UNIVERSAL
LIBRARY

OU_154456

UNIVERSAL
LIBRARY

About the book

Professor Boulding's approach to the subject in *A Reconstruction of Economics* represents his attempt to overcome two maior weaknesses of present

other assets are circulated, and the processes of production, consumption, income, and outgo.

In Part II, Professor Boulding attempts to rectify these defects and to present a system of macrocconomic analysis which is relatively free from conceptual confusions. The central concepts of this system are the balance sheet and the balance sheet identity, the same concepts on which the author bases his theory of the firm in Part I.

OSMANIA UNIVERSITY LIBRARY

Call No. 330.1 B76 R Accession No. 31664

Author Boulding.

Title Re-construction of Economics.

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

A Reconstruction of Economics

A Reconstruction of Economics

KENNETH E. BOULDING
Department of Economics
University of Michigan

154456

JOHN WILEY & SONS, INC., NEW YORK
CHAPMAN & HALL, LIMITED, LONDON

Copyright, 1950, by John Wiley & Sons, Inc.

All rights reserved. This book or any part thereof must not be reproduced in any form without the written permission of the publisher.

Printed in the United States of America

To

J. R. B.

Preface

The principal advantage of a preface is that it gives the author a chance to review his own book, a privilege usually denied him by the journals A preface, therefore, should fulfil the function of a good review, in that it should give the reader—or even the occasional reviewer —enough of an idea of what is in the book so that he may decide for himself whether it is worth the trouble of reading It also gives the author himself a chance at a few afterthoughts.

This work is directed mainly at serious students of economics who have already had some moderate exposure to the subject It is not therefore a complete and exhaustive treatise, though most of the topics of pure theory are at least mentioned in it. Its title is perhaps too pretentious—mainly because I have not been able to think of a good substitute It is not particularly revolutionary; the building blocks are arranged rather differently, but the student will find most of the old blocks here. Nevertheless it is the product of my growing dissatisfaction with the present state of economic theory as generally received and taught, and an attempt to establish some patterns of theory which will be more consistent conceptually and more useful in interpretation than the existing corpus of doctrine.

This dissatisfaction has at least three focal points The first is the failure of economics to integrate itself into the general body of social science. The outstanding success of what may still be called the "Ames school" in agricultural economics has been somewhat facetiously attributed to the discovery that there was no such subject—that there was only economics applied to agricultural problems. In somewhat similar vein I have been gradually coming under the conviction, disturbing for a professional theorist, that there is no such thing as economics—there is only social science applied to economic problems. Indeed, there may not even be such a thing as social science—there may only be general science applied to the problems of society. We are, I believe, on the threshold of a new attempt at integration in the social sciences, perhaps even in general science. Past failures in this respect—notably those of the Spencerians, the Marxists, and the Institutionalists, who integrated a number of specialized errors into erroneous generalities,

should not discourage us. The integration is clearly coming, not by the erection of a vast and tenuous superdiscipline, but by the reaching out of all the specialized disciplines towards one another. Just as the physical sciences developed physical chemistry, and later biochemistry and biophysics when the time was ripe, so the social sciences are developing social psychology, and show promise of developing economic sociology and even perhaps political economy! The sensitive specialist in the pursuit of his problems everywhere finds himself on the frontiers of other disciplines. This is as true of economics as of any other discipline, and in Chapters 1 and 2 I have tried to outline the general setting in which economics, as a separate discipline, lies. It consists, like all theory in whatever subject, of two parts: a theory of the individual and a theory of interaction. All science studies individuals of some kind, whether electrons, atoms, molecules, cells, mice, men, families, firms, churches, or states. All science studies the interaction of individuals and populations of individuals one with another. In Chapter 1 therefore I have sketched the outline of a theory of interaction, and in Chapter 2 a theory of the individual. These two chapters are somewhat separate from the main body of the work and are not particularly complete in themselves: I have nevertheless included them because of the importance of the problem with which they deal.

The rest of the work is for the most part very pure economics. I am convinced that before any adequate integration can be performed the various separate disciplines must be improved: we cannot expect the products of integration to be any better than what is integrated. The first focus of my dissatisfaction with economics is in the theory of the firm, or the economic organism, and its immediate relationships and interactions. This discussion occupies Chapters 3–9. My main cause for dissatisfaction with the existing theory of the firm lies in its deficiency in capital theory. The usual marginal analysis treats the firm as if it had nothing but an income account; it has no balance sheet, no capital problems, and no dynamics; it maximizes an odd variable called "net revenue," of which no accountant ever heard, and presumably lives happily ever after. Consequently the economist has been able to give only fragmentary accounts of inventory changes, investment structures, liquidity positions and like problems. Only in the discussions of banking and corporation finance do we find economists considering the firm as if it had a balance sheet, and these discussions are in no way related to the theory of the firm as it appears in price and value theory. What I have tried to do, therefore, is to introduce the balance sheet as the central analytical concept of the theory of the firm, and to introduce asset preferences, especially as measured by preferred asset ratios, which

hitherto have been confined almost exclusively to banking theory, as a central concept in explaining the behavior of all economic organisms. The asset preference concepts also turn out to be very useful in explaining the nature of price and market equilibria in general, at the microeconomic level. The older theory of "maximizing profits" is seen clearly as a special case of the more general theory of preference, valid only when markets for the individual are perfect. It should be understood, however, that asset preference theory is not a departure from the general principle of maximizing "utility": it merely represents a useful, if rather arbitrary, simplification of the complex preference functions. Much of what the reader will find in these chapters is a restatement and generalization of the old conclusions, which almost always appear as a special case of the more general theory. In one respect, however, my approach represents a real conceptual revolution which the reader trained in the older theory may find difficult to accept: this is in regard to the concept of consumption. Once the emphasis is laid on assets rather than on income it becomes clear that there is a vital distinction between the enjoyment of assets and their consumption—that is, destruction. Consumption, and therefore production and income, are then seen as quantities to be minimized rather than to be maximized in the interests of maximum enjoyment.

The second focus of my dissatisfaction with economics is the condition of the Keynesian "macroeconomics." Although I should certainly accept the title of a Keynesian and, with all modern economists, owe an enormous debt to Keynes's brilliance of insight and imaginative sweep, his system shows a number of weaknesses which have not been corrected by his followers. One of these weaknesses is a general failure to distinguish between two very different processes in economic life, the exchange or payments process on the one hand, by which existing assets, including money, are circulated among various owners, and the processes of production, consumption, income and outgo on the other, by which assets are created, destroyed, and accumulated. These two processes are, of course, related, but are entirely different, and it serves no good purpose to confuse them, as Keynes does, for instance, when he identifies "consumption" with "consumers' expenditure." In Chapter 12, therefore, I develop the theory of payments, or the circulation of existing assets, and in Chapter 13 develop a separate theory of output as a whole and of employment.

Another weakness of the Keynesian economics is its failure to provide any "macroeconomic" theory of distribution commensurate with its theory of employment. This deficiency I have endeavored to correct in Chapter 14, which is perhaps the most original and controversial

section of the work. It is based solidly on the concept of the profit-making process as the gross growth in the value of net worth developed at the microeconomic level in Part 1. The balance sheet is again the basic concept: from the separate summation of business and household balance sheets, identities for gross profits (non-labor income) and for the total wage bill are derived, the component parts of which are in some degree "parameters of behavior" and which can therefore be used to construct simple models of economic equilibrium. The distribution of national income between wages and gross profits is then seen to be determined by several distinct, though related, sets of decisions; gross profits are greater the more business increases its stocks of real capital (investment decisions), the more it finances investment out of profits, the more it distributes in interest, rent, and dividends, and the more consumer credit it extends (the financial decisions) and the more the money stock shifts out of household into business accounts (the liquidity decisions). The total (real) wage bill is greater the greater is the volume of household purchases, the more the money stock shifts out of business into household accounts, the less is the increase in consumer credit and in the sale of securities to households, and the less is distributed to capitalists in rent, interest, and dividends. The distributional structure is thus seen to be the result not of wage bargaining (except in so far as wage bargains react on the prime determinants) but of various decisions in regard to investment, consumption, finance, and liquidity.

These conclusions sound so strange that I expect, and indeed have already encountered, a good deal of resistance to their common acceptance, especially as they seem to introduce an element of economic indeterminacy into the distributional pattern which the current marginal productivity and general equilibrium theory apparently denies. I am convinced, however, that this macroeconomic approach to the theory of distribution is absolutely necessary. The marginal productivity theory, even modified to take care of liquidity and other asset preferences, is strictly valid only at the microeconomic level. The "Walrasian" theory of general equilibrium assumes that distribution is determined by the processes of exchange through the pricing system, determining the price of factors of production. What it overlooks is that payments to capitalists are in the nature of transfers rather than of exchanges. This is particularly true of dividends, which represent, from the point of view of the balance sheet, a simple reduction in assets, whereas an exchange represents a transfer of assets without reduction in their total value. This fact is recognized in national income accounting in the case of interest on the national debt: it should be equally recognizable in the case of private interest and dividend payments. It

is this fact which introduces the apparent indeterminacy into the distributional structure.

In spite of the strangeness of some of the conclusions, therefore, both in the theory of distribution and in the closely related theory of taxation which is developed in Chapter 16, I believe the reader who approaches the argument without prejudice and with some sympathy can hardly fail to be convinced of its truth. The theory I am propounding represents a return to a more "classical" type of theory. In some sense it represents a refinement and rehabilitation of the wage-fund theory, in that it emphasizes the significance of the decisions of capitalists in regard to the disposal of their capital and liberates the laws of distribution from the physical determinism of the laws of production. In this it is closer to Mill than to Marshall, and represents an attempt to set the classical "macroeconomics" on a firmer foundation. The very success of the Marshallian "microeconomics" diverted the attention of economists for two generations from some of the vital problems in macroeconomics which the classical economists raised but did not succeed in solving; I would not be so presumptuous as to claim to have solved them, but I hope at least to call attention to their existence.

My thanks are due to Iowa State College, under whose auspices this work was begun, and to the University of Michigan, under whose auspices it was completed. Thanks are also due to my friends, colleagues, and students, who have suffered patiently under my ideas and have contributed most helpfully to their development. I am particularly indebted to Dr. Albert G. Hart, whose penetrating criticisms, many of which are still valid, have delayed publication, I hope fruitfully. This is a work of ideas rather than of scholarship, and I confess to both negligence and ignorance in the inadequate acknowledgment of the many sources of my ideas. The influence of J. M. Keynes is almost too obvious to mention; I should also mention the work of J. Reinold Noyes, whose *Economic Man* has affected my thinking in ways unintended by its author. Finally, especial thanks are due to my wife, who has borne patiently with the domestic strains of intellectual labor, and by her work as a sociologist has contributed in no small measure to the broadening of a specialist in economics.

KENNETH E. BOULDING

Ann Arbor, Michigan
July, 1950

Contents

Part 1· Microeconomics

CHAPTER

1. An Ecological Introduction 3
2. The Individual Social and Economic Organism 26
3. The Theory of the Economic Organism 39
4. The Marketer in an Imperfect Market 60
5. The General Theory of Preference 79
6. The Theory of Production and Inventory 95
7. Expectations, Gambling, and Insurance 117
8. The Theory of Consumption 135
9. The Equilibrium of Production and Consumption 155

Part 2· Macroeconomics

10. The Nature and Significance of Economic Aggregates 171
11. The Structure of Macroeconomic Aggregates: I. Capital and Income 189
12. The Structure of Macroeconomic Aggregates: II. The Theory of Payments 206
13. The Theory of Output and Employment 229
14. A Macroeconomic Theory of Distribution 243
15. Money, Debt, and Interest 270
16. The Economics of Government 284
17. A Concluding Note 303

Index 309

Part 1

Microeconomics

Chapter 1

An Ecological Introduction

The background of economics

The subject of study known as economics falls conveniently into two parts, which have generally come to be called "microeconomics" and "macroeconomics." Microeconomics is the study of particular economic organisms and their interaction, and of particular economic quantities and their determination. It is built up first of a theory of the individual economic organism (firm or household), the object of which is to determine certain functional relationships between the environment of the organism (those quantities or functions which are given to it and over which it has no control) and its "behavior" (those quantities over which it does have control). By developing a theory of the interaction of organisms through exchange microeconomics also develops a theory of the determination of the main quantities of the system—prices, outputs, consumptions, and so on.

Macroeconomics, which forms the subject matter of the second part of this book, deals not with individual quantities as such, but with aggregates of these quantities—not with individual incomes but with the national [1] income, not with individual prices but with price levels, not with individual outputs but with the national output.

It must not be thought, because of the difference in their methods, that macro- and microeconomics study different phenomena. They are merely different ways of studying the same set of phenomena. Underlying both of them is the real world, from which both are abstracted. It must never be forgotten that economics itself is an abstraction, useful and necessary as it is. We cannot divide the universe of organisms and quantities into those called "economic" (e.g. firms) and those called "non-economic" (churches, states) and so develop a theory of economic life confined to "economic" organisms. All social organisms, in so far as they participate in processes of the production, consumption, and exchange of assets, have economic aspects. Families buy commodities

[1] The word "national" in this connection merely indicates a large aggregate: it is not confined to national states, but may refer to any region or to the world as a whole.

and sell services; churches hire labor and buy equipment. Labor unions negotiate the conditions under which labor is sold. Governments indulge in almost every conceivable kind of economic activity. We must develop a concept of economic life, therefore, not as a *segment* but as an *aspect* of the whole complex of social relationships. This does not mean, of course, that we must refrain from constructing strictly economic models. Nevertheless these economic models should be constructed as special cases of general social models if their true significance is to be assessed.

What this means is that the distinction between the "exogenous" and the "endogenous," useful and necessary as it is in constructing any kind of abstract system, must be used deliberately as a tool for creating a first-approximation analysis and must not serve as an excuse for intellectual laziness. Too often the system-maker (the economist is not the only delinquent here) uses the convenient distinction between the endogenous factors which constitute his system proper and the exogenous factors which operate as disturbing forces from the "outside" as an excuse for not thinking at all about the inconvenient "exogenous" factors. So often in economic discussions, for instance, we find the statement that such and such a policy is "economically" advisable but "politically" impossible used with an air of moral rectitude, as if the "economical" were righteous and the "political" sinful. In fact, however, such a statement should have no more or fewer moral implications than the statement that a certain policy is agriculturally desirable but industrially undesirable. Especially in the consideration of policy it is not legitimate to separate the economic from the political aspects of society with an air of separating the sheep from the goats. The forces which operate to determine the action of government are not wholly dissimilar from the forces which operate to determine the action of a firm, and there is no sense in having an elaborate theory of the behavior and interaction of firms and at the same time regarding government as a *deus ex machina*, capriciously descending at unpredictable moments to confuse—or unravel—the endogenous threads of the purely economic drama. I am not, of course, denying the usefulness of constructing partial systems: indeed, without the ability to construct partially endogenous systems the analysis of social life would be impossible. I am merely pleading for a recognition of the fractional nature of any such model, and for a habit of mind which will enable the specialist to be at least dimly conscious of the other pieces in the intellectual jigsaw puzzle.

It is not the purpose of the present work to present an integrated analysis of society. Such an analysis should be possible. It should be

possible to develop a continuum of theoretical models which will include, say, the state or the church at one end, and proceed through the college, the labor union, the farm organization, and so on to the business corporation and the household. It should be possible, also, to construct models of the interaction of all these organisms in the determination of all the social variables. Such an analysis, however, must wait for the future. A synthesis of inadequate parts may be worse than no synthesis at all,[2] and it is the main task of this volume to prepare the ground for the coming synthesis by attempting to improve the abstraction known as "economics." Even the economist, however, should have some notions as to the setting of his subject, and as to the basic reality from which his abstraction is made. The present chapter, therefore, while somewhat in the nature of an excursion from the main task of the book, is an attempt to develop a concept of the structure of reality, and in particular of the structure of social reality, which may serve as a background to the later abstractions of pure economics.

Reality as an ecosystem

Reality, in its quantitative aspect, must be considered as a system of *populations*, a population being a number of items each of which conforms to a common definition, and including all the items conforming to that definition. We need not discuss the philosophical question whether there are "quanta" or atoms of every population, or whether there is real continuity in the universe: even if there are continuous populations made up of infinite numbers of indefinitely small units, these can be quantified in terms of arbitrary units which can be regarded as "quanta." Thus the total stock of wheat can be described as so many "tons," the "ton" being regarded as a quantum or individual of the population of wheat, even though it is an arbitrary quantity. A ton of wheat, however, is usually homogeneous enough so that its structure—at least for purposes of economics—can be neglected.

Populations act and react upon each other, and the equilibrium size of any given population is a function of the sizes of all others. The general study of the equilibria and dynamics of populations seems to have no name; but as it has probably reached its highest development in the biological study known as "ecology," this name may well be given to it, especially as sociologists have already appropriated the name for social phenomena, though often in a rather narrow geographical sense. In

[2] "Institutionalism" in economics may be regarded as a premature attempt at synthesis of the social sciences, an attempt to synthesize bad economics, bad sociology, and bad anthropology in a medium of subconscious emotional bias.

economic literature the classic expression of the concept is Marshall's comparison of the firms of an industry to the trees of the forest.[3] The Walrasian general equilibrium, however, is essentially a static ecological equilibrium—the idea of mutual determination being fundamental in ecological systems.

The forest, however, does not consist only of trees. It is a complex pattern of organisms, trees, grasses, flowers, birds, mammals, insects, reptiles, bacteria; subsisting, growing, propagating, dying in a maze of complementary and competitive relationships, all founded on the physical environment of earth, air, sun, water. Everything eats and is eaten, breathes and is breathed, competes and complements, meets competition and receives aid. Every organism has a law of growth and of survival of its own; the size or existence of any particular individual is a result of its inner law of growth and life pressing against a complex hostile-friendly external environment, consisting partly of its inorganic environment and partly of the other organisms with which it comes into contact. Organisms may be classified into species, though it is important to note that this classification is not essential to the concept of an ecological system: it would be quite possible to have a system in which each individual, like Dante's angels, constituted a species in itself, or in which the individuals constituted a continuum of characteristics so that no classification could be more than arbitrary. One wonders sometimes whether taxonomy is not more a habit of the human mind rather than a property of nature; as taxonomy is necessary for the existence of language (for we could not have a different name for every individual blade of grass!) we are strongly predisposed towards it, and perhaps sometimes attempt classifications which are too arbitrary to be useful. It may be, for instance, that the concept of an "industry," which in economics corresponds somewhat to that of a "species" in biology, is a classification which exists more in conversation than in reality. Not, of course, that we should despise or reject aids to conversation; we should merely be on our guard against granting them more respect than they deserve.

It must be emphasized that in calling society an ecological system we are not merely using an analogy; society is an *example* of a general concept of an "ecosystem," that is, an ecological system, of which biological systems—forests, fields, swamps—are other examples. It is instructive, therefore, to consider the differences between social and biological ecosystems. The most important difference is a genetic one: there is very little in social ecology which corresponds to the production of new

[3] A. Marshall, *Principles of Economics*, 8th Ed. (Macmillan), p. 457.

organisms by sexual reproduction. There are social phenomena somewhat akin to cross fertilization—Liberia might perhaps be described as by American idealism out of the African jungle—but they occur usually at more complex levels than the individual social organism; the sub-ecosystems that we call cultures sometimes interact to produce a hybrid, but we do not find the First Baptist Church having intercourse with the Second to produce a Third. The genetic processes in the social system are perhaps somewhat more akin to asexual reproduction—the Second Baptist Church may easily be a split from the First—but in general, social genetic processes seem to be more complex than any genetic processes in biology. This seems to be a result of what is perhaps the most basic difference between social and biological ecosystems: that in a human society the "atoms" which comprise the organisms are individuals whose ends the social organisms are at least supposed to serve, whereas in biological ecosystems the ultimate atoms are physicochemical and presumably indifferent to the organism which includes them, though with affinities and relationships of their own which in part determine the organism's power of survival.

The "genes" of social organisms are ideas in the minds of men, and ideas have the remarkable property of propagating by mere contact. When *A* "gives" *B* an idea *A* does not give it up as he would if he gave a commodity to *B*: as a result of the contact two people have the idea instead of one! One may perhaps liken the missionaries and propagandists of society to spermatazoa, specialized carriers of the genes of churches, parties, cooperatives, lodges, colleges and the like, but specialization in this respect does not seem to go as far in social as it does at least in the higher forms of biological life. There is clearly a social phenomenon corresponding to mutation: the person who conceives a new idea with propagating power—the founders of religions, sects, and faiths; even perhaps in a minor way the philosophers and scientists. Vast organizations have sprung from an idea in the mind of one man; as someone has said, all organizations are the lengthened shadows of their founders.

The study of social life might well begin, then, with a consideration of the nature of an ecosystem.[4] An ecosystem may be defined as one in which there are a number of populations, each of which consists of in-

[4] It is not proposed to go very far into the mathematics of ecological systems: Vito Volterra, in his *Leçons sur la théorie mathématique de la lutte pour la vie* (Paris, 1931) has developed an elaborate theory for biological populations, without, apparently, realizing its more general applications. The present chapter is intended to do little more than indicate the broad lines on which such a general theory might be constructed.

dividuals which are identical for purposes of this level of abstraction, i.e., individuals whose differences can be neglected. The size of each population is measured by the number of individuals in it, and the *equilibrium* size of each population is assumed to be a function of the

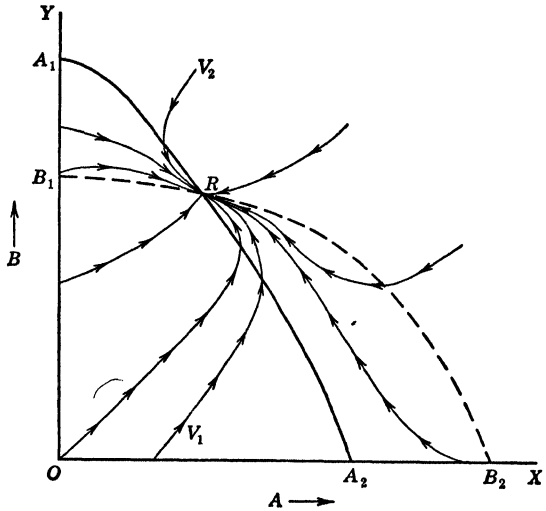


FIG. 1. Equilibrium of two mutually competitive populations.

sizes of all the other populations. Thus, if $a, b, c, \cdots, z$ represent the equilibrium size of the various populations, we have

$$\begin{aligned} a &= F_a(b, c, \cdots, z) \\ b &= F_b(a, c, \cdots, z) \\ &\cdot \\ &\cdot \\ &\cdot \\ z &= F_z(a, b, \cdots) \end{aligned} \tag{1}$$

This is a familiar "general equilibrium" situation with n equations and n unknowns, and clearly a solution is possible, where the size of each population is consistent with the size of every other. This solution need not, of course, be unique, and there may be cases in which there is no solution within the significant range of the variables concerned.

We may begin to study the properties of such a system by considering a system of two populations only, A and B. In Fig. 1 we measure the population of A along OX, the population of B along OY. The solid line A_1A_2 is the A-curve, which shows what would be the equilibrium

population of A for each population of B. In Fig. 1 we have assumed that the relationship between the two populations is a competitive one, so that the less the B population, the more chance A will have to thrive and the greater the A-population. Similarly we have a B-curve—the dotted curve B_1B_2—which likewise assumes a competitive relationship. It is clear that the point of intersection of these two curves is one point

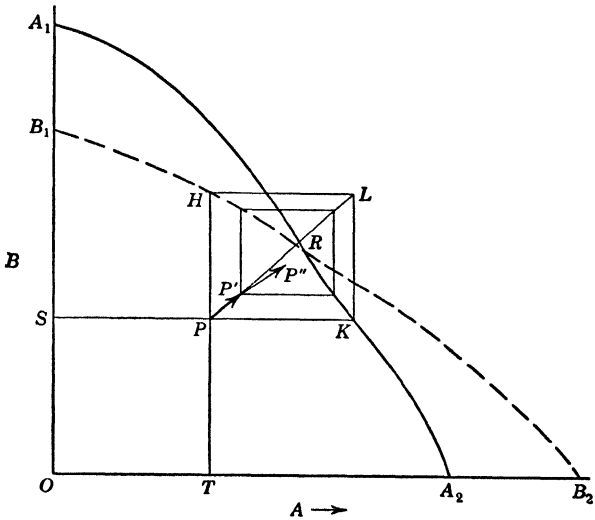


FIG. 2. The construction of vector lines.

of equilibrium, for at this point the two populations are consistent—i.e. each is at a level which is consistent with the level of the other.

The dynamics of ecological equilibrium

It is necessary, of course, to investigate the stability of this equilibrium, and in order to do this it is necessary to discuss the dynamics of the situation. Suppose, for instance, that the point P in Fig. 2 represents the sizes of the two populations at time zero, there being OT of A and OS of B. Neither of these is an equilibrium population: by projecting the horizontal line SP to meet the A-curve A_1A_2 in K, we find the equilibrium population of A (SK) corresponding to the B-population OS; projecting the vertical line TP to meet the B-curve B_1B_2 in H gives us the equilibrium population of B, TH, corresponding to the A-population OT. In the diagram it is clear that from the point P both populations will increase, as both are below their equilibrium levels. If we suppose that the extent of the increase is proportionate to the difference between

the actual and the equilibrium population, i.e. to PK and PH respectively, the resultant movement of the two populations can be represented by a shift along the line PL. If we suppose that this persists for a unit time period, it brings us to P': the movement from P to P' may be called a "vector movement." Starting from P', we can derive a similar vector movement in the next time period to P'', and so could trace the whole course of the two populations from the point P. The unit time period may be made indefinitely small, in which case the vector lines become continuous curves of infinitesimal adjustment. In some problems, however, we may wish to assume that an initial impulse persists for a finite time, after which a new impulse is received: in that case the vector lines will be discontinuous as in the figure.

Whatever the assumption, the vector lines both provide a statement of the potential dynamics of the situation and also give a clear picture of the stability of the equilibrium: a stable point of equilibrium is one towards which vector lines converge from all sides. It should be observed that as long as *some* functional relationship between the rate of change of population and the divergence of the actual from the equilibrium population is postulated, vector lines can be drawn; we do not have to confine ourselves to the limited assumption of the construction in Fig. 2. If, for instance, population A was more responsive to a disequilibrium situation than population B, the vector line at P would have a smaller slope than PP', as A would increase faster than B under the same stimulus. Unless, however, there is a "perverse" relationship between the excess or deficiency of equilibrium over actual population, so that when the actual population was less than the equilibrium population the population would fall, and when greater would rise, the vector line must lie between PH and PK.

In Fig. 1 some of the vector lines are drawn. It will be observed that vector lines will never cross one another, for that would mean that two different sets of population changes would be possible at the same point. In Fig. 1 it is evident that the equilibrium at point R is stable, for all the vector lines converge to that point: no matter from what populations we start, the dynamic course of the populations will end up at R. It will be observed that wherever the vector lines cut the B-curve they are parallel to the A-axis, as B is in equilibrium then, and only A will be increasing. Similarly when the vector lines cut the A-curve they are parallel to the B-axis. It follows that if a vector line cuts the A-curve, as the line V_1 in Fig. 1, in the course of moving to the equilibrium point from a low level, the A-population may first rise, reach a maximum beyond the equilibrium population, and then fall. In moving to the equilibrium point from a high level the population may first fall, reach a

minimum below the equilibrium level, and then rise. A similar proposition holds for B.

In Fig. 3 we exhibit a very interesting case of unstable equilibrium. The point R where the two curves intersect (the B-curve now cutting the A-curve from above instead of from below as in Fig. 1) is a point of unstable equilibrium, as the vector lines clearly show. The points B_1 and A_2 are points of stable equilibrium. In this case the field is divided

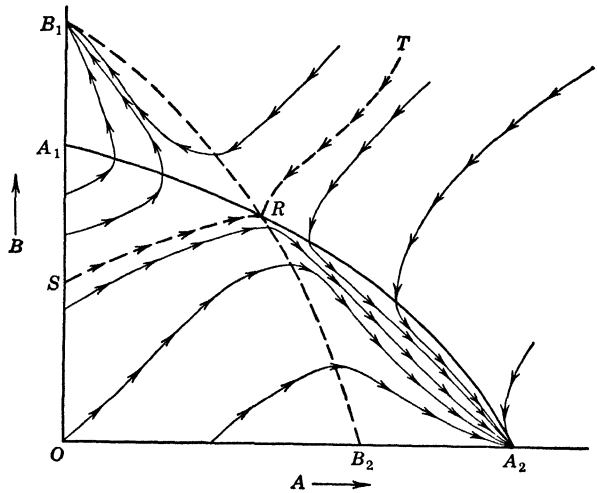


FIG. 3. Unstable equilibrium of two mutually competitive populations.

by a line SRT, which is a line of unstable equilibrium. From any point on the line SR the vector line leads along SR to R: similarly at any point on TR the vector line leads along TR to R. From any point upward and to the left of SRT, however, the vector lines all lead to B_1. In a system starting from any such point, population A is eliminated entirely, and the maximum B-population, OB_1, occupies the field in lonely majesty. From any point below and to the right of SRT all the vector lines lead to A_2, where B is eliminated and A holds undisputed possession. This clearly corresponds to many important situations both in biological and in social ecology. Grass eliminates trees entirely on the Great Plains; trees eliminate grass in the dense forests. Doctors eliminate magicians in western culture; magicians eliminate doctors among the Dobuans. It is evident that even though the points on SRT represent an equilibrium, it is highly unstable: the slightest deviation will send the system off towards either B_1 or A_2.

In Fig. 4 a stable equilibrium is exhibited in which the relationship between the two populations is complementary rather than competitive. Thus the A-curve A_1A_2 shows that a certain population of A (OA_1) will exist in the absence of B, but that the more B there is, the more A can exist: similarly a population OB_1 would exist in the absence of A, but the more A there is, the more B. It might be thought offhand that this situation would lead to indefinitely rising populations, but the diagram

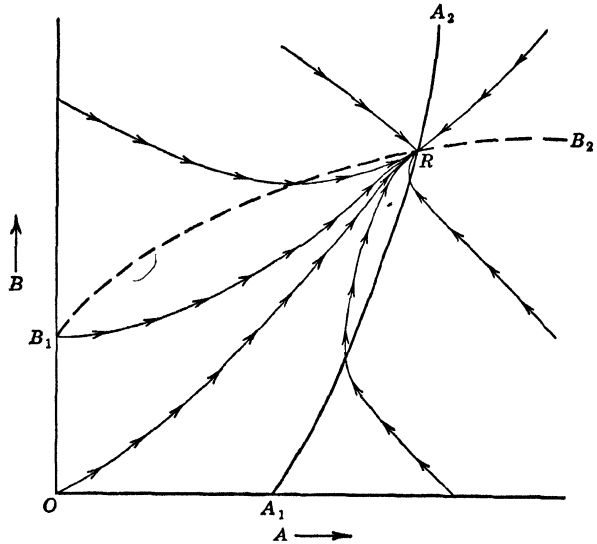


FIG. 4. Stable equilibrium of two mutually cooperating populations.

shows clearly that such need not be the situation: if the curves intersect, as they do at R, the vector lines indicate that this is a position of stable equilibrium.

An interesting case is shown in Fig. 5, where the A- and B-curves exhibit maxima, indicating that the two populations are complementary at lower values and competitive at higher values. In the figure there is a point of unstable equilibrium in the range of complementarity at R: below the line SRT *both* populations die out, and a stable equilibrium is reached only at O, where both populations are zero. Anywhere above the line SRT the populations move towards a stable equilibrium at R'.

The case illustrated in Fig. 6 is interesting, in that it presents an example of "one-sided instability." The A- and B-curves touch each other at R. Below R the equilibrium is unstable: the vector lines all move down to O. Above R the vector lines move towards R. If there is any

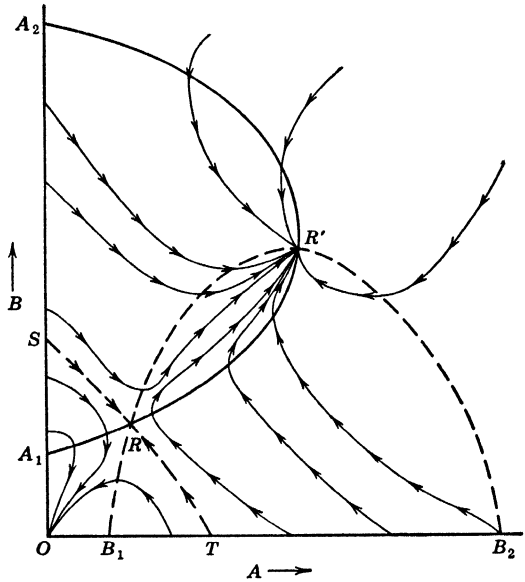


FIG. 5. Equilibria of two populations exhibiting both competition and cooperation.

