50 billion than when it is $150 billion. For to collect the amount required to meet interest payments would require much higher tax rates during depression than during prosperity. And unless there were a change in the tax structure itself, higher tax rates would reduce the propensity to consume and investment even more than lower rates. A high debt may, therefore, exert a mild deflationary pressure upon the economy during prosperity, but during depression the deflationary pressure is likely to be much stronger. When the debt is high, it therefore becomes even more important to avoid depression.

Inflationary Aspects of Public Debt

But a high public debt may also have certain inflationary consequences. When the debt is high, a large volume of government bonds are in the hands of private individuals, commercial banks, investment companies, trust funds, and endowment funds. The individual who owns government bonds may spend at an abnormally high rate on consumers' goods for two reasons. Having government bonds gives him a feeling of security which may lead him to spend more. And since government bonds can be turned into cash much more easily than almost any other kind of wealth, the man who owns them may not only want to spend more money upon consumers' goods; he may be better able to get the money. Therefore, when the public owns a

large quantity of government bonds, we can expect a relatively high propensity to consume. In fact, many economists expect the propensity to consume in the United States to be abnormally high over the next few years because so many individuals possess government bonds.

There is still another way in which a large public debt may be inflationary. When commercial banks acquire the bonds, they pay for them by creating demand deposits, as we have seen,[1] so that the volume of money in circulation is increased. A large debt gives commercial banks an opportunity to expand their investments with perfect safety and thus to increase their demand deposits. Such an increase in demand deposits and in the volume of money means, of course, that individuals and corporations must be persuaded to increase their holdings. This normally leads to a reduced rate of interest. If at the same time the Federal Reserve Banks increase their holdings of government bonds, the reserves of member banks are increased. This, we have seen, allows them to lend and invest more freely, thus forcing the interest rate down still further. The lowered interest rate encourages private investment and tends to raise income, employment, and possibly prices. Hence, if the banking system purchases some of the large volume of bonds that are outstanding when debt is high, the increased amount of money in circulation means lower interest rates and thus some tendency toward higher employment, income, and prices.

On the other hand, a high volume of private debt is unlikely to have any of these inflationary consequences. Individuals who hold corporate bonds cannot usually turn them readily into cash to finance the purchase of consumers' goods. Nor do they have the sense of security afforded them by the ownership of government bonds. Hence the propensity to consume is less likely to rise on this account. Furthermore, commercial banks are not so likely to increase the supply of money by acquiring corporate bonds. Finally, the Federal Reserve Banks do not purchase corporate bonds, and hence the reserves of member banks are not affected when the volume of corporate bonds outstanding is increased.

Summary

In short, whether high debt is inflationary or deflationary is difficult to decide. If the bonds are owned chiefly by wealthy people, the

[1] See Chapter 23.

debt is probably deflationary, and the higher it is, the more deflationary it is likely to be. The transfer of income from the taxpayer to the bondholder would reduce the propensity to consume; private investment would be discouraged; and the boost to consumption that might come because consumers own bonds would be relatively unimportant, since the consumers are assumed to be well-to-do. On the other hand, if many of the bonds are owned by those in the lower income brackets, the net effect may be slightly inflationary. The transfer of income from taxpayer to bondholder may raise the propensity to consume, and the possession of liquid assets by those in the lower income brackets would also raise it. Finally, in spite of the effect upon business confidence, private investment may be stimulated by the relatively high level of consumption. In either case, if banks hold a large amount of bonds, the lowered rate of interest would stimulate expansion. Thus, whether the public debt is inflationary or deflationary depends very much upon who owns the bonds.

There are no grounds for believing that a high public debt destroys the nation's credit or leads to a marked fall in the value of the dollar. Nonetheless, the existence of a high public debt is not a matter for indifference. It may, as we have seen, create deflation, a commodity with which our economy is already too well supplied. If so, the higher the debt, the greater is the importance of financing government expenditures suitably. When the debt is high, we must be especially insistent on maintaining the national income at the maximum. For if it falls, either we shall need extremely high tax rates to collect money for interest or we shall have to borrow. Neither of these alternatives appeals to most of us, the first because it is deflationary, and the second because it is unconventional.

37

Invention and Employment

EVER SINCE THE COMING of the Industrial Revolution and the introduction of modern methods of production, social attitudes toward invention have been contradictory. Both in Great Britain and in the United States there have been outbursts against the machine, during which workmen, fearing that it would take their jobs from them, have rioted against its use and destroyed it. From time to time there has been widespread support for proposals to call a halt to invention. New methods of production have been looked upon as enemies creating more unemployment than productivity. And the inventor has not always been in the happy position of Thomas Edison. Yet not all interests have been hostile to invention. The governments of both this country and Great Britain have endeavored to encourage invention through protective patent laws. As an inducement the inventor, or rather the holder of a patent, has been permitted to exploit the invention monopolistically for a number of years. Thus, while one part of society — generally that part whose livelihood was most directly threatened — has held that inventions are detrimental to the welfare of society, another part, whose livelihood was in no way endangered by invention, has supported the view that technical progress is essential to our social well-being.

Technical progress, as we have already pointed out,[1] may be of two kinds: improvements in methods of producing commodities already known and the development of new products. We shall first consider the former, since it is against this type of invention that most of

[1] See Chapter 28.

the hostility has been directed. Technical developments that make for more efficient production of a commodity already known have been extremely numerous. Some of them have been spectacular, such as the combine for harvesting wheat and the continuous-strip rolling mill used in the steel industry. Others have been much less striking, but the cumulative effects of large numbers of them have been, nonetheless, very great. During the nineteen-thirties, for example, many industries extended the use of semi-automatic control equipment. While the introduction of this equipment may have had a relatively small effect upon costs in any single process, nevertheless, when applied widely it has had a very great effect on industry as a whole.

Inventions and Productivity

A firm adopts an invention that improves the method of producing an established commodity when it expects to profit from so doing through a reduction in costs. Typically, technical improvements of this kind have reduced labor costs by permitting a saving of labor, although some have resulted in other kinds of savings. Since less labor is normally required for the production of any specified output than formerly, we may measure the effect of invention either in terms of reduced labor costs or, more generally, since wage rates may change, in terms of increased labor productivity. It will be worth while to examine the data on the increase in labor productivity in specific industries, though first it should be noted that the efficiency of labor may change even when methods of production have not altered. For, as we have already seen, a change in the level of output is likely to affect labor's productivity,[1] and it is therefore important to distinguish between these different effects. This can be done by comparing productivity at two dates when output was approximately the same.[2]

In most general terms, labor productivity increased by roughly 70 per cent between 1920 and 1940. Let us consider some over-all figures. For all manufacturing industries, output was about the same in 1923 and 1938, but to produce this output required about 35 per cent less labor — measured in terms of man-hours of work — in the

[1] See Chapter 7.

[2] Cost may also alter if the size of firms (the amount of their capital equipment, for instance) alters. Unfortunately, it is not possible to distinguish, in the statistical data, the results of such a change from the results of using an improved technique of production.

later year. Hence, output per man-hour increased by about 50 per cent in that fifteen-year period. Again, the volume of manufacturing was the same in 1926 and in 1936, though the amount of labor required was about 25 per cent less in 1936. Thus, output per man-hour increased by 35 per cent in these ten years. There is no doubt, therefore, that in manufacturing industries generally, labor productivity increased rapidly. In mining industries the results have been substantially similar. Mineral output in 1916 was about as high as it was in 1936, but man-hours of work fell by about 40 per cent. Hence output per man-hour was about 70 per cent higher in 1936 than it was twenty years earlier.

Figures for specific industries show much the same results.[1] The output of the rubber products industry in 1923 was about the same as it was in 1933, but output per man-hour in that period increased by almost 75 per cent. In the sole leather and harness leather industry, output per man-hour went up by about 50 per cent between 1921 and 1935. In the cotton goods industry, it rose by roughly 40 per cent between 1925 and 1936; and in the electric light and power industry, by 43 per cent from 1929 to 1935. In industry after industry the story is the same: comparing two years when output was at about the same level, the output per man-hour was invariably higher in the later year. This reflects in part the higher productivity brought about by the introduction of improved production methods.

It is precisely because invention normally means a sharp increase in labor productivity that invention has been so often the object of criticism. For when labor productivity is increased, less labor is needed than formerly to produce any given output. This fact is obvious — especially to the men who have lost their jobs as better methods have been introduced. And certainly it is natural for them to feel that they have been displaced by the machine.

The Effects of the Invention: Preliminaries

We must distinguish between the consequences of an invention at two stages in its history. In the first stage, the economy experiences the effects of its *introduction;* in the second, the economy faces the consequences of its *use.* To put this more concretely, the invention and the adoption of the continuous-strip rolling mill by firms in the

[1] These statistics have been taken at random from *Production, Employment, and Productivity in Fifty-Nine Manufacturing Industries,* The Works Progress Administration, 1939.

steel industry did not itself bring about the displacement of men. A fairly long period of time had to elapse before the mills were built and in operation, and it was not until then that the firms could dispense with labor. In the first period, the most notable effect of the invention was to stimulate the production and fabrication of the new rolling mills. Hence obviously the immediate result of the invention was an increase in employment. But once the rolling mills are in operation, labor is no longer needed to produce them. Moreover, the maintenance of these mills usually requires relatively little labor, and when they go into operation, therefore, labor may be displaced. Indeed, unless it becomes profitable to expand the output of the mill, less labor will certainly be needed because, as we have seen, the new method of production raises labor productivity. Thus the two consequences of the introduction of the continuous-strip rolling mill were almost exactly opposite. The effect of its introduction was to increase employment, while that of its use was to reduce it — assuming no change in the output of the industry.

Invention and Investment

As we have already seen, inventions stimulate private investment. Business firms must often spend a great deal of money to purchase the equipment needed to introduce new methods of production. Normally they will have to spend a great deal more to scrap the old equipment and to manufacture and install the new than they would have had to spend merely to maintain the old equipment. The chief effect of the exploitation of an invention is, therefore, to raise private investment while the new equipment is being constructed. This fact was brought out very clearly by Mr. Charles R. Hook, the president of the American Rolling Mill Company, in evidence presented to the Temporary National Economic Committee:[1]

> . . . I should like to comment briefly on some of the striking changes in the steel industry which have taken place since the invention of the continuous sheet rolling mill. Since the first continuous sheet rolling mill was put into operation, 27 such mills had been installed by 1937, representing a total investment by the industry of approximately $500,000,000. Building that equipment has provided work for thousands of workers in the construction and equipment industry.

[1] *Investigation of Concentration of Economic Power. Hearings before the Temporary National Economic Committee* (Washington, 1939–41), Part 30, p. 16393.

In referring to his own firm, Mr. Hook said: [1]

> When we were talking about the question of displacement of workers, we have got to take into consideration the number of men that were employed in the manufacture and the installation, the building of the plants, and the installation of the machinery. . . . Between '27 and '37 when we introduced the continuous mill in that plant we spent $20,492,778 actually. In other words, we estimate that the man-years of employment amounted to 9000, or we employed during that time an average of 819 men at full time at approximately $7 a day. You have got to take that into consideration when you are talking about displacement, and by the way, we did use on that construction a number of the men who had formerly worked in the mill when we were running the old style full time.

Now let us see what happens to employment as these new techniques are adopted. There is no reason to suppose that the propensity to consume is affected during the period in which the new equipment is being built. For this reason, and because investment is raised, we may conclude that the adoption of the new technique brings with it higher employment and increased income. But it takes only a short time to replace old equipment with new, and it is only during this period of replacement that investment itself is abnormally high. Hence this favorable effect upon employment is likely to last for only a short period.

Once the new equipment has been installed, investment will return to approximately its old level. In fact it may for a time even fall short of the original level, for, since most of the equipment is new, expenditures for maintenance and repairs will probably at first be abnormally low. Thus, aside from other factors, the adoption of the invention will increase investment and therefore raise employment and income for a period, though only while the equipment is being fabricated. Thereafter they will fall again to near their original level. And, if the amount of investment needed to maintain the new equipment is lower than that needed to maintain the original equipment, employment and income will fall even below the original level.

Thus, aside from the influence of other factors, the effect of invention on investment is such that employment and income can only be maintained at a *permanently* higher level by a relatively steady flow of new inventions. The effect of just one invention, as we have seen,

[1] *Ibid.*, p. 16411.

is to raise employment for a while and then to let it fall again. A succession of inventions, however, will provide a succession of stimulants to investment, so that as one wears off the next begins to be felt. Thus, as investment falls toward the low level needed for maintenance in one industry, it rises when a new investment is exploited in another, and so the high level continues. In other words, if a succession of inventions is adopted, employment and income will rise with the first invention and will remain high. But this will happen only if new developments come in a fairly steady stream.

The Long-Run Effects of the Invention

Inventions are adopted, as we have seen, because they make possible a reduction in costs. In most cases, that reduction comes about because of a saving in labor. In most cases, then, labor productivity is increased once the new methods of production come into use. But as we have seen, an increase in labor productivity means that the same physical output can be produced with a smaller expenditure of labor. If output had been 100 before the new methods of production were adopted, and if the investment needed to install the new methods of production raised output to 110, we might expect employment to increase by about 10 per cent. But unless this invention were followed by others which stimulated investment equally, output would fall again toward its original level.[1] If this should happen, it is clear that the amount of employment needed to produce this output would be lower than originally. If, for instance, an improved method of production makes labor in an economy 50 per cent more productive, we should need only two-thirds as much labor to produce any given output. Hence, unless output is maintained at a higher level, less labor will be needed once the new methods of production are actually in use. In short, if output returns to the pre-invention level, employment, which was high during the period when investment was abnormally high, will now fall to a level below its starting-point.

The question then arises whether output will remain at the original level. This depends, as we know, upon what happens to investment and to the propensity to consume. We have already seen what happens to investment.

[1] We will discuss shortly the changes that can be expected in the propensity to consume. We shall see that these changes are more likely to reduce output below the original level than the reverse.

The Long-Run Effects: Price Reductions

But what about the propensity to consume? There are several factors to be considered before this question can be answered. First, the propensity to consume may change if the price of the product is reduced, though there may be some question as to the direction of the effect. Let us suppose for the moment that the effect is favorable. If so, the physical output will rise even though investment falls to its original level. But since labor productivity is higher, employment will not necessarily rise above the original level. Let us illustrate this more concretely. We shall suppose that after the invention has been adopted, investment returns to its original level, and that because of reduced labor costs the invention brings about a reduction in the price of the product. Although spending by business firms, government, and foreigners remains the same as before, consumers may increase their total spending, not necessarily in money terms, it is true, though probably in terms of the amount purchased. They will certainly buy more of the product whose price has fallen. However, this does not necessarily mean that their total purchases of all commodities measured in physical terms would increase, though this may happen. At any rate, it is quite possible that if prices fall as a result of the adoption of new methods of production, the propensity to consume will increase. This would at least partly offset the unfavorable effect on employment of improved labor productivity. Whether employment increased or fell would depend upon whether investment returned to the pre-invention level, and on whether the propensity to consume rose enough to offset the unfavorable effects on employment of the increase in labor productivity.

If prices fall because of invention, we should expect employment to rise for a short period, and then to fall again to about its original level as the rise in the propensity to consume led to a higher output which could be produced with higher efficiency.

The Long-Run Effect of Invention: No Price Change

If an invention does not lead to lowered prices, employment will probably be affected unfavorably. We have seen many reasons why prices do not fall when costs are reduced.[1] Certainly an examination of the relation between prices and costs in a very large number of American industries does not justify the expectation that prices will

[1] See Chapters 13 and 16.

fall at the same rate as costs. Consequently, when inventions are introduced and costs are lowered, the price may not be changed at all. In this extreme case we should have no reason to expect an increase in the propensity to consume; rather, as we shall see, we may expect the opposite. And if the propensity to consume does fall as the result of an improvement in processes, and if investment does not change, then income will fall below the pre-invention level. If at the same time labor's productivity is raised, employment will fall even faster than income. Let us examine this possibility in detail.

Why should an invention lead to a reduction in the propensity to consume? Mainly because of what its adoption may do to the distribution of income. The adoption of the invention is expected to lead to reduced labor costs per unit of output, and therefore to a reduction in total wage payments. Since the outlay for wages is labor's total income, obviously a reduction in labor costs gives rise to a reduction in labor's income, if output remains constant. But at the same time that labor is getting less, the firm's owners will be getting more. For with output unchanged, total receipts will be the same; and with labor costs lowered, profits will be greater. Labor's share in the total income of the economy will go down as the firm's profit goes up. Hence there will be a shift of income from labor to employer. And as we have seen, when income is transferred from wages to profits, the propensity to consume is adversely affected. For labor generally is in a lower income bracket than the employer, and therefore when $1 billion in income is transferred from labor to the employer, the reduction in labor's consumption is greater than the increase in the employer's, and accordingly the propensity to consume falls. Hence we may conclude that if prices remain the same — or indeed if they do not fall at the same rate as labor costs — the propensity to consume will fall.

In an economy whose prices tend to remain stable, employment is likely to be lower after an important invention than before it, for it is not likely that investment will be stabilized above the pre-invention level, and the propensity to consume will probably be reduced. If investment is unchanged and the propensity to consume is reduced, the physical output would also fall. This in itself would mean reduced employment. But there is a further adverse effect to be considered. Because of increased labor productivity, it takes less labor than before to produce the same output. And with output even

lower, the number of jobs would shrink still further. Thus, as output falls from 100 to 90, the number of jobs may fall from 100 to 70. Employment, therefore, declines not only because of increased labor productivity, but also because of the decline in the propensity to consume.

It may seem paradoxical that our economic well-being should suffer because we are able to enjoy a higher standard of living than ever before, but it can happen just the same. An invention, of course, makes us more productive; our plant and equipment are more efficient, and we should be able to produce more output. But we must not confuse the ability to produce a larger output with the output itself. In 1936 we were able to produce a considerably larger output than in 1929, but we failed to do so by a substantial margin. Again in 1932 we were able to produce a larger output than in 1929, but our failure to do so was even more striking. Unfortunately, to possess the means for high production does not automatically create all the conditions for getting that production. Unless the other conditions — a high level of investment and a high propensity to consume — are also satisfied, we fail to produce at peak capacity, and unemployment is high.

A restatement of the problem may help to clarify the issues. Suppose that improvements in methods of production enable us to produce 10 per cent more than formerly with the same labor force. What conditions must be satisfied in order that we may increase our income by this much? The answer is, obviously, that we must maintain full employment. But how is this to be done? Since full employment is now associated with a higher real income, we must have either an increase in investment or an increase in the propensity to consume, or both, in order to keep full employment. But after an important technological development, investment or the propensity to consume will be substantially increased only if inventions continue to flow at a faster rate than before, or if prices come down enough.

If for any reason the flow of invention is interrupted, however, investment will fall toward its original level, and one of the factors upon which we might have relied for achieving a higher output and a steady level of employment will have disappeared. And, if the fall in price does not keep pace with the fall in costs which comes from increased labor productivity, the propensity to consume will not rise. Indeed, it will probably decline, and thus the other condition of a

higher output and steady employment will not be satisfied. If neither condition is satisfied, unemployment will increase. Therefore, unless inventions continue to be made, or unless the reduction in price is great enough to raise the propensity to consume, improved methods of production bring unemployment.

We may note in passing that if it proves difficult to lower prices when costs are falling, approximately the same effect can be secured by raising wages.[1] If wages are raised and prices are constant, the employed worker receives a higher real income. This will have the same favorable effect upon the propensity to consume as a price reduction with wages fixed. Furthermore, with higher wages and constant prices, there will be no transfer of income from the wage earner to the employer, even though labor productivity is increased. The distribution of income will be unchanged so long as the rise in wage rates keeps step with the rise in labor productivity. With a higher propensity to consume, we should have higher income and steady employment even though investment did not rise.

To depend upon wage increases seems to be a more hopeful method of combating the unemployment caused by improvements in technique than to hope for price reductions. Firms may not want to lower their prices when their costs are reduced, and the consumer, who has the chief incentive for demanding lower prices, is relatively unorganized to bring pressure to bear. Thus there may be no practicable method by which to bring about price reductions. Wage earners are much better able to bring pressure for higher wages since they are already organized in unions, and when costs are falling because of improvements in the technique of production, they may be in a good position to demand higher wages.[2]

Improvements in methods of production are favorable to the economy because they permit us to enjoy a higher standard of living. But they are only favorable if we have the wit to take advantage of them. Otherwise they may do us more harm than good. For, as we have seen, unemployment may increase as productivity rises, unless positive steps are taken to maintain the flow of new inventions or to reduce prices or raise wages as costs fall. The workers who resisted the introduction of the machine which put them out of work may not have

[1] If piece-rates prevail, labor will automatically receive an increase in pay, provided that the rate itself is not reduced.

[2] A more detailed consideration of the effects of changes in wage rates on unemployment is given in the next chapter.

seen the whole issue clearly; they may have been unaware of the advantages that society could reap from the improved methods of production. But they were correct in seeing that the adoption of improved and more efficient methods of production could cut them out of their jobs.

Summary: Inventions That Reduce Costs

When new methods are introduced for producing a commodity already known, employment normally rises while firms are adopting the new processes, since to do so means relatively heavy investment expenditures. Once the new process is in use, employment may fall below its original level, for the increased productivity of labor makes it possible to produce a given output with less labor than was formerly required. In order to avoid an increase in unemployment, inducements must be offered to business firms to persuade them to expand their output. Such expansions in output will be forthcoming (*a*) if the rate of invention is maintained and new processes are adopted at a steady pace; or (*b*) if prices are reduced as costs are; or (*c*) if wage rates are advanced to maintain the ratio of labor cost to price. Otherwise the economy will produce no more than originally, and it may produce less. And if it produces no more than before, unemployment will be higher than at first, because, with the increase in labor productivity, less labor is needed to produce the same output. The short-run effects of inventions such as those we have been considering are favorable to employment and income; but it is easy to see that in the long run, improved methods of production may cause unemployment.

The Development of New Products

The development of new products is likely to favor employment both in the short run and in the long run. The short-run effects upon employment have already been discussed. A good deal of investment may be required in order to put these products on the market. It was necessary, as we have seen, to construct many large plants and install much new equipment in order to build up the automobile industry to a point where it could produce seven million units a year. Likewise, the development of the nylon industry and of the radio industry required substantial investment in new plant and equipment. During the period in which this investment is undertaken, employ-

ment and income are favorably affected. Indeed a substantial part of the prosperity of the nineteen-twenties can be attributed to just such developments, particularly that of the automobile and of various kinds of electrical equipment.

The long-run effects on employment that follow from the development of new products are also likely to be favorable. It is true that once the capacity of an industry has been built up to an adequate level, investment will fall, and the amount of investment needed to maintain the new plant and equipment will be negligible compared to the amount required to build it in the first place. But the propensity to consume may be boosted permanently, for with the development of new products, spending habits may be changed and consumers can be expected to spend a larger sum out of any given income than formerly. Certainly the development of the automobile coaxed consumers into spending more freely. There are certain offsetting factors, however, to be considered. The automobile is a substitute for other methods of travel. While more of the consumers' income is spent upon acquiring and servicing automobiles, less is spent upon travel by horse and buggy and by train. Furthermore, since travel for pleasure becomes available to a large number of families, their expenditures on other kinds of recreation may be expected to decline. Thus the propensity to consume does not necessarily increase by a sum equal to the amount spent on automobiles, but it almost certainly increases to some extent.

Once the capacity to manufacture the new product has been built up, investment returns to approximately its original level, but the propensity to consume probably remains at a higher level than originally, and consequently the national income may be expected to grow. Since an invention of this kind does not directly increase labor productivity, more labor is required to produce a higher income. In conclusion, therefore, we see that the development of the new product will certainly have a favorable effect upon employment in the short run, and that the favorable effect will probably be maintained. At first, as the industry is being built up, investment will be abnormally high. Later, as investment falls toward the original level, the propensity to consume will be stimulated, and thus both income and employment will be maintained at a level higher than before the product was developed.

Summary

We must distinguish between inventions of two types, those which permit industry to produce an established product more efficiently than before and those which put a new product on the market. In the short run, each of these types of invention favors income and employment, for it is necessary to increase investment expenditures in order to exploit either kind of invention. And this, as we have seen, leads to a higher level of income. Cost-reducing inventions may lead to higher unemployment in the long run, though they will not do so if prices are reduced with the fall in costs, or if other inventions become available whose short-run favorable effects upon employment offset the long-run unfavorable effects of the invention first considered. The long-run effects of the development of new products are much more likely to be favorable to income and to employment. But since inventions of both types make possible a higher standard of living, they should obviously be encouraged; and since the adverse effects of the cost-reducing invention upon income and employment can be offset as long as the flow of invention is steadily maintained, prices are brought down, or wages are raised, society obviously gains an advantage from providing as strong a stimulus to invention as possible. The troubles that inventions may cause can only arise if we permit them to do so. If we prevent them, we are all the gainers, for we are able to produce a larger and more varied output.

38

The Money Wage, the Real Wage, and Employment

IN THE EARLY MONTHS OF 1946, the newspapers were filled with little lessons in economics presented by some of the largest corporations and some of the largest labor unions in the United States. Needless to say, the conclusions drawn were often as different as night from day. The advertisements sponsored by the corporations argued that a general rise in money wages would force profits down; that a fall in profits, the motivating force in the economy, would lead to a fall in employment; and that the worker would consequently be injured rather than helped by higher wages. The unions, in their advertisements, stressed the point that the wage earner is a very important part of the consumer market; that only if his income is kept high can business firms find markets for their products; that rising wages are needed to increase the size of the market; and that only by raising wages, therefore, can unemployment be prevented. It is difficult to conceive of views more widely divergent than these. Yet, in a sense, they both sound right. It seems that wage increases cause unemployment; and that wage increases reduce unemployment. Which view is correct? We shall try to find out in this chapter.

The relation between wages and employment is one of the most baffling problems in economics, and for a number of reasons. For one thing, a change in wages affects almost every determinant in the economy, and as we have already seen, it is very difficult to estimate the final result of simultaneous changes in a large number of interrelated determinants. Second, the subject is one about which our emotions are likely to be strong, and our capacity for objective thought,

therefore, very weak. Third, economists themselves disagree as to the method of analysis which the problem requires. But in spite of all these difficulties, the problem obviously demands a solution, or at least, earnest consideration.

Before embarking upon this analysis, it is necessary to define a few terms. Thus in this chapter the term *money wage* means the earnings in money for a unit of work, say an hour; as, "the hourly earnings in manufacturing industries stood at $1.03 in March, 1945." The term *real wage* refers to what can be bought in the way of consumers' goods with the proceeds from one hour's work. We could measure this, if we wanted, as so many slices of bread plus so many cigarettes plus so much steak, and so on. But this measure would be very inconvenient, since we should have to list all the kinds of things that wage earners buy. For that reason, the absolute level of the real wage cannot be measured readily, but changes in that level can be measured easily by the use of index numbers. Suppose, for example, that money wages have increased by 15 per cent within a certain period of time and that the cost of living, or, more accurately, the prices of goods purchased by wage earners, have increased by only 5 per cent in the same period. It then follows that at the end of the period one hour's work buys 10 per cent more than at the beginning, and the real wage has risen by 10 per cent. To illustrate: between June, 1939, and June, 1942, the average money wage in American industry rose from 63.6 to 77.9 cents. In the same period the cost of living rose from 98.6 (the base is 1935–39 = 100) to 116.4. Hence the real wage rose by about 4 per cent, since the money wage rose by 22.5 per cent while the cost of living rose by 19.1 per cent.

Now, let us set the problem: We are to investigate the effect of a general change in the money wage upon both employment and the real wage. The money wage is assumed to change, not necessarily in every firm in the country, but in most firms.

The Simple View

To the unsophisticated, the solution seems obvious. Everyone knows, so we are told, that if a firm has to pay higher wages, it is forced to reduce the number of its employees; on the other hand, if the firm can secure workers at lower wages, it can afford to hire more of them. This view simply embodies what every businessman knows about his own firm, and in accordance with it, we should expect that

a general change in the money wage would lead to a change in the opposite direction in employment.

This type of reasoning gives us the same results we obtained when we investigated the way an individual firm would alter its output (and hence employment) when there was a change in the wage rate it had to pay. A rise in wage rates brings about an increase in marginal costs. But this will lead a firm to reduce its output and to raise prices, *provided that the demand for its product has not altered.*[1] Why, then, is this result, which seems so obvious, and which holds for the individual firm in the conditions assumed, subject to question when we treat, not a rise in wage rates in only one firm, but rather a general increase in money wages? The answer is to be found in the assumption noted above: that *the demand for the product of the firm remains constant* and is not affected by the change in money wages. But this assumption, while realistic enough for an analysis of the effects of an increase in money wages in a single firm, is highly questionable when we are dealing with the effects of a general increase in money wages. Since wage earners buy more than one-half of the total output of consumers' goods, demand cannot be expected to remain constant when their money wages are altered. And if this happens, the assumption which embodies an unchanging demand is obviously inappropriate.

Evidently most of us — trade-union leaders, corporation officers, economists, and so on — believe that the assumption has no basis in fact. This was clearly illustrated during the period from 1943 to 1945 when there was virtually unanimous agreement that wage increases should be limited in size and confined to groups whose wages were out of line with the general level. It was argued that if money wages were raised, the demand for consumers' goods would increase. Since such an increase in demand could lead only to greater inflationary pressures on price, the policy of checking further wage increases was generally supported. By implication, this argument rejects the assumption that a general change in the money wage would leave the demand unaffected. Hence, we shall be going along with the widely accepted view in rejecting the assumption that when money wages increase, each firm is faced with higher marginal costs but with *no change in the demand* for its product. This assumption is realistic when we have to deal with a change in the wage rate in one firm or even in a small sector of the economy. It is not adequate otherwise.

[1] See Chapter 16.

The statement that one effect of a general change in the money wage is to change the total demand for our output implies a change in total spending. We have already found that a convenient way of determining the nature of a change in aggregate spending is to divide the spending stream into two: consumers' purchases and investment purchases. We may then proceed to our analysis of a change in the money wage by determining its effect upon investment and the propensity to consume.

A Wage Increase and Investment

Let us first consider the effect of an increase in the money wage on investment. The first and most obvious consequence of an increase in the money wage is a deterioration in business confidence. Business firms, for obvious reasons, do not like to grant increases in wages. Each firm sees it only as an increase in its own costs. Any gain it makes through increased sales results, not from its own wage increases, but from those given by other firms. Because of reduced business confidence, businessmen will be less optimistic about the expected yields from investment projects. Hence, if there were no other forces set in motion by the rise in the money wage, investment would tend to decline.