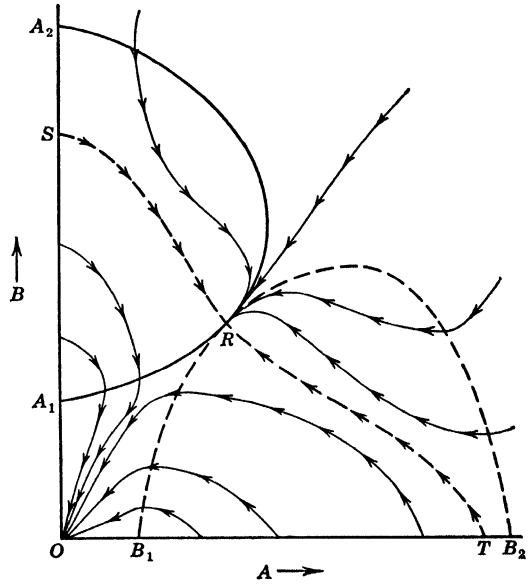


FIG. 6. Equilibrium with one-sided instability.

instability in an equilibrium, however, it must be regarded as unstable: in Fig. 6, although a movement upwards from the equilibrium line *SRT* will result in a path towards *R*, a movement downwards from *SRT* results in a path away from *R*.

Another very interesting case is that of "parasitism," where the "parasite" (*A*) competes with the "host" (*B*) but where the host is complementary with the parasite: the larger the population of parasites, the smaller the population of hosts; the larger the population of hosts, the larger the population of parasites. This situation is shown in Figs. 7, 8 and 9. In Fig. 7 we assume that the parasite is entirely dependent on the host, so that with a zero population of hosts there will be a zero population of parasites: as the host (*B*) population increases, the parasite population increases according to the solid *A*-curve; with no parasites the host population would be *OH*, and with increasing numbers of parasites the host population declines along the *B*-curve. An equilibrium is reached at *R*. The vector lines indicate that this equilibrium is very stable, though there may be spiral paths towards the equilibrium point. The equilibrium not only is stable but is also not very "shiftable": small changes in the *A*- or *B*-functions will not change the equilibrium position very much.

Interesting special cases are shown in Figs. 8 and 9. In Fig. 8 the parasite *can* exist without the host, and even when the host population is zero the parasite population (*OK*) is so great that it does not permit any population of hosts. As the vector lines show, in this situation the host population will be exterminated and the parasites left in possession of the field. In Fig. 9 it is assumed that the parasites must have a certain host population (*OM*) before they can exist at all. If *OM* is greater than *OL*, the maximum host population, it can be seen again from the vector lines that equilibrium will be reached at *OL* with the maximum host population, and that the parasite will be exterminated. It is interesting to observe that in this, as in some other cases, frequently the vector line shows that a population which is destined for ultimate extinction may rise for a considerable way to reach a maximum from which it falls to zero. A rising population at a given moment, therefore, is no necessary criterion of ultimate success. The rise and fall of the Nazis may be an example of precisely this phenomenon.

The "shiftability" of ecological equilibrium

Some consequences of interest can be drawn even from the very simple two-population system. In the first place it is clear that where competitive relations exist between the populations—i.e. where both the *A*-curve and the *B*-curve have a negative slope—the equilibrium posi-

tion, even if stable, may be subject to large changes even if the shift in the population functions is small. Thus in Fig. 10 a slight shift in the B-curve from BB to $B'B'$ shifts the equilibrium from R, with a lot of B and a little of A, to R', with a lot of A and a little of B. Situations can

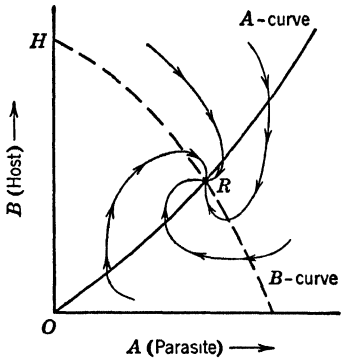


FIG. 7. Parasitic equilibrium.

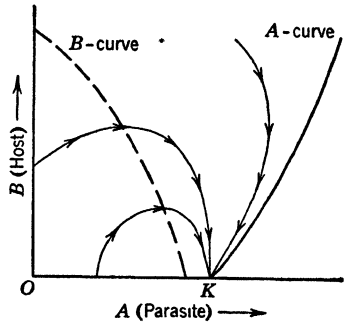


FIG. 8. Parasitic elimination of the host.

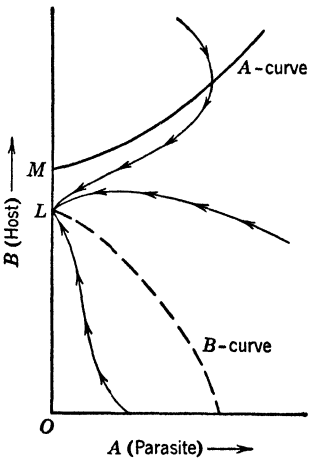


FIG. 9. Parasitism: elimination of the parasite.

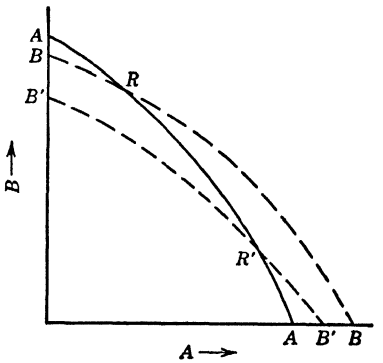


FIG. 10. Shiftability of equilibrium.

easily be pictured in which a very small worsening of the "strength" of a population, as measured by the distance of its curve from the origin, leads from relative success to total extinction: conversely a small improvement in conditions may cause an enormous growth of a particular

population. Several cases, both biological and social, come to mind. Rabbits, which are pleasant small game in England, become devastating pests in Australia; political parties, whose natural enemies are strong and whose vitality is low in the apathetic and smug political climate of the Anglo-Saxon world, increase like Australian rabbits in the genial atmosphere of European political life and threaten to make democracy unworkable. The control of one set of pests often produces a worse one: starlings, brought in perhaps to control the English sparrows, became a worse pest than the sparrows themselves. The suppression of saloons by the political weed-spray of prohibition led to a deplorable proliferation of the more police-immune speakeasies and gangs. The social reformer, like the biological reformer, needs to have some idea of the ecological system with which he is dealing; otherwise his reforms may have unexpected and undesired consequences. This is not to say, of course, that we have no business meddling with ecological systems: indeed, all conscious human activity involves such interference. The production of a lawn or of a wheatfield is the result of deliberate human interference with the ecology of nature; the whole political and social system is an attempt to do for society what a farmer does for a field: to add certain elements to the ecological system that will increase those populations which we regard as desirable (wheat, happy families) and diminish those which we regard as undesirable (tares, armies, criminals).

As we expand the system to include more and more populations, the complexity grows, and it becomes more difficult to define the conditions of stability, but the fundamental principles do not change. There will always be some position of equilibrium, even if it involves the extinction of some populations. The introduction of complementary as well as competitive populations seems likely to make the equilibrium position itself less responsive to shifts in the population functions, so that there is less likelihood of "revolutionary" changes in the equilibrium position as a result of slight changes in the functions. The chain of connection, however, becomes long and complex, and apparently unrelated populations turn out to be connected by long chains of competitive and complementary relationships. Darwin's famous example of spinsters and clover comes to mind: spinsters keep cats, cats eat mice, mice eat bees, bees fertilize clover: hence the more spinsters, the more cats, the fewer mice, the more bees, the more clover. It is even possible to have closed chains which return on themselves: the more turnips, the more cattle, the more manure, the more turnips! It may well be that these self-returning chains are of great importance in interpreting the phenomena of social and economic progress. They are, of course, simply complex cases of complementarity. The analysis of Fig. 5 indicates that there may be

certain critical levels of various populations above which they grow and below which they decline. There may, for instance, be a stable equilibrium at low levels, but if the equilibrium is disturbed sufficiently, the critical point of unstable equilibrium may be passed and the populations will grow to a high-level equilibrium. From the point of view of the theory of economic and social progress, therefore, the study of complementarity in populations may be of great value.

Determinants of the population functions: mortality and fertility

We next turn to the question of what determines the population functions—the *A*- and *B*-curves of our two-population examples. It is at this point that we have to develop a theory of the *individual*, i.e. of the unit of which the populations are composed, for clearly the explanation of the equilibrium size of the population, external conditions being given, must be derived from a knowledge of the way in which the individuals comprising the population react to changes in their own environment. What we need, therefore, is a "dynamic description" of the individual organism: by this I mean the functional relationships between the birth, growth, reproduction, and death of the individual and the environmental conditions with which it may be surrounded. If the population is to be measured only by the *number* of individuals in it (variations in the character or magnitude of the individuals being neglected), the factor determining its equilibrium level is simply the rate of births and the average length of life. In an equilibrium population, in fact, the number in the population is equal to the rate of births multiplied by the average length of life.[5] If, for instance, we have one million automobiles added to the automobile population by production each year, and if the average length of life of the automobile is ten years, then when the automobile population is in equilibrium there will be a total population of ten million automobiles. In equilibrium, of course, the birth and death rates are equal and the population stationary: in the dynamic phase of population change the relations are more complicated, and it is no longer true that the total population is equal to the rate of births[6] multiplied by the average length of life.

If now we can establish functional relations between the environmental conditions (i.e. other populations, including the inanimate populations

[5] See my article, The application of the pure theory of population to the theory of capital, *Quarterly Journal of Economics*, vol. XLVIII, p. 64, for a simple proof of this proposition.

[6] The rate of births is the total number added to the population in a given time period: it should not be confused with what is usually called the "birth rate," which is the proportion of annual births to the total population.

constituting the physical environment) and the rate of births of a population, and corresponding relations between these same environmental conditions and the average length of life of the individual, from these sets of relationships we can establish by simple multiplication the equilibrium population corresponding to each set of environmental conditions.

In a biological population probably the overwhelmingly most significant factor determining the rate of births is the size of the population itself or of its reproductive age groups. If there were no factors affecting the average length of life, therefore, a biological population (e.g. rabbits) would grow indefinitely. In such a population, therefore, the principal limiting factor is the decline in the average length of life of the individual as competing populations press in on its means of subsistence, and as predators, disease, and parasites increase their toll. By contrast, in populations of commodities the growth of numbers is checked more by a decline in births than by any shortening of the average age. Thus the rate of births of automobiles does not depend in any direct way on the number of automobiles, but rather on the number of auto workers, the capital invested in auto production, and so on. As the population of automobiles rises, there is not necessarily much change in the average length of life—increased accidents due to traffic density may shorten the length of life, and the development of specialized repair shops may prolong it, but both effects are likely to be small. The main factor creating an equilibrium automobile population is the slowing up of production (births) as the population increases. In the case of social organisms such as drug stores, gospel tabernacles, or local unions it is probably true that an increasing population produces some adverse effect on the length of life: were there fewer drug stores, the expectation of life of each might be greater. But here also a major factor is the stifling effect of the growth of the population on the number of births.

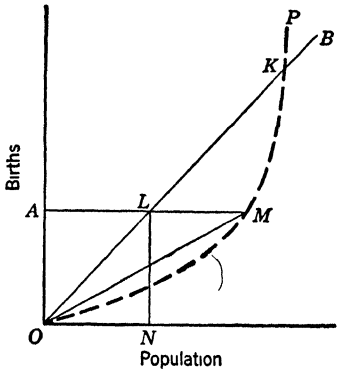


FIG. 11. Biological population equilibrium.

These principles are illustrated in Figs. 11 and 12. The solid line *OLB* represents the number of births which corresponds to each level of the population: thus at a population *ON* there will be a number of births *NL*. The dotted line *OMP* represents the *equilibrium* population which

corresponds to each rate of births. Thus at a rate of births equal to OA the equilibrium population would be AM. This implies that when the equilibrium population is AM the average length of life is AM/OA, or the tangent of the angle AOM. The equilibrium population is at K, where these two curves intersect. The equilibrium in the case of these figures is clearly stable. If we start with a population ON the rate of births will be OA: the population is below the level which can be reached with this rate of births and the existing average length of life, so the population will grow; as it grows the rate of births will increase. If the ad-

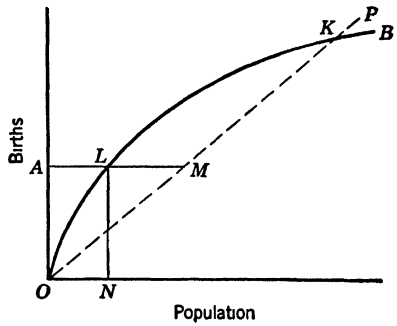


FIG. 12. Social population equilibrium.

justment of births to population is instantaneous whereas the adjustment of population to births is slow, the vector line will be LK, and the population and rate of births will move up to the point K. If the P-curve cut the B-curve from above, the equilibrium would be unstable.

Figure 11 would be an extreme case characteristic of a biological population: the rate of births is proportionate to the population, as indicated by drawing OB as a straight line from the origin. The average length of life is a decreasing function of the population, indicating increasing "pressure of population on the means of subsistence." This is shown by the concavity of the P-curve towards the vertical axis, indicating that the angle AOM declines with increasing population. In this case, therefore, the population is checked by declining longevity (or increasing mortality) rather than by a diminishing birth rate. The birth rate (ratio of the rate of births to the population) is constant and equal to the slope of the line OB. In many actual cases, of course, the effective birth rate may decline with increasing population, and the line OB bend downwards.

Figure 12 shows an extreme case of a "social" population, e.g. of automobiles. The average length of life is assumed to be independent of the

population, as shown by drawing the P-curve as a straight line from the origin: the average length of life (tan AOM) is clearly constant. The B-curve, however, is drawn concave to the population axis, indicating that the birth rate declines with increasing population. In this case the equilibrium population is reached not because of an increase in "mortality" but because of a decline in "fertility." We can look at this, if we like, in another way. The equilibrium population corresponding to the rate of births OA is AM: at this population the birth rate and death rate will be equal, both being measured by OA/AM. At populations below K the birth rate is above the death rate: the population will grow; at populations above K the death rate is above the birth rate: the population will decline. Many populations will, of course, adjust to an equilibrium both by a decline in fertility and by an increase in mortality with expanding population.

It should be observed that competitiveness or complementarity among populations of different kinds may be established either through the impact on fertility or through the impact on mortality. Where two biological populations compete for a common food supply the impact of one on the other is likely in the main to be through the effect on mortality (length of life): thus a rise in the population of squirrels in a wood is likely to diminish the population of chipmunks because the length of life of chipmunks will diminish. There may, however, be effects on biological fertility, particularly if "infant" mortality increases and a smaller proportion of the population survives to the reproductive age. The competition of dandelions against grass on a lawn may be of the twofold kind: the dandelions absorb nutrients and hence increase the mortality of grass blades, and by occupying ground they prevent grass seeds from germinating. Complementarity (symbiosis) among biological populations may also operate either through fertility or through mortality or both. Where two populations each provide nutrition for the other the impact is mainly felt through the favorable effect on the average length of life. There are many cases, however, where one form of life assists another in its reproductive process. It is quite possible, of course, for population A to be helpful to B while B is hurtful to A: this is the parasitism case, previously examined. It is also possible for population A to be complementary or competitive with B while B is independent of A; the B-curve would then be a straight line parallel to the A-axis. In this case the equilibrium is always stable: A simply adjusts itself to the constant level of B.[7]

[7] Interesting relationships may be traced when there are three populations interacting. Thus there is superparasitism: host A being injured by, but helping, parasites B and C, while B is itself host to parasite C and is injured by C while helping it. There

The significance of this framework of ideas for economics can be seen immediately if we observe that the "population" of a commodity is its total stock, its "rate of births" is its rate of production, and its "rate of deaths" is its rate of consumption. Economists have neglected in their analysis the influence of the total stock of a commodity both on its consumption and on its production. It is true, of course, that these functions are dependent on many other things, e.g. demands and techniques. A rise in demand for a commodity will raise its B-curve and permit a larger equilibrium stock: a cheapening of production costs will do the same thing. Nevertheless, on the ordinary assumptions of static economics, the B- and P-curves may be expected to be stable enough to be useful analytical tools.

Heterogeneous populations and imperfect competition

Up to this point we have assumed that the populations consist of large numbers of individuals, so that their functional relationships can be described by approximately smooth curves. In the world of nature, especially in the biological world, this is frequently true—in nature, that is to say, competition is often "perfect," in that the individual is so small a proportion of his total population that his idiosyncrasies can be left out of the picture. The economic model of perfect competition is a similar case: the number of firms in the industry is assumed to be large enough so that changes in their number can be regarded as virtually continuous. When, however, we have populations composed of only a few individuals—or, even more extreme, of only a single individual—problems of "imperfect competition" (and, of course, imperfect cooperation) among populations inevitably arise.

Consider first the situation (Fig. 13) in which one of the populations (B) consists of a small number of large, indivisible individuals, whereas the other population (A) consists of a large number of small individuals so that variations in the population can be regarded as continuous. We cannot now assume that the population functions are continuous, as one of the populations can exist only at levels OB_1, OB_2, OB_3, etc. The A-curve now degenerates into the series of points a_0, a_1, a_2, etc., corresponding to integral values of B. Thus B_1a_1 is that amount of A which

may also be competitive parasitism, where A is host to parasites B and C, B and C being mutually competitive. Or we may have A and B hosts to a single parasite, C, and so on. There are many different possible combinations of competitive, complementary, and independent relationships among three populations, and the number of combinations increases very rapidly as we increase the number of populations. It would take us far beyond the scope of this work to pursue these complexities further.

is in equilibrium, given OB_1 of B. Similarly the B-curve degenerates into a series of points b_0, b_1, b_2; B_1b_1 represents that amount of A at which the population OB_1 of B would be in equilibrium. Let the lines on which these points lie intersect at R: this would be the equilibrium if B were perfectly divisible. The continuous A- and B-curves have a certain significance even in this case: if either population is below the corresponding value on the A- or B-curve, it will tend to rise; if above,

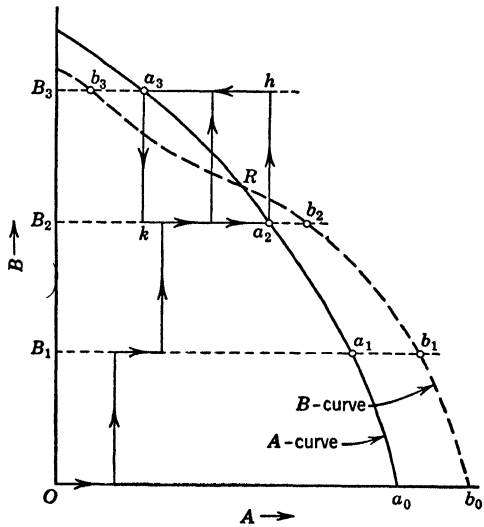


FIG. 13. Competition with one population imperfectly divisible.

to fall, even though the "equilibrium" values cannot be realized. The limiting condition is, of course, that B can rise or fall only in single units. With this limitation we can still draw vector lines; the shape they will take now depends very strongly on the dynamic conditions of the problem. If we start from O both A and B will certainly expand, A continuously and B in jumps. It may be that when B gets to OB_2 there will be no further expansion of B, the pressure towards the expansion of B not being sufficient to create an extra unit. In that case A will expand to B_2a_2, and an equilibrium will be established at the point a_2. At this point there is still some pressure for an expansion of B, but the fact that it can expand only to OB_3 prevents this expansion, which would occur if B were divisible. If the limits to expansion are set by "subsistence" in some sense, this will mean that the B-population is usually prosperous.

It may be, on the other hand, that the pressure to increase B produces an expansion. B will shoot up to OB_3, and the system moves from the

point a_2 to h: from this point population A will decline to B_3a_3. The point a_3 may be a stable equilibrium if the pressure for a reduction in B is not sufficient to drive the unit B_2B_3 out of existence. If, however, this pressure is sufficient B will decline to OB_2 again at the point k, A will expand to the point a_2, and there will be a circular equilibrium following the course a_2ha_3k. Circular equilibria could also be established at any

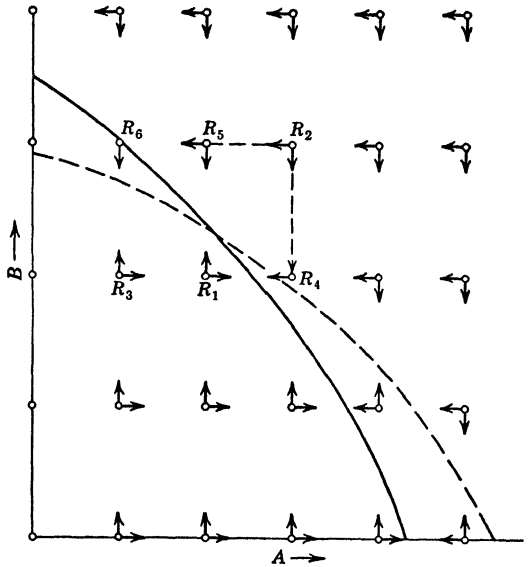


FIG. 14. Competition with both populations imperfectly divisible.

rectangle bounded by the lines B_3h and B_2b_2, depending on the dynamic assumptions made. Thus from a point such as k the pressure on B may produce an increase before A gets to the point a_2.

In Fig. 14 even the assumption that one population is capable of continuous variation is dropped, and both populations are now assumed to be capable of existing only in integral quantities. It is still possible to draw A- and B-curves on this figure, though it will only be by accident that either of them will pass through an integral point. Nevertheless for any given value of A all values of B below the B-curve will be subject to an upward pressure, and all values of B above the B-curve will be subjected to a downward pressure. Similarly, for any given value of B, values of A to the left of the A-curve exhibit upward pressures and all values to the right downward pressures. The direction of these pressures is shown by the little arrows at the points of the field representing pos-

sible combinations of the two populations. Here again stable equilibria or circular equilibria are possible, depending on the dynamic assumptions. At the point R_1, for instance, there is an upward pressure on both populations, but if the pressure is not sufficient to produce a unit jump in either, R_1 will be a stable equilibrium. Similarly R_4, R_5, or R_6 might also be stable positions of equilibrium. If at R_1, however, pressures on A are sufficient to make it take the jump to R_4, and if at R_4 pressures are great enough to take A back to R_1 (the pressures on B being insufficient to move it), there will be a circular equilibrium jumping back and forth between R_1 and R_4. We might have similar circular equilibria between R_1 and R_2, or R_3 and R_5, or R_3 and R_6, depending on the relative strengths of the pressures. It should be observed that the circular equilibrium path does not necessarily include the point of intersection of the A- and B-curves, R, though it is likely to be in the vicinity of this point.

It should be observed that these circular equilibrium paths may exist even when the populations are continuously variable, if the variations themselves are likely to take place in large and sudden jumps which "overshoot" the equilibrium, or if the movement towards equilibrium develops momentum which also takes it past the equilibrium. The cyclical movement observable in some natural populations, e.g. the famous Norwegian lemmings, is probably of this kind. There are many examples also in economic life, e.g. the hog cycle. It is interesting to note that the most extreme model of perfect competition, in which all the firms are exactly alike and the supply curve perfectly elastic, is likely to give rise to these circular equilibria. In this case if there are more than the equilibrium number of firms in the industry all will be equally unprofitable, and there is no reason to suppose that one firm more than another will leave the industry. Indeed, the firm is presented with a problem in strategy surprisingly like those found in oligopoly: the firm that gets out helps those which stay in. Hence the question arises for each firm: "Should I get out, and suffer the inconveniences of transferring my occupation or output to another industry, in which case those who are left will reap the rewards of my trouble, or should I stay in and wait for someone else to get out?" What may happen is that nobody gets out at all, each waiting for the other, until the situation becomes quite intolerable: then, as the firms are all supposed to be alike, a large number of them will get out, leaving the ground bare! This will create an enormously profitable opportunity, but firms may hesitate to enter the industry for fear that a large number will do likewise; however, as the profitable opportunity persists this resistance will gradually be overcome, and at some point everybody's resistance will crumble and all the firms will jump back into the industry again! In this extreme form, of course,

the model sounds almost ridiculous, but the hog cycle is actually an example of a less extreme form of this phenomenon. If the various firms are *not* alike, i.e. if there is a "cost ladder," with firms ranged on an ascending scale of minimum average costs, and consequently a less than perfectly elastic supply, this phenomenon will not occur. Then it is quite clear which firms will leave the industry—those which have the highest costs—and adjustments will take place continuously and without any tendency towards circularity. Similarly in the expansion of the industry those firms which have the lowest costs will come in first.

There is another reason, besides the possibility of discontinuities, why the size of the individual unit of a population may be important in ecological equilibrium. The significance of a population may depend not so much on the number of individuals in it as on the product of the number with some measure of the *size* of the individual. A large number of small individuals may have the same significance as a small number of larger individuals of the same kind. Thus what is significant in considering a population of firms of one kind, i.e. an industry, is not merely the number of firms, but also their total outputs and inputs: 100 firms producing an average of 1000 tons each may have the same significance for the other populations of the system as 200 firms producing 500 tons each, for what affects the other populations may not be the number of firms but their output. This phenomenon is not perhaps so noticeable in natural ecology, where the individual is more standardized, but even there it is the inputs and outputs of populations which are most significant for other populations.

Chapter 2

The Individual Social and Economic Organism

Any theory which is set in an ecological framework—as an economic and social theory must be—must develop a theory of the individual organism. The behavior and relationships of populations, i.e. of aggregates, are made up of the behavior of the individuals which compose them. We cannot therefore properly understand the behavior of populations without knowing something about the behavior of individuals, and the determinants of their birth, growth, size, and death. We are not proposing a "general theory" of the organism which will apply in undiluted generality to such diverse creatures as rye grass, rabbits, tapeworms, professors, Seventh-Day Baptists, milk producers, gas stations, government bureaus, and sonnets. Nor is it even desirable to have such a "general theory" applying to all the diverse organisms of society. We need rather a continuum of theory; otherwise there is danger of generality degenerating into mere vagueness, and we do not want to put organisms as diverse as, say, the family, the corporation, the church, and the state into a single strait jacket of theory. Nevertheless all organisms follow certain broad patterns of form and change in spite of the illimitable variety in detail. All organisms exhibit a "life cycle" of birth, growth, decay, and death, and all organisms exhibit characteristic patterns of short-run behavior in periods too short for the life cycle to be apparent. The first is the long-run, the second the short-run, problem. In the long run, as Keynes says, we are all dead. In the short run we eat, drink, and are occasionally merry. The theorist, however, must contemplate death —or at least transfiguration—as well as merriment.

The homeostasis of the balance sheet

The simplest possible short-run theory of the organism is the theory of "homeostasis," to use a term from physiology. This is the view that there is some "state" of the organism which it is organized to maintain, and any disturbance from this state sets in motion behavior on the part

of the organism which tends to re-establish the desired state. Thus the human body has a complex physicochemical equilibrium, involving, for instance, a constant temperature and certain concentrations of various substances in its various parts. If the equilibrium is disturbed, "wants" are set up which induce the body to act so as to restore its optimum condition.

The simplest theory of the firm is to assume that there is a "homeostasis of the balance sheet"—that there is some desired quantity of all the various items in the balance sheet, and that any disturbance of this structure immediately sets in motion forces which will restore the status quo. Thus, if a customer purchases finished product, this diminishes the firm's stocks of finished product and increases its stocks of money. In order to restore the status quo the firm must spend the increased money stock to produce more finished product. On these assumptions the firm's production, and indeed the production of the whole society, is a necessity imposed on it by the fact of consumption. Because moth and rust corrupt we must weave more cloth and smelt more iron. Every act of production, of course, disturbs the stocks of raw materials and goods in process, which also have to be replaced. Hence consumption directly "causes" production in such a system, and, be it noticed, without any intervention of a price system or any assumption about maximizing profits or any other variable.

The theory of homeostasis and the concept of an "asset structure" or balance sheet can be applied with varying degrees of significance to any social organism. A college, for instance, must replace its graduating seniors by juniors, its juniors by sophomores, its sophomores by freshmen, and its freshmen by a new class of entrants. It must maintain its prestige and position by replacing faculty members as they retire or advance. It must maintain its plant as it depreciates. Similar principles apply to churches, trade unions, and all other social organisms.

Thus the balance sheet is to a firm what its "state" is to a body: it describes the structure of the firm in terms of the various quantities of its parts. Anything that happens to a firm can be described in terms of a "dynamic" balance sheet—i.e. a movie of the balance-sheet changes—for everything that affects the firm in any way will effect changes in the balance sheet. The economist's concept of the balance sheet will, of course, be considerably wider than that of the accountant, will include many intangible items which the accountant's conventions do not permit him to include, and will also break down the items into a much greater detail than will the accountant.

Thus the "assets" side of the economist's balance sheet will consist of a detailed list of all the quantities of the firm's assets of various kinds

together with their values. Strictly, every homogeneous population of assets should be listed separately, and everything which reflects the "position" of the firm should be included. The list will include cash, bank deposits, securities of various kinds, accounts receivable, the quantities of all the kinds of goods, finished, in process, or raw, that it possesses, the quantities of all different kinds of buildings, land, machinery, etc. The accountant's balance sheet usually stops at this point, though occasionally an item for "goodwill," reflecting the value of the firm's market position, is included. In the economist's description there should perhaps be found some account of the firm's market position and also some account of its potential resources. Thus in a non-slave society laborers do not appear in the balance sheet of firms, as they would in a slave society: nevertheless, opportunities for hiring labor, and opportunities for hiring management, may represent positive assets to a firm. This is recognized sometimes in accounting when, for instance, a firm takes out an insurance policy on the services of an able manager, indicating that his death or removal would represent a positive loss of assets to the firm. Similarly, high morale among the firm's actual or potential labor force may equally be an asset in an economic description even though the market does not usually set a value on such factors as these.

Each physical inventory of assets is *valued* according to some *price*. Essentially a price is a transformation ratio at which a physical quantity of assets can be transformed into a value quantity (dollars). It is necessary to do this in order to sum the inventory of assets. We cannot add tons of steel, cans of paint, pounds of rubber, and numbers of automobiles: if, however, we reduce the physical quantity of these things to a dollar equivalent, or to their equivalence in some other "measure of value," we can add the "dollars" together to get a total.

The "price" of cash and bank deposits is unity, as the physical quantity of these assets is measured in "dollars." The same may be true of some securities, though there is frequently a transformation ratio between their "volume," which is measured in conventional dollars, and their value in "real" dollars. Inventories may be valued according to various principles—e.g. "cost or market, whichever is the lower," "first in, first out," or "last in, first out"—but whatever the principle of valuation there will be some "price" or transformation coefficient, i.e. the ratio of the value to the corresponding physical quantity. Buildings and other fixed capital are generally valued as units; the "price" and the value are the same figure. Here again there are many diverse methods of valuation. Whatever the method, however, the *meaning* of valuation is always the same: it is to discover a "dollar equivalent" for each phys-

ical unit of assets, representing at least a potential rate of transformation of non-liquid into liquid assets.

The "liabilities" side of the balance sheet shows the distribution of the total value of assets among the various claimants; looked at in another way, it identifies the providers or sources of the resources represented by the assets. Some claims may be contractual, in which case they are in fixed dollar values (bonds, accounts payable, notes payable, etc.). The excess of the total value of claims (which is, of course, equal to the total value of assets) over contractual claims is the residual claim, or net worth. There seems to be practically nothing in the world of living organisms which corresponds to the liabilities side of a balance sheet: living organisms have "assets" clearly enough, but all their assets are net worth and they have no debts! This again is a reflection of the fact that the constituent atoms of living organisms are not superior to the organisms which they compose, and are, as far as we know, quite passive in their relations with them: the "atoms" of social organisms are human individuals who are very far from passive and which are—or should be—superior to the organisms which they compose. Hence in the world of society an organism may draw resources from a great many different individuals, often removed by many intermediate channels and intermediate organisms. Ultimately the claims on the assets, say, of a bank or any other firm, must be apportioned to the net worths of individuals, but the liabilities of one firm may consist partly of assets of another, whose liabilities in turn consist partly of assets of another, and so the chain is followed until we come to the assets of individual persons.

The dynamics of the balance sheet

The next question is, *What changes* can take place in the position of the firm? Everything that happens to a firm can be expressed by changes in the balance sheet: if we had a moving picture of the complete balance sheet we would know every *objective* change and every *event* in the history of the firm. These changes are of several kinds: (i) purchases and sales of goods; (ii) purchases and sales of labor; (iii) transformations of one asset into another through production (e.g. wheat into flour); (iv) consumption of assets, not directly attributable to transformations (depreciation apart from use); (v) revaluations of assets, i.e. changes in the prices or valuation ratios by which physical quantities are expressed as value (dollar) quantities; (vi) creation or destruction of claims of various kinds (e.g. by borrowing or by repayment of loans). Many changes consist in simultaneous combinations of one or more of these, but all objective changes can be analyzed into one or more of these six types.

Purchases and sales consist of the transformation of assets through exchange: a purchase represents the diminution of money assets and an increase in non-liquid assets. Thus, if a firm buys 1000 bushels of wheat at $2.00 per ton the "cash" or "bank deposits" item in its balance sheet is diminished by $2000; the "wheat" item on its balance sheet is increased by 1000 bushels. It is the usual accounting convention to value purchases at the price paid for them, i.e., it is assumed that equal values are exchanged, so in the above case the wheat would be valued at $2000, and in value terms the transaction consists in a shift of $2000 from one item (cash) to another item (wheat), the total value of assets being unchanged. Similarly a sale represents a shift from non-liquid to liquid assets (money): if 1000 bushels were sold at $2.00 per bushel the transaction would represent an internal shift within the balance sheet of $2000 from the "wheat" to the "cash" item. Thus purchases and sales represent *asset transformation through exchange.* Exchanges normally consist of a transfer of non-liquid assets into money, or vice versa: it is quite possible, however, to have barter exchange, in which one kind of non-liquid asset is transformed into another kind.

Profits are created in exchange, as we shall see more clearly in the next chapter, by growth in the value of assets while they are being held. Thus suppose a man buys 1000 bushels of wheat for $2.00 per bushel and sells them later for $2.50. It still remains true that the purchase involves the giving up of $2000 of money for $2000 worth of wheat, and the sale involves giving up $2500 worth of wheat for $2500 of money. In the interval between the purchase and sale the value of the wheat has risen, because of the rise in its price, from $2000 to $2500. It is this rise in value, not the exchanges themselves, which constitutes profit. This profit is, of course, only "realized" at the date of sale, if by "realization" we mean transforming the asset structure into its original form. Realization, however, is not the same thing as earning.

Exchange is not the only method of transformation of assets: they may also be transformed through *production.* Production generally means the simultaneous creation of one or more kinds of assets through the destruction of other kinds, the net effect being the replacement of the destroyed assets by the produced assets, i.e. the transformation of destroyed assets into produced assets. The destroyed or consumed assets constitute the "cost," from the point of view of the firm, of the assets produced. Thus in the production of 1000 tons of flour 1250 tons of wheat may be consumed, a certain amount of machinery and equipment may be consumed, and a certain quantity of money must be spent in the purchase of labor. The shifts in the asset structure as a result of this productive operation therefore consist of (i) an increase in the

flour item (the "production"), and (ii) decreases in the cash, wheat, and machinery items (the "costs of production"). These increases and decreases may, of course, be valued in dollar terms. If the operation has been profitable the value of the increase in flour should be greater than the value of the decreases in the other items. This is the essence of the profit-making process—the manipulation (transformation) of assets in order to achieve an increase in their total value.

The peculiar nature of purchases of labor should be noticed. From the point of view of the firm the purchase of labor consists in a transformation of liquid into non-liquid assets. In the first instance, of course, as labor is performed the worker builds up a wage claim against the firm which should logically appear on the liabilities side of the balance sheet: the worker in the mere act of labor is providing resources for the firm, is creating assets, and has a generalized claim on the value of these assets created, or at least on part of this value. A wage payment extinguishes the wage claim; from the point of view of the firm it represents a diminution of liquid assets and a corresponding diminution of contractual liabilities. The use of labor plus a wage payment, then, resolves itself into a diminution in liquid assets (money) to the amount of the wage payment, and a rise in other assets which are the product of the work.

Generally we should expect every diminution in an asset to result in a corresponding increase in some other asset. Sometimes, however, the effects of the consumption of one asset are so diffused throughout the process that it is impossible to specify what assets are increased as a result of the consumption. In accounting terms such consumption of assets is said to be "expensed." The salaries of management are frequently in this class: so are some of the "overhead" and maintenance expenditures, e.g. for night watchmen. In principle, however, we may suppose that assets will not be consumed unless their value is at least replaced in the asset structure as a whole, so that all voluntary consumption of assets in principle should be regarded as transformation. There are some difficult problems in connection with unplanned or undesired consumption of assets, e.g. when a fire burns down a building. These losses can usually be spread out over time by insurance, the cost of which represents a kind of depreciation of the goods insured. There are difficult problems also connected with depreciation, which are encountered in the discussion of national income: depreciation is a form of capital (asset) consumption which frequently takes place almost independently of the use of the asset. When wheat is transformed into flour there is no doubt about the connection between the diminution in wheat and the increase in flour. There is considerable doubt, however, whether the depreciation of the flour mills is connected with the increase

in flour, for the mills might very well depreciate just as much even if no flour were produced.

The fifth category of changes in the balance sheet—revaluation of assets—does not affect the "physical" balance sheet, i.e. the list of physical quantities of assets. It does, however, affect the "dollar" balance sheet and affects the total value of assets and also, on the other side of the balance sheet, the net worth. Thus, if a balance sheet contains 1000 bushels of wheat at $2.00 a bushel, this is registered in the dollar balance sheet as $2000. If now the price at which the wheat is valued rises to $2.50, the dollar value of the same quantity of wheat is $2500. If there are no other changes in the balance sheet the net worth will be increased by $500 even though no changes whatever have taken place in the physical balance sheet. It is a difficult problem whether these changes due to revaluations are to be regarded as "spurious." If all prices changed in the same proportion, and all the values on a balance sheet changed in the same proportion, clearly all we would be witnessing would be a change in the size of the unit of account: it would have no more significance than the shifting of a decimal point. If, however, there are changes in the relative prices of various items on the balance sheet, this change is significant, for a price in the sense of a valuation coefficient is a "weight" by which the *significance* of a physical item is assessed. Suppose, for instance, that firm A holds 2000 tons of coal and 100 tons of wheat, while firm B holds 200 tons of coal and 550 tons of wheat. If the price of coal is $10 per ton and of wheat is $40 per ton both firms have assets valued at $24,000. Now suppose that the price of coal rises to $20 and the price of wheat falls to $20 per ton. Firm A's assets are now worth $42,000, while firm B's assets are worth only $15,000, yet there has been no change in the physical structure of their respective balance sheets. The change, however, is significant: A is better off because he has held his assets mainly in a form which has risen in significance, whereas B is worse off because he has held his assets in a form which has fallen in significance. This is not to say, however, that market prices are the true measure of the "significance" of assets, but we cannot rule out the possibility of relative changes in "real" significance which will affect the significant asset-value totals.

The sixth category of changes, creation and destruction of contractual obligations or "securities," may result in changes on both sides of the balance sheet. If a firm borrows $1000 the result is to add an asset of $1000 cash and a liability of a $1000 promissory note. Repayment of the loan likewise involves a subtraction from cash and an equal subtraction from promissory notes. The making of a loan (lending), however, results immediately only in changes on the assets side: there is a

diminution in cash and a corresponding increase in securities of some kind, e.g. notes payable.

Active and passive changes

Changes in the structure of the balance sheet can be described as either *active* or *passive*. Active changes result from what the firm *does*, and are initiated by the firm itself: purchases of raw materials or of labor, borrowing money, initiating an advertising campaign are examples. Passive changes result from what *happens* to a firm. Some of these are initiated by other organisms. Sales are usually of this character, although there may be some activity involved here. Some occur through external changes, such as depreciation and obsolescence.

Active changes frequently involve passivity on the part of some other organism: purchases, for instance, usually involve passivity on the part of the seller, though there may be exceptions to this rule when there is bargaining. The "activity" of a firm (or of any other organism) need not be confined to "active" behavior. It may also consist of setting the terms on which passive changes take place, and so indirectly controlling the volume of these passive changes. Thus a firm may offer a product for sale, leaving it up to buyers how much of the product will be sold (that is, from the point of view of the firm, transformed into money) in a given time. The rate of transformation, however, will be determined in part by the price, or transformation ratio: the higher the price, in general, the less will be sold in a given time (the law of demand). A firm may therefore be able to affect the rate of passive transformations by manipulating the price—also perhaps the selling cost—of its product.

In an organism operating under the pure homeostasis principle every "passive" event will be offset by an "active" event. If all passive events could be suppressed, homeostasis could be maintained by doing nothing whatever. The miser brooding over his store of gold is perhaps the closest approximation to this condition. The thing that forces most organisms into activity is that there are certain unavoidable passive events, such as depreciation or internal consumption, and therefore activity of a highly complex nature may be necessary in order to offset these unavoidable depletions of assets.

The static theory of the firm

Though the homeostasis principle is a useful one, it is clearly only a very first approximation to a theory of the organism, and breaks down at several points. In the first place it says nothing about what determines the equilibrium state itself. In biology this can generally be assumed to be given by the genetic constitution: in social organisms,

however, the equilibrium position of the organism itself is to a considerable degree under the control of the organism's director. Thus in the case of the firm a first-approximation theory can be constructed on the basis of an assumed homeostasis of the balance sheet, but we must go on further to discuss the problem of what determines the "optimum" or equilibrium balance sheet itself, as this is also to some extent under the control of the firm. This should bring us directly into "life-cycle" theory, and indeed one would have expected Marshall's famous analogy of the trees of the forest again to have led economists to a discussion of the forces which determine the birth, growth, decline, and death of a firm. In fact the theory of the firm, and of the economic organism in general, has not developed along these lines, but along the line of static equilibrium theory of "maximizing behavior." The concept of the balance sheet, unfortunately, has not been employed to any extent in developing the static theory of the firm, so that as generally presented in the textbooks the firm is a strange bloodless creature without a balance sheet, without any visible capital structure, without debts, and engaged apparently in the simultaneous purchase of inputs and sale of outputs at constant rates. In spite of this weakness, as we shall see later, much of the static theory of the firm can be salvaged and put into balance-sheet form: nevertheless, even when this has been done we still do not have a life-cycle theory, but only a theory of an equilibrium position of all the variables of the firm.

Static theory visualizes the firm as a process of transformation of input values into output values. The input values and output values are functionally related, first through the physical production function, which relates quantities of physical inputs to physical outputs, and then through the market functions, which relate the quantities of inputs and outputs to their values. These market functions may also include certain other inputs and outputs, as in the theory of selling cost. The combination of all these functions, however, gives us one grand function relating all the variables concerned—inputs, outputs, prices, etc.—in what may be called an "opportunity function" because it shows all the opportunities or possible combinations of all possible variables which lie open to the firm. One of these variables—in elementary theory the net revenue, or the difference between the value of outputs and the value of inputs—is assumed to be maximized, and the position of all the other variables on the opportunity function where this "maximand" is maximized is regarded as the equilibrium position of the firm. The theory of the household follows exactly similar lines, except that in this case the maximand is not an objective variable such as net revenue but a vector of a subjective variable (utility). The principle of maximizing a max-

imand under the limitations imposed by an opportunity function is basic to what is usually called the "marginal analysis." Indeed the whole marginal analysis, elaborate as it is, is simply the spelling out in detail of this fundamental principle of maximization.

The static theory of the firm, in spite of not being a true "life-cycle" theory, can throw some light not only on the equilibrium size of the firm but also on its conditions of birth and death. It is generally assumed that a firm will not be born if at the maximum net revenue of the potential firm the net revenue is negative. On the other hand it is assumed that, wherever there is an opportunity for a firm to exist in which the net revenue is positive, then that firm will be born. Similarly, if an existing firm, through some unfavorable change in its opportunity function—say, through a fall in demand for its product, rising prices of its inputs, or adverse changes in its production function—finds that its net revenue is negative, it may continue to exist for a while but will eventually go out of business ("die") because the resources which are organized into it can find more profitable or advantageous employment elsewhere. The economists never seem to have applied this principle to the household, but clearly it can be applied. Thus a household is born when two young people decide that the net advantages of living together are positive. This may occur either because of a favorable change in the market functions (the young man gets a better job) or because of a favorable anticipated change in the combined utility-production function (they fall in love) or simply because of chance meeting of productive factors (man and woman) whose combination is adjudged to be advantageous. The household usually is broken up (dies) either when one of the partners dies (exhaustion of natural resources) or when the partnership is judged to be unprofitable and dissolved voluntarily by separation or divorce. The static analysis does not perhaps throw much light on the size of the household, this being frequently determined by chance. Increasingly, however, we find that in modern "economic" society the number of children is planned by a balancing of advantage against disadvantage, and under those circumstances we even tend to find the size of the family, as of the firm, rising with the demand for its product.

Insufficiencies of "maximizing behavior" theories

The type of theory represented by the static marginal analysis does not, however, throw a great deal of light on some other social forms. In the attempt to construct a model of a labor union, for instance, labor economists have not found the principle of maximization particularly useful. It does not seem to be a useful principle of interpretation of the

actual behavior of unions to assume, for instance, that they wish to maximize some objective maximand such as the total wage bill of their members, or the average wage of employed members. The reason for this seems to be the weakness of any forces tending to bring the policy of a union into line with these equilibria. Maximization theory is useful only as long as reasonably strong dynamic forces are set up when the variables concerned diverge from the values which yield a maximum of the maximand. It is no use postulating any equilibrium unless we know something about the strength of the dynamic forces which tend to bring the variables back to the equilibrium values, once a disturbance has taken place. In the theory of the firm, and even of the household as an economic unit, there is some reason to suppose that a significant maximand can be found—significant, that is, in the sense that divergences from the optimum position set in motion forces which will bring the organism back to it.

It should be noted that it is not necessary to have an *objective* maximand in order to satisfy this condition. The relative success of the indifference-curve analysis in dealing with problems of individual economic behavior and the behavior of households is proof that significant propositions can be derived from the theory of maximizing behavior even when the thing maximized is a purely subjective vector of choice and when therefore the maximizing behavior is simply a way of expressing the patterns, actual and potential, of human behavior. It may be, however, that the maximum itself is not a "peak" but a broad plateau, where wide variations in the variables under the control of the organism can occur without diminishing the maximand. It may happen also that when the maximand declines below its maximum value little *strain* is placed on the organism, and hence the forces tending to change the various variables are too weak to overcome the resistances to change. In both cases there is not a "point" of equilibrium but a wide *range* of neutral equilibrium, within which the equilibrium is quite indeterminate. This is probably why the feeling is so strong among labor economists that a labor union does not maximize anything, and that different criteria of its behavior must be sought.

Every theory of a single organism can, of course, be brought *formally* under the principle of maximizing behavior by assuming that what is maximized is "utility," meaning by utility simply a quantity, or vector, which is below a maximum at any position of the organism where it is not in equilibrium. In this sense, however, even trees, birds, and flowers "maximize utility," and it may be doubted whether the formal theory has much interpretive value. The reason why the principle of maximization of profits in the firm, and of utility in the household, has

produced valuable results is that divergence from the maximand is *consciously recognized* as a stimulus to action and hence as a force tending to bring the organism back to equilibrium once it has departed from it. In biology, however, and in the equilibrium of many social organisms, the "maximization of utility" is a purely formal solution, incapable of any interpretive power, because the divergence of the maximand from its maximum value is not an operative force bestirring the organism to change.

It is clear that the concept of an *equilibrium* position of an organism is a much more general concept than that of a *maximum* position, particularly if the divergence from equilibrium is related to the dynamic forces which bring about change in the organism. Thus in the case of a labor union it might be quite proper to postulate as an equilibrium wage not that which maximized a wage bill, or which maximized anything, but that which bore a definite proportion to some other wage in some other occupation. It might well be that what drives union leaders is the fear of dissatisfaction among their members, and that what creates dissatisfaction is not any absolute level of wages but rather the invidious comparison with the wages of the man next door. Or we might assume that the dissatisfaction of the membership was related to the length of time that has elapsed since the last wage increase, so that the "equilibrium" wage continually increased, and if wages were stationary the divergence between the actual and the equilibrium wage also continually increased. It is clear that many such models could be constructed, some of which might prove useful in the interpretation of actual situations, but none of which involved any "maximizing" except in the most formal and contentless sense.

I do not mean to minimize the importance of the theory of maximizing behavior; indeed, we shall spend the next few chapters in elaborating it. Nevertheless it is at best a special case of the more general theory of homeostasis, and it cannot throw more than certain gleams of light on the dynamic problem of the "life cycle." A theory of the course of an organism through time must be a theory of disequilibrium rather than of equilibrium. Indeed, true equilibrium is unknown in the world of nature, for all things are subject to the irreversible processes of entropy. We need perhaps to develop a theory of "pressures for change" operating against certain resistances, which give way at certain critical "pressures," something like the "all or none" principle of the neuron or the vacuum tube. It may well be that the great bulk of human behavior does not follow the patterns of sober, reflective maximization of advantage, but rather follows first the principle of inertia (nobody does anything unless he has to!) and secondly the principle of least resistance (if you have to

do anything, you do the thing that is easiest to do!). There is nothing which says that the line of least resistance is the same as the line of greatest advantage except the long, slow retribution of natural selection. Even this type of dynamic theory, however, does not necessarily give us a "life cycle," and does not explain the inexorable and irreversible movement towards the equilibrium of death. Individual, family, firm, nation, and civilization all follow the same grim law, and the history of any organism is strikingly reminiscent of the rise and fall of populations on the road to extinction, as in the unstable equilibrium case of Fig. 3.

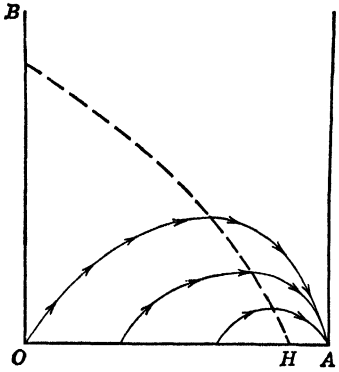


FIG. 15. The life cycle.

Here again the concept of a vector line may be useful in explaining the life history of an organism, or even of any partial population. Suppose that A in Fig. 15 represents the whole population of the universe, B a relatively small population within it. The equilibrium value of A may be supposed to be independent of the value of B, and the A-curve will be a straight line perpendicular to the A-axis. We suppose B to be competitive with A, so that the B-curve (dotted) has a negative slope. We suppose further that the B-curve cuts the A-axis at H, at a value of A less than the equilibrium value. Then, from any point on the A axis between O and H, B will be "born," will first rise, will reach a maximum where the vector line crosses the B-curve, and will decline to zero at the point A. Single organisms are always likely to be in this kind of relationship to the whole universe.

Chapter 3

The Theory of the Economic Organism

We shall now find it valuable to descend from the height of general principles to the detailed discussion of some simple models of purely "economic" organisms in order to cultivate depth as well as breadth.

The pure marketer

The simplest model of an economic organism is that of a "pure marketer." This is an organism whose activity is confined to buying, selling, and holding assets of various kinds; its transformations of assets are solely those of exchange, not of production. In order to simplify the discussion still further we shall assume an organism with a balance sheet containing only three items: on the assets side, two kinds of assets or "exchangeables," which for the sake of convenience we label "money" and "wheat," and on the liabilities side a single item of net worth. The net worth, of course, is equal to the sum of the values of the two exchangeables. At this stage we shall not consider the general economic properties of these exchangeables, and in particular we shall not be concerned with the peculiar properties of money. All that is assumed is that the exchangeables can be transformed one into the other through exchange in a "market," and that there is a definite transformation function, or opportunity function, showing the various combinations of quantities of the two assets which can be attained through exchange. If the market is "perfect," i.e. if each asset can be exchanged for the other at a rate or price which is independent of the quantity exchanged, the transformation function will be linear. If the market is "imperfect" the transformation function will not be linear, but will be curved, indicating that the ratio of exchange is a function of the amount exchanged. In either case, however, it is assumed that the transformation function itself is stable during what may be called the "trading day," i.e. the period of time during which advantage is taken of all opportunities open, and the organism moves to a position of rest or equilibrium rela-

tive to the existing opportunity situation. We must also assume that the preferences of the marketer do not change during the trading day; otherwise he will have to move to a new position of equilibrium.

The closest approximation to such a model in reality is the speculator in the commodity or security market. Thus a wheat speculator holds his assets mainly in the form of wheat and money,[1] and starts out his "day" with certain holdings of both these assets. He finds himself also presented with certain opportunities to buy or to sell. If the market is perfect there will be a single price at which he can either buy or sell unlimited quantities of wheat, or, what is the same thing, sell or buy unlimited quantities of money, subject, of course, to the limitations imposed by his own resources and credit position. In practice, the buying price and the selling price may differ somewhat, giving a "broken-backed" transformation curve, and the price may depend on the amount bought or sold. In the light of the whole opportunity open to him the speculator makes his decision, buys, sells, or does neither, and so moves to a position of equilibrium relative to his existing preferences and opportunities. This closes our "trading day"—which may, of course, be a period of time much shorter than the calendar day—and the speculator will remain inactive until there is a change either in his opportunities or in his preferences, thus opening a new "trading day." An important assumption made here is that change in opportunities takes place entirely as result of forces outside the marketer himself, and that such changes in the future are not affected by his decisions now. Under some forms of market imperfection (e.g. oligopoly) this assumption does not hold, and the problem becomes much more complex. We shall not concern ourselves with these complexities, however, at this point.

The advantage of a simple two-asset model is that the whole balance sheet, and all the objective events in the history of the organism, can be represented in a single two-dimensional diagram. Thus in Fig. 16 we measure "money" along the vertical axis and "wheat" along the horizontal axis. Any point on the physical assets side of the balance sheet can then be represented by a point P_0, OM_0 being the quantity of money and OW_0 the quantity of wheat, money being measured in "dollars" and wheat in "bushels." The same diagram could be used, of course, no matter what the exchangeables.

In order to calculate a net worth it is necessary to reduce the heterogeneous assets to some common denominator. This is done by multiplying the quantity of each by a "valuation coefficient" and so transforming the physical quantity into a value quantity. We could, of

[1] As we shall see later, futures contracts can be regarded as "negative wheat" and debts as "negative money."

course, use some third quantity as a measure of value. Thus suppose the valuation coefficient of wheat were 600 francs per bushel and of dollars was 300 francs per dollar. The total value of, say, 300 bushels of wheat plus $200 would be 180,000 + 60,000, or 240,000 francs; this would be the net worth. In order to represent this diagramatically, however, we would have to introduce a third dimension, and at this stage it will be simpler to take one of the two assets (dollars) as the measure of value. Then only one valuation coefficient is necessary—

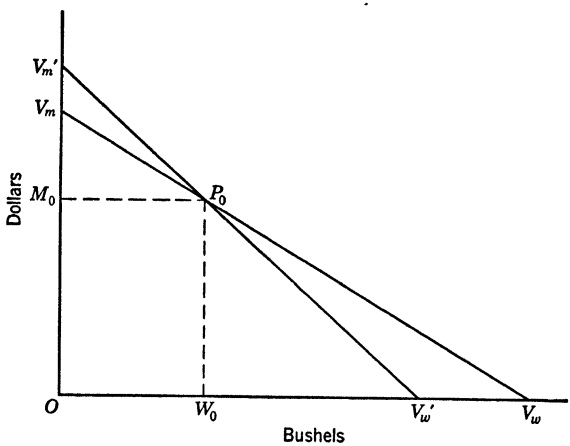


FIG. 16. A two-asset balance sheet.

the value ratio of the two assets. This can be expressed in the figure by the slope of a line through the point P_0, such as $V_mP_0V_w$. M_0V_m is then the dollar value of the quantity of wheat M_0P_0. The valuation coefficient is M_0V_m/M_0P_0 dollars per bushel, or tan $V_mP_0M_0$. Similarly, of course, W_0V_w is the wheat value of the quantity of money W_0P_0, and the valuation coefficient could equally be expressed as tan $W_0P_0V_w$ bushels per dollar. Then the net worth of the organism, in dollar terms, is $OM_0 + M_0V_m$, or OV_m; this is the dollar value of the money plus the dollar value of the wheat. The net worth could also be expressed as OV_w bushels; this is the bushel value of the wheat (OW_0) plus the bushel value of the money (W_0V_w). A "rise" in the dollar value, or "price," of wheat is the same thing as a fall in the wheat value of dollars; it is reflected in a shift of the valuation line from, say, V_mV_w to $V_m'V_w'$. This results in a rise in the net worth measured in dollars, and a fall in the net worth measured in bushels. We shall see later that this apparent paradox leads to some serious difficulties in the interpretation of accounting procedures.

The profit-making process

Various possible successive positions of the organism in time as it takes advantage of changing opportunities are now shown in Fig. 17. We assume a perfect market, to simplify the exposition, so that the transformation (opportunity) lines are straight. With imperfect markets these lines would be curved or broken, but the essentials of the analysis would not be changed. Then we suppose that the marketer starts at the position P_0 and has an opportunity to exchange, repre-

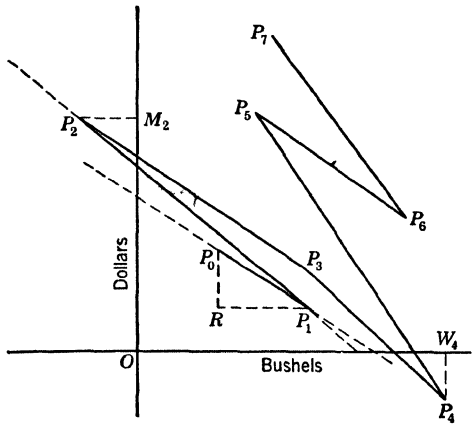


FIG. 17. The profit-making process.

sented by the line P_0P_1 and its extensions. He can move to any position on this line: actually we suppose that he moves to P_1, indicating that he has bought—and therefore added to his holdings—a quantity of wheat equal to RP_1 and has paid for it (thereby subtracting from his holdings) a quantity of money equal to P_0R. The price of wheat is RP_0/RP_1, or tan P_0P_1R. The next day we suppose the price of wheat to have risen. His transformation line is now P_1P_2 and its extensions: the higher price of wheat is reflected in the greater slope of the transformation line. The marketer moves to P_2 in the second quadrant. This means that he has sold wheat futures; i.e. he has sold, for money, a promise to pay wheat in the future: he has incurred a "wheat debt." His physical balance sheet now consists of an amount of money OM_2, greater than his net worth, and a negative item P_2M_2 representing the wheat debt. In order to calculate his net worth the dollar value of the wheat debt must be subtracted from his money holdings. The next day we suppose the price of wheat falls; the transformation line is now

P_2P_3, and the marketer moves to P_3, purchasing wheat and paying off his wheat debt. We now suppose a rise in the price of wheat; this, however, is taken as indication of a further rise, so the marketer moves into the fourth quadrant, indicating that he has bought so much wheat that his whole money stock is exhausted, and he has borrowed money. At P_4 his wheat holdings are OW_4, and his money holdings are a negative amount W_4P_4, indicating a money debt, or a promise to pay money. His net worth is the value of his wheat holdings, less the money debt. We now suppose that his expectations are satisfied and the price of wheat rises on the next day: he sells wheat, pays off his money debt, and moves to P_5, and so on to P_6, P_7, etc.

This analysis throws a great deal of light on the profit-making process. It is clear that, as the various transformations are made and the speculator traces out the path $P_0P_1 \cdots P_7$, he reaches certain positions which are quite obviously superior to his starting point. Thus if we compare the position P_5 with P_0 it is clear that, unless one or both of the assets are actually discommodities, P_5 is unquestionably a preferable position to P_0, for it represents a greater quantity of *both* assets than does P_0. Similarly P_6 is unquestionably a "better" position than P_3, P_7 than P_5, and so on. What is not clear at the moment is whether P_1, say, is better than P_0, for although P_1 represents a greater quantity of wheat than P_0, it represents a smaller quantity of money. Similarly, though P_2 represents a larger quantity of money than P_0, it represents a smaller quantity of wheat. If, then, we are to say whether *any* given change represents a gain or a loss, and even more if we are to be able to measure the gain or loss, we must have a system of valuation of the assets which will enable us to derive a figure for net worth from the physical balance sheet.

The valuation of assets

There are, however, a number of possible systems of valuation, and a difficult question arises over which to select. If we follow the rule of equating the valuation coefficient with the transformation coefficient, i.e. if we value our wheat stocks at the market price of wheat, we run into several difficulties. This assumption, as we have seen, necessitates the proposition that no values are created and no profits made in the pure act of exchange, but that profits and losses come only when stocks are revalued because of a change in price. It may be argued, however, that this convention is unrealistic, because clearly the act of exchange is performed in order to put the speculator in a better position to take advantage of subsequent changes in the transformation function. Thus suppose that in Fig. 18 P_0 is the initial position, and that the organism

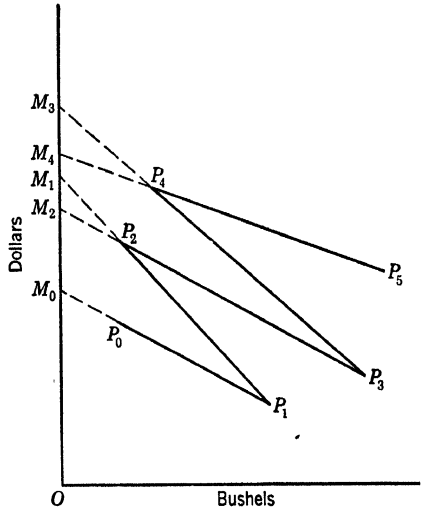


FIG. 18. Valuation at market prices.

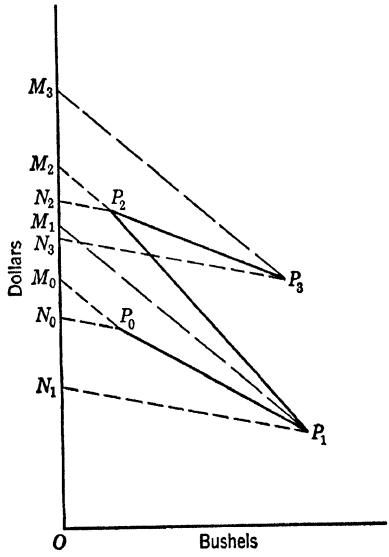


FIG. 19. Valuation at constant prices.

follows the path $P_0P_1P_2 \cdots$. The net worth both at P_0 and at P_1 is OM_0: the exchange makes no difference to it. The revaluation of assets, however, at the beginning of the next day—raising the price, and therefore the valuation coefficient, to the slope of P_1P_2—raises the net worth to OM_1; the profits of the period are the increase in net worth, M_0M_1. The revaluation (fall in price) at the beginning of the next day lowers the net worth, however, to OM_2; there is an apparent loss of M_1M_2. Nevertheless it is upon the foundation of this apparent loss that subsequent profits are made, for if it were not for the *changes* in price the net worth could never be increased through exchange: moving to P_3, to be able to take advantage of the subsequent rise in price, brings the net worth on the fourth day to M_3. So the process goes on, with an apparent alternation of losses and profits, decreases and increases in net worth, even though the general process is clearly a profit-making one.

A possible method of escape from this dilemma is to perform all valuations at a constant valuation ratio, independent of the market price. This is shown in Fig. 19. P_0, P_1, P_2, P_3 are successive positions of the balance sheet. If now the valuation of wheat is performed always at a constant valuation coefficient ("price") equal to the slope of the parallel lines P_0M_0, P_1M_1, P_2M_2, etc., the net worth will rise continually through OM_0, OM_1, OM_2, etc. Unfortunately, however, the mere procedure of valuation at a constant valuation coefficient is not sufficient to insure a steady rise in net worth. Thus we see that if a smaller valuation coefficient is selected, equal to the slope of the parallel lines P_0N_0, P_1N_1, P_2N_2, etc., the net worth again exhibits a back-and-forth movement, ON_0, ON_1, ON_2, etc. There is clearly a range of constant valuation coefficients within which the net worth grows steadily without any ups and downs, and there seems to be no particular principle by which we can select any one of these valuation coefficients within the range as better than any other. All valuation thus seems to possess a certain unavoidable arbitrary element, as long as the asset structure remains heterogeneous. Only in processes which begin and end with a homogeneous quantity of some single asset, e.g. money, can the amount of profit be measured unequivocally. These considerations are important in connection with the definition of the aggregates of macroeconomics also, and we shall return to this topic later.