But other forces are set in motion. The higher the wage, the greater is the inducement for business firms to adopt labor-saving methods of production. If it were possible to hire coal miners for $5 a week, firms would not find it profitable to use coal-cutting machinery. With wages at $55 a week, it is profitable to use various kinds of capital goods as substitutes for labor. And the higher the wage, the larger is the number of the substitutes whose use becomes profitable. To make these changes in methods of production obviously requires new equipment — and the installation of this equipment of course means more private investment. This result then — the tendency to substitute capital for labor as wage rates rise — favors investment.[1]

Another force which tends to stimulate investment when wages are rising is the fear that they will continue to do so, and the consequent belief that it will be wiser to invest before the further increases take place. Since wage increases generally come about through union pressures, and each victory for the union strengthens its position for

[1] This favorable effect is but temporary: the long-run effect is likely to be harmful to employment. Compare Chapter 37.

the next encounter, businessmen tend to feel increasingly at such a time that investment should be made sooner rather than later. This effect can be most easily seen in connection with inventory investment. With rising wage rates, business firms are led to expect rises in the prices of the goods they purchase. A retailer who carries a stock of, let us say, men's shirts, can profit by expanding his inventory at a time when their price is low rather than waiting until later to buy them when the price is higher. Since, when firms accumulate inventories rapidly they make large investment purchases, the inducement to expand inventories during periods of wage increases means abnormally high investment for a time.

What happened late in 1936 and early in 1937 illustrates this situation perfectly. The CIO was organized in the summer of 1936. In the autumn, wage rates began a rapid advance and strikes became more numerous. Average money hourly earnings had been falling slowly through the last months of 1935. By April, 1936, they were only about 1 per cent higher than in 1934, and they stayed constant until September, 1936. Then they began to rise swiftly, and by June, 1937, they were 10 per cent higher than they had been the September before. There were many strikes in the period, and deliveries of goods were often held up. The price of goods at wholesale rose by approximately 7 per cent. The threat of further increases in wages and the difficulties that firms anticipated in getting goods from their suppliers led many firms to expand their inventories rapidly. And investment was accordingly high.

The expectation of further wage increases stimulates other kinds of investment as well, for example, investment in new equipment. Business firms, faced with the problem of buying capital goods now or buying them later at a higher price, will naturally prefer to buy before the price rises. Thus private investment of all kinds will be encouraged while wages are rising if the rise is expected to continue.

Once the price of capital goods has gone up, the marginal efficiency of any investment project falls, unless, as we have noted above, prices are expected to increase still further. Thus, if wage rates in the building trades are increased and the cost of building a house is therefore raised, house-building tends to decline, other things being equal. Or if goods for inventory cost more than they did, the marginal efficiency of investment of that kind declines, unless the prices are expected to rise still further, or other factors offset the effect of the rise in the cost

of investment goods. The rise in the price of investment goods that can be expected as a consequence of the rise in wage rates is then a factor unfavorable to investment.

An increase in wage rates also affects foreign investment. If, as wages are raised, there is a rise in the price of goods sold to foreigners, exports tend to decline. For instance, about the time Congress authorized a large loan to Great Britain in 1946, prices in this country began to rise rather rapidly, and the British hesitated to buy our products even though dollars were available to them. Of course, if the rise in wages and prices is expected to continue, foreign countries may be induced to purchase more heavily at once in order to beat the price rises that they anticipate for the future. The anticipation of further wage increases would therefore be expected to stimulate foreign investment just as it does home investment.

When money wages rise, the interest rate tends to increase, and this also discourages private investment. A rise in money wages forces firms to keep more of their wealth in the form of money in order to take care of the increased outlay for wages, raw materials, and so on. But to increase money holdings means that firms must increase their borrowing. When they do so interest rates rise, unless the banking system is willing to accommodate the increased demand for loans. And higher interest rates also tend to cut down private investment.

In summary, then, a change in wage rates has a complex effect on investment, and no general rule can be laid down by which the consequences in any specific situation can be surely predicted. The probable rise in the interest rate, the probable increase in the price of investment goods both for business firms and for foreign buyers, and the almost certain adverse effect upon business confidence, all tend to reduce investment. On the other hand, the pressure to substitute capital goods for labor, and the inducement to telescope future investment projects into the present if further increases in wage rates are expected, encourage investment. The net effect, whether positive or negative, can be forecast only if the situation is fully known.

A Wage Increase and the Propensity to Consume

The effect of the increase in wage rates upon income and employment depends, not only upon what happens to investment, but also upon what happens to the propensity to consume. There are several

ways in which this factor can be affected. If further increases seem likely, the effect would probably be at least a temporary rise in the propensity to consume, since higher wages will lead people to expect higher prices. Everybody likes to buy at the lowest possible price. And if prices are expected to be higher tomorrow than they are today, there will be a rush to buy today. Furthermore, if higher wages lead to a rise in the interest rate, this also may affect the propensity to consume, though probably only slightly, and not necessarily in the same direction every time. Again, since the propensity to consume depends in part on the distribution of income, a shift in distribution generally has a substantial effect upon the propensity to consume. Higher wages usually add to labor's share in the national income, so that the distribution becomes more nearly equal and the propensity to consume rises.

The analysis of the effects on the distribution of income is complex and must be postponed for a more advanced course in economics, but we can explain in general terms why a rise in wage rates normally increases labor's share in the national income. This result depends upon the fact that prices normally do not rise as quickly as wages. This is true for a variety of reasons. Sometimes oligopolistic considerations hold prices down, and sometimes firms hesitate to raise prices because they know their customers do not like frequent price changes. Whatever the reasons, statistical evidence extending over a number of years shows that when wages rise, prices rise, but by a smaller amount.

If the wage a firm must pay increases by 10 per cent, its payroll increases by the same amount, so long as it does not alter its methods of production or its output. But if the firm raises its price by only 5 per cent, its total receipts go up by only 5 per cent. If wages and prices change this way throughout the economy, and if employment does not change, payrolls — the income of labor — will rise by 10 per cent while the receipts of business firms will rise by only 5 per cent. Therefore, the national income — the sum of the sales receipts of all firms after allowing for double counting — has increased only 5 per cent, and the physical output is not changed. Obviously, then, since labor's income is increased by 10 per cent in money terms, its share in the national income is raised. It will be seen that labor's share in the national income is determined essentially by the ratio of labor costs to prices. A rise in wage rates in an economy such as ours,

where there is some price inflexibility, thus brings about a transfer of income from profits to wages which tends to equalize the distribution of income and to bring about an increased propensity to consume.

In summary, then, a rise in the wage rate almost certainly leads to an increased propensity to consume, and may also lead to a rise in investment. If so, it is clear that income and employment will respond favorably to increases in wages. On the other hand, if investment falls, it tends to cancel the favorable effect on employment of the rise in the propensity to consume. Our conclusion, therefore, must be indefinite: depending upon the circumstances, wage increases may either stimulate or reduce employment, and every situation has to be analyzed independently.

A Wage Reduction, Investment, and the Propensity to Consume

If the effects of a cut in wages are worked out along similar lines, it will be clear that a reduction usually lowers the propensity to consume and, again, that investment may change in either direction. Not only is business confidence likely to be heightened if a wage cut is forced upon labor, but also profits may be expected to increase and the interest rate to fall. And all these things will encourage private investment. Moreover, a reduction in wage rates will normally encourage foreigners to purchase our products, so that foreign investment will also tend to increase. On the other hand, if a cut in wages looks to be only the first in a series of reductions, or if it provokes serious strikes, investment may be reduced for a time.[1] The reduction in the propensity to consume resulting from labor's reduced share of the national income would also be a factor of considerable importance. But the final effect can be estimated only when the situation is fully known. In short, it is no more certain that lowering wages will increase employment than that raising wages will decrease employment. In either case, the whole situation must be known before any safe prediction can be made.

The Money Wage and the Real Wage

In part, at least, we have already suggested the effects on the real wage of a change in the money wage. The real wage increases when

[1] But to illustrate the complexity of the situation, if strikes lead retailers to accumulate inventory in anticipation of difficulties in acquiring goods later, investment would even be encouraged.

the ratio of the money wage to the cost of living increases, and falls off when that ratio declines. Prices are influenced in two ways when wages rise; directly through the increase in business costs even when output and employment are not changed, and indirectly through the induced change in output, omitting the effect of any change in the money wage. The total effect of a change in the money wage rate upon prices and hence upon the real wage is, then, a combination of two influences: first, what happens to the real wage when money wages change, assuming no change in employment; and second, what happens to the real wage because of any consequent change there may be in output and employment. We have already said that when there is no change in employment, the rise in the money wage rate normally leads to a smaller increase in the cost of living because of price inflexibilities. Consequently the real wage tends to rise when the money wage rises, and vice versa. When employment increases and money wages remain the same, there is usually some increase in prices:[1] a small one if output is low to begin with, but a much greater one if output is already near the peak level. Hence, when money wages are not altered and output increases, the real wage tends to fall — slowly if firms are operating well below capacity, rapidly if they are operating close to their peak. For with rising prices and no increase in the money wage, the real wage declines. Conversely, when employment decreases and the money wage remains the same, prices normally fall and the real wage rises, though the reduction in price and therefore the rise in the real wage will be minor except when firms are operating close to their capacity level.

To estimate the total effect of a change in the money wage on the real wage, we must add together these two forces: the effect of a change in the money wage on the real wage, assuming employment is not altered; and the effect on the real wage that follows from any change in employment that the changed money wage brings about. If employment is reduced when the money wage rises, the real wage will rise for two reasons: the increase in the money wage will bring about a rise in the real wage when prices are not perfectly flexible; and in addition, the real wage will rise in association with the decline in employment. If, on the other hand, employment is raised when the money wage rises, the net effect upon the real wage is more difficult to assess. Obviously, it will depend on whether the favorable

[1] See Chapter 32.

effect of the rise in the money wage is offset by the unfavorable effect of the increase in employment. Unless employment is already very high and firms are producing almost to capacity, we should expect the real wage to rise. Similar considerations lead us to expect that a decline in money wage will normally be associated with a reduction in the real wage. Hence we should expect generally to find the money wage and the real wage changing in the same direction. There is considerable statistical data for the United States from 1933 to 1945 to support this conclusion. Price inflexibility is a powerful factor in accounting for this relation. The effect of the reduction of the money wage on employment, and the consequent indirect effect on prices, is usually relatively small; and of course it would in any case operate in the same direction if the change in the money wage were to lead to a change in the opposite direction in employment.

Summary

The analysis of changes in wage rates is complex and needs all the economics we have learned up to this stage — and unfortunately, more, though the method of analysis should be clear. To determine how a change in the money wage affects employment, we must discover how such a change modifies investment and the propensity to consume. We have seen that an increase in the money wage may cause either an increase or a decrease in investment; some of the determinants of investment are favorably altered and others are unfavorably altered by the wage rise. The propensity to consume is almost certainly raised when wages are increased. Thus, there is a presumption that a rise in the money wage will increase employment, but it is by no means a certain conclusion. Generally a rise in the money wage will bring an increase in the real wage. The effects on employment of a cut in the money wage are equally uncertain, but such a reduction would almost certainly reduce the real wage. The uncertainty of our results should not, of course, be surprising. The economy is enormously complex, and any general change in the determining factors is bound to set in motion a large number of forces, some pushing in one direction and some in the other. In the case of a change in wage rates, the forces that favor employment seem to be roughly balanced by the forces that discourage it.

39

The Theory of Employment: Summary

IT IS CLEAR that an economy can lift itself by its own bootstraps. If output is low and unemployment is high, depression can be remedied by increased spending. And the greater the spending, the larger, within limits, is the output of goods and services. In short, we can spend our way into prosperity. Indeed, every period of prosperity we have had has been based upon just such a high level of spending. Conversely, when spending declines, depression and growing unemployment inevitably follow. The search for a cure for unemployment thus becomes a search for ways to increase our total spending. This, in barest outline, is the theory of employment.

The Theory of the Firm and the Theory of Employment

The analysis of fluctuations in employment was built upon two piers, the analysis of the firm and the analysis of the banking system. In Part Two of this book it was shown that the output produced by an individual business firm depends in part upon the demand for its product, and that, ordinarily, the larger the demand, the greater its output will be. Since variations in the total output of the economy are only the sum of variations in the output of the individual firms, this important conclusion follows: when the demand for their products increases, business firms in general are led to increase output and raise prices; when demand declines, they are led to reduce output and lower prices. For the term *demand* in the above statement we may thus substitute *volume of purchases*, since the two things move together. To rephrase this conclusion: the greater the volume of purchases, the

larger is the output of the economy and the higher the prices charged, and vice versa.

To account for variations in employment and output, it is necessary to examine the factors that cause variations in total purchasing. At first glance, this seems a hopeless task. Purchases are made for many reasons and by many types of buyers, and it seems impossible to frame a useful generalization about so wide a range of phenomena as, at the one extreme, buying a ticket to the movies, and at the other, ordering a new rolling mill for a steel firm. We purchase so many kinds of things, and for such different reasons, that a unifying principle to explain them all seems, on the face of it, inconceivable. It was found, however, that a useful analysis can be made when purchases are classified by the four chief types of buyers: consumers, business firms,[1] government purchasers, and foreign buyers. For convenience, we refer to consumers' purchases as consumption and to all other types as investment. In analyzing variations in consumption and investment, we found it helpful to rephrase our formulation as follows: total purchases depend upon the propensity to consume and investment. Since the national income is equal to total purchases, it too depends upon the propensity to consume and investment; and since employment and the national income fluctuate together, we may finally conclude that employment varies with the propensity to consume and investment. This, of course, does not *explain* variations in employment or in income; it merely classifies in a convenient fashion the phenomena and changes which must be explained.

The Interest Rate

The other pier on which our analysis of the theory of employment rested was the discussion of the banking system.[2] The institutions that make up the banking system determine the rate of interest, which is a price like other prices — the one a person pays for holding money. But money is unlike other commodities, since its production is subject to very strict government control and since it is handled by institutions of a very special nature. Because the rate of interest was found to be a significant factor in causing variations in income and employment, it was necessary to digress in order to point out how the rate of interest is determined.

[1] This category includes the purchase of new houses by ordinary individuals.
[2] See Part Three.

Investment, the Propensity to Consume, and the National Income

The statement that income depends upon investment and the propensity to consume does not, of course, explain variations in income; rather, it merely helps to define the task ahead of us. Our chief concern must be to analyze and account for variations in investment and in the propensity to consume. But before doing this, it is desirable to see how such variations affect the level of income. When investment changes, it causes a change in income in the same direction, but even greater in amount. For when more investment goods are produced, the income of those who produce them increases by the same amount, with the result that they spend more on consumers' goods and thus increase the incomes of those in consumers' goods industries by the amount of their increased spending. The increase in incomes resulting from this second rise in spending causes still more consumers' spending, and so on and on. In this multiplier process each change in income induces a change in consumers' spending in the same direction, and each change in the amount spent on consumers' goods produces a further change in income. While the succession of reactions continues almost indefinitely, each is somewhat smaller than the one before. And though there are an infinite number of links in the chain, the fact that each is smaller than the one preceding means that the total increase in income will be finite. However, the increase in income will exceed the increase in investment. Thus, if investment changes by $1 billion, we generally find that income changes by between $2 and $3 billion.

A change in the propensity to consume sets in motion a similar chain of events. When the propensity to consume is raised, there is an increase in the amount of spending upon consumers' goods. Producers of consumers' goods consequently enjoy higher incomes, and hence they too buy more consumers' goods. This in turn leads to still higher incomes for the producers of consumers' goods. And so the process continues, step after step. And because with each increase in income, consumption increases, but by a smaller amount, each link in this chain of events is smaller than the one preceding, and therefore the total change in income is finite. If the propensity to consume so increases that when income is constant, consumption is raised by $1 billion, then we should expect an increase in income of between $2 and $3 billion. The expansion in income would be even greater if increased consumers' spending led business firms to invest more.

Variations in Investment

To carry the analysis a step further, it may be asked why changes occur in investment or the propensity to consume. Private investment, we found, is done when business firms expect to profit from expansion and therefore build more plants, install more machinery, and acquire more inventory. On the other hand, when they see no profit in further expansion, they stop building plants, installing equipment, and acquiring inventory — in other words, when this happens, private investment is very low. It is scarcely possible to list all the factors that influence the expectations of profit from expansion, but some of the more important ones are the rate of interest, the stock of capital goods on hand, the expected growth in market for the products of business firms, the cost of capital goods, expected changes in the cost of capital goods in the future, development of new techniques, development of new products, and taxes. These and many other factors play some part in determining whether business firms will expect maximum profits in rapid expansion, in slow expansion, or in none at all.

Variations in government investment are somewhat more difficult to analyze. Government bodies do not invest for reasons of profit, and it is difficult to make any meaningful generalization as to the motives that prompt them to invest. It is hard, and perhaps unnecessary anyway, to classify the reasons that lead them to hire more school teachers, to build more airports, or to do any of the thousands of things that government bodies undertake. Variations in foreign investment will be considered later, in Part Five of this book, for we cannot account for them until we have studied the mechanism of international trade.

Investment, and especially private investment, is the unstable factor in our economy. At times, business firms anticipate a profit from expanding at a very rapid rate. At other times, as in 1932 and 1933, most of them expect no profit from expansion. Not only is private investment unstable, but our analysis of its determinants suggests that it may tend to vary cyclically as well. For with high investment, capital goods accumulate rapidly. This itself is an unfavorable factor for further expansion, so that after a period of high investment we have reason to expect a slack period, although offsetting forces may come into play to prevent it.

Because business firms earn high profits in periods of prosperity and

may suffer heavy losses in periods of depression, it is obviously to their interest that we enjoy prosperity just as it is to the interest of the farmer, the wage earner, and all other members of the economy. And since prosperity generally requires a high level of private investment, we may wonder why business firms ever permit investment to fall. Why should not each business firm, aware that prosperity can be maintained if private investment is kept high, and aware, too, of its own stake in prosperity, continue to order new plant, new equipment, and goods for inventory? The reason is that our economy is relatively atomistic. No one business firm, not even the biggest of them, does enough investing to guarantee prosperity for the whole economy. By the same token, each firm feels that it may reduce its investment without plunging the economy into depression, for depression comes when investment falls from, let us say, $50 billion to $2 billion a year. The decision of one firm, even a very large one, to reduce its investment expenditure from $200 million a year to $50 million cannot by itself cause depression. Hence, even if each businessman were thoroughly familiar with the theory of employment, it would not be to his interest to maintain investment at a high level unless the objective factors were favorable. The difference his own investment would make to the prosperity of the country would be so slight, and his own share in the difference so infinitesimally small, that he could not be expected to consider this in making his own investment plans.

Variations in the Propensity to Consume

Investment is the unstable factor in our economy; on the other hand, the propensity to consume has been relatively stable. Our habits of consumption change slowly. This is partly because our spending habits as individuals do not alter quickly, but it is also because the distribution of income among individuals does not change rapidly. A list of all the factors that affect the propensity to consume would be extremely long, like the list of all the factors that affect investment. Some of the more important determinants of the propensity to consume are the distribution of income, the tax structure of the economy, expectations as to changes in income and prices, the state of the stock market, the rate of interest, the policy on imports, the financial policies of corporations, and so on. A change in any one of these will bring about a change in the propensity to consume,

and such a change, if investment does not alter, will affect the level of income and employment. While normally we do not experience sharp fluctuations in the propensity to consume, there are reasons why at times we should want to modify it.

Remedies for Depression

We analyze variations in employment in order to learn how to control them. But controlling them, and preventing depressions, is a very complex task, as we have seen. The general objectives are clear: to increase employment, we must either increase the propensity to consume or increase investment. And, as a matter of fact, there is no reason why we should not try to increase both. The difficult problem is to bring about such increases. The only leverage for influencing private investment is to try to make conditions for expansion seem more profitable. Some control can be maintained over the interest rate, and reductions in the interest rate appear likely to help. But control over other factors is practically nonexistent. There is no way to reduce the stock of capital goods in existence — or rather, there would be no use in doing so; it is hard to persuade businessmen that the markets for their products are going to grow very rapidly in the future; it is not easy to stimulate invention. Thus the problem of raising private investment is very difficult indeed.

But raising public investment is quite simple; or rather, it is simple if our legislators think it is desirable. Public investment is as directly under social control as private investment is free from it, and it may be expanded to almost any level desired. Whether it should be so expanded or not is of course a different matter. For it is sometimes argued that while an expansion in public investment, considered by itself, may lead to an increase in income and employment, it may discourage private investment by injuring business confidence. Therefore, so the argument goes, private investment will decline when public investment is raised, in spite of the fact that the market for most goods is increased. It is impossible to assess this argument, since data are lacking, but there is every indication that it should not be given much weight. Certainly the experience of the war years gives us reason to suppose that a sufficient increase in public investment will normally raise income and employment. The long-run effects of an increase in public debt must also be considered, for a high level of public investment means a rapid increase in the public debt. How-

ever, debt also increases when private investment is high. The question then is whether one kind of debt is preferable to the other. And there is no very clear evidence that it is.

Public versus Private Debt

A high public debt has two effects which tend to cancel each other. It exerts a deflationary force, partly because it is likely to discourage private investment by threatening higher taxes, and partly because the transfer of income from taxpayer to bondholder reduces the propensity to consume. But a high national debt also has inflationary effects. When private individuals hold large numbers of bonds, the propensity to consume is likely to be unusually high. Moreover, when the commercial and Federal Reserve Banks hold bonds in large quantities, there is likely to be a great deal of money in circulation, and when this is the case the interest rate is kept low and private investment is encouraged.

While the deflationary effects generally seem to be the stronger, they can be minimized by altering the tax structure and spreading the ownership of government bonds as widely as possible. The difficulties which a large debt imposes upon the economy are due more than anything else to an unsatisfactory tax structure, the adverse effects of which are sharpest when the debt is high. Hence one remedy for the difficulties of a large debt is to improve the tax structure. While there is no reason to suppose that a high debt is desirable, it seems clear that it is not necessarily a very heavy burden. But so much depends upon the rate of interest, upon the tax structure, and upon the distribution of government bonds within the community, that generalizations about the effect of the government debt are impossible.

If a high public debt has certain deflationary effects upon the economy, so too has a high private debt. For it also implies a transfer of income within the economy which has deflationary consequences since the consumer supplies business firms with the receipts from which they pay the interest on their bonds; and it appears that with private as with public debt, the transfer of income is from the lower income groups to the higher. If so, a high private debt also reduces the propensity to consume. Moreover, when private debt is high, the favorable offsets which appear with a high public debt are missing, since corporation bonds are much less readily turned into cash than

government bonds, and therefore do not provide the same stimulus to consumption; and since corporation bonds cannot be purchased by the Federal Reserve Banks to increase commercial bank reserves.

While any method of achieving prosperity creates difficulties, it is hard to believe that the difficulties can be any worse than those brought on by depression. The difficulties of a growing debt can be partly offset by changing the tax structure. The effects of a very rapid accumulation of privately owned capital goods are adverse only if they stand in the way of continued accumulation. Otherwise they are advantageous, since a large stock of capital goods means great wealth and productive capacity. Compared to losses experienced in depression, none of these objections has much validity. The losses due to depression are measured in hundreds of billions of dollars. Any harmful effects of growing debt, rising prices, higher interest rates, and the other accompaniments of prosperity amount, in the final analysis, just to this: it grows somewhat more difficult to maintain prosperity. But this is hardly a good reason for not enjoying it in the first place.

Employment in a Rich Economy

Our economy faces very special difficulties in maintaining prosperity. Our good fortune in possessing so much wealth, and being able, therefore, to enjoy so high a standard of living, is our bad fortune when we seek to avoid depression. Because our stock of capital goods is so large, we are able, with full employment, to produce an immense output. This has three consequences, all unfavorable to the continuance of prosperity. In the first place, we do not want to buy nearly as many consumers' goods as we can produce when all our labor and wealth are employed. We want only about $135 billion worth of consumers' goods when our incomes amount to $200 billion. Consequently, if prosperity is to continue, we must buy $65 billion worth of investment goods every year. The first difficulty, then, is that we have to do an enormous amount of investment each year in order to have full employment.

The second is that the number of profitable opportunities for new investment is relatively low. This is not to say that a sufficient number can never be found. At times they can be, as in 1929 and 1941. But they are far less frequent than they would be if our stock of capital goods were smaller. When there is a large amount of industrial

plant already in existence and industry has the capacity to meet even the large requirements of a wealthy economy, business firms find relatively few profitable opportunities for expansion. While the amount of investment required for full employment is very high, the incentive to invest is relatively low.

Finally, even when investment reaches the full-employment level, the effect of our high productivity is to make prosperity the more difficult to maintain. We enjoy prosperity only when investment is very high — $65 billion a year. But at such a rate of investment, capital goods pile up in staggering amounts. If the goods are owned by the government, they cause no trouble. But if they are owned by private business firms, a saturation point is reached all too soon.[1] It is important to realize just how rapidly this can happen. It took only $1 billion of investment to expand the capacity of the aluminum industry by more than five times between 1940 and 1944. The total value of all the capital goods owned by all the railroads in the United States runs to something like $24 billion. If private firms invested at the rate of $40 billion a year, it would mean that even after making provisions for repairs and maintenance on all existing equipment, they would in a single year's time be adding more than the value of all our railroad properties. From one point of view this situation is obviously a good thing. It means that we are able to build up our productive capacity very swiftly, and hence to raise our standard of living very rapidly. But in an economy in which a large amount of investment is done for reasons of profit, it is dangerous. For as capital goods accumulate, the opportunities for further profitable investment decrease, unless there are offsetting forces at work, and therefore depression becomes inevitable unless public and foreign investment can be raised. For this reason it is not only difficult to attain prosperity in a very wealthy economy such as ours, but it is extremely difficult to keep it.

To put it a little differently, an economy can prosper only as it expands, and the richer it is, the more rapid the rate of expansion has to be. But the richer it is, the less inducement there is to expand, and even though that inducement should exist for a while, the very act of expansion brings it more quickly to an end. Hence there is

[1] This, of course, does not mean that the firms hold more capital goods than they want, or that society has too many. It means only that firms no longer wish to acquire new capital goods at a very high rate.

likely to be an ever-widening gap between the required amount of expansion and the incentives to expand which are normally available. In so far as we can use public investment wisely, we can break this vicious circle. Since public investment is not done for profit, it may still be carried on, even though the stock of capital goods — highways, public parks, hospitals, schools, and so on — is very large. Moreover, such goods do not necessarily add to our productive capacity.[1] Finally, it is not more difficult to build a thousand miles of highway in one year just because a thousand miles of highway were built the year before. And if we run out of suitable highway projects, as undoubtedly we should after a time, there are all sorts of other useful projects we can undertake, such as clearing slums and rebuilding housing, providing facilities for medical care, improving schools in areas where they are now deficient, and if necessary, dispersing cities to protect them from our improved methods of waging war.

It would be unwise to conclude from this that our economic future is black. We do not know enough about what the future holds in the development of new products, the search for new markets, and changes in our consumption habits to give us any confidence in such forecasts. But until important new projects arise which will demand enormous investment, or until we consume a much greater part of our full-employment output, we have reason to worry, though worries of this sort are often unfounded. A striking example of this is to be seen in a report on the prospects for the future made by a most gifted observer in 1886. He saw that railroad construction had been enormously important in maintaining the prosperity of Britain, Germany, France, the United States, and other countries, and he believed that the era of railroad expansion was coming to an end. Since he saw no new industries developing, his outlook was not optimistic. The following extracts from his report should be read not as showing how far wrong he was, but rather as showing how fallible even the best informed prophet can be.

> In England, Belgium, and France the railroads and canals that are really needed have been built. There remain only to be constructed

[1] While these goods will, if the projects are wisely chosen, add to the output of goods and services we want, they may not add to the output of goods sold on the market for profit. The productive capacity of privately owned industry may not increase. Moreover, a good deal of what we have called government investment may go to purchase services which obviously do not pile up at all.

feeding and competing lines. . . . In Holland the great works are completed; Amsterdam is united to the sea, international communications have been well established, and there are no longer urgent works to be undertaken, and the reward of capital to be invested now is not sufficient to tempt lenders. In Italy and Spain the great arteries are provided with railroads. . . . Harbors and rivers are sufficiently developed, and warehouses, water and gas works, tramways, etc., are largely provided for. The Pyrenees and the Alps are tunneled, and a sufficient network of international communication established. In England railroad building cannot be extended to a sufficient degree to absorb much capital or much labor. In Russia the principal lines of railroad have been built with the aid of the Government, and it is not likely that further construction will take place except for strategical purposes. Germany is provided with a full network of railroads, and the facilities for transportation are in excess of actual needs. Austria is in much the same condition as Germany, and Turkey also has as many railroads as can be used. In the United States the mileage of new railroads constructed has been out of all proportion to the increase of products to be carried.

The Suez Canal has been built, terrestrial and transoceanic lines of telegraph have been laid, and the merchant marine has been transformed from wood to iron. Today the carrying service of nations, and especially of the great marine nation, England, is overstocked to a far greater extent than the industries. On all sides one sees the accomplished results of the labor of half a century. . . . Whatever may have been the financial results, industry has been enormously developed, cities have been transformed, distances covered, and a new set of economic tools has been given in profusion to rich countries, and in a more reasonable amount to poorer ones. . . . This full supply of economic tools to meet the wants of nearly all branches of commerce and industry is the most important factor in the present industrial depression. It is true that the discovery of new processes of manufacture will undoubtedly continue, and this will act as an ameliorating influence, but it will not leave room for a marked extension, such as has been witnessed during the last fifty years, or afford a remunerative employment of the vast amount of capital which has been created during that period. . . . The day of large profits is probably past. There may be room for further intensive, but not extensive, development of industry in the present area of civilization. . . . Supplying themselves with full facilities for industries and commerce will give to each of the great nations of Europe and of America something to do, but the part of each in this work will be small and far from enough to insure more than temporary activity. It may

help to keep away stagnation and modify the severity and the duration of industrial depressions.[1]

Perhaps this analysis was made forty-five years too soon?

Inventions and Employment

Invention often makes it difficult to maintain full prosperity. During the period when industry is adopting new processes, investment reaches a high level. New machinery has to be built and installed, and the machine-tool and building industries are given a strong boost. Thus, the years in which a very important invention is being exploited are likely to be years of prosperity. But unless other inventions shortly become available, the aftermath may be unfortunate. The improved methods of production increase the productivity of labor so that fewer men are needed to produce a given output. If investment returns to its former level, as it will probably do unless new inventions come along, and if the propensity to consume is not altered, income also returns to its original level. However, because of the increased productivity of labor that improved processes secure, employment falls when income comes back to the original figure. Thus, while the short-run effect of invention is to raise employment, the long-run effect is probably to lower it beyond what it was in the first place. The effect is intensified in our economy because the use of more efficient processes is likely to reduce the propensity to consume. Those who argue that the machine creates jobs, and those who argue that it destroys jobs, are both right. In the short run the machine does create jobs, though in the long run it may destroy them. However, if new processes and improved methods of production become available in a steady stream, we never have a long run; we have only a succession of short runs. And this to some extent is what has happened, but perhaps not always to an extent sufficient to do away with the unfavorable long-run consequences.

Wages and Employment

It is not certain how changes in the wage rate affect income and employment. A rise in money wages may lead to a rise in income and employment. It may also lead to a decline. Under some circumstances a rise in money wages may reduce investment and may

[1] *First Annual Report of the Commissioner of Labor, 1886*, pp. 256–258.

not increase the propensity to consume. Under other circumstances, it may lead to an increase in investment and in the propensity to consume. It is impossible to predict the effect of a movement of money wages without knowing the circumstances in which the change occurs. We should generally expect an increase in the propensity to consume as money wages rise. This factor is, of course, favorable to income and employment and may itself stimulate private investment. But since an increase in wages is likely on balance to discourage investment, the total effect on income and employment is not certain.

Summary

To say that income and employment depend upon investment and the propensity to consume is only to mark out a path for useful analysis. In every problem that concerns employment and the national income, it is necessary to determine the effects of the projected change upon investment and the propensity to consume, but the determination of such effects is often very difficult. Investment — even private investment — is determined by such a multitude of forces that the analysis of a rise in the wage rate, a change in a particular tax, or a change in government regulations of prices or the securities market, is bound to be complicated. This does not mean that the analysis is incorrect; it means that we have to know much more about the economy than we do. We have to know as much as we can about the forces that determine private investment. We have to know, for example, the results on private investment of a 1 per cent change in the interest rate. We have to know how a change in business confidence really affects private investment. And so on, and so on. If this analysis of employment and the national income suggests anything, it should make clear the following points: that we can have prosperity; that we can get out of the deepest depression if we are willing to reason instead of following maxims that came to us from an economy as different from our own as it, in its turn, was from that of the Stone Age; and that we need to know more about our economy in order to refine our controls over it. We must be willing to learn from observation and experience until, as our knowledge grows, we are able to solve our most important economic problem.

SUGGESTIONS FOR FURTHER READING

Reflecting the great interest economists now take in these matters, there are now hundreds of articles and many good books written about the national income. The reader is urged to go through the pages of recent issues of *The American Economic Review*, *The Economic Journal*, *The Review of Economic Statistics*, *The Survey of Current Business*, and other periodicals for articles about his special interests in this field. The following list is representative of what can be found:

Bowles, Chester. *Tomorrow Without Fear.* New York: Simon and Schuster, 1946.

Intended to show that full prosperity can be maintained and how to do it.

Gilbert, Milton, and Jaszi, George. "National Product and Income Statistics as an Aid in Economic Problems." Appeared in *Dun's Review*, 1944. Reprinted in *Readings in the Theory of Income Distribution* (Philadelphia: The Blakiston Company, 1946).

A simple account of the meaning of national product and income concepts.

Haley, Bernard. "Economic Consequences of Deficit Financing," *Proceedings of the American Economic Association* (1940).

Hansen, Alvin H. *Full Recovery or Stagnation.* New York: W. W. Norton & Company, Inc., 1938.

——— *Fiscal Policy and Business Cycles.* New York: W. W. Norton & Company, Inc., 1941.

Many of Dr. Hansen's important studies on fiscal policy are gathered in these two books. Particularly recommended are Chapters 16–20 in *Full Recovery* and Chapters 1, 2, 4, 6, 9, 12, and 17 in *Fiscal Policy.*

Harvard and Tufts Economists. *An Economic Program for American Democracy.* New York: Vanguard Press, 1938.

A short and simple argument for government investment (in 1938).

Keynes, J. M. *The Means to Prosperity.* London: Macmillan & Company, Limited, 1933.

——— *How to Pay for the War.* London: Macmillan & Company, Limited, 1940.

——— *The General Theory of Employment, Interest and Money.* New York: Harcourt, Brace and Company, 1936.

All of Keynes's writings are of first-rate importance. *The*

Means to Prosperity argues in favor of public works. *How to Pay for the War* with exceptional lucidity sets forth the details of his plan, later adopted in Great Britain. *The General Theory* is of course a classic. It is not at all easy to read, but Chapter 18 gives a clear summary, and Chapters 22–24 are of great interest and easier than the rest of the book.

Robinson, Joan. *An Introduction to the Theory of Employment.* London: Macmillan & Company, Limited, 1937.
A useful summary, short and not difficult.

Smith, Dan Throop. "Economic Consequences of Deficit Financing — A Review," *Proceedings of the American Economic Association* (1940).
A clear statement of why some economists oppose raising government investment in depression.

Terborgh, George. *The Bogey of Economic Maturity.* Chicago Machinery and Allied Products Institute, 1945.
The author believes that private investment can be kept high enough and that we should not rely on high government investment or government measures to raise the propensity to consume.

United States Government: Temporary National Economic Committee. *Investigation of Concentration of Economic Power. Hearings before the Temporary National Economic Committee, Congress of the United States.* Washington: Government Printing Office, 1939–41. Part 9 (1940).
See especially the testimony of Currie (pp. 3520–38) and Hansen (pp. 3495–3520, 3538–59, 3837–59).

United States Government. *Hearings before a Subcommittee of the Committee on Banking and Currency, United States Senate, on the Full Employment Act of 1945.* Washington: Government Printing Office.
See especially the testimony of Paul G. Hoffman, Beardsley Ruml, and Henry A. Wallace.

United States Government: Federal Reserve Board. *Post War Economic Studies.* Washington: Government Printing Office.
Read especially articles in Reports No. 1 and No. 3.

PART FIVE

International Trade, Output, and Income

Introduction

ALTHOUGH the economy of the United States exports only about 5 per cent of its total output, and imports even less, our economic relations with other countries are of very great importance to us, as well as to the rest of the world. For one thing, we can acquire certain commodities only from foreign countries. Such things as natural rubber, bananas, tin, and nickel have to be imported unless we are willing to get along without them. And foreign countries are an important market for some of our products, such as our raw cotton, tobacco, wheat, automobiles, and machine tools. Moreover, the indirect influence of economic affairs abroad upon our own economy is also of great and sometimes of critical importance, most obviously so when economic forces lead countries to war or to active preparations for war. Some of these matters are considered in the three following chapters.

40

The Basis for International Trade

International Economic Transactions

Whatever may have been the history of political isolationism in this country, we have never been able to achieve complete economic isolation from the rest of the world. The clothes you are wearing, the book you are reading, and what you eat for breakfast provide concrete illustrations of this truth. You may not be dressed in English tweeds, but there is a good chance that the raw wool for your suit came from England or Australia. The pulp wood used to manufacture the paper on which your book was printed probably came from Canada. The coffee or tea you had for breakfast may have come from Brazil or India. The sugar was probably imported from Cuba, and bauxite for making the aluminum from which the kettle was fashioned came from Dutch Guiana. The list can be enormously extended. We depend upon the rest of the world for many of the things that we consume. We also depend upon foreigners to buy many of our products: our wheat, cotton, and tobacco, our machine tools, automobiles, and gasoline.

It might be expected, therefore, that since trade with foreign countries enters into our life in so many ways, it would have an important effect upon the operations of our economy. Up to this stage, we have paid only slight attention to this effect; that is, we have been treating our economy as though it were closed and had no economic relations with the rest of the world. In the next three chapters, we shall be concerned with the ways in which economic transactions with other countries influence the operations of the economy. As we shall see,

the nature of our economic relations with other economies is of considerable significance: they play a part in determining what kinds of goods we produce, how they are divided among us, how efficiently we produce them, and how much we produce. Furthermore, through economic ties of this kind, developments in this country affect other countries economically, and what happens in foreign countries influences conditions here.

The Extent of International Trade

The transactions between the United States and other countries include the purchase and sale of a wide variety of marketable items. We sell commodities to foreign countries and they sell goods to us. In addition we buy and sell such services as the use of ocean shipping, insurance, brokerage facilities, and warehouse accommodations. We buy and sell gold and silver. Foreign economies provide for the requirements of American tourists when they are traveling abroad, and our restaurants, stores, hotels, and railroads supply goods and services to foreign tourists. There are shipments of gifts between our country and others, sometimes in the form of money and sometimes of commodities. Finally, there is an active trade in loans and securities between the United States and foreign countries.

The value of our trade in commodities, services, securities, gold, and so on, has varied enormously. In 1929 it amounted to roughly $16 billion; by 1932 it had fallen to less than $5 billion; but it rose again with recovery, and in 1939 it reached approximately $8 billion; while by 1944 it exceeded $23 billion, even if we count lend-lease shipments only as an export of goods and not as setting up a claim by the United States against the recipients. In the next chapter we shall break these figures down to determine the importance of their various components; for the present, this brief over-all view will suffice.

Except when war prevents it, this trade is normally carried on with all countries. In 1938 Great Britain, Canada, Japan, France, Germany, and the Netherlands were our most important customers, but even Greenland bought from us. In that year our most important sources of supply were Great Britain, British Malaya, the Dutch East Indies, Japan, Brazil, the Philippines, Cuba, and above all, Canada, but we even got some things from the Falkland Islands. Thus every foreign country is at once a customer for us and a source of supply.

Let us see what kinds of things we ordinarily buy from these countries and what they get from us. Here is a list of our more important commodity exports in 1938:

grains	trucks and buses
tobacco (unmanufactured)	copper and copper products
raw cotton	machine tools
crude petroleum	electrical machinery and apparatus
motor fuel and gasoline	new passenger cars

The items in this list accounted for about 40 per cent by value of the goods we sold to foreign countries in that year. Our chief commodity imports for that year were as follows:

cane sugar	silk
distilled spirits	wood pulp and newsprint
crude rubber	copper
tobacco (unmanufactured)	tin
jute	fertilizers

These items cover about 49 per cent by value of our total purchases of goods from foreign countries in 1938. A careful analysis of the kinds of goods included in these lists will show how international trade takes place, and will indicate some of the conscquences of that trade.

Specialization and International Trade

Our exports seem to fall into two broad classes, the highly fabricated products of our great mass-production industries, and raw materials that are produced with great efficiency in this country. Our chief imports are raw materials which cannot be efficiently produced in our climate, and mineral resources not found in this country. Thus, our commodity trade with foreign countries consists essentially in an exchange of automobiles for tin, wheat for coffee, machine tools for bananas, and even, as may be seen, tobacco unmanufactured for tobacco unmanufactured.[1] We produce more automobiles and wheat than we consume in this country, and the difference is sold to foreign

[1] This is not to say that the seller of automobiles buys tin, or that the countries that buy automobiles from us sell us tin; the details of trade between countries are much more complicated than this, as we shall see.

countries. Other countries produce more coffee, rubber, and tin than they require for their own use, and part of their surplus is sold to us.

Trade of this kind implies specialization. To the extent that we produce certain goods in sufficient quantities to meet not only our own demands but also certain foreign requirements, we specialize. In other words, we concentrate a relatively large proportion of our own productive resources in export industries, and, to the extent that foreign countries produce certain commodities for the United States and other markets, they specialize in the production of those goods that we import. The existence of foreign trade is in itself an indication that there is some specialization in production. To put this in another way, in the United States a relatively large amount of labor and capital is employed in the automobile, machine tool, wheat, cotton, and other export industries, and a relatively small amount is employed in growing coffee and tropical fruits, mining tin, preparing wood pulp, or providing other things which we import. Thus, while we specialize in producing electrical equipment and growing wheat, other countries specialize in producing nickel and wood pulp and in growing coffee. To use a term introduced earlier,[1] the pattern of production reflects the nature of our international trade.

Obviously we do not specialize in producing things for which our resources or climate are unsuitable, but rather in supplying goods for the production of which we are well equipped. Sometimes we are able to produce efficiently because of our highly skilled labor, and sometimes because of our climate or natural resources. Other countries likewise seem to devote their productive energies to those industries for which their efficiency is high because of special skills and aptitudes, or because of suitable climate and resources. A country tends to specialize, as an inspection of the list of its exports would suggest, in those fields where it has certain advantages in production. This means that we sell other countries those goods which we produce with exceptional efficiency, and they sell us things which it is very difficult for us to produce. In a sense we exchange machine tools, which can only be produced with highly skilled labor, a highly developed steel industry, skilled engineering, and so on, for coffee, which grows best where there is tropical heat and moisture.

Let us try to picture what would happen if, instead of importing,

[1] Compare Chapter 20.

we tried to grow all our own tea, coffee, sugar cane, and tropical fruit, while countries which had been supplying such goods to us were to produce their own machine tools, wheat, and automobiles, or if, in other words, international trade were to stop. If it were necessary, we probably could produce these commodities, but the amount of labor we should have to devote to the task would be very great indeed; and certainly if Nicaragua, Ceylon, Colombia, or Brazil were to attempt to manufacture such complicated articles as automobiles and typewriters, they would have to devote a great part of their resources to these industries. An automobile might be forthcoming after months of work, but its production would require so much labor that little else could be produced. If there were no specialization of the kind that foreign trade makes possible, the total output from a given amount of employment in any country would be considerably lower than it is.

Specialization based upon foreign trade or, as it is often called "the international division of labor," promotes productive efficiency. By taking advantage of the opportunities provided by foreign trade, we are able to acquire products with the minimum expenditure of labor and materials, since we can employ resources to produce more of certain goods than we need, and thus can use the surplus to acquire goods that we cannot produce easily. Instead of devoting labor and materials to the production of coffee in this country, which would be a wasteful way to use our resources, we employ them to manufacture such things as automobiles and machine tools. At the same time, other countries produce the coffee that we want and in exchange are willing to take automobiles from Detroit or perhaps textiles from Lancashire. This raises two questions: First, what drives a country to produce those goods for which its efficiency is relatively very high? And second, what do we mean by the statement that "a country concentrates its resources in those fields where it is relatively efficient"? Does it imply that a country will supply only those commodities in the production of which it is more efficient than the rest of the world, or rather, that it will specialize in producing those commodities for which its own resources are best fitted? To determine which is correct, we shall need to examine more carefully the reasons why international trade takes place.

How Specialization Comes About

The buyer of any commodity will generally choose the cheapest possible source of supply. If he can purchase the kind of shoes he wants in either of several different stores, he will, unless deflected by considerations of service or prestige, buy where the price is lowest. Naturally he will do the same thing if some of the possible sources of supply are foreign. Other things being equal, if a purchaser can get something more cheaply from a foreign source than from a domestic one, he will buy the foreign product, though he must, of course, include in the cost of getting the goods from a foreign source — or, for that matter, from a domestic one — shipping charges, brokerage fees, and tariffs or other taxes. Let us see the significance of the fact that for the buyer the lowest price is the best price, and what it implies about the nature of international trade.

Suppose we are concerned with only two countries, that we have the prices at which a representative list of commodities are available in each country in its own currency, and that we know the cost of transferring each article from one country to the other. The basic data are shown in Table 62. Let us consider the significance of these data.

TABLE 62

Comparative Prices of Foreign and Domestic Commodities

Commodity	Price per Unit in United States, if Produced in United States (in U.S. Dollars)	Price per Unit in France, if Produced in France (in francs)	Cost of Shipping per Unit *	
			From U.S. to France (in U.S. Dollars)	From France to U.S. (in francs)
A	6.00	240	.25	10
B	9.00	480	.75	24
C	3.00	180	1.00	72
D	4.00	360	20.00	1565
E	4.50	240	.30	12
F	6.50	600	.15	12
G	1.50	60	.15	18
H	.12	2	.03	2
I	.82	120	.18	12
J	1.25	60	.06	6

* This figure includes transportation charges, tariffs, insurance charges, and so on.

An American buyer can purchase commodity A from a domestic source for $6 a unit. It would not pay him to import it unless he could get it more cheaply that way than by buying it from a producer in this country. He can import it by paying the dollar equivalent of 250 francs, the price he would have to pay in France (240 francs) plus the cost of transfer to the United States (10 francs). Hence, only if he can buy 250 francs for less than $6 will he import this commodity. A French buyer can make a similar calculation. If he can buy $6.25 for less than 240 francs, it would pay him to import the article, but if the United States currency should be too expensive in francs, he would buy at home.

If the rate of exchange between the dollar and the franc is known, we can calculate the price of the imported goods in each country. By comparing these prices with the prices of the goods produced in the home country, we can then determine whether the goods would be imported or not. If the exchange rate is set at $2 for 100 francs (a franc then being worth 2 cents), we can derive, from the data provided in Table 62, the figures shown in Table 63. Inspection of Table 63 discloses the following. American buyers get a better bargain by importing A and H than by purchasing them at home.

TABLE 63

Prices of Imported and Domestic Product

(When $2 = 100 francs)

Commodity	In the U.S. (in dollars)		In France (in francs)	
	Price of Home Product	Price of Imported Product	Price of Home Product	Price of Imported Product
A	6.00	5.00	240	312.50
B	9.00	10.08	480	487.50
C	3.00	5.04	180	200
D	4.00	38.50	360	1200
E	4.50	5.04	240	240
F	6.50	12.24	600	332.50
G	1.50	1.56	60	82.50
H	.12	.08	2	7.50
J	.82	2.64	120	50
K	1.25	1.32	60	65.50

Buyers need to pay only $5, inclusive of shipping charges, for A if they import it, while they must pay $6 if they buy it from an American supplier. Likewise, the domestic price of H is 12 cents a unit, while it can be imported and made available at only 8 cents a unit. But obviously no one would import C, which costs $5.04, when they can acquire it from American firms for $3. The same considerations show that French buyers would import F and J, and that they would be indifferent to the source of E. Thus, the United States would specialize in producing F and J and perhaps E, and France would concentrate on A and H. Each country would produce enough of the other items in the list to meet its own requirements.

A glance at Table 63 will indicate that we produce F and J very efficiently compared to the French, while they produce A and H very efficiently compared to us. Thus, at the assumed exchange rate, United States producers have a great advantage over French producers in producing F and J. Likewise, French producers have their greatest comparative advantage in the production of A and H. But notice that the term *comparative advantage* must be interpreted carefully so that it covers not only the price of the good in the country of its origin, but also the cost of transferring it to the other country. For even if American producers charged $30 instead of $4 for D — a figure more than four times as high as the price charged by French firms in their own country — it would not be imported into the United States because of the very high cost of transfer; that is, French producers would be far more efficient in *producing* D than in *supplying* it to us. Hence, even though one country is able to produce something more efficiently than another, and even though the comparative advantage in respect to that country is greater than for any other, it would not specialize in producing it if the cost of shipping it abroad is too high. Countries apparently specialize in producing those things which they can supply to others most advantageously, allowing for the costs of transfer.

If the exchange rate were different, the items entering into trade would also be somewhat different. Table 64 has been prepared from the data contained in Table 62 on the assumption, this time, of an exchange rate of $1.50 for 100 francs. Inspection of Table 64 shows that American buyers would find it cheaper, at this exchange rate, to import commodities A, B, E, G, H, and K than to purchase them from domestic producers. French buyers, on the other hand, would

TABLE 64

Prices of Imported and Domestic Product: Different Exchange Rate

(When $1.50 = 100 francs)

Commodity	In the United States (in dollars)		In France (in francs)	
	Price of Home Product	Price of Imported Product	Price of Home Product	Price of Imported Product
A	6.00	3.75	240	416.67
B	9.00	7.56	480	650
C	3.00	3.78	180	266.67
D	4.00	28.87	360	1600
E	4.50	3.78	240	320
F	6.50	9.18	600	443.33
G	1.50	1.17	60	110
H	.12	.06	2	10
J	.82	1.98	120	66.67
K	1.25	.99	60	87.33

find it cheaper to import F and J than to buy them at home. Hence, commodities A, B, E, G, H, and K would now be produced in France in amounts that exceed French requirements, while in the United States the output of F and J would exceed domestic needs. As before, each country would concentrate on the production of those commodities in which its comparative advantage was highest, interpreting this to cover not only relative domestic prices, but also the costs of transfer.

Factors on Which Specialization Depends

Now let us examine commodity A in the light of all this. Whether it is imported into the United States or not is determined, in the first instance, by whether American buyers can get it more cheaply from foreign than from domestic sources. But this depends upon three factors: the exchange rate, the cost of transfer, and the ratio of the domestic price of A in the United States to that in France.

Let us see how each of these factors affects the situation. If the exchange rate is such that the cost of the franc exceeds 2.4 cents, it will not pay to import the good into the United States at the prices given in Table 62, for the American buyer who would have to pay 250 francs — or in American money 250 × 2.4 cents, or $6 — could

get the article as cheaply from an American firm. Indeed, if the exchange rate stood at 100 francs = $2.61 or more, American firms would export A to France. Thus, the more expensive the foreign currency, the lower are our imports and the higher our exports. Like the exchange rate, the cost of transfer is also important in determining which way goods will flow. Again referring to Table 63, if shipping costs were high enough, commodity A would not be imported into the United States. For with an exchange rate of 100 francs = $2, it would not pay to import commodity A if transfer costs exceeded 60 francs. If they were, say, 70 francs, Americans would have to pay the dollar equivalent of 310 francs, or $6.20, to buy the imported good, whereas they could buy the same thing at home for $6. Hence the higher the cost of transfer, the lower are the imports — and, of course, the less the specialization. Finally, the ratio of the American to the French domestic price affects the nature of trade, for the higher the ratio, the more likely Americans will be to import goods from France, other things being equal. Thus, if the ratio stood at $\frac{\$4.90}{240 \text{ francs}}$ — or, with an exchange rate of $2 = 100 francs, at $\frac{\$4.90}{\$4.80}$ — American buyers would not buy from France because, since they must also look after transfer costs if they import, they can buy more cheaply from a domestic producer. But if the ratio of American to French prices stood at $\frac{\$5.10}{240 \text{ francs}}$ — or, with an exchange of $2 = 100 francs, at $\frac{\$5.10}{\$4.80}$ — Americans would import. If the ratio of the domestic to the foreign price is high enough, we import; otherwise, we do not. Hence we see that the direction of trade depends upon the exchange rate, transfer costs, and the ratio of internal prices in the two countries involved.

But we must examine the ratio of internal prices more carefully. At first glance, this ratio may seem to depend only upon the relative efficiency of production in the two countries — and in part it does, but not altogether. Certainly, greater efficiency of labor and capital in producing a certain commodity will be a factor making for a reduction in the ratio of the home to the foreign price, thus increasing the probability that a country will export rather than import. But relative efficiency in physical terms is not the only factor to be considered

in determining this ratio. Relative rates of pay are also influential in determining the ratio of domestic prices, for if, to take an example, wage rates in this industry in France are especially low, then, even though French efficiency may not be great, the cost of production may be relatively low.

These determinants account for differences in cost, but not for differences in price. The final consideration, then, is the relation between cost and price in the two countries. If, for instance, the American price were twice the marginal cost, and the French price were only 50 per cent above marginal cost, the ratio of the American to the French domestic price would tend to be high, even though efficiency in the United States were great. When firms charge the most profitable price, the relation between cost, especially marginal cost, and price, depends upon the elasticity of the average revenue function — in other words, upon the degree of monopoly.[1] The lower the elasticity of demand, and hence the higher the degree of monopoly, the higher is the ratio of price to marginal cost. Thus, if the degree of monopoly is particularly high in the United States as compared to France or other foreign countries, the ratio of the United States price to the French domestic price will also be high, other things being equal. And when this ratio is high, it reduces the probability that we will export the good, and hence that we will specialize in its production.[2]

The question as to which commodities will enter into international trade is thus seen to be a complicated one. In general, a country will export, and thus specialize in, the production of those goods which it can provide to foreigners on the most advantageous terms. Thus, whether a commodity will be traded, and if so which way, depends on a combination of factors: transfer costs, relative efficiency of production in the countries concerned, the rate of exchange, relative rates of pay for the factors of production, and the degree of monopoly in each country. An economy generally specializes in producing commodities for which its resources are particularly efficient, transfer costs are low, the degree of monopoly is relatively low, and the rates of pay are not abnormally high. And it will generally secure from other economies those products which are produced with high effi-

[1] See page 199.

[2] Prices for foreign buyers may be less than prices charged domestic customers. However, we shall not be able to discuss this point.

ciency, and hence low costs, by firms which have a low degree of monopoly, provided that transfer costs are low.

We conclude, therefore, that the international division of labor generally will be such that each country will concentrate in the production of those commodities which it can produce most efficiently, but this result is obviously conditioned and modified by the existence of transfer costs, by variations in the degree of monopoly between industries and countries, and by the existence of abnormally low rates of pay in certain countries and industries. Not only does the pattern of specialization reflect relative productive efficiencies; other factors help to determine it too. Nevertheless, there is certainly a tendency to concentrate in fields where productivity is highest. Hence each country produces a larger output *with a given level of employment* when international trade takes place freely than when it is limited. And the greater the output from a given volume of employment, the higher is the standard of living corresponding to that level of employment.

Trade When a Country Has No Advantages

Suppose there is one country that is less efficient than all other countries in the world. Would it not, if trade were free, be forced to import all its goods from other countries? Could a country without specialized resources or a unique climate or very skilled labor engage in world trade except as a buyer? It is most unlikely. Such a country would export if it could supply something on terms better than those at which other countries could get it anywhere else, including, of course, their own producers. Since its productivity is low, it could give better terms only if the rate of exchange, its wage rates, or the degree of monopoly were low enough. In other words, it might export, even though its efficiency were low, if these factors were favorable to it. And if these factors were favorable to exports, they would limit its imports also. While we cannot say anything about relative wage rates or the degree of monopoly, we shall see in the next chapter that it is almost inconceivable for its exchange rate to be maintained at such a level that trade would be entirely or even preponderantly one way. The price of its currency is almost certain to be low enough to permit it to export something, and at the same time to limit its imports. In short, the country which is a relatively inefficient producer will export something, and even the fact that it

is less efficient in every kind of industry does not mean that it will import everything it needs. For the rate of exchange is normally established at a level which permits it to export those commodities for which its comparative disadvantage in production is least, and to import only those for which it is greatest. Moreover, wages are likely to be unusually low in such a country, and for this reason, too, other countries may be able to buy some of its products cheaply, especially those where its disadvantage in production is least marked. Thus, in the real world, as distinguished from the dream world which some politicians talk about, there is no danger that a country will be flooded by imported goods.

International and Interregional Trade

In summary, each country tends to specialize in the production of those commodities for which its margin of superiority is greatest, or, if it excels in no industry, where its margin of inferiority is least; each country, in other words, does what it can do best, when international trade is free. Hence, each country tends to devote a relatively large fraction of its resources to those industries for which, by comparison with the rest of the world, it is best fitted. A country which engages in international trade produces articles which cost it less to make than to buy, and it buys from others what it can acquire in that way at a smaller cost than if it were to make them itself. The specialization made possible by international trade is essentially no different from that occurring within a country, though some of its effects may be different. Thus, within the United States, the South concentrates on the production of raw cotton, the Northeast produces machine tools and other manufactured goods, and the Midwest produces foodstuffs and many manufactured goods; and there is considerable trade between these areas. The textile mills of New England purchase large amounts of cotton from the South, and consumers in the South buy cars and household appliances from the Midwest and New England. Thus, the specialization that trade makes possible is no different when the trade is between countries from that occurring when it is within them. The trade of New England with the South makes possible the same kind of specialization as occurs with foreign trade.

What, then, is the difference between domestic and foreign trade? The difference consists in the fact that trade between countries takes place between sovereign economies. Foreign trade is trade between

parts of the world over which economic controls are differently administered: where policies differ on monetary matters, wages and prices, banking procedures, and so on. Furthermore, this generally implies that labor is not free to move from one trading unit to another. On the other hand, interregional trade is carried on between areas subject to common policies on fiscal matters, money, labor, and so on. New England and the South are subject to the same economic controls. The use of the dollar issued by the Federal Reserve Bank of Atlanta is just as free in New England as the use of the dollar issued by the Federal Reserve Bank of Boston. No immigration laws check the movement of labor from one region to the other. No supervision is exercised by one region to prevent the inflow of loans or the making of loans to other regions. Trade between countries is vastly different. Money used by one country cannot be used directly for purchases in another. Ordinarily there are restrictions on the movement of labor between countries. Tax policies differ widely. And finally, there are often restrictions on the lending of money by one country to another.

But these differences between international and interregional trade should not obscure one basic similarity: whether trade is carried on between two countries or within a country, the specialization which it makes possible means higher productivity in terms of physical output, and hence, for a given level of employment, a higher level of income. And any obstruction to that trade reduces productivity, and therefore the level of income, unless employment is increased because of the obstruction.

Trade in Securities

Besides the trade in goods and services, there is normally a flourishing international trade in securities. Since securities are purchased for reasons of profit, there are a number of factors involved in such a trade. First, the rate of interest on the securities of one country may be higher than on those of another. Making due allowance for the risks of default and so on, it is clear that lenders purchase the securities that promise the highest returns. Second, when securities are purchased for speculative reasons, those are most attractive that are expected to rise fastest in price. Thus, the stock-market boom of 1927–29 in the United States made the purchase of American securities appear very attractive to many foreigners, even though the dividend yield was often very low. Next, since what the foreigner receives

from dealing in American securities depends in part on what he has to pay for the United States dollar in terms of his own currency, and on what he can expect to get for the dollar when he wants to take his profit, he will be more inclined to buy our securities when he believes that the price of the dollar, expressed in terms of his own currency, will rise. Similarly, the American lender finds the purchase of foreign securities particularly attractive when he expects the price of foreign currency to rise in terms of the dollar.

When carried on by private individuals rather than by central banks, the purchase and sale of gold and silver between two countries also rests upon business considerations. Thus, if an individual can buy gold in England and sell it in the United States at so high a price that he can make a profit when he reconverts his money into pounds, there will be a movement of gold from England to the United States.

Multilateral and Bilateral Trade

While our foreign trade in commodities is simply an exchange of such goods as automobiles for coffee and sugar, the exchange need not be direct. The countries which supply us with the bulk of our imports may not want our tobacco and machine tools, and hence we may not trade directly with them. Consequently trade is generally multilateral. We may, for instance, buy tin from British Malaya but sell very little to that country. On the other hand, we may sell more to Great Britain than we buy from her, and the British may sell much more to Malaya than they buy from that country. In effect, then, we have a three-way transaction involving a purchase of tin by the United States from Malaya, a purchase of raw cotton by England from the United States, and a purchase of textiles by Malaya from England. In fact, many more than three countries participate in such exchanges. With multilateral trade there are no tying devices by which each country purchases only from its own customers.

Multilateral trade makes possible a much more thorough specialization of resources than there could be if trade consisted only in direct swapping. If trade is bilateral, a country specializes in the production of only those commodities that are wanted by the countries from which it imports. If we traded this way, it would mean that we could get only a small supply of Malayan tin, since Malaya does not want many of our goods. But if we got only a small amount of tin from there, we would presumably have to supply the greater part of our

own requirements by mining it in this country. While bilateral trade permits of some specialization of resources, it ordinarily does not allow nearly as much as multilateral trade permits.

Obstacles to Trade

Obstacles to the free flow of trade are of many kinds. Government imposes some of them in the form of tariffs or quotas. It also sometimes controls the foreign exchange market in order to limit the amount of foreign currency available to importers. Bilateral trade agreements also reduce the volume of foreign trade. These forms of public control of trade are to some extent practiced by all countries. But there is also a kind of private control which is perhaps no less important. The most important method of private control is the cartel arrangement, to which firms in several countries are parties. A common provision in such an arrangement is for the members of the cartel to divide the world market on a prearranged basis. For example, in 1936 the General Electric Corporation and the Carboloy Company, one of its subsidiaries, signed an agreement with the German Krupp interests to control dealings in tungsten carbide. According to the terms of this agreement, Krupp was to export no tungsten carbide to the United States, and the American companies were to ship none to Europe or even to Latin America. Such an agreement is obviously much more effective in controlling trade than a simple tariff. It receives less attention merely because so much less is known about it. Cartel arrangements of this kind are, for obvious reasons, kept secret.