Exchange limited by uncertainty

In spite of the difficulties of measuring profit in the profit-making process, some important propositions can be derived. The first is that for any *given* series of transformation functions, provided that a *con-*

sistent method of valuation is used, the wider the "swings," i.e. the more is bought at low prices and sold at high, the greater the profit. This is shown in Fig. 20. Starting from the point P_0, we suppose two successive transformation opportunities (market prices), represented by the slopes of the lines P_0P_1 and P_1P_2. Starting from P_0, the marketer can reach an unequivocally preferred position P_2, following the course $P_0P_1P_2$, buying cheap and selling dear. By buying still more, however, and selling still more, he can follow the course $P_0P_1'P_2'$, which places him in a position P_2', which is clearly preferable to P_2. Similarly, by increasing the amount he buys at the low price and sells at the high, he can always reach a preferred position.

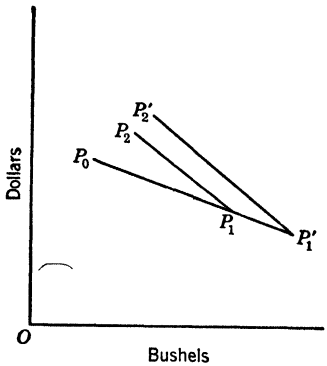


FIG. 20. Greater "swings" mean greater profits.

The question naturally arises then, What limits the "swings" of the speculator? Why does not he get rid of all his money—and, indeed, get indefinitely into debt—when the price of the commodity is low, and get rid of all the commodity and extend himself into commodity futures when its price is high? The principle of profit maximization seems to break down here, for there is no maximum. We seem to have a problem analogous to that of the producer under perfect markets and constant costs. One possible way of escape lies, of course, in the assumption of market imperfection: certainly as the marketer gets into his second or fourth quadrant the market imperfection of his credit position will become important, for there are sharp institutional limits on how much an individual can borrow, in the form either of money loans or of commodity futures. Nevertheless a force will be found to operate probably long before these limits are reached, and there is no need to drag in market imperfection to explain the limitation on the speculative swings. By far the most important limitation is the presence of *uncertainty* in regard to future opportunities (prices). It is true that if the speculator knew in advance what the future course of prices was going to be he could plan his "swings" as widely as the imperfections of the market permitted. In fact, however, he does not know for certain what the future will hold, and if he is not careful he may find himself making losses.

Thus the profit-making process such as the path $P_0P_1P_2 \cdots P_5$ in Fig. 18 is by no means automatic. If the speculator sells at low prices

and buys at high prices he may find himself following the path $P_5P_4P_3 \cdots P_0$, with a declining instead of an increasing net worth. If a speculator in a given situation buys it is because he is expecting a price rise, and hopes to take advantage of it by having large holdings of the commodity which is rising in value. The price rise, however, is not certain: there is always a possibility of a fall, and the more he purchases, and the greater his holdings of the commodity, the more dangerous does this possibility of a fall become. As he moves down the transformation line, increasing his commodity and diminishing his money holdings, the chance both of gain and of loss increases. Here, however, the principle of diminishing marginal significance comes into play. As the speculator moves down the transformation line the significance of a marginal loss increases—no man wants to risk being wiped out, no matter how great the potential gain. As potential gain increases, however, the marginal significance of potential gain declines. At some point the speculator feels that the chance of marginal gain is just barely worth the chance of marginal loss. At this point he will stop and will not proceed further down the transformation line. Similar considerations apply to restrict his sales when he expects a fall in price.[2]

The description of expectations themselves, and of the degree of uncertainty which they involve, is a matter of extreme difficulty. The attempts that have been made to describe uncertain expectations in terms of probability analysis have not been altogether successful, because of the fact that the probability concept involves the idea of a universe of like events, an idea which is not present in the concept of uncertainty. In this connection Shackle's criticisms have a great deal of point,[3] and yet his alternative formulation in terms of single "focus outcomes" may seem too much simplified. Fortunately, however, for our present purposes these difficulties can be by-passed, for at this level of analysis all that is needed as an analytical instrument is some measure of *asset preferences*. These asset preferences will themselves be an expression and a direct result of the marketer's expectations: if, for instance, his expectations move towards a greater degree of confidence in a future rise in price, this will be reflected immediately in a greater preference for the commodity as opposed to money; if he becomes more confident that prices will fall, commodity preference will decline and money preference rise—the preference always shifting towards that asset which is most confidently expected to rise in value or to rise most sharply in value.

[2] This principle is similar to Kalecki's "principle of increasing risk" and is discussed in more detail in a later chapter (pp. 126–127).

[3] See G. L. S. Shackle, *Expectations in Economics* (Cambridge University Press, 1949), Chapter VII.

The asset-preference structure thus reflects indifferently both the "dimensions" of expectations—the size of the expectation itself (e.g. the extent of an expected price rise)—and the degree of confidence which is placed in it. A rise in the size of an expected event may be offset by a decline in the degree of confidence in the expectation, and so may produce no change in asset preferences.[4] From the point of view of the behavior of the marketer, however, it is solely the asset preferences which are significant, no matter how they are determined. The determination of asset preferences becomes important mainly in dynamic systems where we attempt to relate the asset preferences of one date to the events (disappointments, etc.) of a previous date or dates.

Preferences measured by preferred asset ratios

The usual method of describing asset preferences is by an indifference-curve system, and the analysis of the marketer and of the exchange phenomenon by this means is familiar to most economists: a discussion of the method and its results is given in Chapter 5. There are, however, several objections to the indifference-curve analysis. In the first place the usual analysis assumes that the asset-preference structure is independent of the asset-price structure. Thus the indifference map is assumed to be unchanged even when the opportunity line changes. This is a defect which can be remedied, as we shall see, by introducing price as an explicit variable into the indifference (utility) function. Then, however, the analysis loses much of the simplicity and directness which have endeared it to economists. In general, the fact that the preference structure is expressed by a whole functional relationship means that it is difficult to express in simple quantitative form. Thus a phenomenon such as a "rise in preference for money" must be expressed by a complex shift in the character of the whole indifference function, the slopes of the indifference curves (the marginal rates of substitution) at each point in the field becoming flatter than before, as it takes more commodity at each point in the field to compensate for a decline in one unit of money holdings.

Fortunately there is a simple device by which the asset-preference structure can be roughly expressed in terms of a simple quantity or ratio. This quantity is the *preferred asset ratio*, which may be defined as the proportion of the value of his total assets which an individual

[4] This fact opens up the possibility of a measure of the degree of subjective confidence, for if asset preferences are constant, a fall in the degree of confidence which just offsets a rise in expected value is in some sense "equivalent" to it; hence the "offsetting" change in expected value which would just leave asset preferences unchanged can be used to measure the change in the degree of confidence.

wishes to hold in the form in question. Thus if an individual wished to hold 25% of the value of his assets in the form of wheat we would say that his "preferred wheat ratio" was 25%; if he wished to hold 10% of his assets in the form of money we would say that his preferred money (or liquidity) ratio was 10%. The sum of all an individual's preferred asset ratios must, of course, add up to 1, or 100%. The great advantage of this formulation is that a change in the structure of asset preferences can be described simply as a shift in preferred asset ratios. Thus, in a two-asset system with only wheat and money, a change in preferred asset ratios from, say, 70% wheat and 30% money to 80% wheat and 20% money would represent directly an increase in the preference for wheat and a corresponding decline in the preference for money, such as might follow, for example, a more confident expectation of a future rise in the price of wheat or a greater expected rise in the price. The assumptions underlying the preferred-asset-ratio method, and its limitations, are discussed in Chapter 5; meanwhile we can assume at least that the preferred asset ratios are useful "parameters of behavior," the changes in which correspond fairly closely to changes in actual or potential behavior patterns.

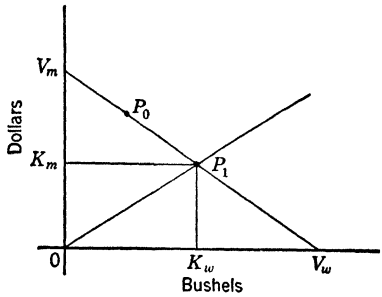


FIG. 21. Preferred asset ratios.

The analysis of the movement of a single marketer along his opportunity line (i.e. his exchange behavior) by means of preferred asset ratios is shown in Fig. 21. The asset combination which represents his starting point is at P_0. Assuming for the moment perfect markets, he will have a transformation line $V_mP_0V_w$, the slope of which is equal to the price of "wheat." At this given price all points on the field with the same asset-preference structure are found on a straight line from the origin, such as OP_1. If the preferred money ratio is r, the preferred wheat ratio is $1 - r$. At any point on the field where these ratios obtain we have, if m is the quantity of money, q the quantity of wheat, and p the price of wheat,

$$\frac{m}{m + pq} = r \quad \text{or} \quad m = \frac{pr}{1 - r}q$$

The intersection of the preferred ratio line with the transformation line at P_1 is the equilibrium point, as this is the only point on the trans-

formation (opportunity) line at which the preferred asset ratios can be satisfied. In this view exchange is regarded as a means for changing the *proportions* of assets held: if the valuation coefficients are equal to the market prices, as we have seen, no accounting values are created in exchange, and exchange does not in itself affect net worth. In order to account for the phenomenon we must therefore interpret exchange as an alteration of the *proportions* of assets within a constant total value. Thus we interpret the dissatisfaction of the marketer at P_0 as resulting from the fact that in view of the nature of his expectations he feels that he is holding too great a proportion of his assets in the form of money and not a sufficient proportion in the form of wheat. He therefore buys wheat: by so doing he diminishes the proportion of his assets in money form and increases the wheat proportion, and he will go on exchanging until he reaches the proportions which satisfy him, i.e. the preferred asset ratios. One advantage of this formulation is that it can very easily be generalized to any number of asset forms: the individual will always have a transformation function showing the various asset combinations open to him through exchange; if the various exchange ratios are consistent, i.e. if there is no possibility of arbitrage, this function will be a "surface" of $n - 1$ dimensions in a field of n asset forms. His opening position P_0 is, of course, a point on this "surface." There will generally be some other point P_1 on the surface at which all his preferred asset ratios are satisfied: the quantities of the various assets exchanged are given by the differences between the coordinates of these two points.

The expenditure-receipts and the supply-demand curves

To return now to Fig. 21, we see that the point P_1 divides the line V_mV_w in the proportion of the preferred asset ratios:

$$\frac{V_mP_1}{P_1V_w} = \frac{V_mK_m}{K_mO} = \frac{1-r}{r}$$

If the preferred asset ratios are themselves invariant with the price, this gives us a very simple formula for the revenue-expenditure curve of the individual—i.e. that curve which shows his position of equilibrium, and therefore how much he will buy or sell at various prices. Thus in Fig. 22 we suppose a family of tranformation lines through P_0, each representing a different hypothetical price. Dividing each of these in the proportion of the preferred asset ratios gives us the locus $P_3'P_2'P_1'P_0P_1P_2$, etc., which is the expenditure-receipts curve. In this case it can be shown

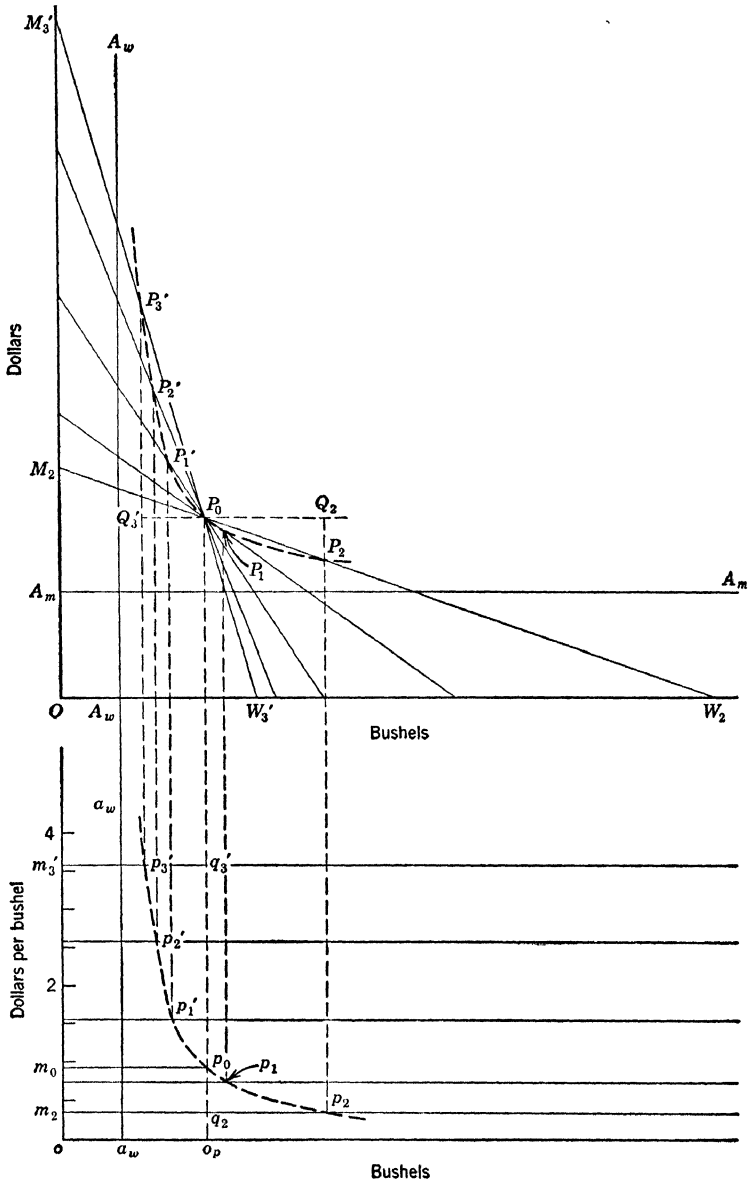


FIG. 22. Perfect markets: the demand-supply curve.

that this curve is a rectangular hyperbola.[5] From this curve the individual demand-supply, or market, curve can immediately be derived, as in the lower part of the figure. This is also a rectangular hyperbola.[5] Thus, take the opportunity line $M_3'P_0W_3'$. In Fig. 22 we suppose the preferred money ratio to be 60%, the preferred wheat ratio to be 40%. The point P_3' is then found by dividing the line $M_3'W_3'$ so that $M_3'P_3'$ is 40% of $M_3'W_3'$, or $P_3'W_3'$ is 60% of $M_3'W_3'$. A perpendicular is then dropped from P_3' to meet $m_3'p_3'$ at p_3', in the lower part of the figure, where om_3' is equal to the slope of the line $M_3'W_3'$, or OM_3'/OW_3'; p_3' is then a point on the individual's supply-demand curve, with its origin at o_p. At a price equal to om_3', or tan $M_3'W_3'O$, the individual will move from P_0 to P_3', and in so doing will sell an amount of wheat P_0Q_3' (or $q_3'p_3'$) and will "buy" an amount of money equal to $Q_3'P_3'$. Similar significance attaches to points (P_2', p_2'), etc. At a price equal

[5] Let m_0, q_0 be the initial stocks of money and of wheat. If m, q are the coordinates of any point on the transformation line at a price p, we have

$$p = \frac{m_0 - m}{q - q_0} \tag{1}$$

This is the equation of the transformation line. If r is the preferred money ratio, we have, at the equilibrium point,

$$r = \frac{m}{m + pq} \tag{2}$$

Eliminating p between these two equations, we get

$$mq - mq_0(1 - r) - qrm_0 = 0 \tag{3}$$

This is the equation of the expenditure-receipts curve. It is a rectangular hyperbola with asymptotes $m = m_0r$ and $q = q_0(1 - r)$. These are the lines A_mA_m and A_wA_w in Fig. 22.

Now, eliminating m between equations (1) and (2), we get

$$pq - pq_0(1 - r) - m_0(1 - r) = 0 \tag{4}$$

This is the equation of the demand-supply, or individual market, curve. It is also a rectangular hyperbola, with asymptotes at $p = 0$ (i.e. the axis oo_p) and at $q = q_0(1 - r)$, which is the line a_wa_w in Fig. 22 If we wish to express this in terms of the quantities of wheat bought and sold, $\bar{q}$, we simply transfer the quantity axis to the new origin o_p by writing $q = \bar{q} - q_0$ in equation (4), which then takes the form

$$p = \frac{m_0(1 - r)}{rq_0 + \bar{q}} \tag{5}$$

The null price, p_n, can be obtained by putting $p = p_n$ and $q = q_0$ in equation (4). We then obtain

$$p_n = \frac{m_0(1 - r)}{q_0r} \tag{6}$$

to om_2, or tan M_2W_2O, the individual will *buy* an amount of wheat equal to P_0Q_2 (q_2p_2) and will "sell" an amount of money equal to Q_2P_2. There will be some price, om_0, at which he will neither buy nor sell. This we call the "null price." The transformation line at this price (not shown in the figure) will touch the expenditure-receipts curve at the point P_0.

Objection may be made to this analysis that it assumes that the preferred asset ratios are independent of the price, and that it therefore suffers from the same defect as the indifference-curve analysis. In the preferred-asset-ratio analysis, however, the defect is easily rectified, for we can readily put into our system a functional relationship between the price and the preferred asset ratios. In the geometrical construction we do not now divide each transformation line in a constant proportion, but in a varying proportion given by the function relating the price to the preferred asset ratios. The expenditure-receipts curve will not now, of course, be a rectangular hyperbola; neither will the demand-supply curve. Depending on the nature of the function relating price to asset ratios, any type of expenditure-receipts or demand-supply function can be obtained. Generally, of course, although past changes in price, by the effect which they have on expectations of future changes, will influence the preferred asset ratios very considerably, there is not much reason to suppose that the absolute level of price will affect these ratios, except in so far as the absolute level is regarded as "high" or "low" relative to some level regarded as "normal." If this is the case a "high" price may predispose the marketer to expect a fall, and a "low" price to expect a rise. Then the preferred money ratio will be higher at high prices than at low: points such as P_3' in Fig. 22 will be moved upwards towards M_3', points such as P_2 will be moved downwards towards W_2, and the expenditure-receipts curve will become flatter than the rectangular hyperbola form. The demand-supply curve will similarly be flattened.

Derivation of market price: the market equation

The derivation of the market price and the quantity exchanged during a "trading day" from all the demand-supply, or market, curves of the various marketers in a market is familiar to most economists. A graphic exposition will be found in my *Economic Analysis*.[6] The market curves of all the marketers are added horizontally: the resulting curve is the total market curve, or excess demand-supply curve, showing the total of offers to buy or sell at each price. The market equilibrium price is

[6] Boulding, *Economic Analysis*, Rev. Ed. (Harper & Brothers, 1948), Chapter 5.

that at which the total excess demand or supply is zero, or, what is the same thing, the total quantity demanded is equal to the total quantity offered for sale. In analytic terms, if $q_1 = F_1(p_1)$, $q_2 = F_2(p_2)$, etc., are the individual demand-supply curves of the various marketers, at the market price we have

$$\Sigma q_1 = \Sigma F_1(p_1) = 0 \tag{7}$$

One advantage of the preferred-asset-ratio analysis is that it enables us to derive a simple and useful identity for the market price. Suppose that $r_1, r_2, \cdots, r_n$ are the preferred liquidity ratios of the various marketers, $m_1, m_2, \cdots, m_n$ their initial money stocks, $q_1, q_2, \cdots, q_n$ their initial commodity stocks, $\bar{q}_1, \bar{q}_2, \cdots, \bar{q}_n$ the quantities of commodity bought or sold by each. Then from equation (5), footnote 5, p. 52, we have, as the equation of marketer 1's individual market curve,

$$\bar{q}_1 = \frac{m_1(1 - r_1)}{p} - r_1 q_1 \tag{8}$$

Similar equations hold for all the others. Then if p is the equilibrium price and the excess demand, $\sum_1^n \bar{q}_1 = 0$, we have

$$p = \frac{m_1(1 - r_1) + m_2(1 - r_2) + \cdots + m_n(1 - r_n)}{r_1 q_1 + r_2 q_2 + \cdots + r_n q_n} \tag{9}$$

This may be called the "market equation," and it is most useful and instructive. The parameters r_1, r_2, etc., as we have seen, are very convenient for expressing the *state of mind* of the marketers: a rise in r_1, for instance, indicates that marketer 1 has increased his "demand for money" and diminished his demand for commodity. The m's and q's are objective data showing the stocks of money and of commodity possessed by the marketers: we have, therefore, a convenient device for separating the effects on price of changes in the objective data on the one hand and the subjective data, or "states of mind," on the other. It will be observed that an increase in the stock of money, a decrease in the stock of commodity, or a decrease in the preferred liquidity ratio of any single marketer will raise the equilibrium price. The greater the preferred liquidity ratio (the r) of any marketer, the greater will be the price impact of a change in his goods stock, and the less will be the price impact of a change in his money stock. The greater the money or commodity stocks of any marketer, the greater will be the price impact of a change in his preferred liquidity ratio. Thus new money that gets into the hands of "hoarders" with high r's will have less effect

on price than if it gets into the hands of "spenders" with low r's. A change in the state of mind of "big men" in the market has a greater effect on price than a similar change in the state of mind of small traders.

Suppose now that the preferred liquidity ratio of all marketers is identical and equal to r. Equation (9) then reduces to

$$p = \frac{M(1 - r)}{Qr} \tag{10}$$

Even if the preferred liquidity ratios of the various marketers differ, equation (10) still holds good if we regard r as in some sense a "market average" of all the individual r's. The market preferred liquidity ratio, r, can be defined as that ratio of liquid assets to the value of total assets at which the market as a whole is content to hold the assets it possesses. If the actual market liquidity ratio is below the preferred ratio, people in the market will feel that they do not have enough money and will try to sell goods in order to acquire more. The attempt to sell goods forces down their price, and with it the total money value of assets, thereby raising the proportion of money to total asset value until the desired proportion is reached. Similarly, if the actual market liquidity ratio is above the preferred ratio, people will try to buy goods to get rid of money; this will raise prices and values until the proportion of money stock to the total asset value declines to the desired level. Price is thus regarded chiefly as a device for adjusting the relative *values* of different kinds of assets until people are content to *hold* the various quantities of these different kinds of assets which exist. The assumption behind this analysis is, of course, that in the price-determining "day" the total stocks of money and of other assets do not change, but are merely shuffled around by exchange among the various marketers; exchange merely rearranges the ownership of assets and does not create or destroy them.[7]

[7] The market preferred liquidity ratio, r, is not a simple average of the individual r_1, r_2, etc., except under rather special conditions. If we equate the two formulae (9) and (10) we have:

$$\frac{M(1 - r)}{Qr} = \frac{\Sigma m_1(1 - r_1)}{\Sigma r_1 q_1}$$

i.e.

$$r = \frac{M\Sigma r_1 q_1}{MQ + M\Sigma q_1 r_1 - Q\Sigma m_1 r_1} = \frac{\dfrac{\Sigma r_1 q_1}{Q}}{1 + \left(\dfrac{\Sigma q_1 r_1}{Q} - \dfrac{\Sigma m_1 r_1}{M}\right)} \tag{11a}$$

The expression $\Sigma r_1 q_1/Q$ $(= \bar{r}_q)$ is the average of the r's weighted by the commodity stocks of the different marketers. The expression $\Sigma r_1 m_1/M$ $(= \bar{r}_m)$ is the average

It is not difficult to extend the market equation to the many-asset market. Suppose that we have many different assets, with total stocks equal to A, B, C, etc., including money, M. Let the price of these assets be p_a, p_b, p_c, etc. Now we can suppose that there is a preferred market ratio for each asset, r_a, r_b, r_c, etc., and for money, r_m. The preferred market ratio for any asset is that ratio of the value of the asset to the total value of all assets at which the market is content to hold the quantity of the asset which it actually finds in its possession. If now the total value of assets is T, we have for commodity A

$$r_a = \frac{p_a A}{T} \tag{12a}$$

Similarly for money we have

$$r_m = \frac{M}{T} \tag{12b}$$

Eliminating T between these two equations, we have

$$p_a = \frac{M r_a}{A r_m} \tag{13a}$$

Similarly we have for the other assets

$$p_b = \frac{M r_b}{B r_m}, \qquad p_c = \frac{M r_c}{C r_m} \tag{13b}$$

Models involving the market equation

The significance of these relationships should be carefully assessed. They are, of course, identities, derived essentially from the definitions of the r's weighted by the *money* stocks of the marketers. We can therefore rewrite (9)

$$r = \frac{\bar{r}_q}{1 + \bar{r}_q - \bar{r}_m} \tag{11b}$$

If the ratio of money to commodity stocks possessed by each marketer were the same, so that $\frac{q_1}{m_1} = \frac{q_2}{m_2} = \cdots = \frac{q_n}{m_n}$, then $\bar{r}_q = \bar{r}_m$, and the market ratio becomes the weighted arithmetic average of the individual ratios, weighted by either the money or the commodity stock of each. If the amounts of money and of commodity stock possessed by each marketer were equal ($q_1 = q_2 = q_n$, and $m_1 = m_2 = m_n$), the formula would reduce to the simple unweighted arithmetic average.

The fact that the market ratio is a complex weighted average of the individual ratios means, of course, that it can change without any change in the individual ratios simply by a change in the stocks held by the individuals. This, however, is quite reasonable: if an individual becomes more or less important in the market, even if his preferences did not change we should expect the market preference ratio to reflect the change in his significance.

of the related variables, and in particular from the value identity which serves to define price as the ratio of the money value of a stock of assets to the physical quantity of the stock. Identities, however, can be extremely useful in analysis if the component variables are *descriptive parameters*, i.e. if they are quantities which are useful in describing significant aggregates or component parts of a system. It may be maintained that all the variables of the market equations (13) are significant descriptive parameters. The total stocks of money and of other assets are historically determined by the processes of production and consumption up to the date of the market, and are equal to the excess of the production over the consumption of each asset since its first production. The preference ratios are significantly stable parameters for describing the general frame of mind of the people in the market. They are not, of course, completely stable, in the sense that there can be changes in the states of mind of individuals which do not affect the preference ratios. Thus the preference ratios of an individual himself may be functions of the prices of assets; also, the market preference ratios, being complex weighted averages of the individual preference ratios, may not reflect changes in the individual ratios because of offsetting changes in the asset stocks held. Nevertheless over substantial ranges these may be regarded as secondary instabilities and do not invalidate the general usefulness of the preference-ratio concept.

If the preference ratios and the asset stocks can be regarded as invariant with respect to prices the system of price determination in equations (13) is, of course, complete. It is interesting to develop the "comparative statics" of such a system. An increase in the quantity of money will increase all prices in the same proportion. An increase in the quantity of any asset, A, will diminish its price, p_a, but will not change the prices of any other assets. A change in the preference ratio for any asset, however, will affect not only its price but also the prices of other assets, because the sum of all the preference ratios must be equal to one, i.e. $r_a + r_b + r_c + \cdots + r_m = 1$. If, therefore, our preferred ratio for A rises, say, from 20 to 25% of our total assets, our preferred ratio for all other assets must fall from 80% to 75%. A rise in preference for A, therefore, will not only raise the price of A; it will lower the prices of all other assets taken together. It need not, of course, affect the price of all assets: if the increase in r_a is wholly compensated for by a decline in r_b, there need be no change in r_c or in the price of C.

If the preference ratios are not independent of the other variables the problem is more complicated, but as long as we can write each preference ratio as a function of the other variables and of other given factors the system is determinate, for with each unknown we shall add

an equation. Thus, if the preference ratios are themselves functions of the prices and quantities of assets, we can write in general:

$$r_a = F_a(p_a, p_b, \cdots, A, B, \cdots, M)$$

$$r_b = F_b(p_a, p_b, \cdots, A, B, \cdots, M)$$

and so on. Complementary and competitive relationships can also be described in these functions. If asset A is complementary to asset B an increase in B will increase the preference for A: i.e. $\frac{\partial r_a}{\partial B}$ will be positive. If asset A is competitive with asset C then an increase in C will lower the preference for A: $\frac{\partial r_a}{\partial C}$ will be negative. If two assets are independent the cross partial differential will be zero. If assets A and B are complementary an increase in the quantity of B will cause not only a decline in p_b but a rise in p_a. If A and B are competitive an increase in B will cause a fall in both p_a and p_b. There may also be reactions on other assets, depending on the way in which changes in one preference ratio affects the others.

Dynamic models involving difference equations or differential equations can also easily be constructed on the basis of the market equations. Thus we can assume that the preference ratio is a function of the price and the rate of change of prices, or the difference between the price of one day and of the day before. We could assume, for instance, that the preferred liquidity ratio was high when prices were high, as then people would expect a fall in prices and would want to hold a larger proportion of their assets in the form of money; similarly, the preferred liquidity ratio would be low when prices were low, as then people would expect prices to rise and would want to hold their assets in the form of goods or securities. If we suppose a lag of one day in the reaction, so that the r_m of one day is determined by the p_m of the day before, a simple two-day cycle of the "cobweb" variety will be set up, a high price one day leading to a high r_m and a low price the next, which leads to a low r_m and a high price again on the third day. This cycle may be explosive, damped, or recurrent, depending on the nature of the function involved. More complex models can be set up on the assumption that r_m is a function not only of yesterday's price but of the day before yesterday's price, and so on. The assumption that r_m is greater the higher the price, and smaller the faster prices are rising leads to the familiar "expectational cycle." Starting with low prices, r_m is low; as prices rise it falls further, accentuating the rise; as prices rise, however, they eventually become "high." Then the absolute height ultimately

more than counterbalances the effect of the rise, and r_m begins to rise; prices fall, which accentuates the rise in r_m and the fall in price until prices are "low" again and r_m begins to fall on that account. Models of any degree of complexity may be constructed, though it may be doubted whether any but the simplest have much interpretive value. These models, however, open up a distinct possibility for future empirical work, especially as information on asset holdings improves. Where stocks of various kinds of assets are known the above analysis gives a simple method for separating the "objective" from the "subjective" elements in price determination, and hence should give a simple method for the study of expectations.

The above examples will have given sufficient indication of the usefulness of the preferred-asset-ratio method. In the next chapter we will develop still further applications to the theory of exchange and production.

Chapter 4

The Marketer in an Imperfect Market

In the preceding chapter we assumed that the marketer was faced by a perfect market, in the sense that he could buy or sell any quantity of his commodity within the range of his interest at a price which was not dependent on his own behavior. This assumption implies, of course, in the analysis of the market, the familiar assumptions of perfect competition—large numbers of traders in close contact, all selling a homogeneous commodity. It is now necessary to remove the limitation of this assumption.

There are two very distinct cases which fall under the general head of market imperfection. One is usually known as monopolistic competition, in which the marketer runs into certain *market resistances* as he expands either his sales or his purchases, forcing him either to take a lower price, or to increase his selling cost, as he increases the volume of his sales, and forcing him to offer a higher price, or to increase his buying cost, as he expands his volume of purchases. In this situation the resistance is a simple function of the volume of sales or purchases and is essentially a result of product differentiation. The second case is oligopoly in its many forms, in which the marketer has to face a problem of *market strategy*. Here we cannot assume that there is any single or simple exchange transformation function of assets in the market. Where the marketer is able to go depends now on the policies and behavior of others *whose policies are consciously affected by his policy*. Hence he cannot regard his future opportunities as independent of his present behavior. The general problems connected with the theory of strategy are highly complex and will not be treated here beyond the simplest cases. For a more elaborate investigation the reader is referred to von Neumann and Morgenstern's *Theory of Games and Economic Behavior*.[1]

[1] J. von Neumann and O. Morgenstern, *The Theory of Games and Economic Behavior* (Princeton, 1944).