While we are not yet in a position to pass judgment upon trade obstacles, one thing we can be sure of. To the extent that these policies check trade, they discourage that specialization of production in each country which makes for maximum productive efficiency. Thus, such obstacles create inefficiency, among other things; and while there may be reasons for encouraging inefficiency in production, it must not be forgotten that the standard of living of the whole world, ourselves included, is lowered by it, provided that the total volume of resources actually employed is the same with and without trade obstacles. It may be a good thing to manufacture inefficiency — either by government, as when it raises the tariff, or by private firms in different countries, as when they agree among themselves not to compete. But it will only be so if we want inefficiency for its own

sake, or if we expect to offset its harmful effects upon our standard of living by creating greater employment. There is no obscuring the fact that, as one economist put it, tariffs and other controls of this sort, both public and private, have the same effect upon our economic well-being as filling in the harbors. But perhaps we want our economic harbors to be filled in. We shall discuss this problem in Chapter 42.

Summary

The existence of international trade implies that countries specialize in production. While the pattern of specialization is complicated, it is basically one in which each country devotes its resources to what it can do best. Even if a country is inferior to some other in every line of industry, it will be led to concentrate in those fields where its margin of inferiority is least. To some extent, however, this pattern of specialization is obscured by other factors; for differences between countries in transfer costs, wage rates, and the degree of monopoly also affect the pattern. When governments and private firms impose other obstacles upon trade, the pattern of specialization is even further obscured. Every deviation of the pattern from that which reflects relative productive efficiencies means a lower total output for a given level of employment. Thus, when government or private firms force an economy to produce commodities for which it is not well suited, they reduce the total output from a given level of employment. This is not necessarily harmful, of course, if such measures lead to higher employment. But if we can get full employment by using other methods that do not reduce efficiency, so much the better; for to do so would result in an even higher output and thus in an even higher standard of living.

41

The Balance of Payments and the Exchange Rate

When a New England textile firm purchases raw cotton, it pays for it with American currency, and when it sells finished cloth to a clothing firm in New York State, it receives American dollars in payment. Even if it sells on credit, the debt is expressed in, and eventually is discharged by, the payment of American money. In the same way, trade within Great Britain is carried on with British pounds, trade within France is conducted with French francs, and so on. The New England textile firm requires payment for its cloth in American money because only American money is acceptable to its wage earners and its suppliers. It cannot use British pounds for its payroll, nor can it pay taxes in Italian lire. Moreover, the firms to which it sells, if they are located in this country, are likely in any case to offer American money as payment, for they are unlikely to have a stock of foreign currency to draw on.

But when the textile firm sells to someone in a foreign country, the situation is different. The purchaser is not likely to have American dollars, nor is the American firm generally willing to accept payment in foreign money. In one way or another, the foreign purchaser of American products must secure American money if he is to be able to pay for the goods. An American who purchases goods from a foreign supplier is subject to a corresponding requirement, since the foreigner normally expects to be paid in the currency of his own country. Thus, the American buyer must somehow get the appropriate kind of foreign money in order to pay the foreign supplier.

Trade between countries is, therefore, more complicated than trade within a country. To the ordinary problem of the buyer, that of having *enough* money, a further difficulty is added — that of having *the right kind* of money. The money must be in a form acceptable to the seller. Thus, the foreign purchaser of American machinery must have not only enough purchasing power, but in addition he must have access to American dollars unless for some reason his American supplier is willing to accept foreign money in payment. But American money is especially hard to acquire outside the United States, because the monetary institutions of foreign countries are not permitted to create American dollars, any more than ours can create British pounds or French francs. Hence the foreign buyer may not readily find American money when he wants to buy our products. How can he get dollars, and how can an American acquire francs or rubles?

The Source of Foreign Exchange

There is, in the final analysis, only one source from which foreigners can get American dollars — from Americans. Likewise there is only one source from which we can get foreign money — from foreigners. All the American money that foreigners use in buying from us must originally have been supplied by Americans, and the same thing is true for the money of every other country; it must have come from there. Why, then, do we supply American money to foreigners? Because we want to buy from them, and to do so we need their currency, which we can acquire only by giving American money in exchange. An American who wants 1 million francs can get them by giving about $40,000 in American money to holders of francs. Hence all the American money that foreigners secure comes to them in exchange for foreign money we want. Or, to look at the transactions from another point of view, all the foreign money we acquire is supplied to us in exchange for dollars that foreigners want.

Thus, whether an international transaction is a sale on credit or for cash, and whether it is for goods, services, securities, or gold, payment must be made in the seller's currency, which he has made available to the buyer in exchange for the latter's money. Thus, when we buy from foreigners they receive American dollars, and when we sell to them they surrender American dollars. Or to look at it from their standpoint, they make foreign money available [1] when they buy

[1] To be exchanged, of course, for American money which is used to pay our suppliers.

from us, and they receive it again when they sell to us. In a sale on credit the same rules apply, except that this kind of transaction involves, not only the sale of a commodity, but also a sale in the other direction of a note or a promise to pay. Thus, when a French firm buys a lathe from an American firm and promises payment a year hence, the lathe moves to France and dollars move from France, and at the same time a security — the French firm's promissory note — comes to this country and is financed by dollars going to France.

Every dollar which foreigners acquire in a certain period may be used in that period in one of two ways. It may be spent upon our products, or it may be stored abroad.[1] This follows because the dollar has to be somewhere at the end of the period — or, more accurately, because it is necessarily owned by someone who must be either an American or a foreigner. If at the end of the period it is owned by an American, we can conclude that it was spent during the period upon our products. If it is owned by a foreigner, it has been added to the stock of dollars held abroad. Thus, if we supply foreigners with $5 billion in a year, the sum of their purchases from us and the additions they make to their own stocks of American dollars must amount to $5 billion. If they buy $4.5 billion worth of goods, services, and so on from us in the year, they have necessarily added $0.5 billion to their dollar holdings. But what happens if their purchases from us equal $5.5 billion in the period? Since we provided only $5 billion in American money, they have to get the other $0.5 billion elsewhere. Obviously they can do so by drawing on their previously acquired holdings of American money. Accordingly we may conclude that, in any period,

1. Dollars made available to foreigners = Foreign purchases from the United States plus the additions to (or minus the subtractions from)[2] their stocks of United States dollars;

or, rearranging slightly; in any period,

[1] Although it is convenient to think in terms of actual currency, most transactions are financed by transferring bank deposits between countries, just as most internal transactions are financed by check. Thus, the phrase "the stock of dollars held by foreigners" is to be understood in the more general sense of deposits held by foreigners payable in American funds. An Englishman who has a $50,000 deposit in the National City Bank holds dollars, in our terminology.

[2] For ease in expression we shall henceforth omit the clause in parentheses. In any case, subtractions are negative additions.

2. Foreign purchases from the United States = Dollars made available to foreigners minus additions to their holdings of United States dollars.

This classifies the *uses* to which foreigners put the dollars they receive.

We may draw up a similar expression to show the *sources* of the dollars made available to foreigners. When we purchase $5 billion worth of foreign money, foreigners acquire $5 billion in American money. We may use the foreign money that we get in exchange either to finance our purchases from them or to build up our own holdings of foreign money, just as they do with ours. Hence, if our purchases abroad total only $4 billion, we have added $1 billion worth of foreign money to our existing stocks. But if our purchases from foreigners amount to $6 billion, say, and we make only $5 billion in American money available (or, to put it otherwise, we buy $5 billion worth of foreign money), this signifies that we have drawn $1 billion worth of foreign money from our holdings.[1] Hence we may conclude that, in any period,

3. The dollars made available to foreigners = Our purchases of foreign money
= Our purchases from foreigners plus the additions to our holdings of foreign money.

Considering this result, together with number 1 above, we have, in any period,

American money used by foreigners = American money supplied to foreigners,

and hence,

4. $\left\{\begin{array}{l}\text{Foreign purchases from the United States plus additions}^{2}\text{ to foreign holdings of our money}\end{array}\right\} = \left\{\begin{array}{l}\text{Our purchases from foreigners plus additions}^{2}\text{ to our holdings of foreign money.}\end{array}\right\}$

[1] This is a negative addition; see preceding footnote.

[2] See preceding footnote.

The Balance of Payments

It is instructive to examine recent data on foreign requirements and supplies of American money in the light of these relations. The figures for 1938 are given in Table 65.

TABLE 65

American Money Used by Foreigners in 1938
(in millions)

Foreign purchases of United States goods	$3,155
Foreign purchases of United States services	1,096
Foreign purchases of long-term securities from United States	1,724
Foreign purchases of short-term securities from United States *	292
Additions to foreign holdings of United States dollars	16
Unexplained items †	532
Total American funds disposed of by foreigners	$6,815

* This is a "net" figure. We bought some short-term securities from them.

† Described by the Department of Commerce as a substantial volume of unidentified capital transactions, as well as the net result of possible errors and omissions in the other estimates.

Let us now see in Table 66 how the funds were made available to foreigners in that year.

TABLE 66

Supplies of American Money to Foreigners in 1938
(in millions)

Our purchases of foreign goods	$2,003
Our purchases of foreign services	1,247
Our purchases of gold and silver	1,864
Our purchases of long-term securities from foreign countries	1,701
Total amount of American funds supplied to foreigners	$6,815

In short, foreigners used $6,815 million in American funds during 1938, almost all of it to finance their purchases of American goods, services, and securities. We provided them with these dollars chiefly in the process of financing our purchases of their goods, securities, services, and gold.

Computations similar to the above may be prepared for any year. In 1938, we were experiencing a moderate depression. By contrast, 1939 was a year of moderate prosperity. In Tables 67 and 68 are

the Balance of Payments, as this kind of account of our economic transactions is called, for that year.

TABLE 67

American Money Used by Foreigners in 1939
(in millions)

Foreign purchases of United States goods	$3,241
Foreign purchases of United States services	1,050
Foreign purchases of long-term securities	1,624
Foreign purchases of short-term securities *	1,185
Additions to foreign holdings of United States dollars	117
Unexplained items	1,037
Total American money used by foreigners	$8,254

* This is a "net" figure. We bought some short-term securities from them.

TABLE 68

Supplies of American Money to Foreigners in 1939
(in millions)

Our purchases of foreign goods	$2,362
Our purchases of foreign services	1,271
Our purchases of gold and silver	3,111
Our purchases of long-term securities	1,510
Total amount of American funds supplied to foreigners	$8,254

Foreigners received $8.254 billion in American money from us, about three-eighths of it in exchange for gold and the rest for goods, securities, and services. They used it mainly to acquire goods, services, and securities. They used all that we gave them, either for buying or for holding. Hence the Balance of Payments always balances.[1] But as we showed earlier, this merely means that every dollar acquired by foreigners is used either to purchase from the United States or to add to dollar holdings abroad. In other words, dollars do not disappear into thin air. The fact that the Balance of Payments *does* balance has no implications beyond those of arithmetic. Thus it does not imply equilibrium or prosperity, nor does it imply that foreign trade is a good thing or that no country benefits from it. It means no more than that whoever made the statement of payments could

[1] That is, it balances if changes in foreign holdings of United States money, American holdings of foreign money, and the unexplained items are included.

add and subtract. If properly estimated, it will always balance, in prosperity and in depression, in normal times and in abnormal times; and so will the Balance of Payments for every other country, even the one we most dislike.

The Balance of Payments and the Exchange Rate

Yet there is much to be learned by seeing just how this balance is maintained. For while an arithmetician can make these figures equal, we shall see that very important economic forces may be set in motion to maintain this equality. An appreciation of the nature of these forces can be gained by analyzing what happens when for any reason the balance is disturbed.

First, however, we must determine the effects of a change in the price of American money in terms of foreign money, or, in other words, of a change in the exchange rate, upon each item in the Balance of Payments. For such changes are normally the means by which the balance is maintained.[1] If the price of the dollar measured in foreign money rises, it takes more pounds, francs, lire, and so on, to buy dollars. When this happens, Americans who want to buy foreign money can get more than before for the same number of dollars. That is to say, when the price of the dollar rises, the price of foreign money drops.

Let us first see how this change in the exchange rate affects foreigners. Since foreigners have to pay more in their own money than before to buy dollars, they must pay more for American goods and services. Hence they can now buy certain things which they formerly imported from the United States more cheaply at home or from other foreign suppliers, and consequently their purchases of goods and services from the United States will fall.[2] They will, however, probably not spend less[3] on American securities, for though these will cost more, in foreign currency, their yield will not alter. Foreigners also reduce their purchases of gold from the United States. And finally, foreign holders of American funds can be expected to release more of these funds to importers, since they earn a higher premium by so

[1] Except for changes in the national income of the countries concerned, a subject to be considered in Chapter 42.

[2] Although they certainly would buy fewer goods, they would spend less on them only if the elasticity of the foreign demand for American goods and services was greater than one. This qualification is also relevant to what follows.

[3] But see qualification below for certain speculative purchases.

doing. Obviously, then, for all these reasons, foreigners will demand fewer dollars when their price rises. It is unnecessary to demonstrate that when the price of the dollar in terms of foreign money falls, the number of dollars demanded by foreigners increases. The demand for dollars is like the demand for ordinary commodities; the higher the price, the less is the amount demanded, and vice versa. We can represent the foreign demand for dollars graphically as in Figure 67. This curve also describes the supply of foreign money.

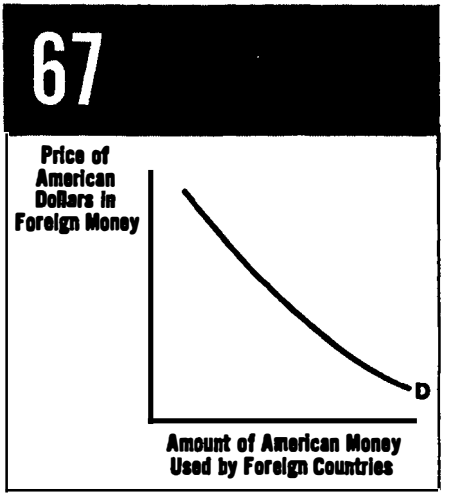

Figure 67. *Foreign Demand for Dollars*

Now, let us trace the American reaction to a fall in the dollar price of foreign money, or, to put it in other words, to a rise in the price of the dollar measured in foreign money. Since, when this happens, Americans can buy francs, pounds, lire, and so on more cheaply, they can purchase goods and services from foreign countries at reduced prices. They therefore spend more upon foreign goods.[1] For the same reason they are likely to buy more gold, though, as we have already noted, the lower price of foreign-owned securities will not necessarily lead them to buy more of these securities. Finally, they may want to add to their stocks of

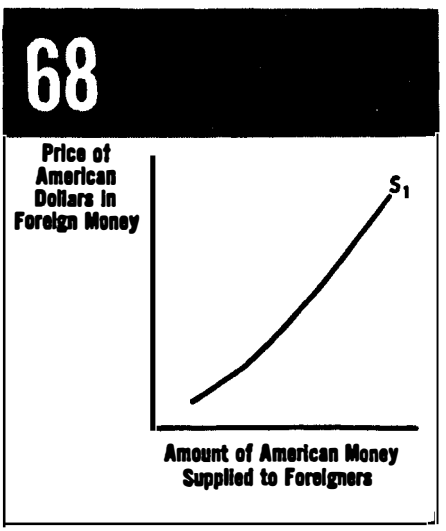

Figure 68. *Supply of Dollars to Foreigners*

[1] Provided that the elasticity of demand of foreign goods is not less than one; this provision is also applicable in connection with purchases of gold and securities.

foreign money against the time when its price will rise again. For all these reasons, we will put more dollars at the disposal of foreigners after the price of our money has risen and fewer after the price has been reduced. The higher the price, the greater is the amount supplied, and vice versa. We can represent the supply of dollars to foreigners graphically as in Figure 68. This curve also denotes the American demand for foreign money; but note that the lower the price of the dollar, the higher is the price of foreign money.

The price of American money in terms of foreign money can now be determined. Figure 69 will show that it must be P_1. If the price were any different, say P_2, we would make available N_2 American dollars, but foreigners would accept only N_3. Because the volume of dollars supplied at P_2 exceeds the volume demanded, the suppliers

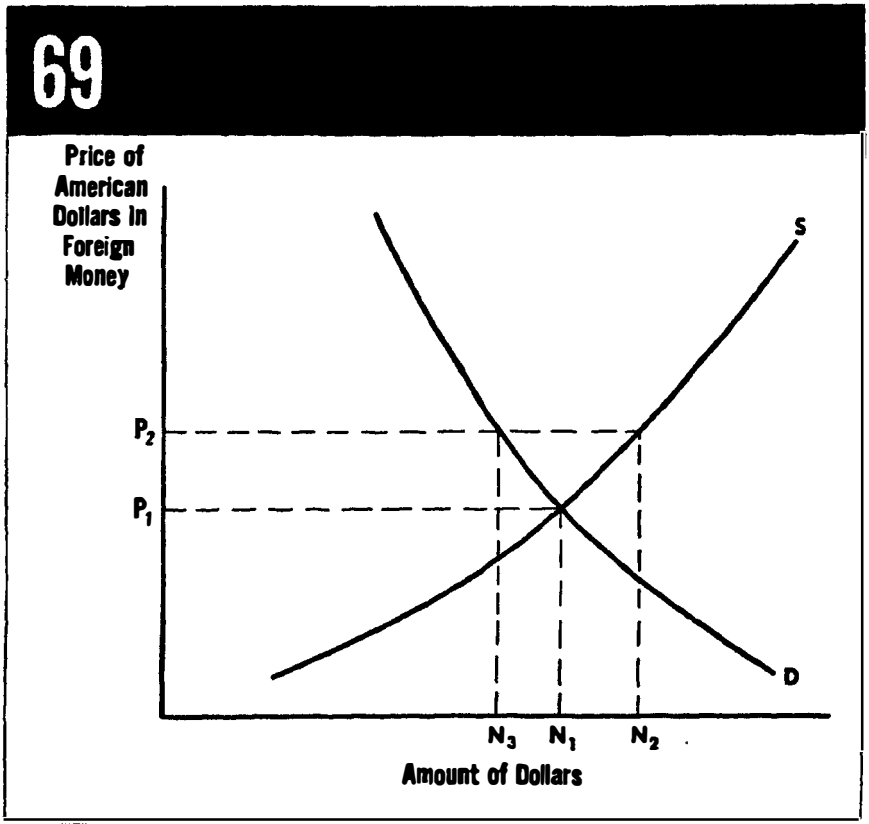

Figure 69. *Demand for and Supply of Dollars*

would be under pressure to lower the price in order to find buyers. Or, from the other point of view, at price P_2, foreigners are willing to make available only $\$N_3$ worth of foreign money, but Americans want $\$N_2$ worth. In attempting to acquire more foreign money, we bid up its price. As soon as the price departs at all from P_1, given the demand and supply situation illustrated, forces are set in motion to bring it back. The price of the dollar on the foreign exchange market would therefore be P_1.

Changes in the Exchange Rate

Now we may return to the problem set out above: what happens when the Balance of Payments is disturbed? Suppose that in the initial situation the Balance of Payments is as shown in Table 69.

TABLE 69 *

Balance of Payments, Before Increase in Demand

(in billions of dollars)

Source of American Money Supplied to Foreigners		Uses to Which Foreigners Put Their Supplies of American Money	
Our commodity imports	$4.4	Our commodity exports	$5.2
Our purchases of services from abroad	1.9	Our sales of services to foreigners	1.9
Our purchases of gold and silver	0.1		
Our net purchases of securities from foreigners	0.3		
Unexplained items	0.4		
Total	$7.1		$7.1

* This table represents the approximate situation in 1929.

Let us now suppose that something happens to disturb the situation, for instance, that the foreign demand for our goods rises and they try to increase their purchases by $2 billion. But foreigners can purchase more only if they have more American money to do it with. From what sources can they get the additional dollars they want? They may be willing to draw more from their accumulated stocks of American money. They may persuade Americans to add to their holdings of foreign money, giving up American money in return. Or they may be able to persuade Americans to purchase more foreign goods

and services, to lend more freely, or to buy more gold. These are the only things they can do, and only if they can increase the amount of American dollars available to them in one or more of these ways can they buy more American goods.

But what about the mechanism by which Americans can be "persuaded" to do any of these things? Very rarely will it be by direct negotiation. When foreigners want more of our petroleum, they do not usually send a representative to this country to urge us to buy more of their gold or securities or woolens. Sometimes this method is employed, but generally the urging is much less direct. Foreigners who want more American oil must first increase their requests for American money. What happens after that depends upon the foreign exchange policy of the governments concerned.

If there are no controls over the foreign exchange rate, and the foreign price of American money is allowed to vary without restraint, an increased demand for dollars will raise their price, for at the original price there will not be enough dollars available to meet the higher

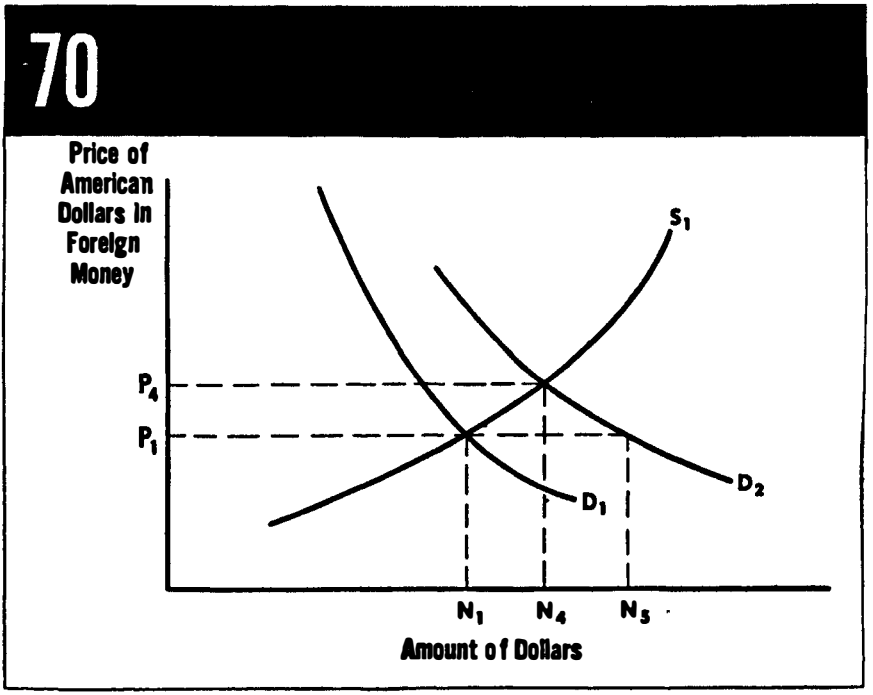

Figure 70. *Effect of Increased Demand for American Dollars*

demand. The only way in which those who want dollars can acquire more of them is by offering a higher price. Graphically, this may be pictured as in Figure 70. D_2 represents the increased demand of dollars. As foreigners attempt to raise their purchases of our products, they demand N_5 dollars. But we are willing to make available only N_1 dollars, or, in other words, we demand only N_1 dollars' worth of foreign money. Foreigners bid more for dollars in order to acquire them, and their price rises. As it rises we make more dollars available, and foreigners reduce their requirements. But the price would increase to P_4, at which level there would be no further pressure to raise it. The new Balance of Payments, to be compared with that shown in Table 69, might be that in Table 70.

TABLE 70

Balance of Payments, After Increase in Demand
(In billions of dollars)

Source of American Money Supplied to Foreigners		Uses to Which Foreigners Put Their Supplies of American Money	
Our commodity imports	$4.7	Our commodity exports	$6.5
Our purchases of services from abroad	2.1	Our sales of services to foreigners	1.8
Our purchases of gold and silver	0.2	Reductions in stock of dollars held by foreigners	−0.2
Our net purchases of securities from foreigners	0.5		
Our additions to stock of foreign money	0.2		
Unexplained items	0.4		
Total	$8.1	Total	$8.1

Thus, we may conclude that when foreigners increase their demand for our commodities, they buy more from us, though the increase in their purchases is smaller than it would have been if the price of the dollar had not gone up. We are persuaded to increase our purchases from them — of goods, services, gold, and securities.[1] It will be clear that in this situation, the change in the rate of exchange serves as the equilibrating mechanism; as the price of the dollar rises, we are encouraged to buy more from foreigners and they are discouraged

[1] We postpone until the next chapter the effects of the increased foreign desire to import upon the national incomes of the countries concerned, and through changes in income upon the items in the Balance of Payments.

from buying our products. Hence the apparent lack of balance in the Balance of Payments is cured.[1]

In the same way we may trace the effects of a loan made by one country to another upon the components of the Balance of Payments. If exchange rates are free, the granting of such a loan by the United States would provide dollars at a faster rate than formerly. Hence, a loan increases the foreign supply of dollars and, other things being equal, thereby causes a reduction in the price of the dollar in terms of foreign money. But a fall in the dollar stimulates American exports; in other words, if we should increase our lending to foreign countries, they would in these circumstances, probably buy more of our commodities. At the same time we should reduce our purchases from foreign countries. What happens when foreign countries attempt to repay their loans is a matter of obvious interest which the reader should be able to analyze for himself.

In the cases already discussed, the change in the exchange rate brings about a new equilibrium. But the rise in the price of the dollar may initiate perverse movements which upset the old equilibrium even further instead of leading to a new one. Equilibrium is normally restored because, as we have seen, a rise in the price of the dollar leads to an increased flow of American money and a decreased demand for it. But if a rise in the price of the dollar leads the market to anticipate a still further increase, then, instead of reducing their purchases of American securities, foreigners would be tempted to buy more of them. And Americans, instead of buying more foreign securities, would probably buy fewer of them. This would happen because the expectation of a further rise in the dollar would promise speculative gains to those who transferred foreign money into dollars or into American assets for later exchange on better terms for foreign money which they expect to become cheaper. Obviously such speculation in securities would tend to drive the price of the dollar even higher, and in that way would add further support to the speculative pressure. Thus, if a change in the exchange rate gives rise to the

[1] This should not imply that the totals could actually be unequal. We have already seen that the totals are necessarily the same. But if, immediately after the increase in demand, the price of the dollar had not risen by enough to maintain the final equality, either there would not be dollars available abroad to pay for all the American products that were wanted, or Americans might find themselves with foreign money which they had hoped to dispose of. Thus, the balance would be maintained, but the situation would obviously not be stable. In either case, pressure to raise the price of the dollar would continue until all those who wanted dollars (or foreign money) at the current exchange rate were satisfied.

belief that further changes in the same direction are to be expected, there will be changes in imports and exports — and more especially in lending — which will prolong and perhaps accelerate the movement in the exchange rate. Thus, the normal tendency toward equilibrium in the Balance of Payments may be upset by heavy speculation on the foreign exchange market, and the adjustments may not tend toward equilibrium but rather away from it.

The Gold Standard

For a number of years before 1914, stable or fixed exchange rates were looked upon as good manners in international economic dealings. There were fluctuations, of course, but they were thought to be evidences of weakness. Stable exchange rates were believed to have enormous advantages for all. The importer or exporter was able to make relatively precise calculations, and the risks of dealings with foreign countries were reduced. Stable exchange rates were also attractive to lenders, for the risks of default or of receiving less than the stipulated interest rate were thought to be reduced. To the generation brought up before 1914, it was unthinkable that the German mark should not be worth twenty cents, just as to our generation it was unthinkable that a package of gum should cost more than a nickel.

The mechanism by which exchange rates were kept stable during these years was the gold standard. Although it was fully operative for only a short period, as measured against the economic history of the United States or England, it was so widely accepted as an institution that long after most countries had left it, statesmen tried to bolster the wavering confidence of their people by proclaiming that their countries adhered to it. The gold standard took on an air of sanctity, and it became very difficult to analyze its workings objectively.

Let us set out in general terms the results which the gold standard was supposed to accomplish. The exchange rate was stabilized, and in place of variations in the price of foreign money which would otherwise come about spontaneously through changes in demand, there were movements of gold. These gold movements were supposed to set in motion forces that restored equilibrium. The mechanism functioned not quite automatically, but certainly with a minimum of human intervention — a feature which undoubtedly commended it to men who distrusted the wisdom or intentions of their fellows.

Membership in the Gold Standard Club required the acceptance of two rules of conduct: each member country had to agree to buy and sell gold at a fixed price, expressed in its own currency; and each had to agree that there should be no tariffs or export duties to hamper the movement of gold into or out of a country. Thus, in the years from 1927 to 1931, the most recent years in which the United States and other countries have followed gold-standard rules, we paid $20.67 for an ounce of gold. Whoever wanted gold could buy it at that price, while whoever had gold to sell was able to get that price for it. We did nothing to restrict the movement of gold into or out of this country. Banks in England in the same period bought and sold gold at about £4–4–11 a fine ounce, and the British government placed no obstacles in the way of gold movements until 1931. Incidentally, the gold-standard mechanism was not formally "organized," and no treaties or international agreement bound the gold-standard countries together.

Now let us see how exchange-rate stability was achieved with the gold standard. When the price of gold is fixed in both dollars and British pounds, a profit can be earned by shipping gold from one country to the other when the exchange rate departs far enough from "parity." If the cost of gold in the United States is $20.67 and in Britain is £4–4–11, and if $4 buys £1, then a large profit can be earned by acquiring gold in Britain, shipping it to the United States, selling it here, and turning the dollars back into pounds. Suppose, for instance, that a dealer buys one hundred ounces of gold in Britain for £424–11–8 (= 100 × £4–4–11). If there are no costs involved in shipping it to the United States, he can sell it here for $2067, with which he can buy £516–15–0 at the assumed exchange rate. Hence he can make a profit of £92–3–4 or $368.67 from the transaction. Obviously gold would be exported from England to the United States at this exchange rate. But if the rate of exchange were £1 = $6 instead of £1 = $4, it would be profitable to ship gold from the United States to Britain. Indeed, at any exchange rate that differed by more than a small amount from £1 = $4.86, it would be profitable to ship gold from one country to the other, and any such exchange rate would therefore lead to gold movements.[1] Thus, if the pound, formerly worth $4.86, rises to $4.90, it would pay to ship gold

[1] The difference need be only large enough to compensate for the costs of shipping gold and to provide a slight margin for profit.

from the United States to Britain, to sell it for pounds, and then to exchange them for dollars.

But the flow of gold is only one part of the process. Equally important is the transfer of pounds back into dollars. Those who sold their gold for pounds, and now wish to use them to acquire dollars again, set up an additional demand for dollars. The gold flow consequently brings about a rise in the price of the dollar, or in other words a reduction in the price of the pound, and the exchange rate moves in the direction of parity (£1 = $4.86). Thus, any significant departure from parity makes it profitable to deal in gold, and these transactions drive the rate back again toward parity. Only when the price of the pound is again quite close to $4.86 does it become unprofitable to buy and sell gold.

This result can be shown diagrammatically. If the price of gold is fixed both in dollars and in foreign money, and if there are no limitations on the movement of gold, the foreign demand for and the supply of dollars become perfectly elastic at certain prices. For instance, when the price of the dollar in terms of pounds falls below £1 = $4.86 or $1 = 4*s.* 1½*d.* (for example, £1 = $4.90), gold exports from the United States become profitable; the amount of American dollars demanded thus responds markedly to the price reduction, and the curve that represents the foreign demand for American dollars can be represented as in Figure 71. Likewise, if the price of the

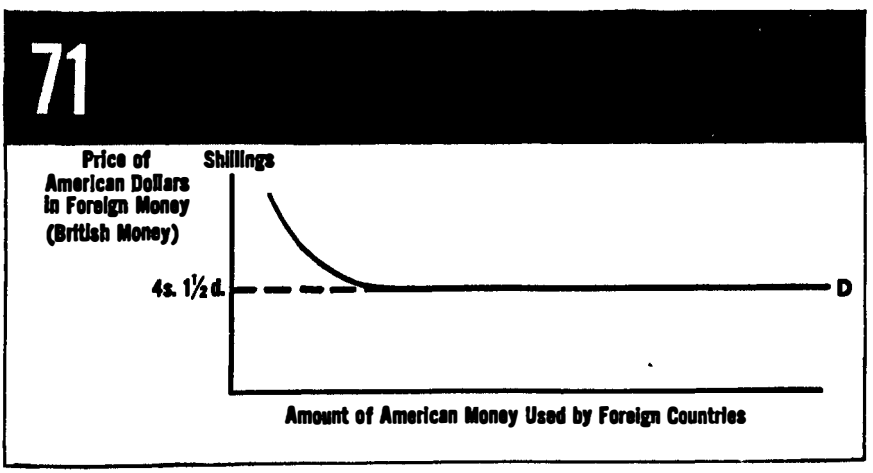

Figure 71. *Foreign Demand for Dollars: Gold Standard*

dollar should rise above £1 = \$4.86 or \$1 = 4*s.* 1½*d.* (for example, to £1 = \$4.80), gold would be imported into the United States and the amount of United States dollars supplied would respond markedly to the price increase. The supply curve thus becomes perfectly elastic for prices of the dollar higher than \$1 = 4*s.* 1½*d.* The supply curve could be represented as in Figure 72. Obviously, the demand and supply curves intersect at \$1 = 4*s.* 1½*d.* or £1 = \$4.86, the equilibrium price or dollar-pound parity. As we have seen, as soon as the price departs from this value, gold movements occur, and the price is restored again.

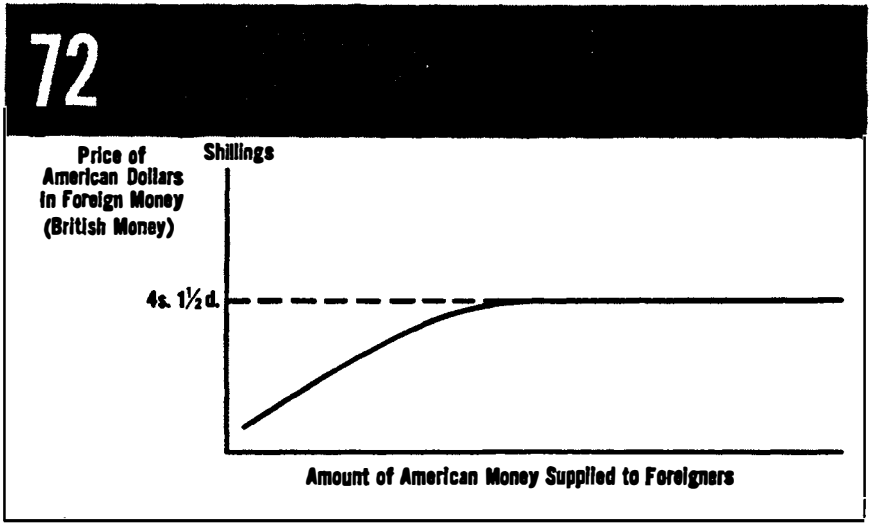

Figure 72. *Supply of Dollars to Foreigners: Gold Standard*

The flow of gold which restored the exchange rate to parity was also interpreted by the monetary authorities of the country losing gold to mean that forces were operating to drive down the price of its currency on the foreign exchange market. If, for instance, the United States were losing gold, it would signify that the price of the dollar in terms of foreign currency would have dropped had the gold standard not been in operation. While the loss of gold would have prevented the exchange rate from falling, it would have done nothing to cure the fundamental difficulty. On receiving the signal, however, the monetary authorities were supposed to take steps to increase the

rate of interest, or, if the country was gaining gold, to lower the rate of interest. These changes in the interest rates were expected to remove the cause of the difficulty.

Now let us see what adjustments we may expect when interest rates are changed in these ways. The country losing gold raised the interest rate. This could be expected to reduce private investment and thus to lower the national income and the price level. With lower prices, imports would fall and exports would be expected to increase.[1] Furthermore, because of the rise in interest rates, foreigners would lend more freely than formerly to the gold-losing country. Lower commodity imports, increased exports, and increased sales of securities to foreigners would all operate to drive up the price of its currency measured in terms of foreign currencies and to bring the gold outflow to an end. Likewise, we may conclude that when the country receiving gold lowered its rate of interest, the opposite effects were to be expected; higher income and higher prices leading to higher imports, falling exports, and probably a decreased willingness on the part of foreigners to buy its securities.[2] Thus, the rate of exchange would alter, and the money of the country which had been gaining gold would become less expensive on the foreign exchange market, and vice versa. Ideally, then, the operation of the gold standard was designed to bring about not only short-period stability in the foreign exchange market, but also to secure long-period stability through its effects on prices in the countries concerned.

But for some countries this stability of foreign exchange rates was an expensive luxury. A country which had been losing gold steadily had to raise the rate of interest to a level high enough to check the gold outflow, which meant that depression had to become sufficiently severe and prices had to fall far enough to balance the Balance of Payments even without important movements of gold. A gold-losing country, therefore, might have to accept depression in order to maintain the price of its money in terms of foreign money, and if the forces making for instability were strong enough, the depression might have to be very deep indeed.[3] Naturally, countries faced with the alternatives of depression or instability of the foreign exchange rate were

[1] We shall postpone consideration of the effect of income changes until the next chapter.

[2] Speculative considerations might lead to increased foreign lending, as already noted.

[3] The severity of depression might be mitigated in so far as the gold-receiving countries lowered interest rates and experienced rising prices, and thus had higher imports and lower exports.

tempted to choose the latter. Thus, in 1931 Great Britain departed from the gold standard because it found that to check the outflow of gold would require the adoption of measures which would simply make a bad depression worse. Moreover, as we shall see in the next chapter, measures to check a gold outflow, even if they intensify depression in the country concerned, may not prove effective. Stability of the exchange rate through the operation of the gold standard is not necessarily attainable; and even when it is, it may be very costly.

The International Monetary Fund

The advantages of exchange-rate stability are real, but countries are no longer willing to press for them if to do so means depression. Under the terms of the 1944 Bretton Woods Agreement — the blueprint for international financial arrangements for the post-war world — it is possible for a country to enjoy stability on the foreign exchange market without being forced into depression to secure it. The flow of gold and of all kinds of foreign currencies can prevent undesirable movements in foreign exchange rates so long as the capacity of the monetary fund is adequate. Variations in exchange rates that may be required by the underlying situation can be made in an orderly fashion. Whereas a country operating on the gold standard had to subject itself to deflation in order to end a gold outflow, it can now, using the facilities of the International Monetary Fund, achieve exchange equilibrium at practically no sacrifice.

The basic principle of the Monetary Fund is simple. Each member country must provide to the Fund an assigned sum made up of its own currency and gold. The Fund starts life with about $8.5 billion of gold and money of all kinds. When a member country finds that it cannot acquire foreign money on the ordinary foreign exchange market, it may get it from the Fund, paying for it in its own money, though this right is subject to definite limitations. Thus, if a country wants more foreign currency at existing exchange rates than foreign countries are willing to supply, it can call on the Fund for the difference. It is not forced to ration an insufficient stock of foreign currency nor to adopt deflationary measures nor, finally, to mark down the price of its money in terms of foreign money. Thus, the Fund offers facilities to countries that are temporarily short of foreign exchange, and by so doing removes from them the incentive to adopt either deflationary or other measures which would injure

other countries. The help is offered on a loan basis, and the rate of interest rises with the size of the loan and the length of time for which it is made.

If a country tends to be chronically short of foreign exchange, it is encouraged to adopt remedial measures under the sponsorship of the Fund. When its currency is "overvalued," its exports are low and its imports are high. The officers of the Fund, on consideration of this problem, may urge it to reduce the price of its currency. Exchange depreciation is not of course a new weapon. In fact, it was used by many countries, including the United States, during the nineteen-thirties. But in the past it has always been used competitively, and once one country reduced the price of its currency, others were often provoked to follow suit. Members of the Fund, however, agree not to depreciate their currency except with the consent of the other members. Thus, exchange depreciation, which may be necessary for a country whose money is overvalued, can be accomplished without inviting retaliation. In this way the Fund not only provides temporary assistance in tiding a country over a period when it cannot acquire an adequate supply of foreign exchange; it also sponsors measures to remedy more fundamental difficulties. And by holding member countries to their agreement not to engage in *competitive* exchange depreciation, it introduces a measure of disarmament into the field of international economic relations.

Countries whose currencies are chronically scarce on foreign markets are, of course, as much out of equilibrium as those that generally have difficulty in finding enough foreign exchange. "Surplus countries" are as unbalanced as "deficit countries," and are just as much in need of corrective policies, if not for their own sake, at any rate for the sake of the rest of the world. The Fund also undertakes to advise "surplus countries" on the steps they should take to restore their equilibrium position in the world economy. Such countries may, for instance, be urged to raise the price of their currency or to reduce their tariffs. For certainly not until they have done something to make their currency more readily available to other countries can there be international economic equilibrium.

The International Monetary Fund was set up to provide a measure of exchange stability, temporary assistance to countries which are short of foreign exchange, and international sponsorship of means to remedy more fundamental causes of disequilibrium. Member coun-

tries agree not to take independent action that may injure other countries without at least consulting them first through the officers of the Fund. In return for this limited sacrifice of independence, each country is assured of some stability on the foreign exchange market, and it is protected from harmful measures that other countries might take. These benefits permit a country to undertake policies designed to maintain prosperity at home without having to fear that its foreign economic position will deteriorate so greatly that it will be forced to give up its internal aims. The gold standard provided exchange stability, but only by exposing every country to the possibility of severe depression. The arrangements drawn up at Bretton Woods also promise some exchange stability, but not at the price of depression. They furthermore set up machinery for the orderly correction of disequilibrating factors.

Summary

When we import, we need foreign money; when we export, foreign buyers need American money. We provide American money when we buy from foreign countries; we are given foreign money when we sell to them. In any period of time, each country either uses or holds for later use as much foreign money as is provided to it. Hence, the Balance of Payments for each country which summarizes the source and amount of foreign money supplied to a country, and the uses to which the money is put, necessarily balances. When exchange rates are unrestricted, any change in the desire for the goods, securities, or services of one country brings about a change in the price of its money measured in the money of other countries. When this happens, adjustments are made in some of the other components of the Balance of Payments. The gold standard provides exchange stability, but often at great cost, for a deficit country may have to adopt deflationary measures and suffer severe depression to maintain the free convertibility of its money into gold. Because the old-fashioned gold standard made external stability its prime objective, internal well-being was often sacrificed. On the other hand, the International Monetary Fund, set up in accordance with the Bretton Woods Agreement of 1944, makes possible stability of foreign exchange rates without forcing depression on member countries. And it provides for member countries other means of achieving stability, both in the short run and the long run.

42

International Trade and the National Income

BUSINESS FIRMS are rarely in the happy position of being able to sell all they can produce. Their markets are usually inadequate, and consequently they generally operate below capacity. Every businessman knows that he can earn higher profits by expanding his market. And everybody knows that there will be more jobs and a larger output when demand grows in size. The whole economy has an interest, then, in finding adequate markets. What could be more natural, in the light of these considerations, than that we typically attempt to reduce foreign competition in the home market and to expand our sales abroad? For when we do away with foreign competition, we presumably increase the size of the domestic market for our own products, and obviously when we sell more to foreign countries we expand our markets too. To discourage imports and encourage exports thus seems to be a logical and direct way of achieving prosperity — higher profits and more jobs. Public support for tariffs and for measures to stimulate exports is easy to understand and to sympathize with. But how much can we actually expect to gain from these policies? Can higher tariffs cure depression? What are the limits on increasing our exports? And how does international trade affect the level of income and employment? These problems are the subject of the present chapter.

Exports, the Desire to Import, and Income

In our analysis of employment in Part Four, we saw that the national income depends upon investment and the propensity to consume, and that the higher these determinants, the higher is the na-

tional income. We also saw that sales to foreigners are a component of investment and that an increase in exports constitutes an increase in investment. Thus, a $5 billion expansion in exports would be expected to lead to a $10 to $15 billion increase in the national income, just as would a $5 billion increase in private or public investment, unless offset by a reduction in the propensity to consume. We may, therefore, conclude that we can raise the national income by increasing our exports, provided that, at the same time, other investment or the propensity to consume does not fall. Cutting imports seems, at first glance, equally effective. Reducing foreign competition in domestic markets exerts its chief effect by raising the propensity to consume, since buyers will purchase more domestic goods when imports become more difficult to get. Such an increase in purchases of home-produced commodities out of a given level of income will, of course, also raise income, provided that the amount of investment does not decline. It appears, then, that by reducing imports [1] and raising exports, we can increase both the propensity to consume and investment, and thereby bring about increased employment and a rise in the national income. Let us examine this matter carefully.

Recent International Trade Policies

It is not clear that the commercial policies followed by most countries between 1930 and 1940 were adopted solely with the aim of raising investment and the propensity to consume, though it is certain that this effect was sought too. As the depression deepened in the United States and Europe, one country after another raised tariffs, established quotas against imports, and reduced the price of its money in foreign money. The tariff was raised to the highest level in our history in 1930, and in 1933 we lowered the price of the dollar in terms of francs, lire, pounds, and so on. In 1932 Great Britain abandoned free trade and imposed a general tariff against imports for the first time in almost a century, and a year earlier she had let the price of the pound fall, thereby making it more expensive for her importers to buy foreign goods, and cheaper for foreigners to buy her products. Other countries raised tariffs, adopted exchange controls which meant rationing foreign exchange to importers, and established import quotas, all with a view to limiting imports. At the same time they made strenuous efforts to expand exports by means of exchange depreciа-

[1] More accurately, the desire to import.

tion, through trade agreements, and in other ways. Each country eagerly tried to persuade foreigners to take its goods, and each took articles from foreigners with the utmost reluctance. Evidently it was thought more blessed to give than to receive. Ironically enough, this all happened at a time when a large part of the population of each country was unable to get adequate food, clothing, and shelter.

The policies followed in these years remind us very strongly of those pursued during the period of mercantilism.[1] According to mercantilist doctrine, exports were regarded as the source of a country's strength, and imports as a cause of weakness. Exports were stimulated by bounties, by prohibiting the development of certain industries in the colonies, and by building up colonial markets for the products of the mother country. Imports were restricted by tariffs or by regulations such as those that limited coastwise shipping to the vessels of the country concerned. But mercantilism, so it is supposed, had by the nineteen-thirties long been dead and discredited. It is generally believed that under the assault of Adam Smith, Ricardo, John Stuart Mill, and others, its basic principles had been shown to be false. The belief that a nation could enrich itself by refusing to buy from foreigners and by pressing its products on them had long ago been proved mistaken — or so at least we thought. At any rate, economists and politicians for long decades had given lip service to the belief in free trade, but in the nineteen-thirties all this changed. Mercantilist policies were not only practiced — that was not novel, for after all they had never been completely abandoned — but they once again became respectable. And there is, so it seems, good reason for adopting such policies; for to keep imports low and to raise exports seems on the face of it a sensible way to achieve prosperity. The new mercantilism was accepted, and it appeared to be a reasonable policy. Yet, even though this policy should prove successful, there may be *better* ways of securing prosperity. Thus, we shall have to examine in some detail the implications of the mercantilist attack on depression.

Tariffs and the National Income — General

Let us begin by tracing the effects of an increase in our tariff.[2]

[1] It must not, of course, be supposed that they were first put into effect during the nineteen-thirties. Countries have always been quick to protect themselves from foreign competition and to stimulate exports, and indeed they still are. But their efforts to reduce imports and raise exports became unusually frenzied in that decade.

[2] Or any other change that reduces our "propensity to import."

Since such an increase persuades us to buy fewer foreign goods, it induces us to increase our purchases of home-produced consumers' goods. This of course means an increase in the propensity to consume, and hence higher employment and income, provided that there is no reduction in investment. Are there, then, any grounds for expecting a change in investment? There is no obvious reason why either public or private investment should decline, but foreign investment may fall. We have already seen that exports may decline if imports do, since the two are not independent of each other. Obviously, then, the nature of their interdependence is critical. If exports fall by as much as imports, income will not rise, but if they fall by less, the national income will increase. Hence an increase in the tariff will raise income if the reduction in imports that it causes is greater than the resulting fall in exports. We must then examine the relation between exports and imports in order to work out the effect of a tariff.

The Relation Between Imports and Exports

Let us review the factors on which exports and imports depend. As we have seen, our exports depend on the attractiveness of our products to foreigners, on the prices at which they can be bought, on the availability of our money in foreign markets, and finally on income in foreign countries. If foreigners can buy attractive imported commodities cheaply, if they are able to purchase the foreign exchange — that is to say, the American dollars — needed for these purchases, and if their income will support a high level of spending, they will buy our products freely and our exports will be high. Thus, our exports depend upon the domestic price of our products, the price of the American dollar in foreign currencies, and the level of foreign income. Exactly the same factors determine the volume of our imports. We import freely when we are able to buy foreign goods cheaply, when their price in our currency is low, and when our income is high. When our income rises, our imports increase substantially, and when it falls, they decline.

This dependence of a country's imports on its income can be illustrated with one or two examples. In a study of the relation between British imports and the British national income,[1] it was shown that between 1924 and 1933 each change of £1 billion in British national

[1] Tse-Chun Chang, "The British Demand for Imports in the Inter-War Period" in *The Economic Journal*, June, 1946.

income was accompanied, on the average, by a change of about £200 million in its retained imports. If, therefore, British incomes should fall by £500 million their imports of foreign goods would decline by about £100 million. Such reduced purchases on the part of the British imply lower exports, not only by the United States, but by other countries as well, since we are not the only source of British imports. Nonetheless our exports would decline to some extent.

The relation between our national income and our imports is also close. We tend to import most freely when our income is highest; when it is very low, our imports are also low. Making allowance for price changes, our national income fell by almost 40 per cent between 1929 and 1933, and in this period the volume of our imports declined by the same percentage. Between 1932 and 1937, our income rose by about 70 per cent, and our imports increased by about 55 per cent. Imports, it may be seen, respond very freely to changes in the national income.

Briefly, then, imports and exports depend on the level of income and prices in both the home country and in foreign countries, and on the rate of exchange. Hence a change in imports can affect the volume of exports only through its effect on incomes in the countries concerned, on their prices expressed in their own currency, or on the rate of exchange. In other words, if there is any connection between our imports and our exports it is established either because foreign incomes depend on our imports or because the rate of exchange or prices, domestic and foreign, vary with our imports.

Now let us return to the question posed above: How will a reduction in our imports affect our exports? We showed in the preceding chapter how a reduction in imports affects the rate of exchange. With a reduced desire to buy from other countries, we require less foreign currency than before. Hence, unless exchange-rate stability is provided by the gold standard or some other form of exchange control, the price of foreign currency expressed in dollars will fall. This means, of course, that dollars become more expensive to the foreigner, and he therefore has to pay more for American products. If his income does not change, he consequently buys less of them and our exports fall. Thus, if our desire to import is reduced, our exports shrink, partly because our dollars, and therefore our products, become more expensive on foreign markets — unless, of course, the exchange rate is stabilized.

There is a second reason why our exports might fall if we bought less from foreigners. A reduction in our imports is equivalent, from the point of view of foreign countries, to a fall in their exports and in the demand for their products. With reduced sales, their income declines, for, as we have seen, a reduction in exports is in this respect like a reduction in any other component of investment. The multiplier effect is set in motion by a decline in exports, and income may fall by two or three times as much as the initiating reduction. When income drops in foreign countries, they buy less from us, even if our prices remain the same. Hence, we may conclude that a decline in our willingness to import affects the level of our exports in two ways: the dollar becomes more expensive, unless exchange rates are kept stable; and the fall in the income of foreign countries leads to a reduction in their imports from us. Hence, a reduction in our willingness to import destroys a part of our export market.

The importance of changes in the income of foreign countries as an equilibrating force cannot be exaggerated. In the previous chapter we showed how reduced purchases abroad lead either to a shift in the exchange rate or, if the exchange rate is controlled, to the shipment of gold or the sale of securities by foreigners. Once disturbed, the Balance of Payments is maintained by a shift in the exchange rate or, if there are exchange controls, by changes in the volume of foreign lending or shipments of gold between countries. But that was by no means the complete story. In fact, it was perhaps not even the more important part of the story. We now see that if the Balance of Payments should be disturbed, it may also lead to changes in the incomes of the countries concerned, and that these changes may be much more important in restoring the balance than the operations of the forces described earlier. Hence, if we raise our tariff to cut imports, our exports may be adversely affected, not only by the resulting change in the exchange rate, if it is permitted, but also through changes in the incomes of foreign countries.

The reader should see that these two forces are related. Whether the increase in the tariff causes a change in the exchange rate or not, our exports would decline. If the exchange rate is not altered when we raise the tariff, the greater part of the adjustment will come about because of a reduction in foreign incomes. Thus, when we reduce imports, foreigners are obliged to send more gold to us if they want to remain on the gold standard. The value of their currency is thereby

maintained, and consequently they cannot avoid the fall in their commodity exports.[1] With reduced exports their incomes fall and consequently they buy less from us. If, however, they do not care to maintain the price of their currency, their incomes would fall by less, and the burden of readjusting the balance would be borne to a greater extent by shifts in the exchange rate.[2] Thus, when we reduce imports, the price of foreign currency measured in United States dollars declines, if exchange rates are uncontrolled. Such a fall affects the components of the trade of foreign countries in two ways: for one, with foreign currency now cheaper, we should buy more of their products than otherwise; and secondly, because the dollar now becomes more expensive for them, they would reduce their purchases of our products. Their exports fall by less than with stable exchange rates, and their imports decline, too, but also by less. Since this constitutes a reduction in investment, and an increase in the propensity to consume, their incomes may not be reduced at all. But notice that if this happens their exports and their imports both decline. They export less because initially we imposed a tariff the effects of which are only partly offset by the fall in the price of their currency. And they import less because the price of dollars in their currency has risen. Hence, our exports would fall, whether exchange rates were controlled or not, as a result of the reduction in our imports. But with unstable exchange rates, our exports would not fall so far, essentially because our imports do not fall so far. For in spite of our intention to reduce them, our imports are partly maintained because foreign currency becomes cheaper for us to buy. Since the fall in our imports is smaller, foreign income would not fall so far, and hence our exports would decline by less, but they would nevertheless fall to some extent.

The Tariff and Income — Details

Raising tariffs increases the propensity to consume home-produced products, but, as we have now seen, it also causes a reduction in exports. Hence raising the tariff raises the propensity to consume and at the same time reduces foreign investment. Whether the national

[1] Gold is exported instead of commodities, but gold exports are not sales of newly created products — of products created in response to foreign demand. Exports of gold do not affect the economy as investment does.

[2] A shift in the exchange rate may not affect the volume of exports and imports significantly, and in that case the fall in income would be an important feature in the mechanics of adjustment.

income is favorably or adversely affected by the tariff increase therefore depends upon the increase in the propensity to consume relative to the decrease in investment. If our commodity imports [1] decline more than our exports, income and employment in this country will increase. Certainly this may occur. However, we must consider an additional factor in this connection. If our imports drop by more than our exports, the implication is that foreign countries experience a larger decline in their exports than in their imports; and this, unless it is countered by an increase in home investment, causes a reduction in their income and employment; for income declines when the reduction in exports exceeds the reduction in imports, and rises when the opposite is true. To avoid reductions in income and employment, foreign countries would be compelled to reduce their imports to keep pace with the decline in their exports. This means, in short, that we are able to increase our income only to the extent that we can force other countries to lower theirs, and naturally, no other country willingly accepts an imposed depression. Hence, other nations would be induced to adopt policies like our own in order to avoid depression.[2] As they raise their tariff and try to expand exports, any favorable effects we might otherwise have enjoyed are taken from us, for the increase in their tariffs reduces our exports. Hence the increased propensity to consume at home may be completely offset by the decline in foreign investment, if we provoke foreign countries to restrict their imports in order to check a decline in their income. Any gain in income which we secure in this way is secured at the expense of foreigners, and we may expect that they will try to prevent it. Since they have control over their own import policies, our gains are likely to be small and temporary.

It must be strongly emphasized that the gains in income which we secure through restricting imports are obtained almost entirely [3] at the expense of foreign countries. We can increase our income provided that our propensity to consume increases by enough to offset the decline in our exports. But if it does, it implies that foreign countries experience a larger decline in their exports than in their imports. Just as a larger reduction in imports than in exports means an increase in

[1] Including imports of services.

[2] They could also prevent income from falling by raising home investment enough to offset the reduction in foreign investment; but most countries have preferred supplying foreigners with goods to building parks, highways, and homes for themselves.

[3] The reason for the qualification will become apparent in the next paragraph.

income, so the reverse means a reduction. If we succeed in reducing imports by more than the fall in exports, the additional income we thus create in this country is taken from foreign countries. Raising tariffs against foreign goods is thus in the first place a remedy which is uncertain in its effectiveness, since its success depends upon our being able to reduce imports by more than exports; and second, it is a remedy which operates only by pushing other countries deeper into depression. World income is not raised [1] when tariffs are raised (or lowered) and a country only succeeds in lifting its own income by reducing that of another.

Changing the tariff may stimulate private investment within the country concerned, but the stimulus is not likely to last for long. When the tariff is raised against certain products which had been imported before, it may become profitable to produce the goods within the protected country. To do so, it will be necessary to build plant and install equipment; thus for a time private investment will be encouraged. And though other industries will be injured, especially those that depended upon foreign markets, it is improbable that they will reduce their investment by enough to offset the increase in investment in the expanding industries. But it should be noted that these are the likely results, not so much of an increase in the tariff, but of a *change* in the tariff. If the tariff had been reduced, the induced investment in the stimulated *export* industries would almost certainly exceed the reduction in investment expenditures on the part of industries that had lost tariff protection. Hence, we may conclude that a change in the tariff will normally lead to a short-period favorable response in private investment. Since this is not, however, an international trade effect, and since, in any case, it is short-lived, we shall not consider it further.

Tariffs and Employment

And yet, though raising tariffs does not affect total world income, it may raise total world *employment*. A tariff makes production less efficient. Therefore to produce a certain output requires the employment of a larger labor force after tariffs have been raised. Higher tariffs may thus cure unemployment in two ways. They create jobs for any country which can improve its position at the expense of its neighbors, and they reduce unemployment because they make labor less effi-

[1] Except very temporarily.

cient. Jobs created out of inefficiency are likely to mean that the real income of the employed worker falls, for with constant money wage rates and a fall in efficiency, prices rise. Hence, while raising tariffs is one way of increasing employment, it can hardly be regarded as a sensible method. Indeed, as a cure for unemployment it is less acceptable than reducing hours of work. For when hours are cut, at least some workers are able to enjoy the benefits of increased leisure. In contrast, the increased employment that comes about because labor efficiency is reduced means neither a shortening of the work day nor an alleviation of the disagreeable features of work. No one gains increased leisure. All that happens is that poverty is spread more widely. It is well to remember that, although tariffs can cure unemployment, they do so only by creating more unemployment in foreign countries or by reducing labor efficiency to such a degree that more employment is needed to produce any given output.[1]

Foreign Lending and the National Income

More positive methods of raising income and employment through foreign trade may be available to a country, as when it lends more freely to other countries. It is interesting to see how this affects income. Does it mean, for instance, that if our income rises, foreign incomes decline, as happens when the tariff is raised? Let us consider the effects of a loan to another country. We shall suppose that the loan is made so that the borrowing country can increase its purchases from us, the lender. The foreign country may need to borrow in order to acquire the foreign exchange with which to pay for more of our goods. If we lend to foreign countries, therefore, we enable them to increase their imports from us.[2] A good example of this is seen in the large loan to Great Britain which was authorized by the United States Congress in the summer of 1946. Under the terms of this loan, Great Britain received a credit on which it could draw up to $3.75 billion. There was, of course, no reason to suppose that the British would spend the whole of this sum in the United States. They might, for example, use some of the funds to purchase goods from France, and if so the French would get more American dollars than they otherwise would have had. And they, in turn, could use some of the money to

[1] There may be other good reasons for raising tariffs, quite apart from any usefulness they may have in creating employment. In any event, a country may be pressed to adopt higher tariffs to protect itself from the economic vandalism of others.

[2] An increase in lending might, of course, lead only to the export of gold.

buy goods in still other countries. Nevertheless there can be no doubt that eventually all of these dollars would return to the United States accompanying orders for American products. By granting this loan, we made dollars available to foreign countries which they could have got otherwise only by increasing their sales to us. But this means that they would not in fact get the dollars, for we are not likely to purchase more goods and gold from foreign countries. Generally, a loan supplies foreign countries with dollars which they can use to increase their purchases of goods from us. Increased spending by foreigners for our products leads to an increase in our income and employment unless we are already enjoying peak prosperity. Hence, when we make a loan, we export more, and investment rises. Consequently, our income rises too, and this causes an increase in our imports. Moreover, this increase in imports does not cause a decline in our consumption of home products; rather, it results from the expansion in our income. Indeed, with this rise in income, consumption of both domestic products and imports would increase. Imports are like saving in that they are a resultant; they do not bring about a reduction in income,[1] but they do mean that foreigners are provided with still more American currency, and thus can expand their imports from us even further. As a result, our exports increase further, investment goes up, and income and employment rise.

In summary, then, increased lending leads to increased exports, income, and employment. It is important for the reader to notice that unlike the situation discussed above, with foreign lending income does not fall because of the increase in imports, for imports rise *because* of the increase in our income. And while our increased imports do not reduce income in this country, they do raise income in foreign countries. Hence, instead of world income remaining unaffected, as it does when one country raises its tariff, it is, in this case, raised. This happens because the increase in lending finances an increase in the propensity to consume [2] in foreign countries. In fact, it is only because they wish to consume more that they find it necessary to borrow from us in the first place. In contrast, an increase in tariffs simply brings about a diversion of consumption, for then we consume more of our own products and less of everybody else's. But the circumstances

[1] Though, as we have implied, an increased desire to import might do so, just as would an increased desire to save.