Exchange in imperfect markets

Let us consider first monopolistic markets, or market resistances. In this case the more the marketer buys, in general, the higher the price he will have to pay, and the more he sells, the less the price he will receive. The transformation function is not, therefore, a straight line but will be curved as in Fig. 23. Instead of the marketer being passive in regard to the price, and active in deciding how much he will sell at the price given to him by the market, as in a perfect market, the marketer is now active in regard to the price, and passive in regard to the amount which will be bought from him, or offered to him, at the price which he himself sets. Thus in Fig. 23 we have the initial asset position P_0, as in Fig. 18. At a price equal to tan $Q_1P_0N_1$ the marketer can sell P_0N_1: in order to sell more, however, e.g. an amount P_0N_2, the marketer must lower his price to tan $Q_2P_0N_2$. In the figure the opportunity (transformation) function exhibits a maximum value of money at Q_m, indicating that the market resistances beyond this point are great enough to make his sales curve relatively inelastic; i.e. in order to diminish his holdings of goods further by sale he will have to lower the price so much that the increased sales will not compensate for the lower price. Similarly we have supposed that the function exhibits a maximum quantity of goods at Q_g', indicating that the market's "sales curve for money" is likewise inelastic beyond this point. Attempts beyond this point to induce the market to part with more of the commodity by offering a higher price will fail, and what is more, may actually result in a diminution of the amount of commodity which the marketer can purchase. These two maximum points need not, of course, be manifested in the first quadrant, though there is nothing improbable about their presence.

The transformation function $Q_2P_0Q_1'$ is drawn on the assumption of *uniform pricing*—that is, it is assumed that the marketer charges or pays the same price for all units which he sells or buys. It may be possible, however, for the marketer to practice price discrimination, i.e. to charge different prices for different units of what he sells, or to pay different prices for different units of what he buys. This is only possible if there is no recontract in the market. If the individuals who buy from the marketer can resell to other individuals who also are his actual or potential customers, then any attempt at price discrimination will result in the resale of the commodity by those who got it cheap to those who are being charged dear. The marketer will then find that the "cheap" part of his market expands at the expense of the "dear" part, and if this process goes to completion he will sell all his commodity

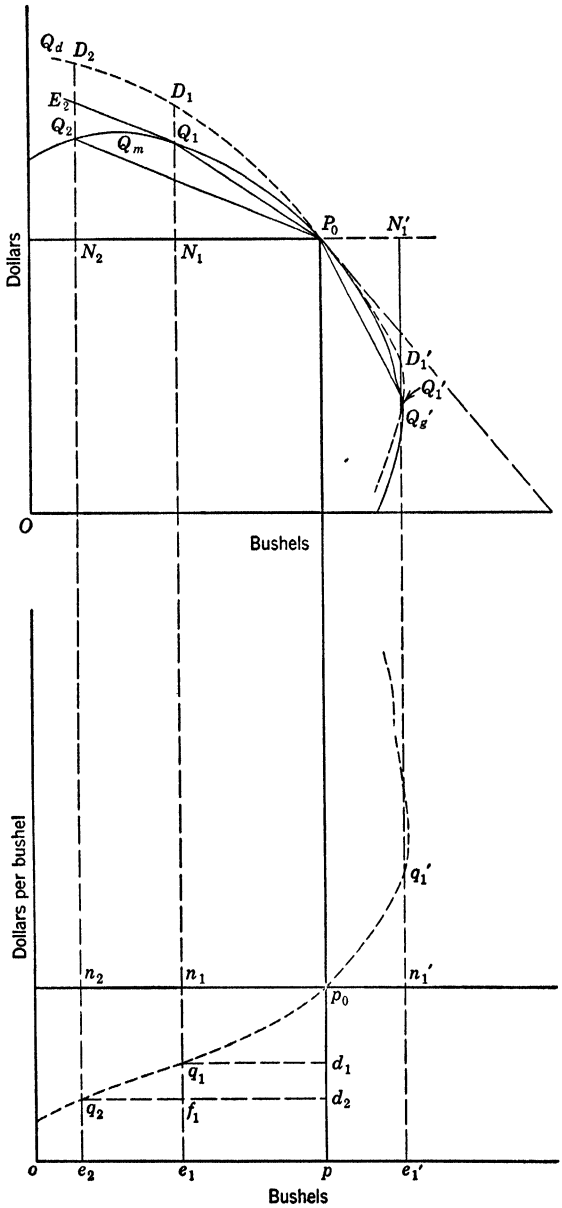


FIG. 23. Imperfect markets: the sales-purchase curve.

at the lowest price which he charges. If, however, the individuals with whom he trades are not all in economic contact with each other, price discrimination will be possible.

Some of these principles can be illustrated more conveniently by means of the lower half of Fig. 23, in which the data of the upper half are presented on a field in which the vertical coordinate is dollars per bushel instead of dollars. Thus e_2q_2, e_1q_1, $e_1'q_1'$ in the lower figure are equal to the tangents of the angles $N_2P_0Q_2$, $N_1P_0Q_1$, and $N_1'P_0Q_1'$ in the upper figure: they represent the uniform prices which must be charged if amounts equal to P_0N_2 (pe_2), P_0N_1 (pe_1) are to be sold, or if P_0N_1' (pe_1') is to be bought. The price pp_0, equal to the slope of the curve $Q_1P_0Q_1'$ at P_0, is the price at which the marketer can neither buy nor sell. The curve $q_2p_0q_1'$ may be called the "sales-purchase" curve. It shows how much the marketer *can* sell or buy at various prices, i.e. how much the market is willing to take from him or to offer to him when he undertakes to sell or to buy at uniform prices. It must be distinguished very sharply from the demand-supply curve of Fig. 22, which shows how much the marketer would be willing to buy or to sell if he were given the chance to buy or sell in unlimited quantities at the various prices.

Now let us suppose that the sales represented by the quantity pe_1 (or P_0N_1) are made to a group of buyers who are not in contact with the buyers of the amount represented by e_1e_2 (or N_1N_2). Then the marketer could charge a price e_1q_1 to the first group and e_2q_2 to the second group. His transformation function would then be $P_0Q_1E_2$ where Q_1E_2 is drawn parallel to P_0Q_2. He would end up with an increase of N_2E_2 in his money stock, equal to the area of the two rectangles $e_2q_2f_1e_1$ and $e_1q_1d_1p$, instead of an increase of N_2Q_2, equal to the rectangle $e_2q_2d_2p$, which is all he would have acquired if he had charged a flat price of e_2q_2 (tan $Q_2P_0N_2$) to all buyers. The finer the price discrimination is shaded, the more money the marketer will be able to get for a given amount of commodity. With perfect price discrimination, in which that price just sufficient to overcome the resistance to the sale is charged for each unit, the amount which the marketer could obtain for the amount pe_2 would be the area $e_2q_2p_0p$, or N_2D_2. The locus of all such points as D_2 is the dotted line $D_2D_1P_0D_1' \cdots$. This is the transformation function under perfect price discrimination, and represents the theoretically most favorable opportunity open to the marketer with the given market. In practice, of course, it is never possible to discriminate so finely as to reach this theoretical maximum, and the limit of the opportunity function will lie somewhere between the no-discrimination and the perfect-discrimination curves.

It should be observed that the transformation function represents the *boundary* of what may be called the "opportunity area," which is the area bounded within the transformation line. The individual, by dint of throwing assets away, can reach any point within the area bounded by the transformation line. Unless one or both of the assets are discommodities, however, for any point inside the boundary of the opportunity surface we can always find a point on the boundary which represents a larger quantity of *both* assets and will therefore be preferred. If, however, one or both of the assets are discommodities within the boundary of the transformation line it is quite possible for the preferred point to be within the opportunity area, rather than on its boundary. In this case the usual assumption that the preferred point must lie on the boundary of the opportunity area does not necessarily hold.

Somewhere on his opportunity field, therefore, and usually somewhere on its boundary, the individual will have a preferred point to which he will move by buying and selling. Given the preferred point, we know what prices he must charge or offer, and how much he will sell or buy. We can, of course, define the preferred point by means of the indifference-curve analysis, and suppose that it is that point where the opportunity boundary touches an indifference curve. In view of the instability of the indifference-curve system, and its possible dependence on price or other market conditions, such analysis remains pretty much at the formal level, and as in the case of perfect markets we need to explain the preferences of the marketer in terms of the magnitude, direction, and certainty of his expectations of future market situations. The profit-making process under imperfect markets is essentially similar to that under perfect markets: it consists in transforming assets into money before a flattening of the transformation curve, and transforming them into goods before its steepening. Thus we will get a zigzag course in the asset structure similar to that of Fig. 18, with the exception that the transformation paths are now curved instead of being straight lines. In this case also the wider the swings the greater the rate of profit: on the other hand also the wider *each* swing, the greater the risk of loss if expectations are not fulfilled. Hence the marketer will again follow each transformation curve up or down to the point where he feels the marginal fear of loss just balances the marginal hope of gain from future changes in market opportunity.

Preferred asset ratios in imperfect markets

The question naturally arises whether the preferred asset ratios can be used with imperfect markets as a convenient "parameter of preference." The answer seems to be that on certain assumptions they may

so be used, but that their use raises acute problems relating to the *valuation* of assets in an imperfect market. The asset ratios at any point of the transformation line depend, of course, on the method of valuation of the various assets—in our simple case of the two-asset firm, on the method of valuation of the "wheat." Thus suppose that in Fig. 24 P_0 is the original asset position, QP_0Q' is the imperfect-market transformation curve without price discrimination. We suppose that Q_1 is the preferred point on this curve, and that the marketer therefore

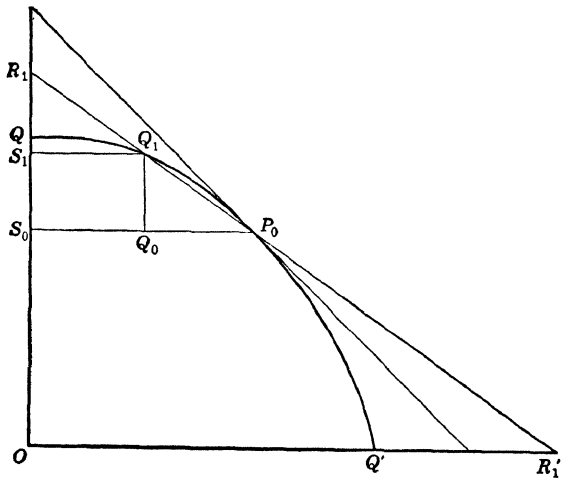


FIG. 24. Valuation in imperfect markets.

sells an amount of wheat P_0Q_0, and acquires Q_0Q_1 of money, at a price tan $Q_1P_0Q_0$. If now the remaining stocks of wheat, S_1Q_1, are valued at the selling price of that "day," their value will be S_1R_1, and the total net worth will be OR_1. If the line R_1P_0 is produced to meet the horizontal axis in R_1', then the "money ratio" at Q_1 is Q_1R_1'/R_1R_1', and the "wheat ratio" is R_1Q_1/R_1R_1'. As at each point on the transformation curve there is a definite asset ratio, the movement along the curve could be interpreted as a search for the *preferred* asset ratio, this being a simple measure of asset preference. As the marketer *sells*, the wheat ratio declines and the money ratio grows, for two reasons: both because he is diminishing his physical stocks of wheat and increasing his physical stocks of money, and also for a reason not present in perfect markets—because in order to sell more he must diminish the *valuation* which is put on a unit of wheat. Similarly as the marketer buys he increases

his wheat ratio and diminishes his money ratio, both because of that change in physical quantities and also because in order to buy more he must increase the value of wheat per unit. He changes his asset ratio towards wheat to take advantage of a future rise in the demand for his wheat or, what is the same thing, a fall in the demand for his money: because he is not certain of this future, however, he will not go to the extreme in transforming his assets into wheat, but will stop at some point where he feels that the wheat and money ratios are "about right" in view of the chance of disappointment. Similarly he changes his asset ratio towards money in expectation of a rise in demand for his money or, what is the same thing, a fall in demand for his wheat: here again he will not go all the way possible, but will stop at some asset ratio which again seems right in view of the uncertainty of the future.[2]

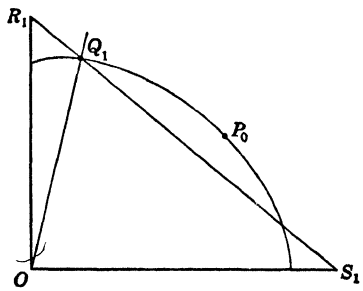


FIG. 25. Valuation at constant price in imperfect markets.

In the above construction it is assumed that inventories are valued at sales price. They are not necessarily so valued, either in economic theory or in accounting practice. If followed consistently, sales-price valuation results in the same type of meaningless fluctuation in net worth (and therefore in profits) on each revaluation that we saw in Fig. 19 (p. 44). Here also many other systems of valuation are possible, each of which may result in a different asset ratio in value terms for each point in the physical-asset field. Nevertheless, provided that the system of valuation is consistent, the asset ratios, and therefore the preferred asset ratios, have a definite meaning. Thus in Fig. 25 the valuation of wheat is done at a constant valuation coefficient, independent of the selling or buying price. Then through any point such as Q_1 on the transformation function we draw a line R_1S_1 the slope of which is equal to the constant valuation coefficient. The wheat ratio is then R_1Q_1/R_1S_1, the money ratio Q_1S_1/R_1S_1. At every point there is clearly a perfectly definite asset or money ratio. The "preferred"

[2] A convenient geometrical construction to find that point on the transformation curve which has a *given* asset ratio is as follows: through P_0 draw a hypothetical receipts-expenditure hyperbola $P_3'P_0P_2$ (as in Fig. 22) corresponding to the given asset ratios. In this case, of course, this is merely a construction line and does not represent alternative possibilities for the marketer. Where this constant-asset-ratio hyperbola cuts the transformation curve is the required point.

asset ratio can now be expressed by a straight line from the origin.[3] Thus all points on the line OQ_1 represent a money ratio of Q_1S_1/R_1S_1, as long as the valuation coefficient for wheat is equal to tan R_1S_1O. In order to find the point on the transformation curve which corresponds to any given asset ratio, we draw the line from the origin corresponding to this asset ratio and find where it intersects the transformation curve.

The analysis with a fixed valuation coefficient applies equally well, of course, to price discrimination: no matter how the transformation curve is defined the above construction enables us to find the point with the preferred asset ratio. Valuation at the "market price," however, becomes somewhat ambiguous under price discrimination; should the valuation be done at the average price received for the total sales (or paid for the total purchases), or should it be done at the "marginal price" received (or paid) for the last unit? There is no "right" answer to this question, in view of the essentially arbitrary nature of the valuation process: we can only state that, provided that one method of valuation is consistently adopted, the concept of a preferred asset ratio will have some descriptive value.

Bilateral monopoly

In the above analysis we have assumed in some sense the "impersonality" of the market, regarding it from the point of view of the individual marketer simply as an exchange opportunity, the character of which does not depend on his own decisions. We must now investigate another aspect of the exchange problem, in which the element of strategy enters, and in which the opportunities facing each marketer are clearly determined by the behavior of others, which behavior in its turn may be conditioned by the behavior of the first marketer. The simplest case of such a condition is "bilateral monopoly," or simple exchange between two parties. Here the path traversed in the exchange of assets is a *mutual* path: exchange is seen clearly as a *redistribution* of assets of various kinds between the two exchangers. This is illustrated in Fig. 26. Here we suppose two exchangers A and B, exchanging two assets, which we will again call "money" and "wheat." As we are considering pure exchange, we will suppose that the total quantity of the two assets held by *both* exchangers together is constant, and that exchange merely results in the rearrangement of ownership. We start therefore with a rectangle, O_aHO_bK, where the side O_aH ($= O_bK$) represents the total quantity of wheat in the possession of the two exchangers,

[3] If k is the slope of a line from the origin, m and w the quantity of money and wheat at any point on the line, v the valuation coefficient, and r the money ratio, we have $r = m/(wv + m)$ and $m = kw$: whence $k = rv/(1 - r)$.

and O_aK ($= O_bH$) represents the total quantity of money. Then any point inside this rectangle represents one possible *distribution* of the total assets between the two marketers. We measure A's quantities from the origin O_a, B's from the origin O_b. Thus the point P_0 represents a position where A has M_aP_0 and B has P_0M_b of wheat, and A has N_aP_0 and B has P_0N_b of money. No matter where P_0 is within the rectangle we have $M_aP_0 + P_0M_b = O_aH$, and $N_aP_0 + P_0N_b = O_aK$. Exchange at a fixed price will be represented by movement

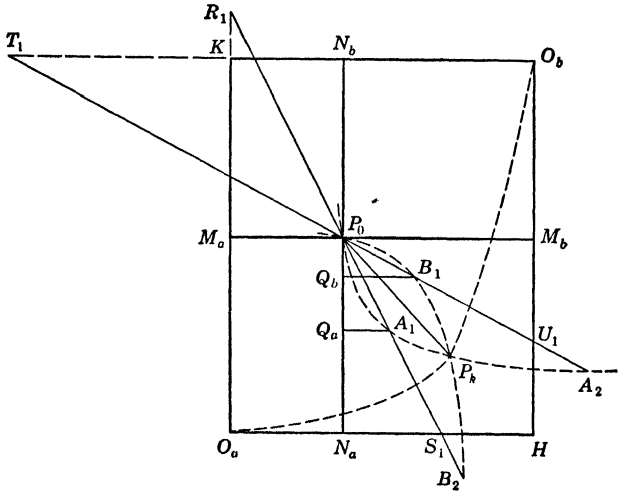


FIG. 26. Bilateral monopoly.

along a straight line through P_0, the slope of which is equal to the exchange price. Thus the movement from P_0 to A_1 would indicate that A had given up, and B acquired an amount of money P_0Q_a in return for an amount of commodity Q_aA_1 at a price tan $P_0A_1Q_a$.

Now let us suppose that each marketer has preferred asset ratios of his own. For marketer A an expenditure-receipts hyperbola, $P_0A_1P_k$, can be drawn, any point on which, such as A_1, shows that combination of assets which yields A his preferred asset ratio at the price tan $P_0A_1Q_a$. In this case we have drawn the A-hyperbola on the assumption of a preferred money ratio, r_a, of 25%, so that $R_1S_1 = 4A_1S_1$. Similarly an expenditure-receipts hyperbola can be drawn for B, with O_b as origin, $P_0B_1P_k$, any point on which, such as B_1, shows that combination of assets which yields B his preferred asset ratio at the price tan $P_0B_1Q_b$. In this case we have assumed that B's preferred money ratio, r_b, is 75%, so that $T_1B_1 = \frac{3}{4}T_1U_1$.

Suppose now that a price is set equal to tan $P_0A_1Q_a$, so that R_1S_1 is the mutual transformation line. Exchange will take place until the point A_1 is reached. At this point A will refuse to exchange any further; and as exchange cannot take place without the consent of both parties, the exchange will go no further. B would still like to continue exchanging, as he has by no means reached his preferred asset ratio: at this price he would like to go on to B_2. In the figure B_2 is outside the rectangle, indicating that at this price B would like to acquire more money than both parties possess together, but this is not the limiting condition: B cannot satisfy his asset preferences at this price because of A's unwillingness to go beyond A_1. Similarly, if the price were tan $P_0B_1Q_b$, A would like to go exchanging up to A_2: he cannot do this, however, because B will reach his preferred asset ratio at B_1, and will go no further.

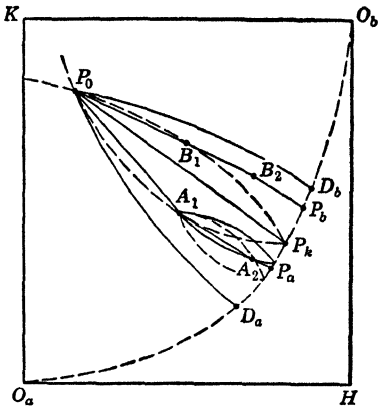


FIG. 27. Price discrimination in bilateral monopoly.

It is clear that there is only one price in these circumstances at which the preferred asset ratios of both marketers can be satisfied simultaneously. This is the price equal to the slope of the line P_0P_k, where P_k is the point of intersection of the two expenditure-receipts hyperbolae. At this price the marketers will move to P_k, where both will be satisfied. It is shown below that no matter where the point P_0 lies in the field, the point P_k must lie on the dotted hyperbola $O_aP_kO_b$. By analogy with a construction of Edgeworth, this may be called the *contract curve*. At any point in the field not on the contract curve no price can be found at which the preferred asset ratios of both marketers are simultaneously satisfied. It can be shown that, provided there is possibility of each party quoting a price, there is no stable position of equilibrium which is not on the contract curve. Thus suppose that in Fig. 27 we start from the position P_0, and B quotes a price equal to the slope of P_0A_1. Exchange takes place until we arrive at A_1, where A will stop exchanging. B, however, is not satisfied; and in order to persuade A to buy more he will lower the price. The transformation line now becomes A_1A_2; A now is willing to go on to A_2, where the transformation line again cuts his new expenditure-receipts hyperbola. If A_2 is not on the contract curve B will still be unsatisfied, and will still

have a chance to improve his asset ratios by lowering his offer price. This process will stop only when the contract curve is reached and both parties are satisfied.

In the above case we have assumed that B is active in regard to price policy, but allows A to determine the quantity exchanged. This is possible because the initial price (under the conditions of Fig. 27, where in the initial position B has an excess of wheat and will be a seller of wheat, and A has an excess of money and will be a buyer of wheat) is greater than the "mutual price" equal to the slope of P_0P_k. We could also assume, of course, that A is active in regard to price and B determines the quantity exchanged. In this case the path to the contract curve will follow a course such as $P_0B_1B_2P_b$. This represents price discrimination by A against B, just as the path $P_0A_1A_2P_a$ represents price discrimination by B against A. The more finely the discrimination is carried out—i.e. the more segments there are in the price-discrimination paths—the further out will the paths lie: the limiting position of perfect price discrimination is given by the curves P_0D_b and P_0D_a.

The final position of equilibrium must lie on the contract curve between D_a and D_b. The lines P_0D_a and P_0D_b are true indifference curves: D_a represents a position which is only just as good as P_0 in A's estimation, and D_b represents a position which is only just as good as P_0 in B's estimation. The closer the final equilibrium point is to D_a, the more B benefits from the exchange; the closer it is to D_b, the more A benefits. It will be observed that the successive price adjustments always take place in the same direction as long as the same party remains active in regard to price. Alternations of activity and passivity on the part of one individual are, of course, possible, but as the active-passive relationship is likely to be determined by the character of the individuals or by the sociological setting of the exchange it is likely to be a fairly constant relationship. Thus in retailing it is an almost universal convention in western society for the seller to be active in regard to the price and for the buyer to determine the quantity taken. In exchange between farmers and dealers or wholesalers it is frequently customary for the buyer to set the price, and for the seller to decide how much he will sell. There are, of course, frequent situations in which sociological, political, or legal considerations intervene to prevent an "economic" adjustment. Thus there are considerable sociological and even legal pressures against the practice of price discrimination. If price discrimination is impossible the quoter of the price must stick by it; if a seller quotes a price that is higher than the mutual price the buyer will stop short of the seller's preferred point, and the seller has no recourse in requoting the price. Similarly if a buyer quotes a price that is less

than the mutual price, the seller will stop short of the buyer's preferred point. In the absence of price discrimination, therefore, there is a strong inducement for either the buyer or the seller to quote the mutual price, equal to the slope of P_0P_k in Figs. 26 and 27. Under these circumstances the problem of simple bilateral exchange is more "determinate" in the economic models than has frequently been supposed.

It is shown in the appendix to this chapter that, if M_a, M_b, Q_a and Q_b represent the initial quantities of money and of wheat held by two marketers, and if r_a, r_b are their preferred money ratios, the mutual price, p_k, is given by the formula

$$p_k = \frac{M_a(1 - r_a) + M_b(1 - r_b)}{r_aQ_a + r_bQ_b} \tag{1}$$

This is essentially the same as the "market equation" [equation (9) on p. 54] showing the market price in a perfectly competitive market. There is no guarantee, of course, that in bilateral exchange the mutual price will be the one actually arrived at. Nevertheless there are pressures tending to establish it as a likely price, especially in the absence of price discrimination.

Exchange with more than two marketers

The next step is to extend the analysis to more than two marketers. This is done in Fig. 28, where within the rectangle O_aHO_dK it is possible to represent all the possible distributions of the total quantity of money (O_aK) and the total quantity of goods (O_aH) among a number of marketers. If there are N marketers $N - 1$ points within the rectangle are sufficient to indicate the division of the total stock of assets among them. In the figure we suppose four marketers, A, B, C, and D: three points, P_b, P_c, and P_d, define the positions of the rectangles marked A, B, C, and D, the sides of which show how much goods and money belong to each marketer. Thus the total amount of money, O_aK, is divided into O_aM_a for A, P_bM_b for B, P_cM_c for C, and P_dM_d for D. Similarly the total amount of commodity, O_aH, is divided into O_aQ_a for A, P_bQ_b for B, P_cQ_c for C, and P_dQ_d for D. It is evident that any combination of asset distributions in the market can be represented by various positions of the three points P_b, P_c, and P_d. Any combination of exchanges in the market can likewise be represented by shifts in these three points along opportunity lines the slopes of which are equal to the prices or ratios of exchange. Consider, for instance, the movement from P_b, P_c, P_d to P_b', P_c', P_d'. Marketer A has bought Q_aQ_a' of goods, and paid M_aM_a' for it. Marketer B has increased his holdings of goods from P_bQ_b to $P_b'Q_b'$, i.e. has bought $R_b'Q_b'$, where Q_bR_b' is drawn

parallel to P_bP_b', and has given up $S_b'Q_b$ of money. Marketer C has not changed his holdings at all: $P_cQ_c = P_c'Q_c'$, and $Q_cP_d = Q_c'P_d'$. Marketer D has sold M_dM_d' of goods and received Q_dQ_d' in money. The total amount bought of each asset must, of course, equal the total

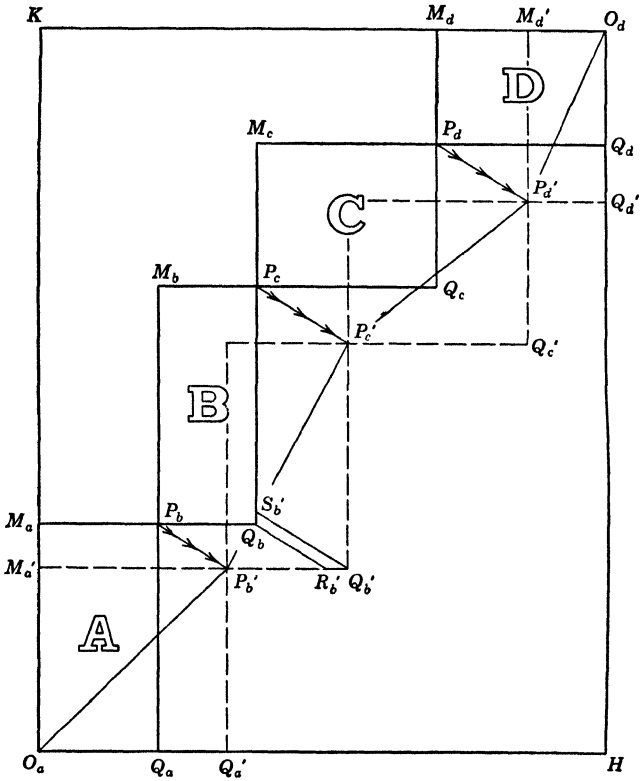


FIG. 28. Exchange with four marketers.

amount sold: in this diagram it is clear that A and B have bought $Q_aQ_a' + R_b'Q_b' = M_dM_d'$ of goods from D, and have received corresponding amounts of money. It will be observed that, although in this particular case it is possible to discern from the diagram exactly who exchanged what with whom, in the general case it is possible to discern only the net exchanges of each marketer, not his individual transactions with particular buyers or sellers.

It is shown in the appendix to this chapter that here also there is normally only one price at which all the marketers can move to the

position which satisfies their preferred asset ratios, on the assumption that the valuation of the commodity is performed at the market price. The "mutual price," p_x, is again given by the familiar formula

$$p_x = \frac{\Sigma M_a(1 - r_a)}{\Sigma Q_a r_a} \tag{2}$$

In the discussion of exchange we have assumed so far that the valuation ratio at which assets are valued is the same as the price of the last unit of the asset bought or sold. This assumption is not, however, necessary to the logic of the argument, and models can be constructed on any assumption as to the valuation of assets. An interesting case is that in which the commodity asset is valued by both parties to the exchange at a constant ratio, which may, however, be different for each party. The solution is shown in Fig. 29 for the two-exchanger case. The rectangle O_aHO_bK shows the total stock of assets and its distribution, as in Figs. 26 and 27.

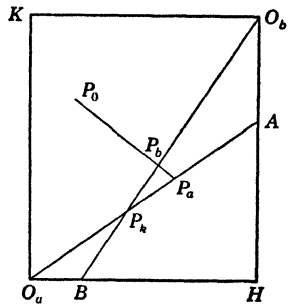


FIG. 29. Exchange with fixed valuation ratios.

If now marketer A has a constant preferred asset ratio, and if he performs his valuations at a fixed valuation ratio, all points at which his preferred asset ratio is satisfied lie on a straight line from the origin such as O_aA, the slope of which depends on the valuation ratio and on the preferred asset ratio, as we have seen on p. 49. Similarly all points at which B can satisfy his preferred asset ratio lie on a line such as O_bB. Where these lines intersect, at P_k, is the only point at which both parties can be satisfied. The somewhat unrealistic nature of this particular assumption is shown by the fact that any path to this point would suit the parties equally well, as their valuations are independent of the exchange ratio. In this case, however, there will be little or no tendency for the mutual price to be an equilibrium price, or for the marketers to move to the point P_k. If, for instance, we start off with A quoting a price equal to the slope of P_0P_a, with A passive as to quantity, B will move to P_b, and then will not wish to go further. A would like to go to P_a, but cannot because of B's unwillingness. Once having arrived at P_b, however, no change in price short of a negative price will induce B to move further. On these assumptions, therefore, any point on the line $O_aP_kO_b$ will be a stable equilibrium. The "contract curve" here is the broken line $O_aP_kO_b$.

Profit-making and the zero-sum game

The above analysis throws a good deal of light on the profit-making process in pure exchange. It is clear, both in the two-marketer case and the many-marketer case, that if the total quantities of assets of all kinds in the market remain constant, and if the valuation ratios are constant, then the total *value* of all assets is constant. Under these circumstances the gain of one marketer is bound to take place at the expense of another: in the language of the theory of games, what we have is a "zero-sum" game. These relative gains and losses occur in two ways. One marketer may improve his position at the expense of another by successive exchanges at different prices at different times. Thus the marketer who buys cheap and sells dear will benefit at the expense of those who buy dear and sell cheap. There will be a zigzag path in the "market rectangle" similar to that of Fig. 18, p. 44, which will result in an increasing quantity of assets in the possession of the successful marketer, and a corresponding diminishing quantity in the possession of the unsuccessful. This is the process of "pure" speculation, or arbitrage through time. The other method of change in relative position is simple arbitrage, i.e. purchase from a marketer who is selling cheap and sale to one who is buying dear. This, of course, can happen only when there are more than two marketers involved. In the case of Fig. 28, for instance, if the lines P_cP_c' and P_dP_d' diverged, marketer C could improve his position at the expense of the others, for he can reach positions in which he has at least as much of one asset as he started with, and more of the other.

The only way in which there can be gains by all parties in a "pure market" with fixed asset quantities circulating among the marketers is through a revaluation of assets. If the valuation coefficients of the non-liquid assets are raised, the total value in money (dollar) terms of all assets rises, and hence the gain of one does not necessarily involve the gain of all. Such a gain, however, from the point of view of the whole market, can certainly be regarded as "spurious," even though to an individual a gain by revaluation may represent a genuine improvement of his *relative* position. The transformation of the "zero-sum" game of the pure marketer into the "positive-sum" game of economic activity as a whole can come about only through the *creation* of assets, i.e. through production. It is to the theory of production, then, that we must turn in continuing the analysis. It must not be inferred, of course, that the activity of exchange is somehow in essence "unproductive." The market in some form or other, i.e. the exchange of specialized products, is absolutely necessary if the division of labor,

i.e. the specialization of production, is to take place, and it is a commonplace of economics that specialization is the principal source of increased productivity. Nevertheless it can be stated with some confidence that exchange which does not in some way give rise to specialized production is a "zero-sum" game, and that the main justification for asset transformations through exchange is the effect which they have on the creation of assets through transformations in production.