[2] Or higher investment.

which lead countries to borrow more heavily from us are usually quite different. It normally means that they choose to consume more in total, and not that they want to consume our products instead of their own. And naturally when the world propensity to consume increases, investment remaining constant, employment and income for the whole world will rise. By lending, we simply secure our share of that increase, for we enable others to extend to us a share in the increased market. The loan itself is not the initiating factor; it simply makes possible the transmission in our direction of favorable developments initiated in foreign countries.

For what reasons would foreign countries want to increase their total purchases? In general, because of an increase in their income. We have already investigated the factors that could bring about such a rise in income. We have seen, for instance, that income increases when private firms spend more on plant and equipment or when government increases its investment. Whatever the reason for it, their total purchases both at home and abroad will rise because of the increase in income. To repeat, the loan itself is not the initiating factor; it merely permits us to secure our share of the favorable stimulus to income and employment which comes into being because of the policy followed in other countries.

The costs involved in such a loan are precisely the same as those in loans made to our own business firms or to the government. Our national or private debt rises and therefore the familiar difficulties of a high debt become slightly more serious. Otherwise the only cost is that which we face when the loan is repaid. For as lending money to foreign countries provides them with dollars which enable them to give us a share of their expanding markets, so requiring foreign countries to repay loans makes it impossible for them to buy as much of our products as they otherwise would. Since foreign countries when they repay debt are not allowed to print United States currency, they have fewer American dollars left with which to buy our products. Hence, it follows that our exports will be lower when the debt is being repaid than when we are willing to carry the loan, and with lower exports and no offsetting changes, income and employment decline. Thus, we reach the paradoxical conclusion that our income rises when we lend to foreigners; and that it is reduced when they repay their debt. And this seems to mean that employment and income are highest when we give goods away.

But is this really a paradox? Does not exactly the same situation exist in the domestic economy? When the government or a private firm borrows money and spends it on the products of American industries, our income and employment are high. When it reduces its purchases of investment goods and repays debt, employment and income fall. The situation in respect to domestic investment and internal debt is thus exactly analogous to that which rules in our relations with foreign countries.

There is one occasion, however, in which this result is inapplicable for both the domestic and the world economy. When we have full employment, then lending more either to foreigners, to American business firms, or to the government does not add to the total real income. All that happens at such a time is that an increased flow of dollars bids for a fixed supply of goods, so that unless deterred by government controls, business firms find it profitable to raise prices. Hence, to encourage borrowing at such a time is to encourage not an increase in real income — that is already at peak — but simply an increase in price levels. And at such a time we should look with favor upon repayment of debt, whether it is debt owed by American firms, by the government, or by foreign countries.

Gold Imports and the National Income

There is another way in which we can provide foreigners with American dollars and so secure the benefits of expansion abroad. If we buy gold from them, we increase the flow of dollars to foreign countries. Furthermore, in this case we do not even have to be concerned about the somewhat unfavorable effects of the increase in our national debt, since the gold we buy is paid for neither out of taxes nor from borrowed funds, but by a deposit created against the gold import.[1] The dollars thus become available to foreigners at no cost to us either in taxes or in increased debt. With these dollars foreign countries can increase their purchases of our goods. Hence, a rapid entry of gold into the United States has the same immediate effect upon income and employment as lending large amounts to foreign countries. With increased exports and no decline in the propensity to consume, our income and employment rise. But unlike the situation when we lend to foreigners, there is no increase in our national

[1] This has been discussed. See Chapter 25.

debt.[1] The difference in results, of course, rests upon the fact that we finance gold purchases quite differently from our purchases of foreign securities. If we were willing to hold foreign securities as a reserve against Federal Reserve Bank liabilities, the two processes would be exactly the same, but, of course, we are not. We do hold gold as a reserve. Our willingness to import gold, therefore, provides us with a painless method of creating foreign markets for our products. The reason we cannot take advantage of this to a much greater extent than we do is that the stock of gold in the hands of foreign countries is quite limited.[2]

Two other characteristics distinguish a gold import from that of ordinary commodities. When we buy woolens from England, competition is increased for the American manufacturer and spending upon our own product is reduced. But when we buy gold from England, American gold mines are completely unaffected, for they continue to get $35 an ounce for all the gold they mine, no matter how much we import from foreign countries. Thus, to import gold freely neither reduces the propensity to consume domestic products nor does it hurt the American gold producer. On this account gold imports have a second characteristic not shared by imports of other kinds of goods: there is no tariff barrier against gold. Foreigners can make gold available to us wherever they want, without any restrictions on our part. Thus, not because of any peculiar physical properties, but merely because of the social institutions we have set up for dealing with it, gold plays a unique role in international economic dealings, so that whenever foreign countries secure large stocks of it, our exports and therefore our incomes and employment are likely to be increased. Hence, if we can devise no better way of raising employment, we do well to buy gold from foreigners. Directly, of course, we acquire only gold for which we have little use. But we secure very important advantages indirectly as a brief consideration of the multiplier effect will show. It is clear that if we can provide more goods to foreigners we also find it profitable to produce more goods for ourselves.

Three unique and important characteristics of gold, derived from our social institutions rather than from its inherent usefulness, allow gold to affect our incomes. In the first place, we finance our pur-

[1] When the lending is done by private individuals or firms, private debt will increase if the funds to be lent are themselves borrowed.

[2] Gold imports also bring about an increase in the reserves of the banking system and thus tend to keep the rate of interest low. See Chapter 25, where this point is discussed.

chases of gold in an extremely unconventional manner; we neither tax nor borrow to secure the funds with which we pay domestic mines and foreigners for it. Second, importing gold does not affect our willingness to purchase the products of our own gold mines, and therefore increased imports of gold do not reduce the propensity to consume our own products. And third, we allow all the gold foreigners wish to send to enter the country at a fixed price of $35 an ounce. We can learn something about the nature of our economy by considering what would happen if we extended these rules to certain other commodities. Suppose, for instance, that the government agreed to purchase at a fixed price above the cost of production not only gold, which is relatively scarce, but also all the limburger cheese offered to it, and that it financed these purchases in the way it now finances its purchases of gold.[1] The government would thus stand ready to buy limburger cheese from American producers at a fixed price, no matter how much it might import from foreign countries. In the mind of this author there is no doubt that such a step, however foolish it may seem, would guarantee full employment in this country and indirectly in the world.[2] The policy is obviously absurd, but not because it would fail to give us full employment. It is absurd because it would encourage us to devote a large fraction of our total resources to producing a commodity of very limited usefulness, for after all, our willingness to consume limburger cheese is about equal to our willingness to consume gold. We do not want full employment with perhaps a fifth of our total resources engaged in mining gold and making limburger cheese. That would be too much of a boondoggle. We should naturally prefer to have everyone working at jobs where they make the kinds of commodities we most need — houses, clothing, automobiles, highways, and so on. But even employment at making the 100 millionth ton of limburger cheese is better than unemployment, for although the labor of those who are producing the cheese may be relatively wasted because our willingness to consume it is restricted, yet if some of our workers are employed (no matter how uselessly), our total output of butter and homes and motor cars and clothing and other useful products will be higher than if they are out of work. Thus, even the adoption of the limburger cheese standard, absurd as it is, would be much

[1] One way in which it could do this would be to permit the Federal Reserve Banks to hold limburger cheese as a part of their reserves against liabilities.

[2] It might be necessary to extend the scheme to one or two other commodities as well.

better than the policy the world followed between 1929 and 1939, for it would give us full employment and a relatively large output of useful commodities, whereas the policies of the nineteen-thirties gave us, at the worst point, perhaps twenty-five million unemployed throughout the world and a relatively small output of useful commodities.

The Spread of Prosperity

We have noted that countries are able to export prosperity and depression to each other. It is perhaps well to gather together our fragments of description of the mechanism by which this particular trade takes place. To illustrate the process, let us assume that income and employment increase in this country — for instance, because we find it profitable to exploit a new invention. As a result, we import more goods from foreign countries, and this, of course, constitutes an expansion in their investment, so that their income increases unless, for some reason quite unconnected with this process, other investment or their propensity to consume declines. In this way, prosperity in this country would be exported to foreign countries. A depression in this country would be exported in much the same way. If for some reason our domestic investment should fall off, our income would decline and hence our imports would diminish. Since this would constitute a decline in foreign exports and investment, foreign income would also fall and unemployment abroad would increase. Thus, both prosperity and depression in this country spread to foreign countries.

Although our imports are relatively small compared to our national income, they are much larger when set against the incomes of most other countries. Thus, in 1939 our imports from Great Britain amounted to 3 per cent of Great Britain's income and our imports from Canada came to 8 per cent of Canada's income.[1] It follows that other countries are profoundly affected by what happens in the United States. We, however, are much less strongly affected by what happens abroad, for our exports make up only a small part of our total investment, and therefore an even smaller part of our total income. Hence, whether we have prosperity or depression is a matter of very great importance to the rest of the world, though the well-being of other nations is less important to us — economically, at any rate.

[1] Our direct purchases from a country are not the only way in which our imports affect that country. It may also find a better market for its products in a third country from which we have expanded our purchases. For instance, when we buy more Canadian produce, our purchase may cause an increase in British sales to Canada.

This does not mean that a foreign country cannot insulate itself against the adverse effects of a depression in the United States. A depression here means a sharp reduction in the foreign investment of other countries, since we are an extremely important customer to most of them. But they could, if they chose, offset the fall in foreign investment by increasing the total of their domestic investment, both private and public. If they succeeded, they might, it is true, have to face a severe decline in the value of their currency in terms of dollars, unless they got help from the Monetary Fund. But there is no doubt that most foreign countries would today prefer a fall in the price of their currency on the foreign exchange market to a depression imported from abroad.[1]

Summary

A country seeking to raise income, profits, and employment is very naturally tempted to raise its tariff and to try to expand exports, but while these steps may succeed in their aim, they are of limited effectiveness. Exports are likely to be reduced if imports are cut, especially if other countries are led to restrict their imports too. Moreover, a part of the gain in income that one country achieves by raising its tariff is got at the expense of others, and the remainder by reducing labor efficiency. Yet in a world where each country sees economic salvation in cutting its imports and pushing its exports, any country which refrains from doing so is at a disadvantage — for its foreign investment is likely to be low. Such a country may be able to avoid depression, but to do so will require a high level of private and public investment.

When a country lends it is likely to raise income and employment, though it may have troubles when the loan is repaid. Loans to foreign countries do not depress their income; rather, they permit those countries to purchase more from the lender, with the result that its own income is raised. Purchases of gold by this country from foreign countries also bring about an increase in income and employment, for in addition to their effect in lowering interest rates, they give dollars to foreigners which can be spent upon our products. Unfortunately, the output of gold is relatively small. If it were much larger, we could

[1] In the nineteen-thirties, many countries seemed to pursue a policy which implied a preference for depression over the immorality of letting the price of their money on the foreign exchange market fall. But generally even these countries gave in after a time.

enjoy much higher incomes. But to rely on gold as a method of achieving prosperity is far from good sense, for we do not especially want larger and larger stocks of it. Nevertheless, prosperity supported by gold, or limburger cheese, or by building houses or factories is a great deal better than no prosperity at all.

Prosperity and depression tend to spread from one country to others. Because of our unique position in the world economy — since we produce almost half the world's output of goods and services — we are an especially important factor in determining whether foreign countries are prosperous or not. If we are prosperous, they are almost bound to be, in normal times, unless we achieve prosperity by measures that reduce foreign income. If we are not, though they can avoid depression, the task is much more difficult.

The most significant conclusion to be drawn from this analysis of the relation between international trade and the national income is that there are two basically different routes to prosperity. One of them is expansionist, and its effect is to raise income and employment, not only in the country that follows it, but in other countries as well. The other is a form of economic warfare in which the successful country gains, but only at the expense of its neighbors. Measures which increase the total amount of investment or the propensity to consume in all countries combined are of the first type. Measures which do not have this effect, but which simply improve the lot of one country at the expense of others, are of the second. Thus, when private investment is raised through, let us say, a reduction in the interest rate, world income and employment are increased. But when a tariff is imposed, nothing happens to total investment or the world propensity to consume, and so total income is not raised. Measures of the first kind increase the size of the economic pie which the whole world shares; beggar-my-neighbor remedies only mean a bigger slice for the victor and a smaller for the vanquished. And while a bigger slice of the same old pie is as good for the victor, he is likely to be much less secure, because those whose slice is cut down will generally fight for more.

SUGGESTIONS FOR FURTHER READING

Halm, George. *International Monetary Cooperation.* Chapel Hill: University of North Carolina Press, 1945.

A straightforward presentation of the essentials of international trade theory (in its monetary aspects) and of the way in which the International Monetary Fund and Bank fit into the picture.

——— *Monetary Theory.* Philadelphia: The Blakiston Company, 1946 (2nd edition).

See especially Chapters 11–15 inclusive.

Hansen, Alvin H. *America's Role in the World Economy.* New York: W. W. Norton & Company, Inc., 1945. (Also, New York: Penguin Books, Inc., 1946.)

A very clear account of the international economic problem and the measures taken to keep it under control.

United States Government. *Hearings before the Subcommittee on Foreign Trade and Shipping, of the Special Committee on Post-War Economic Policy.* Washington: Government Printing Office, 1945.

Read especially the testimony of Lauchlin Currie (p. 1769) and Stacy May (p. 1037).

United States Government: Federal Reserve Board. *Post War Economic Studies: No. 7* (1947). Washington: Government Printing Office.

PART SIX

Interest Groups in the Economy

Introduction

Some of the most serious issues that arise in the economy have to do with the distribution of the total output among different sections and classes within the economy. We are familiar with the problem of the relative shares of labor and the employer; a strike gives us a very clear picture of what happens when there is a dispute about it. We are also familiar with this problem in connection with the farmer. But obviously the struggle for a larger share of the economic pie is not limited to these several groups and classes. Doctors, grocers, professors, bakers, and men and women in nearly every other walk of life naturally try to improve their lot, and their struggle to do so inevitably throws them into economic conflict with others. In Part Six we shall consider some of these problems of income distribution. We shall also consider one of the weapons employed by special groups to gain an increased share of the national income.

43

Labor

IN HIS ECONOMIC LIFE the "average" American is a person of varied interests and mixed loyalties. He is a wage earner and an employer, a farmer and a city dweller, a small businessman and a wealthy capitalist, a man whose annual income is $500 and $500,000, a consumer and a producer, someone whose income depends upon foreign markets and someone whose income is subject to the stress of foreign competition, and many other things besides. To consider him only as a statistical average obscures some of his most important characteristics. As a wage earner he has certain interests that are at variance with his interests as an employer. His desires as a small businessman and as a leader of big business are often quite different; as a seller of wheat, cotton, and corn, his concern is in some measure in conflict with his interests as a producer of bread, clothing, and canned meat. In each of these special categories, he has peculiar and special interests. He is not simply an arithmetical average; there are elements of conflict in his soul.

Every person in the economy has certain interests which he shares with others. The farmer, although he competes with other farmers in selling his product, has a common interest with them in trying to get a good price. The employer, although he competes with other employers, sometimes makes common cause with them in resisting demands for higher wages. The doctor, although he is in competition with other doctors, leagues with them to look after the interests of his profession, and the employee, although he is in a sense a rival of other

employees for a limited number of jobs, joins with them in unions to fight for higher wages and better working conditions. It is a natural tendency for people whose economic interests are in some measure identical to band together to promote these interests. There are scores of organizations with this avowed [1] purpose. To name only a few, the American Medical Association, the American Bankers' Association, the National Association of Manufacturers, the National Wool Growers' Association, the American Bar Association, the American Bottlers of Carbonated Beverages, the National Wooden Box Association, the American Producers of Domestic Inedible Fats, the American Federation of Labor, the Congress of Industrial Organizations (CIO), and countless others — all are designed to foster the interests of their members and thus to promote the interests of the United States. Or so it is alleged, for naturally the members of such groups tend to identify their own interests with those of the nation. The number of interest groups is, of course, very large. In this chapter we shall deal with the problems of one of the most important of them, labor; and in the next with agriculture. We shall also devote some attention to monopoly, which in a sense is the objective of any interest group.

The Employee in the Economy

About 55,320,000 people went to work one day in May, 1946.[2] Only about 8,880,000 of them worked on farms, which meant that 46,440,000 were working in non-agricultural industries — manufacturing, trade, government, transportation, and so on. Of this group, approximately 9,005,000 were engaged as employers — proprietors and self-employed persons — or as domestic servants. The remaining 37,435,000 persons went to jobs as wage earners and salaried employees in non-agricultural industries. There were in addition 2,310,000 persons who were unemployed on that date, and hence our total employee force was 39,745,000. Wage and salaried workers, including the unemployed, thus accounted for about 82 per cent of all those who were engaged in non-agricultural occupations in May, 1946.[3] Since most of those who work are wage earners or salaried employees,

[1] Or unavowed purpose.

[2] In addition, there were 3,430,000 in the armed forces.

[3] The figure for this date is unusually low; it was about 88 per cent in 1944, 85 per cent in 1929, and 88 per cent in 1932.

and only a relatively small number are businessmen and employers, the labor problem, as it is called, concerns most Americans directly.

Of course, employees differ enormously in what they do and in what they earn. Most of them do not earn much, as we shall see. There are a few, however, such as the salaried managers of large corporations, whose earnings are entirely adequate. The rest, hired employees in the ordinary sense of the term, depend on a business firm or on the government for their jobs, and ultimately work under the direction of an employer. These are the factory workers, the clerical help in large offices, the salespeople in department stores, the postmen and policemen, and so on. These are the 80 to 90 per cent of the country's non-agricultural labor force.

While they differ widely in the work they do and the incomes they get, they have one common interest: they are all employees, and they all receive an income known as a wage or a salary which is properly a "work" income, not a "property" income. Yet they also differ widely. Some are white-collar workers — about 11,000,000 according to an estimate for 1940. The remainder are manual workers. Some are employed in industries where pay is low, others in industries where it is good.

Employees' Income

Generally speaking, employee incomes are not high compared to some other classes of incomes. One good measure of this is seen in the comparison between the annual income of wage-earning families and others, in Table 71. It is apparent that the families of wage earners do not receive as much as the families of their employers.

We may also estimate labor's annual earning, not on a family basis, but for the individual employee, whether working or not. In 1929, there were about 31.1 million employees who were not engaged in agriculture, and total non-agricultural wages and salaries came to $51 billion, or to $1640 per employee. In 1932, the average annual earning per employee, including those then unemployed, came to only $888; in 1939 it had risen to $1274, and in 1945 it stood at approximately $2500.[1] The enormous variability in these figures is striking: by 1945 the average was almost three times what it had been in 1932.

[1] The figure for 1945 is a very crude estimate. A rough allowance had to be made for government payments to the armed forces.

TABLE 71

Average (Mean) Income per Family * (1935–1936)

Occupational Group †	Income
Wage-earning	$1289
Farming	1259
Clerical	1901
Business	
Salaried	4212
Independent	2547
Professional	
Salaried	3087
Independent	6734

* Families that received any relief are excluded. Since they would almost all be in the top two classes in the table, all wage-earning and farming family incomes (relief as well as non-relief) would have been a good deal lower.

† *Consumers' Incomes in the United States*, 1935–1936: National Resources Committee.

Labor's total income also shows great variability. In 1929, labor received about $51.4 billion, and in 1932 only $30.9 billion. By 1939, the figure had climbed again to $47.4 billion, and by 1945 [1] it was about $95.5 billion. Even when allowance is made for the rise in prices between 1932 and 1945 (about 30 per cent in the cost of living), the increase is noteworthy.

Since labor accounts for more than four fifths of those who work, omitting agriculture, it is not surprising that it receives well over half the national income. In 1929, labor got about two dollars out of every three; in 1932, about four out of every five; and in 1939 and 1945, about the same share as in 1929.[2] In good years, labor receives about 65 to 70 per cent of the income; in bad years, up to 80 per cent. But a high percentage of a low income still does not provide a high standard of living.

Generally labor is not paid by the year, but by the week or hour. Hence data on hourly and weekly earnings are of special interest. In

[1] See previous footnote concerning the 1945 figure.

[2] The actual figures were 67.1 per cent, 81 per cent, 67.2 per cent, and 65.2 per cent in the four years. They are arrived at after subtracting agricultural income, and, in 1945, the income of those in the armed forces. But see the previous footnote concerning the 1945 figure.

Tables 72 and 73 are a few figures selected from the enormous number available. The general rise in wages between 1914 and 1929, the decline to 1932, then the rise to 1939 and the steep rise thereafter are notable.

TABLE 72

Average Weekly Earnings

Year	Manufacturing	Railroads	Retail Trade	Bituminous-Coal Mining
1914	$11.01	$11.40	—	$12.22
1929	26.40	28.26	—	27.56 *
1932	17.86	23.13	$20.71	13.91
1939	24.58	30.71	21.17	23.88
1944	46.08	46.89	27.09 †	50.95 †

* Bituminous, anthracite, and metalliferous. † September.

TABLE 73

Average Hourly Earnings

Year	All Industries *	Manufacturing	Railroads	Bituminous-Coal Mining
1914	—	23.3¢	25.2¢	34.4¢ *
1929	—	56.6	63.0	69.3 †
1932	49.0¢	45.8	59.1	58.9 ‡
1939	65.7	64.4	70.7	89.3
1944	92.0	102.2	96.6 §	118.7 §
Dec. 1945	94.7	99.7	95.7	128.1

* Manufacturing, mining, public utilities, trade, railroads, some services, and building; hence almost all industries.
† Bituminous, anthracite, and metalliferous. ‡ 1934. § December.

It is also worth observing that weekly and hourly wages vary rather widely between different industries. In January, 1945, average weekly and hourly earnings in selected industries were as shown in Table 74.

Not only between industries, but even within an industry, wages may vary widely. For instance, in 1937, the average annual earnings of employees in the knit-goods industries whose work extended over a year ranged from $1348 for skilled workers to $869 for unskilled work-

TABLE 74

Earnings In Selected Industries

Industry	Average Weekly Earnings	Average Hourly Earnings
Blast furnaces, steel mills and rolling mills	$55.04	119.1¢
Forgings, iron and steel	61.95	129.1
Machine tools	60.21	117.2
Automobiles	59.42	131.4
Sawmills and logging camps	32.43	77.3
Pottery	35.92	89.5
Cotton manufacture, except small wares	27.78	65.3
Hosiery	29.81	77.6
Work shirts	21.17	56.3
Cigarettes	35.07	78.2
Printing, book and job	45.10	104.9
Fertilizers	30.58	68.3
Rubber tubes and tires	64.29	132.0
Retail trade — General merchandise	22.31	63.2
Private building construction	52.98	136.4

ers. Earnings varied from 75 cents an hour for skilled workers in the knitting department to 42 cents for packers, 37.6 cents for brush-machine operators, and 23.7 cents an hour for female learners and apprentices. And in August, 1937, average weekly earnings in electrical manufacturing industries were $50.33 for skilled Boring-mill operators, $34.22 for insulation cutters, and $20.30 for rackers and unrackers.

Employees today work fewer hours in a week than they did in 1914, as is shown in Table 75, for weekly hours of work in manufacturing

TABLE 75

Hours Worked per Week

Year	Manufacturing	Bituminous-Coal Mining	Retail Trade
1914	49.4	35.2	—
1919	46.3	35.5	—
1929	45.7	40.2 *	—
1932	38.2	28.7 *	41.5 †
1939	37.7	27.1	43.3
April 1946	40.6	27.0	40.3

* Bituminous, anthracite, and metalliferous. † 1934.

and several non-manufacturing industries. Thus, in both manufacturing and bituminous-coal mining, hours were reduced by about 20 per cent between 1914 and 1946. Since the number of people in the labor force increased by about 50 per cent in the same period, there was a net increase in working capacity of only 20 per cent. Yet in spite of the relatively small size of this increase, we were able by 1946 to produce approximately three times as much in physical terms as we could in 1914. This simply means that the average worker was able by that date to produce very much more in an hour than he could in 1914.

Two conclusions seem to stand out among all that can be drawn from this mass of data. The first is that labor was very much better off in 1946 than in 1914 or even 1929; the second is that labor is not as prosperous as most other classes in the economy. Wages and earnings have increased markedly, but even in 1946 labor still lived on "the other side of the railroad tracks." Let us attempt to account for these findings.

The increase in weekly, hourly, and annual earnings has been striking, though part of it is illusory. The cost of living rose by 80 per cent between 1914 and 1946, and by 8 per cent between 1929 and 1946. Even so, there was a large rise in real earnings; for after all, weekly earnings in manufacturing were almost four times as high in 1946 as in 1914, they were 60 per cent above the 1929 level; and the increase in real hourly earnings is even more striking. In short, the employee is able to buy much more today than in earlier decades for an hour's or a week's work.

Each of these measures of labor's well-being is significant, for each answers a different question. The amount which labor receives for a unit of work — an hour, day, week, or year — obviously relates only to the employed worker. The amount which the employee takes home in his pay envelope each week allows for the effects of variations in the length of his work-week, but does not consider the unemployed. The most basic measure of all, perhaps, is how much labor receives in total: the aggregate earnings of all labor, employed and unemployed, in the course of, let us say, a year. We shall in this chapter consider this measure, the total amount earned by employees in a period of time.[1]

[1] Although it is not convenient in this book to cast the analysis in terms of marginal productivity as properly interpreted, the economist will recognize that the factors under-

Determinants of Labor's Income

For convenience we can divide the forces which determine labor's income into two classes: those which determine the size of the total income and those which determine the percentage that labor receives of that income. Thus, if labor's share remains fixed at 70 per cent, any factor which changes the size of the income — for example, from $120 billion to $150 billion — brings a large increase in labor's earnings; in this case from $84 billion to $105 billion. And if the national income is $120 billion, any factor that raises labor's share from, let us say, 70 per cent to 75 per cent will increase labor's total earnings from $84 billion to $90 billion.

The National Income

The total income depends in the first instance upon the amount of employment and the productivity of labor. If 40 million men are at work and their average output per man-year is $3000, the national income will be $120 billion. If only 30 million are at work and their average output per man-year is $1500, the national income will be $45 billion. Thus, we need to consider both the factors that affect employment and also those that affect labor's productivity. We have already analyzed explicitly the factors that determine the amount of employment.[1] It is only necessary to stress at this stage that the changes in output are much greater than the variations in employment which bring them about; that is, when employment rises or falls by 10 per cent, output expands or decreases by a considerably larger percentage. This is a corollary of the observation [2] that the average labor cost in any firm falls as output rises, and vice versa. Both derive from the fact that since a plant is designed to operate with maximum efficiency when its output is high, efficiency is much reduced when output is lower. The level of employment is thus a very important factor in determining the total output and hence labor's real income.

Labor's productivity has increased enormously in the last several decades. In 1929 there were about 30.6 million non-agricultural employees working roughly 45 hours a week, and the gross national product in 1929 prices was $99.4 billion. By the first quarter of 1946, there were 35.8 million employees who worked roughly 41 hours a

lying that approach have simply been grouped somewhat differently. The effect of each of the factors to which attention is directed can be interpreted in terms of the marginal productivity analysis.

[1] See Part Four. [2] See Chapter 7.

week, and the gross national product, measured in 1929 prices, was approximately $150 billion.[1] Hence, physical output increased by about 50 per cent while the amount of employment in terms of man-hours of work increased by only 4 per cent. Thus, labor productivity increased by about 45 per cent [2] in less than twenty years.

The productivity of labor at any given level of employment depends partly on methods of production and partly on the volume of capital goods that labor uses.[3] There is no doubt that new methods of production have been a most important factor in making possible our vast increase in output, especially since the increase in the size of our labor force has been relatively small. But not only have we adopted improved methods; we have extended their use more and more widely. A crude measure of this may be obtained from data on the installed horsepower capacity of manufacturing establishments. In 1914, firms engaged in manufacturing had power equipment on hand which provided about 21 million horsepower when operating at capacity. By 1929 the figure was about 42 million, and by 1942 it had risen to approximately 59 million. Thus, each employee had the use of about 40 per cent more machine power in 1942 than in 1929, and almost three times as much as in 1914. It is not surprising that labor's productivity has grown.

Labor's Share in the National Income

The volume of employment and labor's productivity determine how much will be produced for the whole economy. The amount labor gets obviously depends, too, on the whole complex of factors which determine its percentage share in the national income; in other words, on the factors which determine whether labor will get 70 or 60 per cent of the total income. In most general terms the fraction of the total income [4] received by labor depends upon the ratio of labor cost per unit of output to the price. If the labor cost of the average commodity is 70 per cent of the price, labor receives 70 per cent of the total proceeds. To illustrate, if the output is 1 billion units and the price per unit is $100, then the total receipts will be $100 billion. And if the labor cost per unit of output is $70, or 70 per cent of the price,

[1] Subtracting the gross national product of those who were in the armed forces.

[2] This is, of course, a very crude measure.

[3] In Chapter 37 we draw attention to the effects of new methods of production on labor productivity.

[4] Strictly, the gross national product.

the total labor costs are $70 billion. Since the total receipts [1] equal the national income [2] while the total labor costs equal labor's total earnings, it is apparent that labor's share in the total equals 70 per cent, or $\frac{\$70 \text{ billion}}{\$100 \text{ billion}}$; in other words, the ratio of the labor cost per unit of output to the price.[3]

The ratio of labor cost to the price of the product varies with each of the following factors:

a. The level of output. In most firms an increase in output is associated with a decline in the average labor cost, but not in marginal cost.[4] If firms adopt the most profitable price, this will mean that their price will be more or less stable if output alone changes, until near-capacity levels are reached. But labor cost per unit of output declines as output expands, and therefore the ratio of labor cost to price tends to fall as output increases, and to rise as output falls. Hence, with increases in output, we should expect some decline in the share of income that labor receives, unless this effect is offset by others.

b. The wage rate. A rise in the wage rate increases the average labor cost of any output. When the money wage rate goes up 10 per cent, the average labor cost also goes up 10 per cent. If there were no factors making for price inflexibility, prices would rise at the same rate. But we have seen that they tend to be inflexible, so that when money wages rise, prices do not increase quite as much. In short, when money wages increase, there is a rise in the ratio of unit labor cost to price and hence in labor's share of the total output. When money wages fall, the ratio of labor cost to price declines.

c. Improvements in methods of production. Improved methods of production and the growing use of capital equipment increase the national income at any level of employment, but they also reduce the labor cost per unit of output. If prices are inflexible and do not fall at a rate that corresponds to the gain in productivity, the ratio of labor cost to price declines, and with it labor's percentage share of a given income. This is why, as we saw in Chapter 37, invention may actually hurt labor, though obviously, if prices fall far enough or wage rates are increased, and if employment is maintained, labor benefits from invention.

[1] As we saw in Chapter 21. [2] Strictly, the gross national product.
[3] See Chapter 37, where this was shown in a somewhat different way.
[4] See Chapter 7.

d. Changes in the degree of monopoly. If firms are able to achieve a higher degree of monopoly either in selling their product or in buying labor,[1] the ratio of their price to their average labor cost is raised. If, for instance, firms achieve a stronger monopoly position, and the average value of the elasticity of the demand for their products falls from 4 to 3, they will raise prices by 9/8, or 12.5 per cent,[2] with no change in their marginal costs. If there is no change in the average labor cost, it follows that the ratio of unit labor cost to price falls by 12.5 per cent, and this, as we have seen, leads to a similar reduction in labor's share of the total income.

e. Changes in price of agricultural products or imported goods. A rise in the price of raw materials brings about an increase in the marginal cost, and when this happens, firms generally raise their prices. But when raw materials rise in price, it does not lead to an increase in the labor cost per unit of output; hence the ratio of average labor cost to price declines, and hence labor's share in the national income falls. We have already taken account of changes in the price of raw materials produced within the non-agricultural sector of the economy; price changes brought about, for instance, by changes in wage rates or in the degree of monopoly. Here we are concerned with changes in the prices of raw materials produced outside that sector; that is, in agricultural industries or in outside economies.

In summary, then, it appears that labor's percentage share and hence its real income depends upon the level of employment, the methods of production, the stock of capital goods, money wage rates, the degree of monopoly possessed by business firms, and the prices of imports and agricultural commodities. Other forces should be mentioned for a complete account of the determination of labor's percentage share in the national income, but they do not appear to be as important as the foregoing. A change in any of these factors is likely to affect the amount of labor's income. But unfortunately for the dogmatist, the effects of such changes are often uncertain. An increase in employment always leads to a higher national income but generally reduces labor's percentage share in it. However, the total effect is

[1] The monopoly power of the firm in hiring may be of considerable importance in interpreting the functions of labor-unionism. Unfortunately, however, it is not possible to develop it properly in the compass of this book.

[2] This follows from the relation noted in Chapter 17 that price $= \frac{e}{e-1} \cdot m$ where $e =$ the elasticity of the average revenue function, and $m =$ the marginal cost.

always predictable when employment rises, for the rise in output that comes with an increase in employment is much greater than the reduction in the ratio of average labor cost to price. Hence, labor's real income goes up when employment increases. The effect of an increase in money wages is perhaps less certain. A general rise in money wages may force a reduction in total employment, though the opposite result is more probable.[1] But if a rise in money wages does lead to a reduction in employment, the total amount that labor earns is subject to two opposing forces: it tends to receive less since less is produced; on the other hand, it tends to receive more since the ratio of labor cost to price increases with the rise in the money wage. To illustrate: national income in real terms may decline from 100 to 95 when money wages rise by 20 per cent, but labor's share of the income may rise from 60 per cent to 70 per cent, following the wage increase. Hence, the total amount received by labor would rise from 60 to 66.5. In this case labor as a whole would benefit even though employment fell. However, if employment increases when money wages rise, labor can expect to get a larger proportion of a higher income. In that case labor's total income would increase sharply. It will be clear that the effect of a rise in money wages is complex. The same is true for changes in other factors. It will be found generally that most changes which occur in the economy affect not one but several of these variables, and before we are able to determine how labor's earnings are altered, we have to take all these changes into account.

The Effect of Labor Unions

It is particularly interesting to see how the growth in the power of labor unions affects labor's real income. The labor union is not a new development, either in this economy or in others. It actually came into existence with the beginning of the wage system and the rise of the employee class. Labor unions are more than a hundred years old in this country and even older in Great Britain, but it is only in the last forty years that they have become a force to reckon with. Between 1900 and 1914, the number of workers organized into unions varied from 2 to 2.5 million; during the nineteen-twenties there were about 3.5 million, and not many more in 1935. The great organizing drives of the unions that made up the CIO began in the summer of 1936, and by 1937, there were more than 7 million union members. By the

[1] See Chapter 38.

close of 1944, this number had grown to 11 million and in 1945 to about 13.5 million. When it is remembered that clerical workers and employees of retail stores are not subject, except in a minor way, to organization, it will be seen that a very significant part of the total labor force is now organized in labor unions. The United States Department of Labor [1] estimated that in 1945 about 48 per cent of all workers in occupations which could be organized were covered by union agreements. In manufacturing industries, the figure was even higher — about 67 per cent of the production wage earners were employed under conditions set out in union agreements. Obviously, the labor union is now an extremely important force in the economy.

It is in a sense pointless to discuss whether labor unions are a good thing or a bad. They are here, and undoubtedly they will stay. To argue that they introduce a monopoly into the economy is absurd. Large firms, often acting in concert in labor matters, introduced that element long before labor unions achieved any status. What unions do is to permit labor to meet monopoly with monopoly. Bargaining between a large corporation and the individual, unorganized worker cannot be called competitive. To make up for the monopoly advantage enjoyed by the modern corporation, labor has naturally taken to unionism. It is perhaps significant that organization is most nearly complete in some of the large-firm industries such as those producing agricultural equipment, aluminum, automobiles, meat-packing, iron and steel, rubber, electrical machinery, coal mining, and railroads; while it is relatively unimportant in retail and wholesale trade, agriculture, beauty shops, servicing, maintenance of building, laundries, and other industries where firms are small. Doing away with unions would not do away with monopoly in the labor market; it would only mean unrestrained monopoly on the employers' side.

May we expect that labor's total income, in real terms, will rise as unions grow more powerful? Labor unions have two important effects on the economy. First, they owe their success to the fact that they are much better able to bargain with employers than are their individual members, for they confront the monopolistic buyer of labor with a monopolistic seller. Their achievements in securing increases in the money wage have been remarkable. Second, by checking or even nullifying the employer's monopoly advantage in hiring, they reduce the degree of monopoly in the economy in this respect.

[1] Bulletin no. 865, Bureau of Labor Statistics, United States Department of Labor.

How do these changes affect labor's real income? As we have seen, when money wages rise, labor's share in the national income also goes up. If, as a result, the national income remains steady or, as we may expect, rises, labor's total income sharply rises. If to this we add the effect of the reduced degree of monopoly in buying labor, it will be seen that labor's real income will advance considerably. There is little doubt that unions are good for labor.[1] It should be noted in passing that this does not necessarily mean that they are bad for the employer. For even though the employer's *share* of the national income may fall, his total return need not fall if the national income itself rises sufficiently.

Other aspects of union activities should, of course, also be considered before a final judgment is rendered, and this is not the place to consider them or to give that judgment. Strikes, the strongest weapon possessed by labor organizations, may reduce the national income, though hardly by a fraction of the amount that unemployment causes. This problem would have to be investigated. The effects of practices allegedly followed by certain unions in hindering the adoption of new inventions or in preventing efficient work would also have to be evaluated. The political consequences of the development of immensely powerful labor unions in an economy in which the large employer is already very powerful would have to be considered. But omitting these considerations, either because they raise problems outside the field of economics or because they seem relatively unimportant, labor organizations appear to have demonstrated their ability to help labor.

Summary

"Labor" accounts for most of those who work. Its income is not high compared to the incomes of some other classes in the economy, though in years of full employment its earnings are relatively good. The interest of labor is, of course, to get as large an income as possible. There are two methods by which its income can be raised; one is at the expense of the rest of the economy, and the other is by raising the total income. These methods are, of course, not necessarily at odds, and it may prove possible both to raise the national income and at the same time to increase labor's percentage share in that income. Maintaining the national income at peak is much the most important single

[1] Unless it is concluded that rising money wages reduce employment; then the total effect is doubtful.

step that could be taken to assure a decent income for labor. Examination of the data will illustrate the importance of keeping income high. We have seen that in 1932 labor earned about 80 per cent of the total income. Even if it had increased its share to 100 per cent, its total income would have come to only $40 billion — that is, its earnings would have been far below what they were in 1944 and 1945. Thus, while raising labor's share in the total income is bound to improve labor's position (unless the total income itself falls as a consequence), keeping the total income at peak is an even more effective step. We have discussed the methods by which the national income can be maintained at the full-employment level.[1] It will be clear that unions can do little to secure this objective except perhaps by exerting political pressure in support of policies that favor employment. It is obvious that labor's interest is most effectively served by the maintenance of prosperity. However, raising labor's share in the total income is something which the unions can accomplish. The wage rate, the degree of monopoly of business firms, the efficiency of production, and the price of agricultural and imported goods are the factors that determine the percentage of the total income which labor receives; and the unions, because they can influence the first two of these determinants, can within certain limits increase labor's share at the expense of the rest of the economy. But while it would be wrong to minimize the importance of such an increase in the percentage of the total income going to labor, it can never be as important in raising labor's total income, even if fully exploited, as an increase in income itself.

Labor's primary interest is to maintain full employment; but that is not at all in conflict with the interests of other groups, for the employer, the farmer, and the investor all gain when we have peak prosperity. If we can keep the national income at its maximum, it will not be necessary for labor to live "across the tracks." Our economy can produce enough to provide a decent livelihood for all if we do away with depression.

[1] Compare Part Four, and especially Chapters 34 and 35.

44

Agriculture

THE FARMER supplies us with our most essential commodity — food — and with many other things besides. It would be hard to get along without the product of any other broad sector of the economy, but it would be impossible to do without the products of the farm. And yet the farmer, necessary as he is to our well-being, has not, according to many observers, received an economic reward that reflects his importance.

About one person out of every four in the United States lives on a farm [1] and one person out of every four who work is engaged in farming.[2] In accordance with these figures, we might expect that one dollar in every four or five out of the national income would be earned in farming. But the actual figure is very much lower than that. In the period from 1935 to 1939, only one dollar in every twelve went to the farmer. With about one-fifth of the working force receiving only one-twelfth of the national income, it is apparent that agriculture did not pay very well. And those years were not exceptional. In 1932, incomes in agriculture were even lower, both in absolute terms and in comparison with incomes in other sectors. Our national income was only $40 billion, but the farmer received a mere $2.4 billion of it, or

[1] The farm population in 1935 was 31.8 million, and our total population was 126.9 million. Between 1939 and 1945 there was a marked reduction in farm population.

[2] In the period from 1935 to 1939, agriculture accounted for about 24.5 per cent of all employment and 20.7 per cent of the whole working force, including the unemployed. The percentage has fallen sharply since 1941. In the period from February, 1945, to February, 1946, farm employment accounted for only 16 per cent of total employment, including, like the figures above, employers, professional people, and so on.

one-sixteenth of the total. Thus, while those engaged in agriculture accounted for 22 per cent of the occupied population, including the unemployed, they received only 6 per cent of all the income. In 1945, however, the situation was happier. Farmers and farm laborers received about $13.8 billion out of a total national income of $160 billion; that is, they got about one dollar out of every twelve. Yet even this high income — though in real terms more than four times that of 1932 — was only a relatively small fraction of the total. Clearly, the farmer's economic position has not been favorable.[1]

The Agricultural Sector

Before we examine the nature of the farm problem in detail, let us sketch a rough picture of the agricultural sector of the economy. There were about 6 million farms in the United States in 1940, about 6.4 million in 1920. In 1943, the value of their output was $21.3 billion dollars,[2] about $2 billion of which was consumed on the farm while the rest was sold for cash. Table 76 shows the measure of the

TABLE 76

Cash Received from Sale of Farm Products: 1943

Product	Amount (in millions)
Wheat	$ 825
Feed grains and hay	1,114
Cotton and products	1,412
Tobacco	557
Fruits and vegetables	2,622
Cattle	2,560
Hogs	2,953
Dairy products	2,804
Chickens and eggs	2,263
All others	2,142
Total	$19,252

importance of the various products of the farm. Cotton was grown on almost five million farms, wheat on more than one million, and hogs were raised on nearly four million.[3]

[1] This is not a new phenomenon. In 1900, the agricultural working force constituted about 35 per cent of the total, but it received about 19 per cent of the national income.

[2] Exclusive of government payments.

[3] See *Structure of the American Economy*, National Resources Committee, 1939.

We have already seen that the agricultural sector of our economy is less important than it was in the eighteenth century and the nineteenth.[1] While in 1870, about 52 per cent of those gainfully employed were connected with agriculture, by 1930, the figure had fallen to about 22 per cent, and by 1945 to 16.4 per cent. Moreover, the position of agriculture has declined, even in absolute terms. In 1910, approximately 11,600,000 persons were engaged in agriculture. By 1940, the number had fallen to 9,200,000, and by May, 1946, to only 8,880,000. It has been estimated by the United States Department of Agriculture that if the economy reaches a high level of prosperity between 1945 and 1950, about 8 million, or about 14 per cent of the 56 million working in civilian occupations, will be engaged in agriculture. Thus, in terms of numbers the importance of agriculture would be only a quarter as great as it was eighty years ago.

Agriculture is an industry in which the small firm is still characteristic. Yet, though most farms are small, the importance of the large farm is surprising. Thus, in 1939, about 100,000 of the 6 million farms, or less than 2 per cent, held about 35 per cent of all farm land and about 10 per cent of the total value of farm land and buildings. About 1 per cent of all the farms had products valued at over $10,000, and these accounted for about 17 per cent of the value of all farm products. The 2.5 per cent of the farms whose product exceeded $6000 in value produced about 36 per cent of the total value of farm products. While agriculture is far from being a large-firm industry, it is by no means one in which only small units exist.

In most other sectors of the economy employees far outnumber their employers, but not in agriculture. For instance, in 1935 there were about 8.7 million family workers — that is, operating owners, tenants, share-croppers, and working members of their families — and only 2.5 million hired workers, according to an estimate prepared by the Works Progress Administration. On most farms there is no hired labor. It is estimated that labor was hired on only 1.5 of the 6.8 million farms that operated in 1935, and 5 per cent of these farms — or about 1 per cent of all farms — gave employment to about 56 per cent of all hired farm labor. This figure also indicates that the large firm is surprisingly important in agriculture — although, of course, not nearly as important as in most other industries.

[1] See Chapter 2.

Incomes in Agriculture

The low total income earned in agriculture and the relatively large number of persons engaged suggests that farm family incomes are low. This is borne out by data from a study of consumers' income made by the National Resources Committee, according to which over half of all non-relief farm families received incomes of less than $1000 in 1935–36, while about one-sixth had less than $500. Of course, if farm families who had received some relief during the year had been included, the fractions receiving less than $500 and $1000 would certainly have been much higher. About 1.4 per cent of all farm families had incomes that exceeded $5000. It was noted in the previous chapter that the average annual income per farm family was $1259 in 1935–36, and that this was the lowest figure for any occupational group. It is obvious that, compared to those of most other occupational groups, farm incomes have been low.

Moreover, farm incomes are variable in the extreme. In 1932 they stood at only $2.4 billion, yet by 1945 they equaled $13.8 billion. It is true that the prices of things farmers had to buy were higher in 1945, for the cost of living was about 30 per cent above the 1932 figure, but even allowing for this, the real value of farm incomes had increased more than four times over.

Farm income is closely related to the national income, and rises or falls with it, as is indicated in Figure 73. Here each entry shows the national income for one year, and the income earned in agriculture for the same year. Thus, one entry shows that when in 1929 the national income was $83.4 billion, farm income was $6.7 billion, and so on. It is worth noting that farm income throughout the period never came to as much as 10 per cent of the national income.

Factors That Explain Low Agricultural Incomes

There seem to be three main reasons why farm incomes are so low. Agriculture is the most nearly competitive industry in the economy. The demand for most agricultural products is relatively inelastic for price; and variable costs in agriculture are low and very flexible, the costs of starting a farm are very small, and furthermore, it is difficult to adjust output quickly.

Let us analyze the significance of each of these factors in turn. It was shown in Chapter 20 that a competitive industry produced more than an industry in which there was an element of monopoly, when

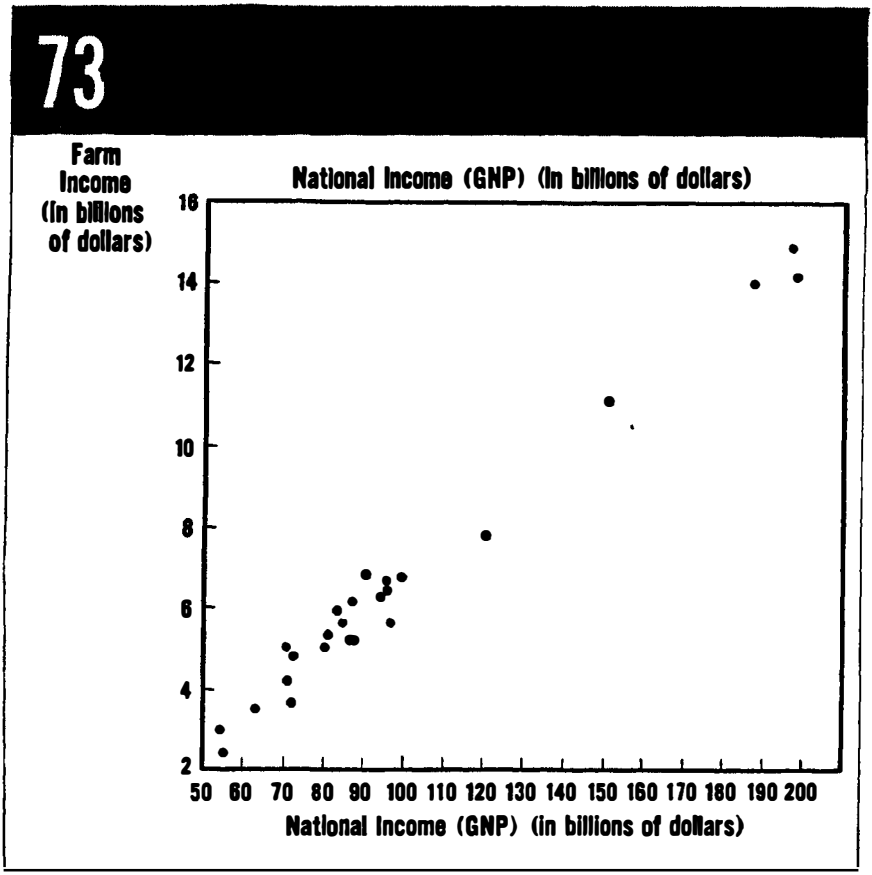

Figure 73. *Relation Between Farm Income and National Income*

both were faced with the same level of demand.[1] If the demand for two products is the same, the output of the one produced under more competitive conditions exceeds the output of the other. A single firm which controls an industry can determine output after taking into account the relatively low elasticity of the demand function. Consequently, its output tends to be low and its price high. But in an industry which is strongly competitive, each firm is faced, not with an inelastic or only moderately elastic demand function, but in the extreme case with one which is perfectly elastic. Hence, their output

[1] As in Chapter 20, it is here assumed that the units in which the commodities are measured are identical in terms of cost.

tends to be high and their price relatively low. The competitive industry produces more and sells at a lower price than a monopolistic industry operating under the same conditions of demand and cost. This does not necessarily mean that the more competitive industry has lower receipts. Let us analyze the situation with the help of the diagram in Figure 74. Here DD_1 is the demand curve for both industries.

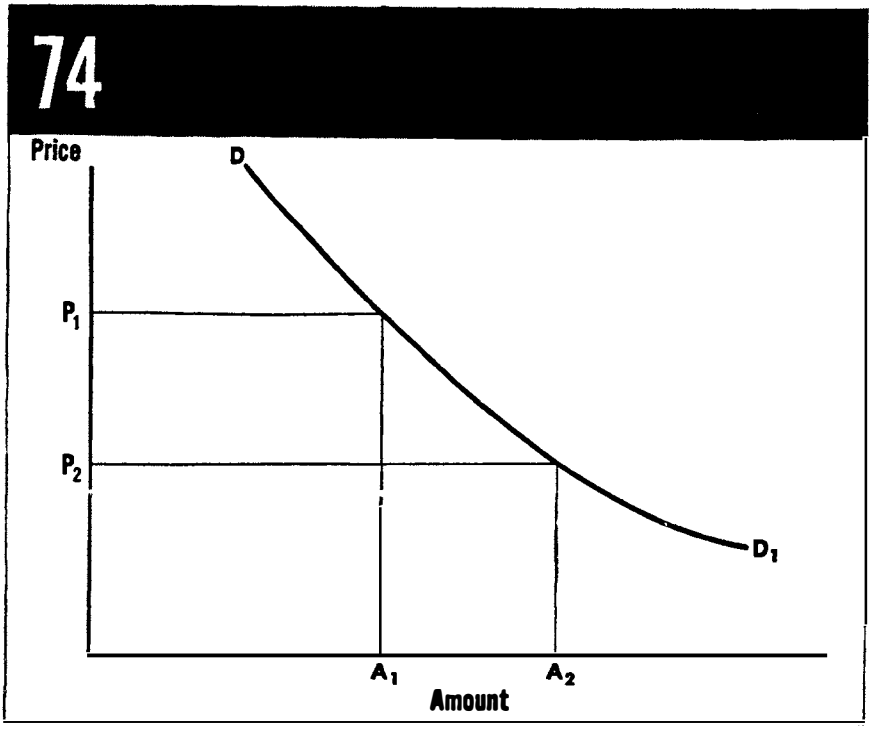

Figure 74. *Receipts with Monopoly and Perfect Competition*

The more monopolistic industry produces A_1 and charges P_1; the competitive industry produces A_2 and charges P_2. If the demand for the product, DD_1, is elastic, the total receipts of all firms in the competitive industry, $A_2 \cdot P_2$, exceed the receipts of firms in the more monopolistic industry, $A_1 \cdot P_1$. But, though the total receipts of the competitive industry are greater, it does not follow that the income earned in the competitive industry is also greater, and, more important, we cannot assume that the income per man engaged in the competitive in-

dustry is higher. For to produce an output of A_2 rather than A_1 obviously entails higher costs. To the extent that these higher costs represent payments to other industries for raw materials, they reduce the income earned in the more competitive industry, and hence that income will not exceed the income earned in the monopoly industry by as much as the difference in their receipts. It is as though the total receipts of the competitive industry were 100, while the total receipts of the monopolistic industry were only 90. But the competitive industry must buy more raw materials from other industries because its output is higher; let us suppose that the outlay for raw materials is 50 for the monopolistic industry and 55 for the competitive industry. Then the incomes earned by those engaged in the monopolistic industry (rent, interest, wages, salaries, and profits) come to 40, and the incomes earned by those engaged in the competitive industry come to 45. The difference in the total income earned is 5, while the difference in total receipts is 10. And the income per head will be lower in the competitive industry than in the more monopolistic industry, for to produce A_2 rather than A_1 requires a larger amount of employment [1] by approximately the ratio A_2/A_1. With considerably more employment in the competitive industry, and receipts greater by a factor which is less than the ratio A_2/A_1, it is apparent that income per head would be somewhat lower in the competitive industry.[2] If the number engaged in the competitive industry were 40, and in the monopolistic industry were only 20, income per head in the former would be 45/40 or 1.125; in the latter, 40/20 or 2. The fact that agricultural industries tend to be relatively competitive is thus one factor that keeps income per head relatively low.

Another reason for the low income level of agriculture is that the buyers of agricultural products often occupy a relatively strong monopoly position. Thus, most of the tobacco grown in the United States is sold to four buyers, who take about 80 per cent of the total crop each year. Likewise, there are not many purchasers of livestock or dairy products. Naturally the farmer in such a market has little bargaining power, and in this respect he is often in much the same

[1] The reader is reminded of the way in which the units of output were defined in each industry.

[2] If the demand for the product were relatively inelastic, the ratio of income per head in the competitive and monopolistic industries would be even lower. This follows from the fact that the ratio of the prices charged in the competitive and monopolistic industries would be lower under the assumed conditions.

position as the laborer. Because of the weakness of his bargaining position, his income tends to be relatively low.

The inelasticity of demand for the products of the farm is well known. Most of them are regarded as necessities, and it is supposed that buyers will expand their purchases by only a small amount when the price is lowered. Hence, in order to sell more, a large reduction in price is necessary, and the larger the output, the lower the total receipts of the supplying firms tend to be. But, as we have seen, firms in agriculture which typically sell in a competitive market tend to produce a large output in relation to the demand — and when demand is inelastic, the larger the output, the lower are the receipts and hence the lower the total income, and even more particularly, the income per head; for the inelasticity of demand for agricultural products strengthens the ill effects of keen competition, so that even in a good year like 1945, agricultural income is low compared to that in other sectors of the economy.

Agricultural income also falls as national income does, and perhaps even more rapidly. There seem to be two reasons why this is true. The first is that, as income drops, the demand for the products of agriculture declines. Since people eat less well and buy fewer new clothes, the demand for wheat, livestock, vegetables, fruits, and cotton responds. Secondly, for reasons which we shall discuss shortly, the supply of agricultural commodities often increases during depression, and it is in any case likely to be very inelastic. Let us illustrate these changes diagrammatically; as in Figure 75. The initial demand is indicated by D_1, and that after the decline in the national income by D_2. The original supply is represented by S_1, and hence the price is P_1, and the income in the industry, not allowing for the purchase of raw materials from other industries, is shown by the area $P_1 \cdot O_1$. When the national income declines, the supply increases from S_1 to S_2 for reasons which will shortly be explained. Hence, the price falls to P_2 and the output changes to O_2. The total receipts of the industry, now $P_2 \cdot O_2$, are much lower than before. Since they may be roughly identified with the incomes earned in the industry, we may conclude that agricultural income falls sharply when the national income declines.

Figure 75 shows that prices fall sharply in depression, but output scarcely decreases at all, and this is exactly what happens in many agricultural industries. For example, the price of wheat fell by

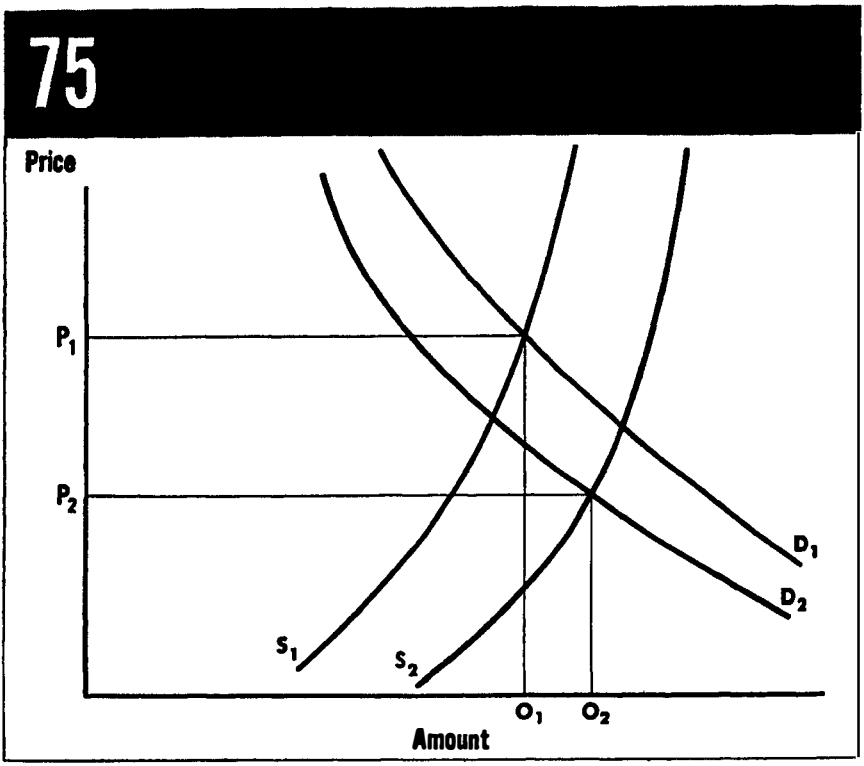

Figure 75. *Demand and Supply: Agricultural Products*

52 per cent between 1929 and 1932, but production fell by only 8 per cent. In the same period the price of fluid milk fell by 33 per cent and output actually rose; corn prices dropped by about 68 per cent, but production increased by 16 per cent; the price of hogs fell by 63 per cent, and marketings fell by only 20 per cent. In contrast, during those years the price of pig iron fell by 23 per cent and output fell by 80 per cent. When the demand for agricultural products declines, price generally falls very greatly and output only slightly. This, as an inspection of Figure 75 will indicate, is the result of the fact that the supply of agricultural products is inelastic and tends to increase in depression.

There seem to be two reasons for this perversity in supply. First, the number of firms tends to expand, and second, their cost conditions change drastically. When unemployment becomes severe, many

people go back to the farms, but when jobs are readily available in factories, the farm population declines sharply. Thus, between 1929 and 1932, the number of persons engaged in agriculture rose by 5 per cent, or from 10.5 million to 11.1 million, and the number of farms increased by 4 per cent. Then, between 1940 and 1944 when jobs became plentiful in other branches of the economy, the numbers engaged in agriculture dropped by almost 20 per cent, from 10.6 million to 8.7 million. Furthermore, most products come from farms where variable costs are very low, since most farm labor is performed by the farm family. Thus, because other jobs are not available to him, the farmer has nothing better to do, when agricultural prices fall, than to go on farming, with the result that the supply of agricultural commodities is relatively inelastic. And since some of the variable costs fall quite sharply during depression,[1] the supply itself grows.

In summary, then, the very nature of farm work makes the supply of the product inelastic. Moreover, during depression the supply even tends to increase while demand unfortunately goes down, and in consequence price falls sharply. In depression, therefore, total receipts and farm income both suffer.

Measures to Raise Farm Incomes

What can be done to ease the farm income problem? The most important thing is to maintain the national income at peak level. If it falls to $40 billion, farmers cannot hope to have a proper income no matter what efforts are made in their behalf. Agriculture, incidentally, can make a contribution to that objective. The investment done by agriculture amounted to over $1.2 billion a year or about 5 per cent of the total, during the nineteen-twenties, and it fell to only $0.6 billion in 1930–34. If farm investment could be maintained at a high level, it would ease the problem of securing adequate investment and thus of maintaining prosperity.

If the national income is maintained at a high level, the demand for agricultural products also remains high and at the same time the supply is kept down. This, of course, spells high income for the farmer. When income is high, there are good job opportunities in other industries for the surplus farm population, and hence the movement of labor from the farm to the city goes on. As a result, the supply of farm prod-

[1] For instance, wage rates for agricultural labor were $51.22 a month in 1929 and only $28.88 in 1932.

ucts is not increased. But if the national income drops, the demand for farm products falls, and at the same time growing unemployment forces many families to return to the farm, bringing about an increase in the supply of farm products. Falling demand and increasing supply slash farm income. Farmers are not, of course, exposed to unemployment, but industrial unemployment reduces their incomes just as surely as if they themselves were unemployed.

The importance of maintaining the national income at capacity level can be judged from the data in Table 77, presented in 1945 by the Department of Agriculture as an estimate of farm conditions in 1950. With full employment, farm incomes are likely to be three times as great as with deep depression.

TABLE 77

National Income and Agricultural Earnings: Different Levels of Prosperity: 1950

	Full Employment	Moderate Prosperity *	Bad Depression †
National income	$150 billion	$105 billion	$57.5 billion
Gross national product	$188 billion	$130 billion	$76.5 billion
Farm prices (1909–1914 = 100)	165	125	75
Net income of farm operators	$9 billion	$6.5 billion	$3.0 billion
Total agricultural employment	8 million	8.5 million	9.0 million

* Unemployment of 7 million. † Unemployment of 17 million.