Appendix to Chapter 4

The mutual-price equilibrium

The equation for the contract curve of Fig. 26 may be derived as follows. Let M_a, Q_a be the initial stocks of money and of commodity possessed by marketer A. Let r_a be his preferred money ratio, and let m_a, q_a be his final stocks of money and commodity. Let M_b, Q_b, r_b, m_b, q_b be the corresponding quantities for marketer B. Let p_k be the mutual price. Then by definition,

$$r_a = \frac{m_a}{p_k q_a + m_a} \tag{1}$$

$$r_b = \frac{m_b}{p_k q_b + m_b} = \frac{M - m_a}{p_k(Q - q_a) + M - m_a} \tag{2}$$

as $m_a + m_b = M$, the total stock of money possessed by the two marketers, and $q_a + q_b = Q$, the total stock of commodity possessed by them. Eliminating p_k between equations (1) and (2), we have

$$(r_a - r_b)q_a m_a + Qr_b(1 - r_a)m_a - Mr_a(1 - r_b)q_a = 0 \tag{3}$$

This is the equation of the contract curve of Fig. 26, measured from the origin O_a. It is a rectangular hyperbola with asymptotes at

$$m_a = \frac{r_a(1 - r_b)M}{r_a - r_b}, \qquad q_a = \frac{-r_b(1 - r_a)Q}{r_a - r_b}$$

If $r_a = r_b$, i.e. if the preferred money ratios of the two marketers are equal, equation (3) reduces to

$$Qm_a - Mq_a = 0$$

This is the straight line joining the origins, $O_a O_b$ in Fig. 26. The contract curve always goes through O_a and O_b, as may be seen by putting $m_a = 0$, $q_a = 0$, or $m_a = M$, $q_a = Q$ in equation (3).

Now, if p_k is the price at which the exchange has been made, the value of money given up (or acquired) is equal to the value of goods bought (or sold). That is,

$$M_a - m_a = p_k(q_a - Q_a) \tag{4}$$

Eliminating q_a between equations (1) and (4), we have

$$p_k Q_a r_a = m_a - M_a r_a \tag{5}$$

Similarly we have

$$p_k Q_b r_b = m_b - M_b r_b \tag{6}$$

Adding equations (5) and (6), as $m_a + m_b = M = M_a + M_b$, we have

$$p_k = \frac{M_a(1 - r_a) + M_b(1 - r_b)}{Q_a r_a + Q_b r_b} \tag{7}$$

In general, in the n-marketer case, we have

$$\sum_1^n m_a = M = \sum_1^n M_a \tag{8}$$

and summing all equations of type (5) for each marketer we have

$$p_k = \frac{\sum_1^n M_a(1 - r_a)}{\sum_1^n Q_a r_a} \tag{9}$$

This is the same formula that we developed for the case of perfect competition on p. 54, equation (9). It is therefore a perfectly general formula for the mutual price. It is only under conditions of perfect competition, however, that there is any clear and positive tendency for this price to be realized.

We can also find the equilibrium holdings, at the mutual price, for each marketer. Thus, solving equations (1) and (4) for q_a and m_a, we have, for marketer A,

$$q_a = \left(\frac{M_a}{p_k} + Q_a\right)(1 - r_a) \tag{10}$$

$$m_a = (M_a + p_k Q_a) r_a \tag{11}$$

Similar equations give the equilibrium positions of the other marketers. Dividing (11) by (10), we have:

$$\frac{m_a}{q_a} = p_k \frac{r_a}{1 - r_a} \tag{12}$$

Similar equations obtain for the other marketers. This set of equations gives us a simple geometrical solution to the problem of the mutual-price equilibrium in the many-marketer case. The mutual price is calculated

from formula (9). The ratio m_a/q_a is then obtained from equation (12). A line from O_a, O_aP_b' in Fig. 28, is then drawn with the slope m_a/q_a, and a line from P_b is drawn with negative slope equal to p_k. Where these lines intersect at P_b' is A's equilibrium position. Similarly the line $P_b'P_c'$ is drawn with slope m_b/q_b to meet P_cP_c' with slope p_k in P_c', which gives the equilibrium position of B. Similarly the whole course of the equilibrium line, $O_aP_b'P_c' \cdots O_d$ can be traced. If all the r's are equal, the slopes of the segments O_aP_b', $P_b'P_c'$ are also equal, and the various equilibrium points lie on the straight line O_aO_d.

The analysis to this point has been conducted on two main assumptions: that the preferred asset ratios are constant—i.e. are invariant with respect to the other variables of the problem—and that the assets are valued at the ratio at which they can be exchanged. Neither of these assumptions are vital to the method of analysis, though, of course, their removal complicates the solutions. If we assume, for instance, that the preferred money ratio is a function of the price, so that $r_a = F_a(p_k)$, $r_b = F_b(p_k)$, and so on, these equations must be solved simultaneously with equation (9) to give a determinate solution. Then, of course, the expenditure-receipts curves and the contract curve in the two-marketer case no longer have to be hyperbolae, but if the preference ratios are related to the other variables by known continuous functions these curves can theoretically be derived. Similarly, if the valuation coefficients are functions of some other variables, these functions can be put into the system and a solution usually obtained.

Chapter 5

The General Theory of Preference

The preferred-asset-ratio method of describing the preference structure of the individual is new to economics in its present application, though the roots of the idea may be found far back in the discussion of the quantity theory of money; indeed, the method may not inaccurately be described as a generalization of the quantity theory. Consequently it seems desirable at this point in the analysis to pause in the general argument in order to make clear the basic assumptions and limitations of the preferred-asset-ratio method, and to relate it to other methods of describing preference structures—notably, of course, the indifference-curve method. The reader who is in haste to follow the general argument and who is familiar with the indifference-curve analysis may well proceed to the next chapter.

Indifference curves

The indifference curve is by now such a familiar weapon of economic analysis that a description is hardly necessary in a work of this nature. Indifference curves represent the "contours" of a hypothetical utility or preference function. A single indifference curve (or function) defines all those points in a field of potential choice—i.e. all those combinations of values of the significant variables—to which the chooser is *indifferent.* It is supposed that for any two points in the field of potential choice, A and B, we can always determine whether one is to be preferred to the other or whether they are indifferent; that is, there are three possibilities: A preferred to B, B preferred to A, or A and B equally preferable, i.e. indifferent. Hence of any two indifference curves in the same system we can always say that one is "higher" than the other, i.e. any point on the "higher" indifference curve will be preferred to any point on the lower. This is the only assumption usually necessary to determine the most preferred point in a field, and in particular, we do not have to assume that utility, or preferability, is cardinally measurable.

The method of use of indifference curves is always the same. In the general field of potential choice, representing all conceptually possible combinations of the relevant variables, we have an *opportunity function* limiting the choice to certain areas of the field. The problem is to find the point within the opportunity area which is on the highest indifference curve, i.e. which is the most preferred point. Unless the preference function exhibits a maximum within the opportunity area, this preferred point will be found where the boundary of the opportunity area *touches*

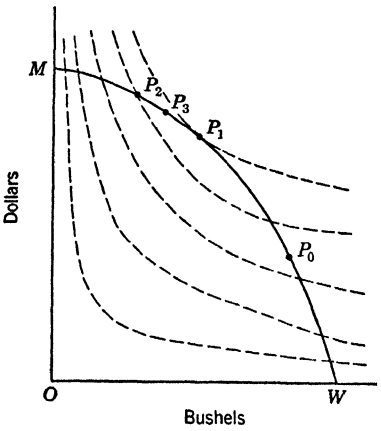


FIG. 30. Indifference curves.

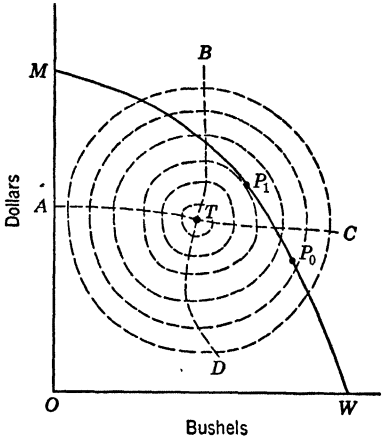


FIG. 31. Indifference curves

an indifference curve. This principle is illustrated in Figs. 30 and 31. We suppose a two-asset field in the case of exchange under imperfect markets. P_0 represents the marketer's initial position; the line MP_0W is the market opportunity line, as in Fig. 23. The dotted lines are indifference curves: they may be regarded as the contours of a "preference mountain" rising above the plane of the figure. The "opportunity area" is that bounded by the opportunity line and the axes (neglecting positions in other quadrants), i.e. the area OMP_0W. The individual can attain any combination of asset quantities within this area but cannot get outside it. We assume that there is a possibility of simple destruction or throwing away of assets to enable him to move towards the origin if he wishes. Then in Fig. 30 the preferred position is P_1, where the opportunity boundary touches an indifference curve: at any other point within the opportunity area it is possible by movement in some direction to reach a higher indifference curve. We may think of the opportunity boundary as a "fence" across the preference "mountain"

which the individual cannot cross: nevertheless he wishes to get as high up on the mountain as the fence permits.

In Fig. 31 the indifference curves are roughly circular, indicating that the preference surface has a maximum within the opportunity area. In such a case the point of tangency, P_1, is not particularly significant. The individual will proceed to the maximum point of the preference surface at T. In the quadrant ATD both assets are commodities; i.e. an increase in either leads to a preferred position. In quadrant BTC both assets are discommodities, e.g. garbage: an increase in either leads to a less preferred position. In quadrant DTC "dollars" are a commodity and "bushels" a discommodity. In quadrant ATB dollars are a discommodity, bushels a commodity. In general we may suppose that any asset becomes a discommodity in a sufficiently large quantity, even though in regard to most of the assets of economic life we are still operating in the quadrant ATD. Hence the indifference curves are likely to be concentric, and roughly circular or oval in shape.

Objections to indifference-curve analysis

The usefulness of indifference curves in economic analysis is beyond question. For some examples of their use, the reader is referred to my *Economic Analysis*.[1] Nevertheless as an analytical instrument they are open to certain objections. One objection is that they say too much: another that they say too little. They say too much because in seeking to cover the whole field of potential preferences they seem to state more than actually exists in the mind. We make choices in particular situations: we do not contemplate making choices in an indefinitely large number of situations. We may perhaps sidestep this objection by saying that it is only that portion of the preference function around the actual position of the individual that is relevant; nevertheless conclusions are often drawn on the assumption that there is a stable system of preferences covering all potential experience.

A shift in preference can be represented only by a complex shift in the whole preference function, and it seems to be impossible to derive from the preference function or the indifference-curve system itself simple parameters which will express simple shifts in preference. Suppose, for instance, that there is a shift in preference towards "dollars" and away from "bushels." This means that the indifference curves will become flatter than before. Thus in Fig. 30 P_2 previously was on a lower indifference curve than P_1, and represents a situation with more dollars and less bushels than P_1. If now dollars become more highly regarded

[1] Boulding, *Economic Analysis*, Rev. Ed. (Harper & Brothers, 1948), Chapter 34.

relative to bushels, the esteem in which P_2 is held will rise relative to P_1, and if this movement is sufficient the new indifference curves will go through both P_1 and P_2. This will shift the preferred point from P_1 towards P_2, say, to P_3, as we should expect: the rise in preference for dollars leads to a shift towards dollar holdings. It is difficult, if not impossible, however, to give meaning to the question of *how great* has been the shift in preference towards dollars in terms of the indifference-curve analysis, because of the fact that what is involved is not a single parameter but a whole multidimensional functional relationship. An answer might perhaps be found in terms of the slope of the indifference curve—an important quantity which is usually called the *marginal rate of substitution*. The marginal rate of substitution is different, however, at different points in the field. Indifference curves cannot possibly have a constant slope unless the assets are perfect substitutes in the mind of the holder. If we could assume that a given shift in preference produced a constant proportionate change in the marginal rate of substitution at each point, such a change might be taken as measuring the change in preference: however, such a constant proportionate change is most unlikely.

A more serious objection to the indifference-curve analysis is that, as usually employed, at least, it assumes that the preference system is invariant with respect to the opportunity function, and in particular is invariant with respect to the prices of the assets. As a first approximation this may be justified; it would be rash indeed, however, to assume that such invariance always exists. There is the familiar case of the "display good," e.g. diamonds, in which the preference function clearly depends on the price of the good—preference for it being stronger at high prices, as the preference arises from the fact that the good is a symbol of wealth. The phenomenon may be much more general, however, than economists have supposed. If there is a strong sense of "normality" in the attitude towards price—that is, if there is some level of price which is felt to be "normal"—prices above that level being regarded as "high" (and therefore likely to fall), whereas prices below that level are regarded as "low" (and therefore likely to rise), then the absolute level of price will exert a marked influence on the preference structure, the preference for an asset being lower at a high price than at a low price. This is not the same thing, it should be observed, as the negative slope of a demand curve, for this is drawn usually on the assumption of a fixed underlying preference function *from which* the "law of demand" is derived.

This objection is not, of course, fatal to the *principle* of maximizing a preference function, though it may well be fatal to a good many of the conclusions which have been drawn, for instance, in welfare eco-

nomics. Thus, if we assume that "utility" depends on prices as well as on the quantities of assets, then even in the two-asset case we can no longer describe the preference structure by a single set of indifference curves in two dimensions, and we must draw indifference surfaces in a three-dimensional figure where, say, quantity of money, quantity of "wheat," and price of wheat are the three axes. Certain possible properties of this system are worth investigating. There are two principal reasons why we may expect price to affect preferences. One is the fact that the *money value* of the total asset holdings, and hence the net worth of the individual in dollar terms, depends on the "valuation coefficient" of "wheat," which in turn either may be equal to, or may be some function of, its price. Thus, if a man has $1000 cash and 500 bushels of wheat, at a dollar a bushel his net worth is $1500, at three dollars a bushel it would be $2500.

This effect is sometimes called the "money illusion," as it depends upon the fact that our accounting is done in terms of "dollars" rather than "bushels." If the net worth were calculated in "bushels" (and there is no reason, other than convention, why it should not) then in the above example the net worth of the individual when wheat was a dollar a bushel would be 1500 bushels, and when it was three dollars a bushel (or, what is the same thing, money was a third of a bushel per dollar) the net worth would be $833\frac{1}{3}$ bushels. A rise in the money price thus raises the dollar value of the net worth and lowers the bushel value: it is for this reason that we refer to the effect as an illusion. Nevertheless the illusion may be a strong one where there is a habit of thinking of the dollar as a unit of relatively constant significance, and, indeed, where the belief in the long-run stability of the value is justified, the "money illusion" is much less illusory, and in fact passes over into the second reason for the effect of price on preferences.

This second reason is that the absolute level of price may affect the expectations of future prices and hence may quite "rationally" affect preferences. This is particularly true if there is some level of price which is regarded as "normal." Prices above this are then regarded as "high," and the higher the price above normal, the greater the expectation of a fall and therefore the greater the preference for money. Prices below normal are regarded as "low" and are expected to rise: this shifts the preference towards goods.

The indifference surface

Consider now a section of the three-dimensional price-money-wheat field in the price-money plane, such as in Figs. 32 and 33, quantity of wheat being held constant. A pure money-illusion effect will give indif-

ference curves as in Fig. 32; a "normal-level-of-price" effect will give indifference curves as in Fig. 33, the indifference curves here being sections of the indifference surfaces in the three-dimensional field. In the case of Fig. 32, as we raise the price along MH, keeping both the quantity of money and the quantity of wheat constant, the individual rises to successively higher indifference curves, indicating preferred positions. This, as we have seen, occurs because of the effect of the higher price on the dollar value of net worth. In the case of Fig. 33, as we raise

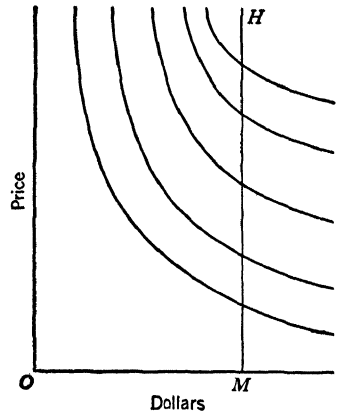


FIG. 32. Price-indifference curves.

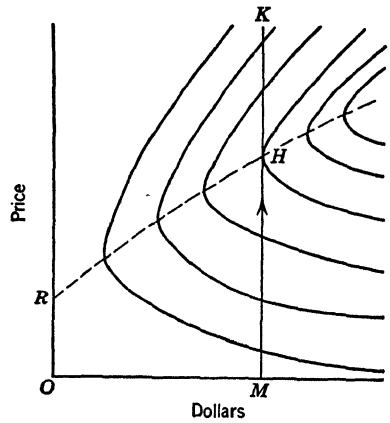


FIG. 33. Price-indifference curves.

price along MK, again keeping the quantity of money and the quantity of wheat constant, we find that "utility" rises to a maximum at H and then falls. This means that there is some price at which the individual is best satisfied with the quantities of money and wheat which he has: at prices higher or lower than this he will be less well satisfied with his existing holdings, because of the effect which the price has on his expectations. The line through the turning points of the indifference curves, RH, may be expected to have an upward slope, as, with a given holding of wheat, the greater the money holdings the more content will the individuals be with high prices, for if high prices produce an expectation of a fall they will cause a high preference for money. The sections of the three-dimensional figure in the wheat-price plane may be expected to have exactly similar properties. If there is a combination of the money-illusion effect with the normal-level-of-price effect the general pattern will still be like that of Fig. 33, but the maximum points will be at higher levels of price. Each three-dimensional indifference surface will be something like a snub nose pointing down towards the origin.

There will also be an opportunity surface in the three-dimensional field, money-wheat-price. It is a helical surface, $p = (m - m_0)/(q_0 - q)$, the contours of which in the q-m plane are the set of opportunity lines through the point S, representing the initial asset position, m_0, q_0, such as A_1B_1, A_2B_2, etc., in Figs. 34 and 35. A line such as A_4B_4 in the three-dimensional figure (Fig. 34) will touch one of the indifference

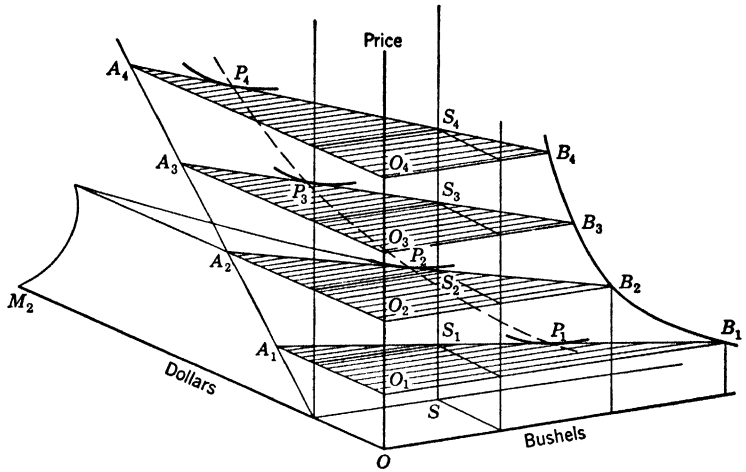


FIG. 34. Indifference and opportunity surfaces.

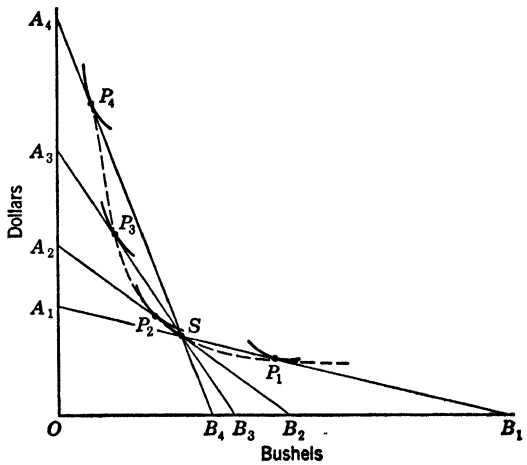


FIG. 35. Indifference and opportunity contours.

surfaces at a point P_4; this is the preferred position at this price. Similarly there are preferred points P_3, P_2, P_1, etc., on the other opportunity lines where these touch an indifference surface. The line $P_1 \cdots P_4$ may be called the preferred-point line. It should be observed that there is no reason to suppose that the points P_1, P_2, P_3, P_4, etc., in Fig. 34 lie on the same indifference surface, though there is nothing inherently impossible in their so doing. Normally, however, we may expect them to lie on different indifference surfaces, and there will be some point on the line P_1P_4 which lies on a lower indifference curve than any other, either above or below it—i.e. which is the least desirable of all the positions on the preferred-point line. The projection of the preferred-point line on the base plane is the line P_1P_4 in Fig. 35, which is the familiar expenditure-receipts curve.

Implications for welfare economics

If now the preferences are independent of price the indifference surfaces in Fig. 34 become vertical semicylindrical surfaces with the indifference-curve system in the m-q plane as base: i.e. the section of an indifference surface is the same at all levels of price. Any such section

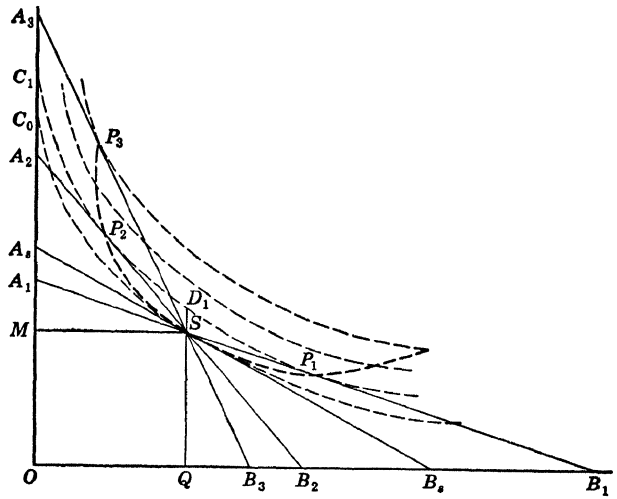


FIG. 36. The compensating payment.

is shown in Fig. 36: the light dotted lines are indifference curves, assumed to be independent of price; the heavy dotted line is the preferred-point line, or receipts-expenditure curve. In this case the preferred-point line touches an indifference curve at S, the initial position of the

marketer, and S is therefore the least preferred point on the preferred-point line. That is to say, the marketer is worst off when the price is equal to his null price, and he will be better off if the price is such that he wants to trade. Trading, that is, always leads to a gain for the trader, as any point on the preferred-point line other than S, e.g. P_1 or P_2, is on a higher indifference curve than S.

If now we take the more general assumption that preferences are a function of price it is no longer true that there is always a gain from trade. The indifference curves at P_1, P_2, P_3, P_4 in Fig. 35 belong to different systems, i.e. to different sections of the three-dimensional indifference surfaces, and there is no reason to suppose that S is the least preferred point on the preferred-point curve. Thus P_2 or P_1 might easily lie on "lower" indifference surfaces than S. If a rise in price, quantities of assets being constant, increases satisfaction (i.e. if we are in the range MH in Fig. 33) then moving from S to P_1 may lead to a decline in satisfaction because it involves a decline in price. If a rise in price diminishes satisfaction then P_2 may represent a less satisfactory position than S. At first sight this phenomenon may seem puzzling, as, of course, trade cannot take place unless the trader feels that he is better off as a result of the trade. At any given price that is not the null price, the point to which the trader moves, e g. P_1, P_2, etc., is a preferred point to S. The asset combination at the point S, however, at the market price may be either more or less satisfactory than the same combination at the null price. The basic reason for this phenomenon is that price has a twofold aspect: it represents on the one hand an opportunity to exchange, and on the other hand a method of valuation. It is only if we assume that the valuation aspects of price can be neglected that the usual indifference-curve analysis applies. This valuation aspect may be, but is not necessarily, irrational. We do not have to assume that valuation is performed at the market price, or that the preference for high asset values (and therefore high prices) is a "money illusion," for valuation depends on expectations of the future, and present price may very legitimately be a guide to these expectations.

This generalization of the theory of preference unfortunately destroys much of the foundation of modern welfare economics, which is valid only in the highly special case where preferences are independent of price. The basic concept of welfare economics is the *compensating payment*. This is essentially a method of valuation by means of preference equivalents. Thus in Fig. 36, if the indifference curve at S cuts the money axis at C_0, the amount of money OC_0 is regarded as equivalent, i.e. equally preferred, to the asset combination represented by the point S. To put the same thing in other terms, the amount of money

MC_0 is the amount that would just suffice to compensate the individual for the sacrifice of his non-money assets, MS. Similarly, if the indifference curve through P_1 cuts the money axis at C_1, the amount C_0C_1 is the "compensating payment" which measures the "gain from trade" to the individual who moves from S to P_1: i.e. it is the payment which would just compensate the individual for the prohibition of trade at the price tan A_1SM. Even in this case there is a possible ambiguity in the concept. If QS is extended to meet the indifference curve from P_1 in D_1, SD_1 represents the addition to the money stock, the goods stock being unchanged, which would compensate for the loss of trade as between positions S and P_1. SD_1 may not be equal to C_0C_1; hence in specifying the compensating payment it is strictly necessary to specify the quantity of the other assets to which the money payment is added. Only if the indifference curves are parallel vertically—i.e. if the marginal rate of substitution (the slope of the indifference curves) is invariant with respect to the quantity of money—is the compensating payment independent of the asset structure.

If now preference is also a function of price, the concept of the compensating payment becomes almost meaningless. The indifference surfaces of Fig. 34 cut the price-dollars plane in a curve. In order to specify a dollar equivalent of any given asset combination, then, we must also specify a price, and there seems to be no particular reason to choose one price rather than another. We may perhaps say that the dollar equivalent of any asset-price combination is found where the indifference surface through that point cuts the money axis, where both price and quantity of goods are zero. Thus in Fig. 34 we suppose that the indifference surface through P_2 cuts the money axis at M_2: then OM_2 is the dollar equivalent (when both price and quantity of goods are zero) of the asset-price combination represented by the point P_2. Such a quantity, however, is scarcely very significant.[2]

Preferred-asset-ratio theory

The next problem is to consider the relation of this more general theory of preference to the preferred-asset-ratio theory of the two previous chapters. We have assumed that in general there is a preference function of the form $U = U(m, q, p)$ or, more generally still in the

[2] The consumer's surplus concept becomes even more dubious as a measure of welfare when we take into account the preference effects of price. The concept is strictly valid only when the marginal rate of substitution is independent of the quantity of money. (See my *Economic Analysis*, p. 762.) When the preference structure depends on the price it is extremely difficult to formulate any meaning for the consumer's surplus.

multiple-commodity case, $U = U(m, q_1, q_2, \cdots, q_n, p_1, p_2, \cdots, p_n)$. Now, if $r_1, r_2, \cdots, r_n$ are the money and the commodity ratios, we have by definition

$$r_m = \frac{m}{m + \Sigma p_1 q_1}, \quad r_1 = \frac{p_1 q_1}{m + \Sigma p_1 q_1}, \quad r_2 = \frac{p_2 q_2}{m + \Sigma p_1 q_1}, \quad \text{etc.}$$

Substituting the value of $m + \Sigma p_1 q_1$ from the first of these identities into the rest, we have

$$p_1 = \frac{m r_1}{q_1 r_m}, \quad p_2 = \frac{m r_2}{q_2 r_m}, \quad \text{etc.}$$

Substituting these values of p_1, p_2, etc., in the general preference function we get a new function,

$$U = \psi(m, q_1, q_2, \cdots, q_n, r_m, r_1, r_2, \cdots, r_n)$$

As the sum of all the r's is unity, one of them can always be derived from all the others and may be omitted from the preference function. In the simple two-asset case we can express the preference function as a series of indifference surfaces in a p-q-r_m diagram, $U = \psi(m, q, r_m)$. This will express exactly the same set of preferences as the indifference surfaces $U = U(m, q, p)$ in Fig. 36. It is evident, then, that we can substitute r_m as a parameter in place of p or, more generally, the whole set of asset ratios in place of the set of prices, and still express exactly the same preference structure.

In the analysis of the previous chapters it was assumed that the *preferred* asset ratios are constant, i.e. that the asset ratios at the preferred points P_1, P_2, etc., of Figs. 34 and 35 are equal. This is, of course, a highly restricted assumption: it leads, as we have seen, to the conclusion that the preferred-point line of Fig. 35 is a rectangular hyperbola. The justification for the assumption is that it is an extremely convenient first approximation, enabling us to substitute a single preference parameter in the description of a preference structure for the complexities of the general preference functions; it thus enables us to say very simply whether the preference for any asset has increased or decreased. It is quite possible now to extend this method further and to suppose that the preferred ratios themselves are functions of the price. This immediately releases the preferred-point line from being a rectangular hyperbola: from any preferred-point line $P_1P_2P_3$, etc., in Fig. 35 a function relating price to the preferred money ratio can be derived. We may reasonably expect to find, for instance, that at high prices there is a greater preference for money and at low prices for commodity. This (in Fig. 35) will push P_4 away from A_4 towards B_4, and will push P_1

away from B_1 towards A_1. These movements may be sufficient to give the preferred-point curve a minimum value for both money and wheat, i.e. will make it circular rather than hyperboloid in shape.

Although the assumption that the preferred asset ratios are a function of price is sufficient to give us the preferred-point curve (or expenditure-receipts curve), and hence the demand-supply curve of the individual, it is not sufficient to give us the relative degree of preference for points on the preferred-point line; i.e. it does not tell us whether P_2 is on a higher indifference surface than P_1. In general this problem must be solved by reference to the general preference surface of Fig. 34. If the asset ratios at the preferred points are equal, the three-dimensional indifference surface can be simplified into a set of curves in the dollars-bushels plane which may be called "equipreference curves," to distinguish them from indifference curves. In the bushels-dollars field of Fig. 37 the lines from the origin OE_1, OE_2, etc., represent combinations with a constant asset ratio at various prices. Thus if r_m is the money ratio, we have by definition

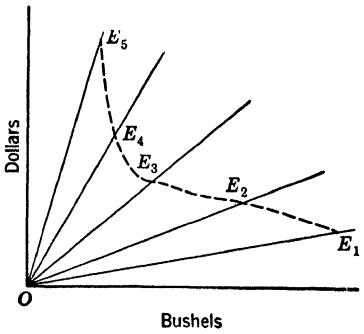


FIG. 37. The equipreference curve.

$$r_m = \frac{m}{m + pq}, \quad \text{or} \quad \frac{m}{q} = \frac{r_m}{1 - r_m} \cdot p$$

The lines OE_1, OE_2, etc., may be regarded as the contours of the surface of constant r_m in the pqm field of Fig. 34. This surface cuts one of the three-dimensional indifference curves in a line which, projected on to the dollars-bushels plane, is the line $E_1E_2 \cdots E_5$. This may be called an *equipreference* curve: each point on it yields equal satisfaction to the individual, under the condition that the price is given by the formula $p = m(1 - r_m)/qr_m$. There will, of course, be a whole system of equipreference curves, one curve for each level of satisfaction. They are to be distinguished from indifference curves, which are sections of indifference surfaces cut by a $p = p_k$ plane, i.e. price being constant. Only if satisfaction is invariant with respect to price, so that the indifference surfaces in Fig. 34 are cylindrical, with a constant cross section in the horizontal plane, will the indifference curves and equipreference curves be the same.

It must be emphasized that the method of asset preference ratios does not involve different basic assumptions from the generalized indifference function method, i.e. it is in no sense "non-Euclidean" economics. It represents merely a rather arbitrary simplification of the *functions* involved—a simplification which can be justified on the grounds of convenience in certain problems. In this sense it is somewhat analogous to the arbitrary simplification of functions which is involved, for instance, in the use of the concept of elasticity, or in the Marshallian assumption of constant marginal utility of money. This latter assumption is, of course, equivalent to assuming that the marginal rate of substitution (the slope of the indifference curves) is independent of the quantity of money, depending only on the quantity of the other asset. This yields indifference curves which are "parallel" in a vertical direction—an assumption which clearly can hold true only over very small areas if the indifference curves are generally circular in shape. In spite of their limitations there is great value in these simplifications of the preference function, especially in so far as they reveal the *significant* properties of the general preference function.