A considerable part of our total farm output is normally sold to foreign countries; about 10 per cent by value for the period from 1935 to 1939. Agricultural prosperity requires this market, and would certainly be encouraged by its further development. While the volume of our exports of agricultural products to foreign markets does not depend upon ourselves alone, our own policy is nevertheless an important factor in determining it. If we maintain a high level of income in this country, we help other countries to prosper; and if, in addition, we make dollars readily available to them, they are better able to buy our goods, as we saw in Chapter 42. Thus, our own policy has a direct effect upon our exports.

But even if our national income is high and the volume of agricultural exports is satisfactory, we shall not have solved the whole problem of agriculture, as an examination of Table 77 will show. If 8

million people engaged in agriculture shared an income of about $11 billion,[1] the individual income would be $1375, far above the usual figure for farmers,[2] but not high compared to the income per head in non-agricultural pursuits, where it would be about $2480.[3] This discrepancy must be due in considerable measure to the fact that agricultural industries are generally competitive while most others are more strongly monopolistic. The remedy for this situation is twofold; to reduce so far as possible the elements of monopoly outside of agriculture,[4] and to implant some monopoly characteristics, or at least some of their fruits, in agriculture. The Agricultural Adjustment Acts, the first of which was passed in 1933, do introduce an element of monopoly into agriculture. By making payments to farmers conditional on their not exceeding a designated output or acreage of particular crops, they give each producer an incentive to set his output at less than the level he would choose independently. To a very slight degree, this achieves the same result as monopolistic determination of output.

Other measures to raise farm income, such as the encouragement of better methods,[5] control of erosion, provision of adequate credit facilities, provision of electric power, and so on, do not require discussion here.

Summary

The agricultural problem is fundamentally an income problem. Even in good years the farmer earns less than the rest of us, and in depression he makes very little indeed. Agriculture suffers during depression not only because of the decline in the demand for its products, but also because it experiences a great deal of what has been called "disguised unemployment"; numbers of people take up farming because there is nothing else for them to do, and in consequence the supply of agricultural products tends to increase. Therefore, to keep farm incomes at a high level it is necessary to keep the

[1] An allowance was made for wages for farm labor.

[2] In 1932, it was about $210.

[3] Since the computations are not intended to be precise, rents, interest, and profits for that part of the population which is not occupied were not subtracted before determining the income per head of the employed population (including, of course, farmers, individual proprietors, businessmen, and so on).

[4] This topic will be discussed in the next chapter.

[5] Which incidentally would also raise the real value of non-farm incomes if any of the reduction in cost were passed on in the form of lower prices.

national income at its peak. But agriculture suffers also for another reason. Because it is more competitive than other industries, its output is relatively high and its price is relatively low compared to those of other industries. This difficulty could be lessened by reducing the strength of monopoly elsewhere and, if necessary, introducing monopolistic elements into agriculture, as recent legislation has already done. But it should be emphasized that even if the farmers' share in the total income could be made to correspond with their numerical importance, this would be a relatively ineffective measure unless the total income were high. For even though agriculture's share of the total income were raised to about 13 per cent (the ratio of employment in agriculture to total employment), if the national income fell to $57.5 billion, agriculture's income would be only $7.5 billion. But if the national income were kept at peak, and nothing were done to raise agriculture's share in the national income, farm income would be about $11 billion. Thus, the best way to keep farm incomes high is to keep the national income high. Measures to increase the farmers' share in it are desirable, too, but not nearly as important; for the effects are bound to be smaller, and, perhaps more important, they only succeed in so far as they injure the well-being of other members of society.

45

Monopoly

"PEOPLE OF THE SAME TRADE seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices. It is impossible indeed to prevent such meetings, by any law which could be executed or would be consistent with liberty and justice. But though the law cannot hinder people of the same trade from sometimes assembling together, it ought to do nothing to facilitate such assemblies; much less to render them necessary." Thus wrote Adam Smith in 1776. The economy he describes was in most respects very different from that of today, but in this respect, at any rate, it was similar: the steady and persistent drive toward monopoly that colored the economy of two centuries ago is still a basic element in our own. It is a drive that affects not only business firms, who want to sell their products under the favorable conditions of monopoly; it is also an objective of almost all economic interest groups: the professions, farmers, labor, and even the whole economy in its relations with others. Why does it exist? How is it realized? What are its consequences? What do we do about it? These are the questions to be discussed in this chapter.

To look no farther back in economic history, the guild structure of industry in medieval Europe gives us a striking example of monopoly. Each guild was organized to protect the rights of its members against competition from outsiders, and only those who had served an apprenticeship were allowed to set up as craftsmen in the town. In Frankfort, indeed, even the beggars were organized into a guild —

although it is hard to imagine the nature of the on-the-job training that it provided — and outside beggars were permitted to ply their trade in the city on only a few set days in the year. Each craft had its own monopoly organization in the towns, and each organization endeavored to protect its members from the competition of out-of-town craftsmen or interlopers. But the drive to monopoly was not bounded by the guilds, for the guild in one craft would often attempt to extend its sway over related crafts. Thus, one such organization which occupied a strategic position in the knife-making trade brought the other knife-makers under its control. Within each guild, too, the struggle for monopoly power went on. As the masters of the guild assumed more and more control over their journeymen and apprentices, the latter organized their own "unions" to bring to bear all the little pressure their monopoly position conferred.

Monopoly certainly did not come into being with big business in the modern form. Nor is the power which monopoly confers sought only by big business firms. As we saw in our examination of agriculture and labor, all those who sell anything for economic gain have a motive for achieving a monopolistic position. Whether they sell oil or oranges, or their ability to work in the coal mines or to look after the accounts of a business firm, they do what they can to strengthen their bargaining position as sellers. Likewise, all those who buy, in a business capacity, have an interest in achieving monopoly power,[1] and whether they buy raw tobacco, the services of steel workers, or poultry feed, they try to strengthen their bargaining position as buyers. Monopoly is thus to be regarded as a means to an end — higher income. No matter what the economic interest group, it will make the same attempt to strengthen its monopoly position so that it can buy cheap and sell dear. Interpreted in this way, monopoly power is sought not only by peculiarly selfish business firms, but by every self-seeking group in the economy.

Why Monopolies Exist

Nevertheless, it is true that in a modern economy the business firm is usually better able to achieve monopoly status than are other economic institutions. In most industries the business firm has a head start on, say, the worker, the farmer, or the consumer. As we have seen, in most industries a very few large firms occupy the dominant

[1] Strictly, monopsony power.

position. This in itself would not mean full-scale monopoly: two or three firms, when they constitute the greater part of an industry, might still engage in price competition, though this kind of competition would not be perfect in the technical sense. But it is obviously much easier to organize a small number of firms to act as a single unit than to organize a million consumers in a large city, or the 500,000 men in the automobile industry, or the 4 million cotton-growers. It is, therefore, only to be expected that monopoly power will appear first and grow strongest among business firms. For that reason, most of our attention in this chapter will be directed to business monopolies. Moreover, the effects of this form of monopoly are likely to be more important for the economy than the consequences of monopoly power in other segments of our economy because of the critical importance of the business firm in a capitalist system.

Extent of Monopoly

Let us first examine briefly some of the situations where strong monopoly power now exists.[1] The dominant firms in the steel industry — though this is certainly not the most monopolistic industry in the country — have achieved strong monopoly status. We have already noted that the three largest have more than three-fifths of the ingot capacity of the industry, and that most of the firms in the industry make a practice of following the lead of the United States Steel Corporation in setting price. Moreover, the firms in that industry use the basing-point system in quoting prices, and as a result a buyer has to pay the same whether he orders from the nearest source of supply or from a distant one. Other monopolistic industries, as we have seen, are the public utilities and the railroads. So are the aluminum and nickel industries. Most plumbing supplies are sold under strong monopolistic conditions. Cigarettes, roofing material, asbestos, and in many local markets, bakery and dairy products, are provided by firms which enjoy considerable monopoly power. There is no point in extending the list: it would require less space to name the industries in which the dominant firms have not achieved monopoly status. Suppliers of goods to consumers or to other firms generally occupy a strong bargaining position by virtue of their monopoly power. Fur-

[1] Needless to say, a monopoly position is just as strong when a number of firms act together as when there is only one firm. Furthermore, many of the effects of the monopoly follow even though it is not absolute; hence the examples to be noted below are not intended to illustrate pure monopoly only.

thermore, some industries have achieved a strong monopoly position in buying. This is particularly true of those that purchase raw materials from unorganized producers — for example, tobacco, raw milk, and livestock. It is also true, although less so today than formerly, in the labor market where firms frequently act together to prevent wages from rising too far.

But monopoly is not confined to the business firm, as we have said. Labor unions bargain collectively for their members. Producers of certain products — for example, citrus fruits — band together to sell at a good price. Occasionally even consumers join together to buy on advantageous terms. The drive to achieve monopoly status — or, as it may be called, to do away with "unfair and cut-throat competition, to protect legitimate business from the chiseler or the price-cutter" — is one that affects all sectors of the economy.

Motives for Monopoly

Why does monopoly tend to grow? What interest is served by its development? We have already seen the answer to this in earlier chapters.[1] Firms that have monopoly power can earn greater profits than they could otherwise, since they can sell at a higher price and probably buy at a lower one. The stronger the monopoly power, the higher is the profit which can be realized under given conditions of demand and cost. This is true for buyers or sellers of all types. Consequently, business firms adopt various devices to strengthen their monopoly position; labor unions seek the closed shop, or in some cases high initiation fees, to benefit their membership; and professional organizations limit the numbers who may be called to practice their profession. Individuals and institutions generally attempt to bolster their monopoly position because they recognize that strong monopoly power helps to get a high income.

Methods of Achieving Monopoly

There are many ways of achieving monopoly, ranging from the agreements casually made at a dinner or a reception to all-out racketeering. A single firm may extend its control over an industry by buying up its competitors. Or a number of firms may join together and pool their assets in one large firm like the United States Steel Corporation. Or a holding company may be organized to own the

[1] See Chapter 20, for example.

securities of a number of operating companies and in that way to insure that they are all subject to one control. The right to exploit patents may give a firm the power to control an industry, and even when the patent-holding firm allows others to use its patented process, it often requires them to charge prices that it sets or to limit their production to a certain figure. Frequently, when the ownership of patents is so divided that each of three or four firms in an industry controls a part of the most efficient process, all the patents are assigned to a patent pool, and each of the assignors is then free to use them all. But generally in such cases the firms enter into an agreement about prices, output, or division of markets. Again, firms are often able to act together because a director of one of them is associated closely with, or may indeed himself be, a director of several related ones. Through interlocking directorships the five largest insurance companies were in 1935 related to 780 other corporations, 100 of them being other insurance companies. In some industries, firms act together without any formal agreement or consultation. Understandings about price leadership illustrate this practice. Trade associations often sponsor co-operation among their members which is actually directed toward monopoly. This is generally done by suggesting to member firms codes of "fair trade" which commonly do away with competitive price policies — whatever their ethical merits may be. Sometimes trade associations sponsor price leadership or the adoption of standard cost-accounting procedures. On the face of it, the latter practice seems innocent enough; but when firms are urged to charge a price based upon standard costs, and when information is given about standard costs in the industry, the effect is to urge all firms to follow the price which the association recommends.

Other methods of obtaining monopoly require the assistance of favorable legislation. Monopoly grants are, of course, given by the state to railroads, airlines, and public utilities. On grounds of public health, restrictions may be imposed on the movement of fruits and vegetables across state lines. In one state milk which comes from outside a certain area cannot be sold unless it is colored red — a ruling which, of course, gives an effective monopoly to the producers inside the area. Tariffs prevent the import of many commodities and thus prevent competition from outside countries. Building codes enforced by local governments often contain provisions that reserve the market to suppliers of certain types of building materials, and con-

sequently architects and engineers frequently find that they cannot introduce improved building methods. The means to monopoly power are almost as numerous as the broad limits of human ingenuity can devise.

The monopoly's control may be exercised directly over price, and all the firms that are party to the system may agree — or at least act as though they had agreed — not to engage in price competition. Or the control may be exercised over output, and firms may agree to produce no more than a certain number of units in a given period of time, as is often required of firms that secure a license to use a patented production process. Finally, control is sometimes directed toward parceling out the market among the participating firms, each agreeing to supply customers in his own market area, but no other. Control through patent ownership is sometimes able to keep out all competitors. Firms have even been able to do this when their patents have not in fact conferred the protection claimed for them. For example, the manufacture of cemented tungsten carbide was confined before 1940 to licensees of the Carboloy Company. Any other firm which undertook to produce either this or related substances was threatened with a suit for infringement of patent rights. Generally the threat alone was enough to bring the rival company to terms, for the expense of fighting such a suit is very often far too heavy for a small company. Finally, however, the courts determined that the patents were invalid after they had been used for years to prevent competition. In some industries, where it is not possible to prevent competitors from entering the field, the new firms are urged to enter into whatever agreement already binds the other firms.

Monopoly and the Whole Economy

The firms that succeed in winning some measure of monopoly power earn higher profits as a result, and the same is true of labor, or professional groups. This, as we have seen, is the incentive to organize monopolistic controls. But what about the rest of the economy? Is a gain for one interest group made only at the expense of another? Does everyone else lose when business manages to increase its monopoly strength?

The answer to this question is an emphatic yes. Generally there is not only a loss for the rest of the economy, but the loss is very often greater than the gain to those who strengthen their monopoly position.

This implies that as monopoly grows stronger, total output falls. There are three reasons for believing that the total output is lowered when monopoly controls are strengthened. A growth of monopoly discourages investment; it generally reduces the propensity to consume; and it lowers the amount of output and employment that corresponds to a given level of spending. Let us examine these three reasons in some detail.

The first reason is that generally a growth in business monopoly discourages investment. In the first place, if the monopoly controls are effective, the participating firms earn higher profits but reduce their output. But when output is lowered, firms have less incentive to expand their plant and equipment, since there is no reason to enlarge when current output is below existing capacity, even though profits may be extremely high. This, of course, does not mean that firms in a monopoly position never expand; but only that they are slower to do so than they would have been if their monopoly strength had been less. Second, when the firms which produce investment goods strengthen their monopoly position, they raise the price of these goods. We have already seen that when the cost of investment goods goes up, marginal projects are not undertaken. Third, a firm in a monopoly position has less inducement to adopt improved methods of production than a firm which has many competitors. For such a firm must bear the full charge of obsolescence and can set against it only the advantage of being able to produce at a lower cost, since it does not compete on the basis of price. In contrast, a firm which must face competition would be able to secure an additional advantage from the adoption of an improved technique, for it could gain at the expense of its competitors, since it would now be able to undercut their price. When price competition is absent, firms have less reason for adopting improvements in the process of production than they would have had in a more competitive situation. Hence, the investment which is based upon the installation of improved machinery is discouraged by monopoly. This does not mean that firms in a monopoly situation never adopt new methods; if that were true, we should not have had the enormous increases in productivity that characterize our economy. And it certainly does not mean that firms in a monopoly situation are not eager to develop — as distinguished from adopting — improved methods of production. Indeed, the urge to invent and develop may be even stronger, since the maintenance

of their monopoly status may require that they secure control of improved techniques which otherwise might fall into the hands of potential competitors. Even so, such firms are under less pressure to adopt new methods of production than are firms in a more competitive situation. Consequently, the inducement to invest grows weaker as monopoly grows stronger.

The second reason is that monopoly — especially business monopoly — is also likely to lower the propensity to consume. We have already seen that firms earn higher profits when they improve their monopoly position either as sellers of product or as buyers of labor or raw materials. Moreover, the increase in profits is made at the expense of labor and the suppliers of raw materials. Hence, when business monopoly becomes more powerful, there is a shift of income from wages and farm incomes to profits and consequently income distribution grows more unequal and the propensity to consume drops.

The third reason is that the growth of business monopoly reduces output and employment, even with the same amount of spending. When prices are higher, as they are when monopoly increases, the same amount of spending buys fewer goods. Or to put it another way, when a given amount of spending is directed at monopolistic firms, they produce less and charge a higher price than would more competitive firms. Hence, even if investment and the propensity to consume were not adversely affected — as unfortunately they are — output and employment would still be reduced.

Since growing monopoly power in business firms lowers investment, the propensity to consume, and the level of output corresponding to a given level of spending, or to a given level of investment and propensity to consume, it obviously reduces the national output.

Because output and employment are lower in a monopolistic than in a competitive economy, the real national income is also lower. We have already pointed out that the business interests which succeed in enhancing their monopoly power make higher profits at any given level of demand and cost. Hence, it follows that even if incomes were maintained, other types of income recipients would lose by the amount these firms gained. But since income itself drops, other interests lose by even more. Indeed, it is quite possible that even the interests which stand to profit from the improvement in their monopoly position will lose out in the end. For they may force the income of the economy so far down that even with their greater percentage

share of that income their own total earnings will fall too. The point has been made elsewhere that a high percentage of a low income is but cold comfort. Indeed, in 1932 even those firms whose monopoly position was most secure did not earn high profits. Efforts to secure a larger slice of pie may end by reducing the size of the pie itself. The whole attempt may, therefore, be self-defeating, even though in the narrower sense it succeeds.

No one interest group is, of course, likely to be influenced by these considerations, for always it can clearly see the gain to be achieved from strengthening its monopoly position, and it naturally acts to secure that gain. The long-run consequences of its action upon the total income are unfortunately less apparent. Furthermore, unless the interest group centers in a basic industry, its efforts will affect the national income only slightly, and thus it is not deterred from seeking the larger slice by considerations relating to the size of the pie itself. Each group that stands to profit from an increase in its monopoly power is under pressure to take action directed to that end, even though in doing so it may not only injure the rest of the economy, but, ultimately, itself as well. For all these reasons, the social interest is often furthered by restricting monopoly; but for obvious reasons these efforts have not met with notable success. The concentration of power is certainly not less marked in our economy today than it was at the beginning of the present century, and it may have grown. As long as so great a degree of monopoly exists, it is probable that output will be below capacity, distribution of the output will be unequal, and the pattern of production will remain unbalanced.[1]

Summary

Monopoly status is sought by all economic interest groups, though the powers secured by business firms are of critical importance. These controls may be directed at maintaining a monopoly price, at limiting output, at sharing the markets, or at restricting the number of firms that produce a commodity, and the methods of achieving these powers range from implicit agreements to formal combination. The chief effects of a growth in business monopoly are an increased share of the total income for the monopoly group and a reduction in the total income of the economy. Thus, growing monopoly is likely to be harmful to the rest of the economy which is left behind in the race

[1] For a discussion of this last point, see Chapter 20.

to secure monopoly status, and it may even injure the groups that achieve it, since their prosperity depends in part upon the prosperity of the economy as a whole. Nonetheless, it is difficult to control monopoly, as the most casual glance at the world about us will show.

SUGGESTIONS FOR FURTHER READING

Arnold, Thurman W. *The Folklore of Capitalism.* New Haven: Yale University Press, 1937.

Chapters 8 and 9 give an amusing account of monopoly control.

Edwards, Corwin. "An Appraisal of the Anti Trust Laws," *Proceedings of the American Economic Association* (1946).

——— "Can the Anti Trust Laws Preserve Competition?" *Proceedings of the American Economic Association* (1939).

Lester, Richard. *Economics of Labor.* New York: The Macmillan Company, 1941.

See Chapters 8–12.

Lynch, David. *The Concentration of Economic Power.* New York: Columbia University Press, 1946.

This is a very useful summary (400 pages) of the data brought before the TNEC.

United States Government: Temporary National Economic Committee. *Investigation of Concentration of Economic Power. Hearings before the Temporary National Economic Committee, Congress of the United States.* Part 25: "Cartels." Washington: Government Printing Office, 1939–41.

Note the testimony of Theodore J. Kreps and Clair Wilcox.

United States Government: Temporary National Economic Committee. *Investigation of Concentration of Economic Power. Monographs for the Temporary National Economic Committee, Congress of the United States.* Washington: Government Printing Office, 1939–41.

Monograph No. 5, "Industrial Wage Rates, Labor Costs and Price Policies" (the Summary); and Monograph No. 16, "Anti Trust in Action" (note Chapters 3, 5, and 6 of Section 2).

United States Government: Federal Reserve Board. *Post War Economic Studies.* Washington: Government Printing Office.

See No. 2 of this series, "Agricultural Adjustment and Income."

United States Government: Department of Agriculture. *What Peace Can Mean to the American Farmer.* (Miscellaneous Publication No. 562.) Washington: Government Printing Office, 1945.

Shows clearly the close relation between farmers' income and the national income.

46

A Resurvey of the Economy

OUR ECONOMY is an institution of the greatest complexity and variety which has evolved out of a multitude of social forms and practices, and which is still evolving. Its basic units — producing business firms, financial organizations, and government bodies — are varied in the extreme. Its resources — labor, capital goods, and "management" — are no less so. The operations of these various structural units and the ways in which they affect the rest of the economy are complicated and difficult to disentangle. Indeed, we cannot avoid a feeling of helplessness and inadequacy when faced with the problems created by this complex and ever-changing organism, for they are bound to be perplexing and intractable and they sometimes seem hopelessly involved. Hence the analysis of its difficulties is not likely to be neat, nor are the solutions likely to be simple. If our economy were no more complicated than an automobile engine, we could diagnose and cure its difficulties easily enough. Unfortunately, it is not so uncomplicated as that — a consideration which should warn us to beware of the "simple," "common-sense" approach of the charlatan.

Yet the complexity of the problems offers us no excuse for delay in handling them. They demand solutions, and quickly, because they affect us most gravely. We cannot afford depressions. We must not permit extreme poverty. We will not allow continued shortages of essential goods and services such as housing, electric power, and food. If the economist cannot or will not give good and sensible advice on how to deal with these problems, someone else with a

louder voice, but even less insight into the workings of the economy, will persuade us to accept his remedies, to the ultimate disadvantage of us all. Hence, we need continuous, patient, and dispassionate investigations of the operations of the economy; for only as our understanding grows can we hope to treat these urgent problems competently. This book does not, of course, pretend to solve these problems. So much can hardly be expected of an introductory survey in any science, and especially in one whose investigators must work in an atmosphere surcharged with the claims of interest groups and preconceived opinions. But if the book can give no final answers, at least it points the way and suggests a sound approach to the problems.

In order to see once more the broad outlines of this approach, let us briefly retrace our steps. For convenience, our analysis of the economy has been divided into six sections, organized around a single aim: to examine the factors which determine the size and composition of our total output and the distribution of that output. To do this, we first surveyed in very general terms the nature of the economy and the materials of which it is composed (Part One). In this process we saw that the critically important unit in the entire structure is the business firm, for on its level are made most of the decisions about output and price. We then (Part Two) discussed in considerable detail the factors that affect the firm's most important decisions, particularly noting the effect on output of changes in demand, costs, and the degree of monopoly. We were then able to appraise the composition of the total output and to isolate the effects of monopoly on the make-up of that output. We then (Part Three) briefly examined the financial sector of the economy — the operations of the commercial and Federal Reserve Banks — with a particular view to seeing how the interest rate is determined. We saw that the commercial banks governed by the Federal Reserve Banks influence the rate of interest through their control over the amount of money in circulation. We also saw that gold is important in the monetary sector of the economy because of the requirement that the Federal Reserve Banks keep reserves in the form of gold against their liabilities. As long as this regulation remains in effect, movements and stocks of gold will continue to be significant.

With these discussions — of the economy as a whole, of the business firm, and of the monetary and banking systems — as background, it was possible to approach in Part Four the most important problem

which faces our economy today — the problem of depression and inflation. Basing our analysis on what we had learned about the business firm and the interest rate, we saw that the factors which determine the level of the national income and employment are the amount of investment, private, public, and foreign, and the propensity to consume. We saw that while an economy as wealthy as ours is liable to depressions, it can avoid them by keeping investment and the willingness to spend at a high enough level. We then turned our attention (Part Five) to our relations with other economies, and we saw that unobstructed international trade adds to our productive efficiency, but that unless we can avoid the adverse effects on employment of freely moving trade, we cannot share in the fruits of this heightened productivity. And finally (Part Six) we briefly surveyed a few of the more important conflicts of interest in the economy.

In the course of this discussion we noted once more a point made earlier in the chapters on International Trade — that there are essentially two ways of increasing the prosperity of any group within the economy or of the economy as a whole. One method, covering a wide variety of measures, seeks the prosperity of the particular interest group or economy by increasing the size of the pie from which the group's own share is cut. The other secures the prosperity of the group by getting a larger slice of a given pie — in other words, by drawing on the well-being of other groups. To illustrate, the farmer can gain prosperity either at the expense of, let us say, labor; or he can get it by sharing with all others a general rise in the national income. Similarly, our whole economy can gain in prosperity either through increasing the size of our own market by reducing those of other countries — in other words by stealing customers — or it can do so by developing a market where none existed before. Raising tariffs exemplifies the first method; encouraging private domestic investment by lowering interest rates illustrates the second. We shall say more later about these alternative routes to prosperity.

Since the economy is in a constant state of flux and change, we can hardly expect that it will always be suffering from the same ailments or that any particular remedy will always be equally beneficial. The most pressing economic problem in the nineteen-thirties was unemployment; the most pressing one in the eighteen-thirties was certainly different — perhaps devising ways to facilitate a rapid growth in the stock of capital goods. And whereas in the nineteen-twenties a cer-

tain medicine for unemployment may have worked wonders, by the nineteen-fifties a completely different one may be required.

To remind ourselves once more how much our economy has changed, it may be well to note again those features which have most greatly altered in recent years. The first, and perhaps most profound change is in the form of the economy itself. The simple handicraft economy of a century ago no longer exists. Where formerly most of our goods were produced for direct use by the producer or his family, most of our output is now directed to a market, for sale at a profit. Ours is no longer a subsistence economy on the Robinson Crusoe pattern. Two consequent developments are the growth of labor as a class and the increased importance of large corporations. Most of our production is carried on today by very large business firms, and most of the men and women engaged in production are hired employees — wage earners and salaried workers. The growing importance of labor and of the large corporation has brought with it many new problems and also many new opportunities; these things too denote a radical change in the form of the economy within the last century. Another important development has been the shift in social attitudes toward the economic difficulties of particular groups. This is not the place to discuss either the reasons for these new attitudes or their merits, but it is essential to realize that they do exist. Today we accept labor unions, unemployment insurance, the regulation of the financial practices of corporations, and many other things which were not known in the economy of a hundred years ago and would then probably have been heartily disapproved. Naturally, developments like these affect the operations of the economy, and any proposals for improving the machinery of the economy as a whole must take these new institutions into account. We cannot simply wish them out of existence. Although our economy is dynamic and shifting, those features of it which have changed have not usually done so at the same rate or even at the same time. For instance, in the nineteenth century population was growing very quickly and methods of production lagged behind, whereas today the growth of population is relatively slow and technical advance has speeded up immensely.

The economy has grown very rapidly in wealth. Our immense stock of capital goods and the highly developed skills of our labor force enable us to produce several times as much as we could even

forty years ago. This has two consequences, one favorable, the other harmful. The welcome result is that we can avoid poverty if we can keep the economy in good working order. With full employment in 1946, the average family could enjoy an income of about $5000 a year. Already we can produce enough to provide every family in the country with a decent standard of living. If our ability to produce continues to increase at the rate of the last twenty years, and if we are able to maintain full employment, the average family should by 1970 have an income of nearly $10,000 a year. We can banish poverty right now, if — and unfortunately it is a big if — we can avoid unemployment and depression. Our great wealth, and as a result the immensity of our potential output, is enough to permit us all to have more than a decent standard of living.

But our great wealth also raises an unwelcome problem; for, as we have seen, it makes the attainment of full employment more difficult, and these difficulties appear to be growing with frightening speed. Other things being equal, growing wealth itself implies fewer outlets for private investment. But even more important, growing wealth implies that higher and higher investment is needed to provide jobs for all. With the wealth at our command in 1929, about $30 billion worth of investment would have given us full employment. By 1949 we shall require approximately $65 billion worth of investment projects in order to provide full employment.[1] By 1970 the figure will be well over $100 billion unless the propensity to consume has meanwhile been very greatly increased. Obviously this is a most significant development, for it means that there will be depression, heavy unemployment, and avoidable poverty unless more and more investment outlets become available as the years pass. The situation we now face is unique in two respects. The amount of investment now needed for full employment is immense, and it is growing very rapidly. To prescribe for our economic ailments without considering these facts is to prescribe for only half of the disease.

The fundamental change in the structure of our economy has given rise to new economic problems. Indeed, one problem we face today is more profoundly difficult and disturbing than any that have fretted us at earlier stages in our national history. A century ago our economy was a relatively static organization in which small firms pre-

[1] Unless the propensity to consume should be a good deal higher than it was between 1920 and 1940.

dominated, production was largely for use rather than for sale, and the wage earner was a rare bird. Such an economy could not have a high standard of living, but neither could it have severe depression.[1] Today the situation is quite different. A relatively small number of very large firms produce most of our output and hire most of the people engaged in production. Great wealth has been accumulated, but now and in future the rate of accumulation must be ever more rapid if the economy is to avoid serious depression. We *can* enjoy a king's lot of goods and services, but we *are liable*, if depression is not prevented, to live very poorly — and, ironically enough, in the midst of idle factories whose products we should at such times sorely need. The situation is obviously far too serious to allow us the luxury of substituting slogans for analysis. We must be prepared to accept new ways of doing things as well as old, for the problems we face are new and alarming.

Once more: as members of one interest group or another — wage earners, businessmen, farmers, and Americans — we are presented with two different ways of improving our economic lot. We can enjoy a higher real income by getting a larger slice of the economic pie at the expense of someone else. But that is not the only way. Whenever the economy is operating below peak capacity and maximum efficiency — and it has almost *always* done so, even during war, let alone in normal times — our economic position can be improved without injuring anyone else. For we can increase the size of the pie. To illustrate: labor can gain if, when wages rise, profits fall; it can also gain if employment and the national income are raised — and if this happens profits will go up, too. It is not necessary that one group suffer for another to become better off. Nor is it necessary for one economy to improve its position by reducing income elsewhere. An economy, just like any interest group in the economy, can grow in prosperity and wealth without injuring any other.

It is unfortunate that the ways by which the whole pie can be made larger are not available to any single group within the economy. Labor or farmers by themselves cannot do a great deal to create prosperity. To do so takes considered, and concerted, action by the whole economy.[2] That is a fact which simply must be faced. But

[1] It could, of course, be impoverished by crop failures or other natural disasters, but that is a different kind of ailment.

[2] Or fortunate developments such as inventions.

though a single interest group can do little to create prosperity, prosperity affects each group. If we can maintain the total income at peak level, we shall automatically reduce the pressure which the various interest groups bring to bear upon each other. For if labor, capital, the farmer — in other words, everyone — gets more of what he wants, inevitably the friction between groups will be eased and the conflict will be less intense. Likewise, economic warfare between groups will be less bitter if all of them can enjoy a high level of prosperity.

While it may be legitimate for each group to try to improve its lot at the expense of competing groups, the struggle to do so becomes anti-social when it causes a reduction in the total output of the economy. The damage done by such a struggle can be most successfully prevented, not by legislative restraint, but by society's adopting measures to keep the total output as high as possible, its composition as nearly ideal as possible, and the distribution as fair as possible. When that is done, the economic problem will be solved.