Bilateral exchange under general preference theory

The complexities involved in the use of the general preference function can be illustrated dramatically by reference to the theory of bilateral exchange developed in the previous chapter under the preferred-asset-ratio simplification. The usual exposition of this problem is in terms of indifference curves which are invariant with respect to price. Thus in Fig. 38 O_aHO_bK is the rectangle within which we can show the distribution of two assets between two owners, as in Figs. 26 and 27, p. 68. P_0 again represents the initial position. We postulate two systems of indifference curves, one for each marketer; then the "contract curve" is the locus of all points at which an indifference curve of one set is tangent to one of the other set, as at D_a or D_b. $P_0A_1P_k$ is part of the expenditure-receipts curve of marketer A; i.e. it is the locus of all points of tangency between A's indifference curves and successive straight lines through P_0, representing alternative opportunity lines at constant prices. Similarly $P_0B_1P_k$ is the expenditure-receipts curve of marketer B. These two curves intersect at P_k, which must, of course, be on the contract curve. The price represented by the slope of the line P_0P_k is what I have earlier called the "mutual price": it is that at which neither party is restrained from continuing exchange by the unwillingness of the other to continue. The line P_0P_k touches indifference curves of both individuals at P_k. A's indifference curve through P_0 ($P_0D_aP_x$) marks the path of extreme price discrimination by B; B's indifference

curves through P_0 ($P_0D_bP_x$) marks the limit of extreme price discrimination by A. At any point within the area $P_0D_aP_xD_b$ both parties are at least as well off as they are at P_0: exchange clearly cannot lead to any point outside this area. Actually exchange will never go beyond the area $P_0D_aD_b$, as to go beyond the contract curve would worsen the

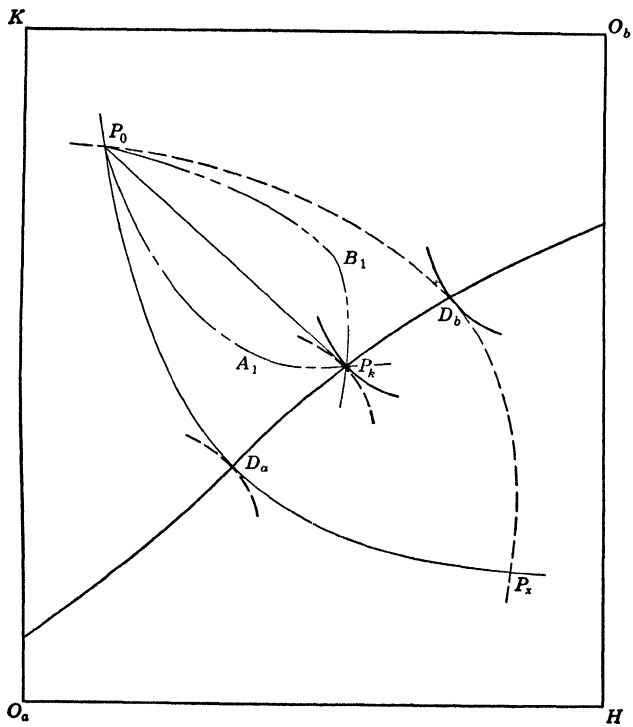


FIG. 38. Bilateral exchange analyzed by indifference curves.

position of one of the parties. If recontract is possible—i.e. if either of the parties can change the price offered once an equilibrium has been reached at some previous price—the parties will move to some point on the contract curve between D_a and D_b: exactly where depends on the relative abilities at price discrimination, on relative "activity" in respect to price, or on relative bargaining power.

The exact formal similarity between the analysis of Fig. 38 and that of Fig. 27 will strike the reader immediately. That of Fig. 38 is apparently more general: thus on the indifference-curve analysis the contract curve does not have to be a rectangular hyperbola, nor does it have to

go through the origins. As we have seen, however, the apparent lack of generality of the preferred-asset-ratio method can be removed if it is assumed that the preferred asset ratio is itself a function of price. The really difficult problem arises, however, when we try to generalize the indifference-curve type of analysis to take account of the effect of price itself on preferences. In Fig. 38 we would then have a different

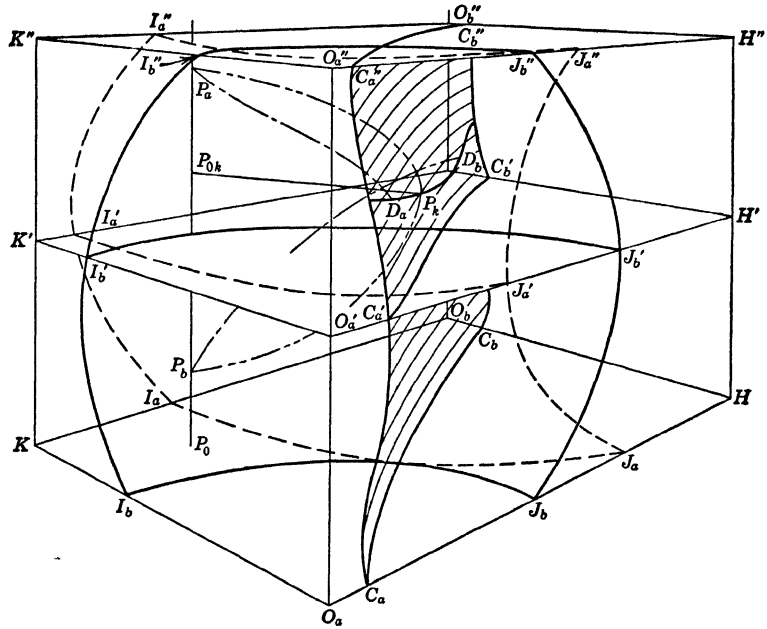


FIG. 39. Bilateral exchange.

system of indifference curves and a different contract curve at each price. We still have the curves of extreme price discrimination, but they no longer represent indifference curves. There may possibly be a number of different "mutual prices," as each price carries with it its own system of indifference curves. The expenditure-receipts curves also have meaning, but it is no longer necessarily true that moving either individual on his expenditure-receipts curve away from P_0 improves his position, as at each point on the curves we have a new system of indifference curves.

The three-dimensional analysis of the problem is shown in Fig. 39. Here we measure dollars along O_aK, bushels along O_aH, and price vertically along O_aO_a''. There will be indifference surface systems for

both marketers: the dotted curves I_aJ_a, $I_a'J_a'$, $I_a''J_a''$ are the sections of one of A's indifference surfaces in planes corresponding to prices of O, O_aO_a', and O_aO_a''. The solid curves I_bJ_b, $I_b'J_b'$, $I_b''J_b''$ are the corresponding sections of one of B's indifference surfaces. In each plane of constant price such as O_aHO_bK, $O_a'H'C_b'K'$, $O_a''H''O_b''K''$ there will be a contract curve such as C_aC_b, $C_a'C_b'$, $C_a''C_b''$, each being the locus of mutual tangency of the indifference curves in that plane. These contract curves lie in an irregular surface, shaded in the figure. On this contract surface there is a contract line $D_aP_kD_b$, which is the locus of all points where the indifference *surfaces* touch. P_0 represents the initial asset position. P_0P_a is the highest price at which A is willing to trade (he will, of course, trade only an infinitesimal quantity at this price). Similarly P_0P_b is the lowest price at which B is willing to trade. The line from P_a to D_a on the contract line is the curve of maximum price discrimination against A. It lies on A's indifference surface through P_a, at points where the slope of the tangent in a constant-price plane is equal to the constant price. The line P_bD_b is a similar curve of maximum price discrimination against B, lying on B's indifference surface through P_b. The volume enclosed within these two indifference surfaces represents all those combinations of quantities and price which are not disadvantageous to at least one party. The lines P_aP_k and P_bP_k are expenditure-receipts curves, the projections of which are the curves $P_0A_1P_k$ and $P_0B_1P_k$ in Fig. 38. Returning to Fig. 39, P_0P_{0k} is the "mutual price." We are assuming here in effect that the vertical axis is the "marginal price," i.e. the price of the last unit bought or sold. The significant price from the point of view of the preference structure may, of course, be different from this, and the analysis would become proportionately more complicated.

The complexity of this general preference analysis hardly seems to be justified by the meager additional results obtained from it. Nevertheless it is necessary to include it for the sake of completeness, and also perhaps to illustrate the justification for shortcut parameters of preference such as the preferred asset ratios. In succeeding chapters we shall employ whatever method of describing preference seems most fruitful, without seeking great exactness or refinement.

Chapter 6

The Theory of Production and Inventory

Up to this point we have treated only of exchange and of asset transformation through exchange. Exchange is not, however, the only method of asset transformation, and the analysis must now be extended to include asset transformation through production.

Production under perfect markets

Production, from the point of view of the firm, is a form of internal exchange, whereby certain assets such as raw materials and money are transformed into finished goods. Production usually involves the purchase of labor, which, again from the point of view of the firm, represents a transformation of liquid assets (the money paid out as wages) into some kind of product. The production or physical cost function, describing what kinds and quantities of assets must be sacrificed in order to produce various quantities of product, is, of course, a many-dimensional affair, involving relationships among many kinds of assets, money, raw materials, machines, finished goods, etc. However, these relationships can be summed up in a *total cost curve*, assuming that the purchase of raw materials, labor, etc., results in a diminution of money stocks, and that we can postulate a certain relationship between the quantity of product and the minimum diminution of money stocks which is necessary—other assets being held constant—in order to produce these various quantities of product.

This relationship is shown in Fig. 40. As before we assume a balance sheet with only two items in it, money stocks (measured vertically) and commodity stocks (measured horizontally). The initial position of the firm is shown at P_0. $P_0P_1P_2$ is the firm's total cost curve, showing how much money must be given up to produce various amounts of commodity. Thus if the firm produces Q_1P_1 of commodity it must give up P_0Q_1 of money: P_0Q_1 is the "cost of production" of Q_1P_1. The total cost curve in this figure is a "production opportunity line," as it

shows all combinations of money and commodity to which the firm can move by production. The marginal cost is the slope of this production opportunity line, showing how much money must be given up in order to produce an extra unit of commodity. In the figure it has been assumed that marginal cost increases continually with output; with a decreasing marginal cost the production opportunity line would be convex to the origin; if marginal cost is constant the production opportunity line is straight.

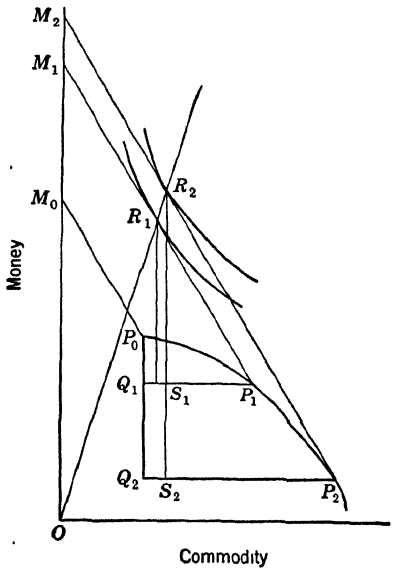


FIG. 40. Production in perfect markets.

We suppose, furthermore, in Fig. 40 that the firm has the opportunity to exchange commodity for money in a perfect market at a price equal to the slope of the lines P_1M_1, P_2M_2. If, then, the firm decides to produce an amount equal to Q_1P_1, it moves from P_0 to P_1 along the production line and then can move to any point along P_1M_1 by exchange. The line P_1M_1 may be called an *exchange path*. Some point on this line, R_1, will be preferred to any other. As the price in this case is not a variable of the problem there are no objections to the simple indifference-curve method, and we can assume that at R_1 the opportunity line touches an indifference curve. If the preferred asset ratios are constant all the preferred points will lie on a straight line from the origin, OR_1R_2, as the price, or rather the valuation ratio, is assumed constant.

If, therefore, the firm decides to produce Q_1P_1 it will sell P_1S_1, moving finally from the point P_0 to the point R_1 by a combination of production and exchange. In the figure, however, it is clearly possible for the firm to reach a higher indifference curve by producing more, i.e. by moving further down the production line: as it does so the exchange path P_1M_1 moves outward, touching successively higher indifference curves. At the point P_2, however, where the exchange path touches the production line, the optimum output is reached: if the firm produces more than this the exchange path will retreat towards the origin. The point R_2, therefore, is the best point which the firm can attain by the combina-

tion of production and exchange open to it; Q_2P_2 is the optimum output, of which S_2P_2 will be sold and Q_2S_2 added to inventory.

The money profit made by the firm is the increase in the total value of its assets. The total value of its assets at P_0 is OM_0, where P_0M_0 is an exchange or valuation path. This assumes that assets are valued at market price. Similarly the total value of assets at P_1 (or R_1) is OM_1, and at P_2 (or R_2) is OM_2. The total profit from producing Q_1P_1 is therefore M_0M_1; from producing Q_2P_2 is M_0M_2. The maximum profit is clearly M_0M_2, where M_2P_2 touches the production curve.

Some interesting conclusions follow from this analysis. This first is that the firm will produce that amount at which the marginal cost (the slope of the production opportunity, or total cost curve) is equal to the price at which the product can be sold, assuming perfect markets. Thus the familiar condition of the elementary marginal analysis is seen to be true even in this more complicated case. The second conclusion is that the amount produced depends only on the total cost curve and the price of the product, and does not depend on the assumptions we make in regard to the indifference curves or preference systems. No matter what assumption is made regarding preferences, the firm will always prefer to be on line P_2M_2 rather than on line P_1M_1, provided that neither money nor goods are a discommodity, for if we take any point R_1 on the line M_1P_1 a range of points on the line P_2M_2 can be found which represent greater amounts both of money and of commodity. Thus the profit-maximizing assumption in the theory of production seems to be independent of the preference system; what the preference system determines is the *distribution* of assets among their various forms. For the theory of production in perfect markets all we need to assume is that the total value of assets is maximized. The position of the indifference curves will determine where, on the exchange path P_2M_2, the point of equilibrium R_2 will lie: the optimum point, however, must lie somewhere on P_2M_2, and no matter where R_2 is—i.e. no matter how much of the product is *exchanged*—the total profit M_0M_2 is not affected, for exchange is assumed to be of equal values and does not affect the total of assets.

Production under imperfect markets

If, now, we remove the assumption of perfect markets the problem becomes more complicated. It is illustrated in Fig. 41. P_0 is the original position of the firm, $P_0P_1 \cdots P_4$ the total cost, or production opportunity, curve, as before. Instead of straight exchange paths we now have curved paths, $P_1K_1R_1$, $P_2K_2R_2$, etc., showing the opportunities for exchange in an imperfect market. In the figure we have assumed that the price must be lowered in order to obtain a larger volume of sales: hence the exchange

paths curve downwards. The shape of the paths $P_1K_1R_1$, $P_2K_2R_2$, etc., is assumed to be constant, indicating that the price depends only on the quantity which the firm wishes to sell, not on the amount of assets which it possesses. The various exchange paths will have an envelope, the dotted line $P_4K_3K_2K_1$; at P_4 the exchange path and the envelope touch the total cost curve. The position of equilibrium clearly must lie on the

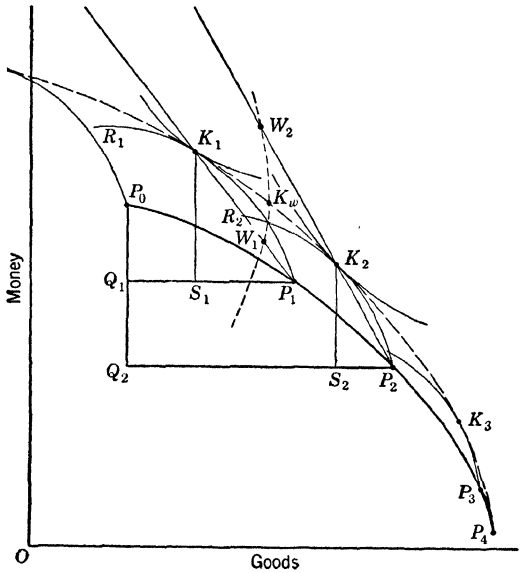


FIG. 41. Production in imperfect markets.

envelope curve, because for any point on a price path which is not on the envelope curve a point can be found on another exchange path at which the quantities of both assets are greater. The equilibrium point K_2 is, then, where the *envelope* curve is touched by an indifference curve, assuming that the indifference curve in question is consistent with the price. If P_2R_2 is the exchange path which goes through K_2 then an amount Q_2P_2 will be produced, the price charged will be tan $K_2P_2Q_2$, and the quantity sold will be P_2S_2.

The important conclusion emerges that when markets are imperfect the total amount produced is *not* invariant with respect to the asset-preference system because the price is now a variable of the problem. A shift in the preference system—e.g. a general fall in the marginal rate of substitution of money for commodity—will shift the point of equilibrium

from K_2 to K_1. Production declines from Q_2P_2 to Q_1P_1. Price falls from tan $K_2P_2S_2$ to tan $K_1P_1S_1$. Sales increase from P_2S_2 to P_1S_1.

It remains true that at the point of equilibrium the marginal revenue of sales is equal to the marginal cost of the amount of output produced. The slope of the envelope curve at K_2 is equal to the slope of the production (total cost) curve at P_2, both being equal to the marginal rate of substitution of money for commodity. The "marginal cost equals marginal revenue" condition is not, however, sufficient in this case to determine the volume of output and sales, as this equality holds at all corresponding points on the envelope curve and the total cost curve.

It is not difficult to apply the method of preference ratios to this problem. Divide the price lines P_1K_1, P_2K_2, etc., at points W_1, W_2, etc., so that, as in Fig. 21 (p. 49) the intercept between the axes is divided in a constant ratio r, the preferred liquidity ratio of the firm. Then the dotted locus W_1W_2, etc., may be called the preferred ratio curve. Where this intersects the envelope curve, K_w, is the equilibrium point. An increase in liquidity preference, measured by a rise in the preferred liquidity ratio, moves the preferred-ratio curve W_1W_2 to the left. It consequently moves the equilibrium point K_w to the left, resulting in a decline in production, an increase in sales, and a fall in the price charged. This is the same as the result obtained previously by assuming a decline in the marginal rate of substitution of money for commodity at each point, which is another way of measuring a rise in liquidity preference: it now takes more commodity to compensate for the loss of the more highly esteemed unit of money. This result is what we should expect. Production involves giving up liquid for non-liquid assets; hence we should normally expect an increase in liquidity preference to make people less willing to produce. Sales involve giving up non-liquid for liquid assets; hence we should expect a rise in liquidity preference to increase the willingness to sell. The lower the price charged, the easier it is to sell, i.e. to acquire liquid assets. Hence it is not surprising to find that a rise in liquidity preference results in a fall in the price.

Limits of profit maximization

The relation of this analysis to the usual assumptions of profit maximization is of great interest. Profit, in a single transformation such as that described here, may be defined in two ways, either as the increase in the net worth, or as the difference between total revenue and total cost. Thus in Fig. 42 at an output Q_1P_1 the total cost is P_0Q_1 and the total revenue (i.e. the value of the output) is Q_1V_1, assuming that the output is valued at its selling price, tan $K_1P_1Q_1$. The difference between total revenue and total cost is therefore P_0V_1. If the initial quantities of

assets at P_0 are valued at the same price as the final quantities at K_1, the increase in net worth is T_0T_1, P_0T_0 being drawn parallel to P_1K_1, for OT_0 is the net worth corresponding to the position P_0, and OT_1 is the net worth corresponding to K_1.

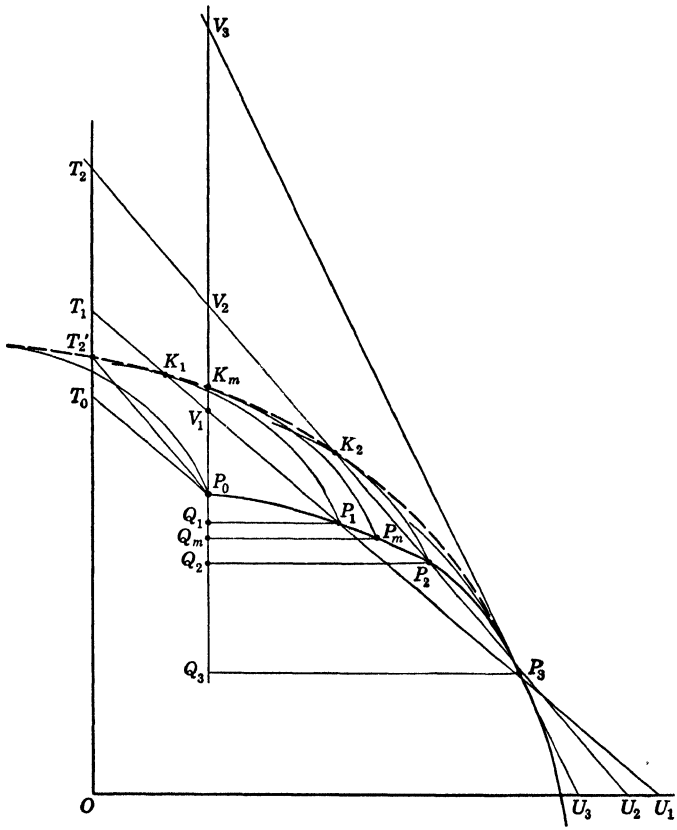


FIG. 42. Profit maximization in imperfect markets.

If there is a change in the valuation coefficient between the beginning and the end of the process it is no longer true that the increase in net worth is equal to the difference between total cost and total revenue (the net revenue). Suppose, for instance, that we value P_0 according to the valuation line P_0T_0, and value P_2 according to the line P_2T_2. We suppose, that is, that in the course of the process there has been a rise in the price of the product. The net revenue at the output Q_2P_2 is then P_0V_2

(= $T_2'T_2$, where P_0T_2' is drawn parallel to P_2T_2). The increase in net worth, however, is T_0T_2. This is comprised of two parts: the net revenue, $T_2'T_2$, and a sum T_0T_2' which is due to the revaluation of the goods in stock at the opening of the period. The LIFO (last in first out) method of inventory accounting is essentially an attempt to prevent these revaluations appearing in the profit-and-loss statement: on this method the inventory at position P_0 would be valued at approximately the same prices as at K_2; hence the increase in net worth would be equal to the net revenue.

What, then, happens to the principle of profit maximization under these circumstances? It is evident that the maximum net revenue is P_0V_3, where P_3 is the point of tangency of the envelope curve with the total cost curve, and the price (tan $V_3P_3Q_3$) is that at which only an infinitesimal amount can be sold! On the LIFO method of calculating net worth this will also be the maximum profit. It is evident that only under very extreme conditions will this be the preferred point. Why, then, does the principle of profit (or net revenue) maximization apparently give a determinate solution in the elementary marginal analysis? The answer is that the elementary analysis assumes as a limiting condition that all goods produced are sold, i.e. that there is no change in inventory during the accounting period. Under this limitation, of course, the maximum net revenue is P_0K_m, K_m being the point where the envelope curve cuts P_0V_3. Carrying the total revenue curve K_mP_m back to the total cost curve at P_m gives us the equilibrium point on the total cost curve. The price is tan $K_mP_mQ_m$; the quantity produced and sold Q_mP_m. We see clearly, therefore, that the "marginal cost equals marginal revenue" condition of the elementary analysis yields a determinate solution only because of the limiting condition that the whole output must be sold. This really implies a special case of the limitation imposed by asset preferences. If all such limitation is removed the principle of profit maximization becomes nonsensical.

Special cost conditions

Some special cases and applications now deserve consideration. In the first case, where production takes place at constant marginal cost, the total cost curve is a straight line, as in Figs. 43 and 44. Under perfect markets, as long as marginal cost is constant and less than the price, an increase in output always results in movement of the price line outwards. Hence the only possible position of equilibrium would be at the maximum point of the indifference system, K_1 in Fig. 43. This is the point at which both money and goods become discommodities. If the possession of assets is troublesome, this position of equilibrium may not

be so far removed from reality as might be thought. The indifference curves, or the preference function of which they are contours, represent the net advantages to be derived from the ownership of assets. A firm might well refuse to expand its operations, even though no obstacles presented themselves in the way of rising marginal costs or imperfect markets for output, simply because it was too much trouble to be big! This would represent a true case of equilibrium at the maximum point of the indifference system. If, however, the firm is to reach an equilibrium short of this maximum point it is clear that with perfect markets

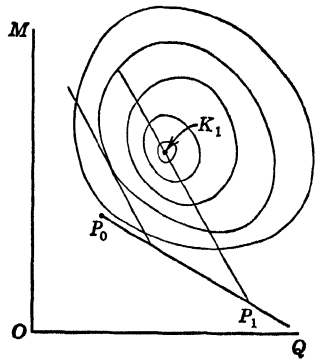


FIG. 43. Constant marginal cost.

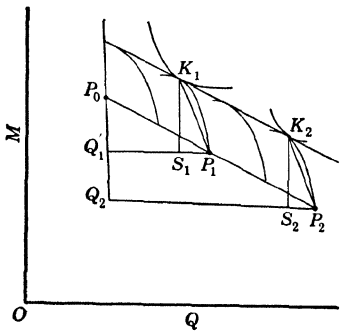


FIG. 44. Constant marginal cost.

for output there must be marginal costs rising with output, for unless the total cost curve bends downward at some point the line P_1K_1 will always move outward as P_1 moves down the total cost curve.

If there are imperfect markets for output, as in Fig. 44, the marginal cost can be constant and there will still be a determinate point of equilibrium which may be short of the maximum point on the indifference surface. Imperfection in the market may be just as effective a factor in limiting the size of the operation as rising marginal costs. Thus with constant marginal costs the envelope curve will be a straight line parallel to the total cost curve, and this will be touched by an indifference curve at some optimum point K_1. In this special case the price charged is independent of the preference system. A fall in liquidity preference, for instance, may send the equilibrium point from K_1 down to K_2. Because the revenue curves all have the same shape, the marginal revenues at K_1 and K_2 will be equal (both being equal to the constant marginal cost) and the price at K_1, tan $K_1P_1Q_1$, will likewise be equal to the price at K_2, tan $K_2P_2Q_2$. In this case, therefore, a fall in liquidity preference will raise output but will affect neither the price nor the volume of sales: S_1P_1 is equal to S_2P_2.

In the above examples we have assumed that there is no fixed cost, i.e. no cost which is independent of the amount produced. It is easy, however, to introduce a fixed cost into our figures simply by dropping the whole level of the production opportunity, or total cost, curve. Thus in Fig. 45 the total fixed cost is P_0R_1, indicating that even if nothing is produced the total of assets will decline by this amount. As the figure is drawn the assumption is implied that the fixed cost is all in terms of

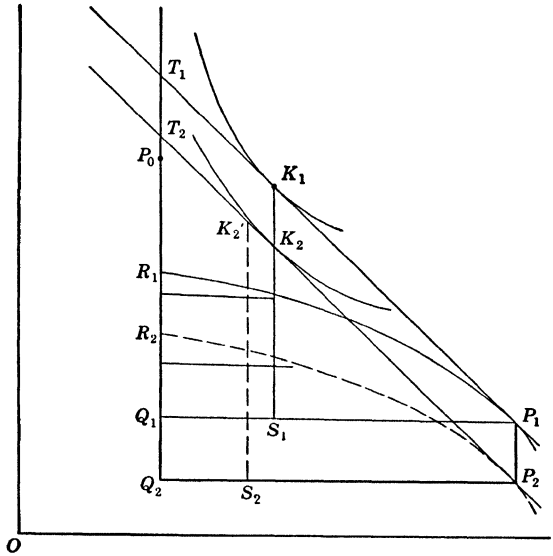


FIG. 45. Fixed cost and a fixed tax.

"money," i.e. assets other than the product. It could be, of course, that inventories of product also depreciated even if no production was undertaken, in which case the fixed cost would be indicated by a movement to some point a little to the left of R_1. For the moment, however, we may neglect this complication.

The total variable cost curve is now R_1P_1: the total cost curve, or the actual production opportunity line, is $P_0R_1P_1$. It will be seen that under these circumstances there is a strong inducement to produce. A firm without fixed costs can at least stay in the same place if it produces nothing. A firm with fixed costs must produce or must at least actively follow some opportunity line, whether of production and exchange or of pure speculative exchange, in order to maintain itself. Otherwise assets will decline by the amount of the fixed cost in each period until the net

worth is extinguished. With perfect markets, in this case as in the other, the firm will produce that quantity at which the marginal cost is equal to the price—i.e. Q_1P_1 in Fig. 45—and will move to a preferred point K_1 on the exchange path. It may be observed at this point that the average total cost is the slope of the line P_0P_1. The total profit (net revenue) is P_0T_1, which is the difference between the price and the average total cost, multiplied by the output. If price is equal to average total cost the profit is zero; if the price is below the minimum average total cost the profit will be negative, even at the optimum point.

Effects of taxation

We can now study the effect of a change in the fixed cost—e.g. through a fixed tax, independent of output. An important proposition of the elementary marginal analysis [1] must now be modified. This is the proposition that a fixed tax (or any other change in fixed costs) will affect neither the output nor the price of a firm, provided that it does not drive the firm out of business altogether by reducing profits below the level at which the firm will operate. For perfect markets the proposition still holds in regard to output, though it does not necessarily hold in regard to sales. Thus in Fig. 45 a fixed tax equal to R_1R_2 will lower the total cost curve from $P_0R_1P_1$ to $P_0R_2P_2$. As, however, the marginal cost is unchanged, and the market price likewise assumed to be unchanged, the price line P_2T_2 will touch the curve R_2P_2 at a point P_2 which is vertically below P_1: i.e. Q_1P_1 and Q_2P_2 are equal; the tax has not charged output. Sales, however, may have changed; the indifference curve system may, for instance, touch P_2T_2 at K_2' and P_1T_1 at K_1. If the indifference curves are parallel, i.e. if the marginal rate of substitution is invariant with respect to the quantity of money, the preferred point K_2 will be directly under K_1, as the indifference curves at K_1 and K_2 must have the same slope, equal to the price of the product. In this case the fixed tax will affect neither the sales nor the output. If, however, the marginal rate of substitution rises with increasing stocks of money, indicating a marginal utility of money rising faster than the marginal utility of product, so that more money is necessary to compensate for the loss of a unit of commodity, then the indifference curve at K_2, immediately below K_1, will be less steep than at K_1 and will cut the price path: the equilibrium will be at K_2', to the left of K_2. The result of the fixed tax is then an increase in sales from P_1S_1 to P_2S_2' and a decline in additions to inventory.

In the case of imperfect markets for output, shown in Fig. 46, the divergence from the simple analysis is even greater, for if the indifference

[1] Cf. Boulding, *Economic Analysis*, 2nd Ed. (Harper & Brothers, 1948), p. 552.

curves are not parallel a fixed tax may affect output, sales, and price. Thus before the tax the equilibrium point is K_1; after a tax equal to R_1R_2 is imposed the cost curve and the envelope curve shift vertically downward. If the indifference curves are parallel the point of tangency between the indifference curves and the new envelope curve, K_2, will be vertically below K_1, for the envelope curve has been shifted vertically downwards, and its slope at K_2 is therefore equal to the slope at K_1. If, however, the indifference curves become less steep at lower money stocks

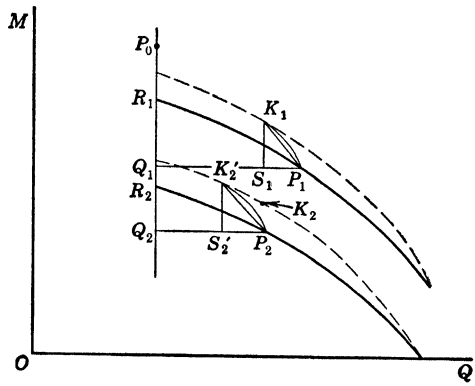


FIG. 46. A fixed tax in imperfect markets.

and the same product stock level, the equilibrium point after the tax will be K_2'. The tax will have lowered output from Q_1P_1 to Q_2P_2, will have raised sales from P_1S_1 to P_2S_2', and will have actually lowered the price from tan $K_1P_1S_1$ to tan $K_2'P_2S_2'$, in order to get the increased sales. This odd result follows, of course, because the tax has an effect in increasing liquidity preference by diminishing the money stock, and hence increasing the desire to sell, the lower price being necessary in order to sell more.

The effect of variable taxes can also be examined with the aid of the above analysis. In the elementary marginal analysis [2] the result is obtained that a tax per unit of output, because it raises the marginal cost curve, will lower output and raise price. We now have to extend the analysis, and distinguish between the effects of a tax per unit of output and a tax per unit of sales, as we no longer assume that output and sales are identical. Figure 47 shows the effect of a tax per unit of output; Fig. 48, the effect of a tax per unit of sales. The tax on output shifts the total cost curve from P_0P_1 to P_0P_2. In the figure, for the sake of

[2] Boulding, *Economic Analysis*, 2nd Ed., p. 553.

convenience, the total cost curve is drawn as a straight line, but this is not, of course, necessary for the analysis. The envelope curve comes closer to the total cost curve as the total cost curve increases its slope. The point of equilibrium before tax is K_1. Even if the marginal rate of substitution is independent of the money stock, so that the slope of the indifference curve at K' is the same as at K_1, it is clear that the indifference curve at K' must cut the envelope curve there, and the point of equilibrium must be further up the curve at K_2. It is evident that the

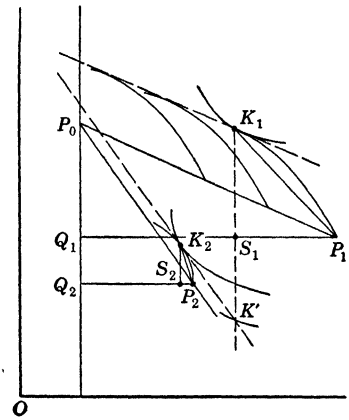


FIG. 47. Tax per unit of output.

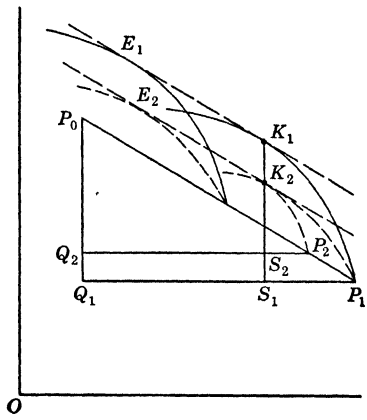


FIG. 48. Tax per unit of sales.

tax has lowered output, has lowered sales even more sharply, and has raised price.

In Fig. 48 we show a tax not on output but on sales. This is treated most simply as a deduction from the total revenue: the result of the tax, therefore, is that the total-revenue-less-tax curves lie below the total revenue curves, and we have a net-of-tax envelope curve E_2 below the original envelope curve E_1. The slope of the revenue-net-of-tax curve at K_2 is equal to that at K_1 if marginal cost is constant, as in the figure. Hence if in addition the indifference curves are parallel the result of the tax is to lower sales by as much as production is lowered, so that the change in inventories is not affected: $Q_1S_1 = Q_2S_2$. If, however, the marginal rate of substitution increases with increase in the money stock, the point K_2 will be further up the envelope curve. Production will be less; sales and price, however, are in this case independent of the preference structure, as the envelope curve E_2 is parallel to the total cost curve. This last proposition holds, of course, only under constant marginal costs. If marginal costs are increasing with output, K_2 will be further

down the envelope curve: production will be greater but sales will be less than in Fig. 48.

The demand for input

The above analysis can also be applied to the problem of the demand of a firm for input. As the total cost of a given output is equal to the value of the inputs which are used up in producing it, a rise in the price of one of these inputs will have much the same effect on the position of the firm as a tax per unit of output. If the marginal physical productivity of the factor is constant, so that a unit increase in product is always obtained with the sacrifice of a given amount of the factor, the two cases are perfectly analogous: a rise in the price of the factor is exactly equivalent to a tax per unit of output. If, however, the marginal productivity of the factor falls with increased output, more of the factor must be used at larger outputs than at smaller in order to get a unit increase in product; hence a rise in the price of the factor is equivalent to a larger tax on larger outputs, i.e. to a "progressive" tax on outputs. With perfect factor markets and independent factors the change in the marginal cost which results from a unit change in the price of a factor is equal to the reciprocal of the marginal physical productivity of that factor.[3]

The effect of a rise in the price of a factor under conditions of perfect markets, for both input and output, is shown in Fig. 49. In this problem the analysis is sometimes a little easier if the relationships are expressed

[3] Let C be the total cost of output x, K the total fixed cost, p_1, p_2, etc., the prices of factors consumed in quantities q_1, q_2, etc. Then

$$C = K + p_1q_1 + p_2q_2 + \text{etc.}$$

If m_c is the marginal cost, and p_1, p_2, etc., are constant (perfect markets),

$$m_c = \frac{dC}{dx} = p_1\frac{dq_1}{dx} + p_2\frac{dq_2}{dx} + \text{etc.}$$

The quantities $\frac{dq_1}{dx}$, $\frac{dq_2}{dx}$, etc., are the reciprocals of the marginal physical productivities of the various factors: they may be called the marginal factor cost of the product and defined as the amount of factor which must be consumed to add one unit to the product. If we assume that the marginal productivities and marginal factor costs of the various factors are independent one of another, and the factors are bought in perfect markets, we have

$$\frac{dm_c}{dp_1} = \frac{dq_1}{dx}$$

That is, the change in the marginal cost of the product per unit change in the price of a factor is equal to its marginal factor cost.

on a price- (or marginal cost or revenue) output field, as in the lower part of Figs. 49 and 50. Thus the total cost curve P_0P_1 is translated into a marginal cost curve MC; each ordinate of the marginal cost curve measures the slope of the total cost curve at the same quantity of product: the increasing steepness of the total cost curve as product is increased is reflected in a rising MC curve. The exchange paths all reduce to a single price-sales line, mp_1, as the price is assumed constant. Equilibrium output Q_0Q_1, or q_0q_1, is given at the point P_1, where the total cost curve touches a price path, or by the point p_1 where the marginal cost curve cuts the price-sales line, indicating that marginal cost equals price. If the marginal rate of substitution is independent of the quantity of money the indifference curves reduce to a single marginal rate of substitution curve, MRS in the lower figure, each ordinate of which measures the slope of any indifference curve at the same quantity of product. The volume of sales is then p_2p_1, and the addition to inventories p_0p_2, where p_2 is the point of intersection of the marginal rate of substitution curve with the price-sales line. If the marginal rate of substitution varies with money stocks there will be not a single MRS curve, but a whole family of them, each indifference curve generating a different MRS curve. The point p_2 is then the intersection of the price-sales curve with the MRS curve corresponding to the indifference curve at P_2.

The effect of an increase in the price of a factor is to raise the MC curve to MC', and correspondingly lower the total cost curve to P_0P_1'. It will be observed that output will diminish from Q_0Q_1 (q_0q_1) to Q_0Q_1' (q_0q_1'). This diminution in output is not affected in this case by the position of the indifference curves, which influence only the distribution of output between sales and inventory. It follows also—as each quantity of output is produced with a definite quantity of the input—that the demand function for the input is unaffected by changes in liquidity preference. If we know (in addition to the data presented in Fig. 49) the production function, showing what quantity of input will be used to produce each quantity of output, the whole demand curve for the input can be traced by considering the effect of various prices of the input on the output, and therefore on the input, quantities. It is clear that the extent of the decline in output which results from a rise in the price of input depends on the slope of the marginal cost curve: if this is steep a rise in the marginal cost curve causes only a small decline in output.

If the marginal rate of substitution is independent of the money stock, so that there is only one MRS curve, a rise in the price of an input will not affect the investment in inventory, p_0p_2. The decline in output will result in an exactly equal decline in sales, both declines being equal to

$p_1'p_1$. If, however, the marginal rate of substitution is directly related to the money stock, the lower MRS curve, MRS', will be the significant one: inventory will be increased by only p_0p_2', and sales will fall less—to $p_1'p_2'$ instead of to $p_1'p_2$. If the decline in inventory investment is

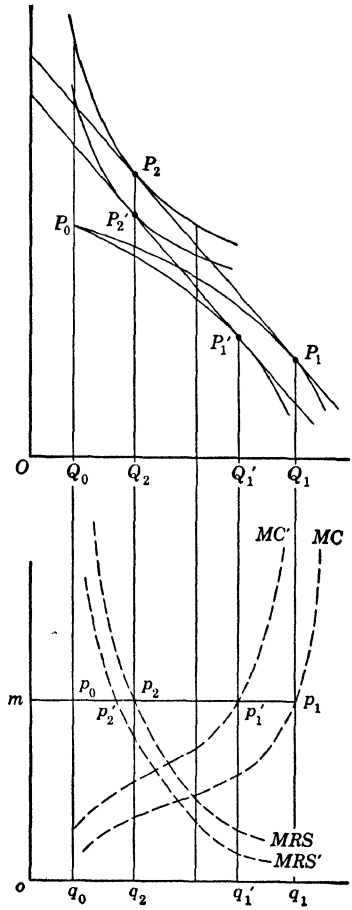


FIG. 49. Demand for input.

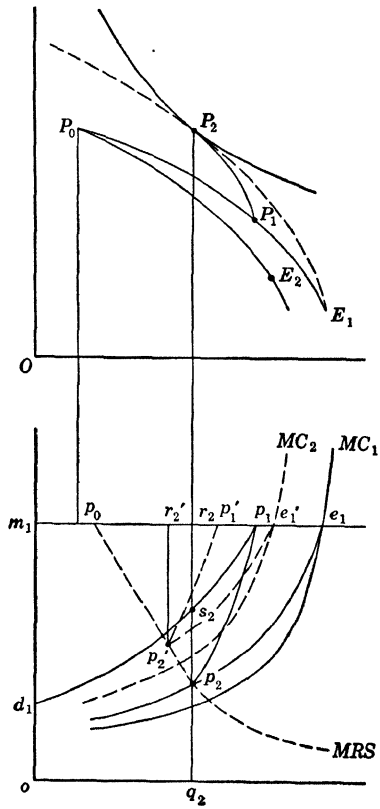


FIG. 50. Demand for input.

sufficient it might be possible, though unlikely, for a rise in the price of an input actually to increase the *sales* of output in the next planning period, in the effort to become more liquid, though it will always decrease production.

When there is an imperfect market for the product the problem is more complicated and is illustrated in Fig. 50. P_0E_1 is the original total cost curve; MC_1 the corresponding marginal cost curve. E_1P_2 is the envelope revenue curve; the corresponding curve in the lower figure is s_1p_2, which may be called the "marginal envelope curve": each ordinate of this curve measures the slope of the envelope curve at the corresponding quantity. The price om_1 is the least price at which the quantity demanded is zero; i.e. it is the price at which the sales curve cuts its price axis. From any point on the marginal envelope curve a marginal revenue curve such as p_2p_1 can be drawn. The corresponding sales curve is p_1d_1. If the marginal rate of substitution does not vary with the stock of money there will be a single marginal rate of substitution curve, MRS, and the equilibrium will be at p_2, where the MRS curve cuts the marginal envelope curve. If the marginal revenue curve through p_2 is traced back to p_1 on m_1e_1, then p_0p_1 will be the output and r_2p_1 the sales at a price q_2s_2, where s_2p_1 is the sales curve through p_1.

If now there is a rise in the price of an input the marginal cost curve rises and moves leftward to the position MC_2. The marginal envelope curve shifts with it to a position where it cuts the MRS curve at p_2'. Output declines from p_0p_1 to p_0p_1'. Sales have also declined, from r_2p_1 to $r_2'p_1'$, as the triangular figures $p_2p_1r_2$ and $p_2'p_1'r_2'$ are approximately similar and p_2' is closer to the line m_1e_1 than p_2. Inventory accumulation, too, has declined. If the marginal rate of substitution falls with decline in the money stock the relevant MRS curve will shift leftwards as in Fig. 49, and the equilibrium point p_2' will move further down the marginal envelope curve. This will make output smaller, but sales larger, than if there were a unique MRS curve.

In the case of imperfect markets for products it is no longer true that the effect of a rise in marginal costs on output is independent of the preference structure. If the slope of the indifference curves changes—if, for instance, there is a rise in liquidity preference, so that the marginal rate of substitution at each point falls—the MRS curves will fall downwards and to the left, and the equilibrium point p_2 will also move downward and to the left along the marginal envelope curve. This will clearly reduce output (as the marginal revenue curves are all parallel) and will increase sales as the point p_2 moves farther from the line m_1e_1 and hence the distance r_2p_1 becomes larger. This proposition has already been established. The elasticity of demand for input depends mainly on the slope of the MRS curve between, say, p_2 and p_2'. This may or may not be affected by a general change in liquidity preference. If the decline in the marginal rate of substitution is uniform for all quantities the MRS curve will move downward parallel to itself, and un-

less the curvature of the *MRS* curve is strong there will not be much change in the elasticity of demand for input, though there will, of course, be a fall in the demand for input.

Dynamics of production

So far we have considered only the theory of production on a single "day," or planning period. We can now move somewhat towards the

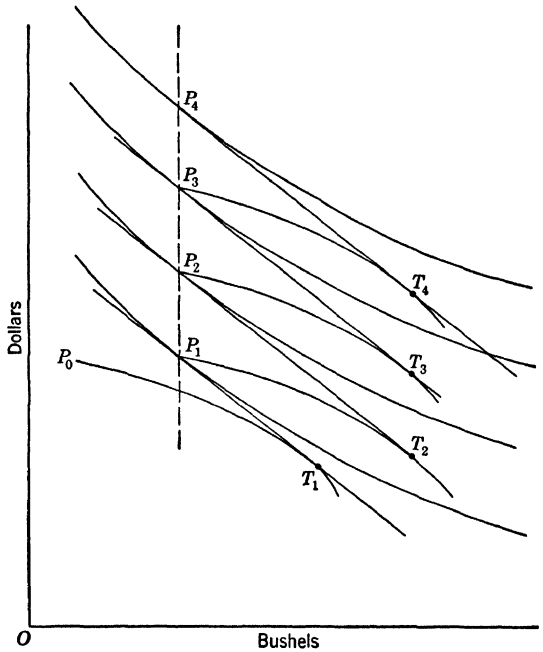


FIG. 51. Dynamics of production.

dynamics of the firm by considering the movement of the balance sheet through a succession of production periods. This process is illustrated for perfect output markets in Figs. 51 and 52. In Fig. 51 it is assumed that the indifference curves are parallel. Starting from P_0 in the first day we follow the production line P_0T_1 and the exchange path T_1P_1 to P_1. The second day we start from P_1 and proceed along a similar path—if production and price conditions have not changed—to P_2. The next day we go to P_3, the next to P_4, and so on. It is clear in this case that the exchange paths P_1T_1, P_2T_2, etc., will be both parallel and equidistant. On the assumptions of this figure, therefore, it is clear that the

locus of starting positions, $P_0P_1P_2 \cdots$, will become vertical at P_1, indicating that the firm has accumulated all the inventory it needs and sells its whole product in each period, adding to its money stock by the amount of the profit.

From this figure it looks as if the firm will go on accumulating money stocks for ever and will never distribute any dividends! This somewhat absurd conclusion can be avoided, however, if we assume that there is a fixed cost. Suppose, for instance, that the firm reaches its optimum inventory at P_1, and that there is a fixed cost equal to P_2P_1. In this case

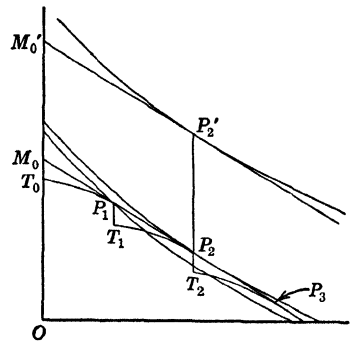


FIG. 52. Dynamic equilibrium of production.

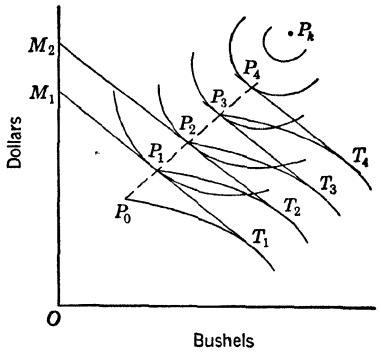


FIG. 53. Dynamic equilibrium of production.

the cost curve will be, on the second day, $P_2P_1T_2$, and the firm will simply repeat the cycle $P_2P_1T_2P_2$ indefinitely. This is essentially the equilibrium under perfect competition. If we think of the fixed cost as a "rent," then the rent plus the other costs at optimum output is so great that there is no "profit" and it is impossible for the firm to expand. If it once deviates from the optimum position it will actually make losses and decline. If the indifference curves are all parallel and there are fixed costs which absorb all profit the firm will simply proceed to the optimum inventory and trace out a cycle at that point. Thus, in Fig. 52, suppose that the firm starts with capital in money equal to OM_0. In the first production period it proceeds to P_1, following its total cost curve or production path $M_0T_0P_1$: in the case of the figure it will not sell anything in the first period, as it has not yet reached its optimum inventory. It will repeat its process in subsequent production periods until it reaches the point P_2, where the price line M_0P_2 touches an indifference curve. From this point it will simply repeat the pattern $P_2T_2P_3P_2$ until conditions change. As it makes no profits it cannot get off the price, or valuation, line M_0P_3.

In Fig. 53 we make the less restrictive assumption of a general circularity in the indifference curves, indicating that there is some quantity of assets beyond which the firm will not wish to go. In such a case, if the firm makes profits in each production period, indicated by the rise in the exchange path, its initial position will trace out a locus $P_0P_1P_2 \cdots$ which will be positively sloped, indicating that as the firm expands it will increase not only its liquid but also its other assets, roughly in proportion. If we make the assumption that the preferred asset ratios are constant, the locus $P_0P_1P_2 \cdots$ will be a straight line from the origin, and the expansion of the firm will result in an equal proportionate increase in all assets. Here also, of course, it is true that if a firm makes zero profits it will simply move along the price line corresponding to its total capital to the point where its asset preferences are satisfied. Thus if the firm started with liquid assets equal to OM_1 it would proceed until it reached P_1, after which it would simply repeat a cycle as in Fig. 52. In this general case, however, the optimum inventory is not independent of the amount of capital with which the firm starts. The more capital a firm has, the larger will be its inventory; from OM_2 it proceeds to P_2. If the indifference curves were parallel this would not be true: no matter how much capital a firm started with, it would always have the same optimum inventory. Thus, in Fig. 52, if the firm started with OM_0' instead of OM_0 it would reach an equilibrium at P_2', where the inventory of product is the same as at P_2.

The dynamic theory of the firm, even at this level of abstraction, would not be complete without some reference to the problem of the distribution of dividends. This is a problem in the theory of the firm analogous to the problem of the distribution of a consumer's income between consumption and saving, and it presents similar analytical difficulties. From the point of view of the balance sheet a cash dividend represents a simple destruction of liquid assets. It is not an exchange, for no asset is created to correspond to the assets destroyed: it is, from the firm's point of view, an act of more or less voluntary consumption. It may be asked, therefore, why should the firm ever perform such an act? It is clear that unless we are at or beyond the summit of the indifference surface—P_k in Fig. 53—a dividend distribution, representing as it does a simple decline in liquid and in total assets, i.e. a movement exactly similar to that occasioned by a fixed cost, say, from P_2 to T_2 in Fig. 52, will always land the firm on a lower indifference curve.

It seems clear, then, that either we must regard dividends as a "fixed cost," like depreciation—a destruction of assets which the nature of the financial mores imposes and which is therefore not the result of a decision by the firm but is an external imposition on the firm, like taxes—or we must assume that dividends will not be declared until the firm

actually overshoots the summit of its preference surface. This assumption, however, soon gets us into grave difficulties, for if the opportunity lines lie in the neighborhood of the preference summit some very strange things can happen! Thus in Fig. 54 we suppose that the firm starts its production "day" at the summit of its indifference surface. If it has no fixed costs there will be no incentive whatever to produce, for any path it takes voluntarily will take it to a lower indifference curve. If, however, the firm has fixed costs it is forced to produce in order to maintain itself. Thus suppose it has a fixed cost equal to PT, and a variable cost

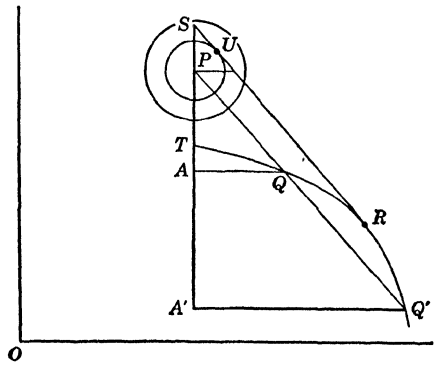


FIG. 54. The theory of dividends.

curve TR. If it abstained from production it would fall from P to T, which is on a lower indifference curve. By producing along the line TR and selling the product it can make profits and so restore its position to the summit again, following the path $PTQP$.

There is no reason to suppose that the point Q represents a maximum profit point: in the figure it clearly does not. There are, however, many ways in which the firm could restore itself to the point P. It could, for instance, produce more, thus making a profit, and then distribute the profit in dividends. If, for instance, it produced at the maximum profit point R, where the marginal cost is equal to the price of the product, it could follow the path $PTRSP$. Note here that there is no particular reason for it to stop at the point U, where the price path is touched by an indifference curve. It is true that U is on the highest indifference which can be reached on the line SR: but the firm can always reach points inside the indifference circle at U which are preferable simply by giving away either money or product. Hence any volume of output between QA and $Q'A'$ will enable the firm to get back to the point P by an infinite variety of combinations of outputs, sales, and distributions of money or

even of product! It is clear that asset preferences alone are not sufficient to yield a determinate solution in this case: we must specify a further system of preferences in regard to the alternative paths to the preferred asset position.

Fortunately this problem is likely to be significant only when the "preference peak" lies to the left of the exchange path, for only then is there any option in the paths taken to the optimum position. Even in the more usual case, however, it must be admitted that the asset-ratio analysis implicitly assumes that the firm is neutral in regard to various paths to its possible asset positions. This may not be true even where the asset-preference peak is beyond the exchange path: the firm might prefer one combination of assets to another *because* of the path taken to reach them, and we could not, therefore, assume an invariate system of asset preferences. These considerations, however, lead us into complications beyond the scope of this analysis, and we may neglect them in the hope that they will not prove to be quantitatively important. We shall not assume, however, that the firm's behavior can be explained by asset preferences alone, and in the next chapter other preference systems will be examined.

In the case of Fig. 54 the simplest assumption is, of course, that of profit maximization; the firm would produce to the point R and would sell all that it produced, proceeding along the exchange path to the point S, and then would distribute its profits, SP, in dividends, bringing it back to the point P. It must be emphasized, however, that there is no necessity for profit maximization: the firm may well prefer a quieter, or even a more active, life than that yielded at maximum profits.

It is clear, in sum, that what emerges from this analysis is a theory of inventory. Such a theory will, of course, be more significant in short-run analysis than in the long run, for we may reasonably expect that the longer the period of time we take into consideration, the less important will be inventory changes relative to the total volume of output and sales, and in the long run we may expect a firm to reach a final equilibrium in the position of Fig. 54. Nevertheless the dynamic changes involved in inventory movements are of such prime importance in explaining the actual movements of the economy, and especially its fluctuations, that it is most useful to have some theoretical insights into their operations. In particular the effect of asset preferences on sales and inventory changes is a matter of the utmost importance for any student of business fluctuations. If, now, we can determine the asset preferences themselves by means of their relation to some previous experiences we will have all the elements of a truly dynamic system of difference equations. The link, of course, between past experience and present asset preferences

is expectations of the future. Asset preferences, as we have seen, are determined by expectations, and expectations are presumably determined by past experience. It is theoretically possible, therefore, to bypass expectations altogether in the development of dynamic systems, and to regard the asset preferences as if they were derived directly from past experiences. There is much to be said for this, as expectations, being excessively subjective and even subconscious, are notoriously difficult to deal with. There are, however, some interesting phenomena connected with expectations—and the uncertainty which inevitably accompanies them in a greater or less degree—which will be considered in the next chapter before the main thread of the argument is resumed.

Chapter 7

Expectations, Gambling, and Insurance

The present chapter is again somewhat of a digression from the main theme, and the impatient reader may omit it. There are, however, many interesting questions relating to expectations, risk, and uncertainty which should at least be considered briefly in a work of this kind, even though an extended coverage of the problem is not contemplated.

Expectations and uncertainty

Asset preferences, as we have seen, are determined mainly by expectations in regard to future events or future opportunities. These expectations are of many kinds. They may be price expectations, in which case the general principle may be laid down that assets or asset combinations will stand higher on the scale of preference the more rapidly their prices are expected to rise. There may also be production expectations, e.g. of yield from a mine, from land, from fixed capital, from an education. All these expectations represent expected opportunities for asset transformation or creation of some kind. They all exhibit in greater or less degree the phenomenon of uncertainty. Because of this, as Makower and Marschak [1] have shown so brilliantly, it is not merely the expected "lucrativity" that affects asset preferences, but also various other qualities or properties of the expectation—safety, chanciness, and so on. We cannot, therefore, deal with the uncertainty problem by computing a "probable value" of an outcome and regarding this as a "certainty equivalent." We are not affected merely by what we regard as the most probable of a set of possible outcomes, or even by any average "expected value" weighted by the various probabilities; we are affected in our behavior by the less probable, or even by the most improbable, of possible outcomes, as witnessed by the demand for sweepstake tickets, wild-cat shares, and, in some cases, matrimony.

[1] H. Makower and J. Marschak, Assets, prices, and monetary theory, *Economica*, N.S., vol. 5, p. 261 (August 1938).

Probably the simplest way to develop a theory of uncertainty is to record preferences on an indifference map in a field with the most probable or expected value of the asset as one axis and some measure of the certainty of this most probable value as the other. Thus in Fig. 55 we plot the most probable value of an asset along OY and the possible range of values along OX.[2] Thus suppose that the point P represented an expected value of \$110 (NP) with a range of \$60 (ON); P would represent an asset the value of which might fall anywhere between, say, \$80 and \$140, or \$70 and \$130, but which would be most likely to fall at \$110. This construction cannot clearly differentiate between probability distributions of different degrees of skewness, but it has nevertheless several applications.

The theory of gambling

It can be applied, for instance, to the theory of gambling. Suppose, for instance, that a man has \$100 cash, with a chance of gambling at

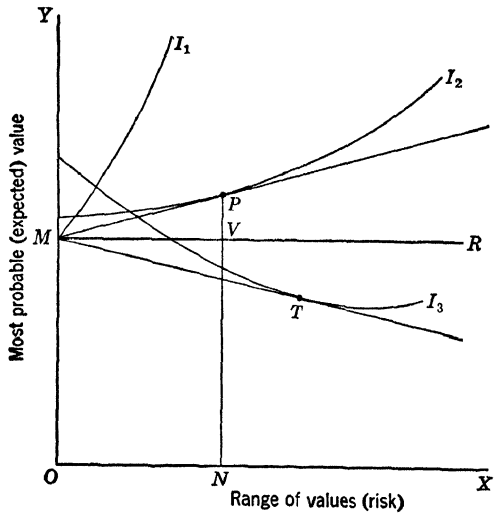


FIG. 55. The theory of gambling.

constant odds, say, of 10 to 1. If he bets \$1 he may have \$99 if he loses, or \$109 if he wins. Suppose that he estimates the true chances of winning at 5 to 1. That is to say, he thinks that he has a chance of 4 in 5 of

[2] This construction is mainly due to M. W. Reder, *Studies in the Theory of Welfare Economics* (Columbia, 1947), Chapter VII.

losing (i.e. having \$99) and a 1 in 5 chance of winning (i.e. having \$109). The expected value of his assets is therefore $99(\frac{4}{5}) + 109(\frac{1}{5})$, or \$101. The point on the field of Fig. 55 representing an expected value of \$101 and a range of \$10 (99–109) is therefore open to him. If he bets \$2 he may have either \$98 or \$118, with a range of \$20 and an expected value of \$102. As he bets more and more he clearly moves up an "opportunity line" such as MP in Fig. 55. If there is a "perfect market" in gambling, i.e. if the odds at which he can bet (the "objective odds") and the odds which he thinks are correct (the "subjective odds") are constant, this opportunity line will be straight. If the objective odds are more favorable to the gambler than the subjective odds, the opportunity line will slope upwards, as MP. If the subjective and objective odds are equal, the line will be parallel to the range axis, as MR. If the objective odds are less favorable to the gambler than the subjective odds, the opportunity line will slope downwards, as MT. Thus if in the previous example the subjective odds were also 1 in 10, then with a bet of \$1 the expected value would be $99(\frac{9}{10}) + 109(\frac{1}{10})$, or \$100; similarly no matter how large the bet the expected value remains at \$100: all that betting does is to increase the possible range of a given expected value. If the subjective odds were 1 in 20, the expected value for a \$1 bet would be $99(\frac{19}{20}) + 109(\frac{1}{20})$, or \$99.50.[3]

We can also postulate a system of indifference curves in the field of Fig. 55. If the individual is averse to risk, these will have an upward slope, indicating that in order to induce the individual to take a longer chance the expected value must be increased. The indifference curves are also likely to get steeper at higher ranges, indicating that the more

[3] Suppose that the gambler starts with a capital sum a, can bet indefinite amounts at x to 1, and has a subjective valuation of the odds of y to 1. Suppose he bets a sum n. Then the extremes of the range of possible values are $a - n$ if he loses, $a + n(x - 1)$ if he wins. The range, r, is given by

$$r = nx \tag{1}$$

The expected value (subjective) of the asset is

$$v = \frac{(a - n)(y - 1)}{y} + \frac{a + n(x - 1)}{y} = a + \frac{n(x - y)}{y} \tag{2}$$

Eliminating n between equations (1) and (2), we get the equation of the opportunity line:

$$v = a + \frac{r(x - y)}{xy} = a + r\left(\frac{1}{y} - \frac{1}{x}\right) \tag{3}$$

If $x = y$, i.e. if subjective and objective odds are the same, the opportunity line becomes simply $v = a$. If x exceeds y, the line will have a positive slope; if y exceeds x, a negative.

risk an individual is bearing, the less he wants to incur more risk, and the more "bribe" in the way of an increase in expected value must be offered to induce him to do so. The slope of such an indifference curve might be called the "rate of risk substitution." It is the increase in the range of values which the individual would be willing to accept in return for a unit increase in the expected value. In this case the "higher" indifference curves are those towards higher expected values and lower risks; the indifference surface is a hill which rises towards the point Y and falls towards X.

If the indifference curves like I_1 are steeper than the opportunity line at the point M the individual will stay right at M and will not gamble at all. If, however, the individual is not much averse to risk-bearing his indifference curves will be flat like I_2 and may touch an opportunity line at P. He will then gamble an amount represented by the point P, i.e. an amount ON/x, where x to 1 is the objective odds. It can easily be seen that if the individual is averse to risk he will gamble only if he thinks that the objective odds are too high, i.e. if he thinks that the chance of winning is better than the "market" odds, so that the opportunity line has a positive slope. If the individual thinks that the chances of winning are equal to, or less than, the betting odds indicate, he will not gamble unless his indifference curves start off, at least, with a negative slope, i.e. unless risk is a "commodity" rather than a discommodity to him, indicating that he is willing to give up expected value in order to have greater risk. Thus the opportunity line MT could be touched only by an indifference curve like I_3. In fact, of course, in any gambling situation where the "bookies" make a living, the actual chance of winning must normally be less than the quoted odds: hence any objective appraisal of the opportunity line would suggest that it must be like MT rather than like MP. Hence the only reasonable explanations of the existence of gambling are, first, that most people have an overweening opinion of their own good fortune and hence rate their chance of winning at more than the actuarial value, which would lead to positively sloped opportunity lines; or, second, we must assume that there is a positive preference for risk on the part of the gambler.

Figure 55 will also enable us to develop the theory of the bookie, the "supplier" of gambling, as well as of the gambler, its demander. A bookie who accepts a bet at 10 to 1, starting with \$100, will have \$91 if the gambler wins, \$101 if he does not. His range is likewise \$10: the expected value of his capital, however, will be greater than \$100 only if his estimate of the chances is more than 10 to 1. Hence, if the objective and subjective odds are such that the gambler's opportunity line is MP, the bookie's will be MT; if the gambler's is MT the bookie's is

MP.[4] It is clear that if the subjective odds are based on an actuarial estimate the bookie's opportunity line will be likely to be positively sloped. There is no need, therefore, to assume that bookies are risk-likers: a bookie might have some aversion to risk, with an opportunity line MP and an indifference curve I_2, and still find a point of equilibrium at P corresponding to how many bets he wished to handle. On the whole, however, it is reasonable to suppose that as the slopes of the opportunity lines, depending as they do on the difference between the objective and the subjective odds, are not likely to be great, anyone with a strong aversion to risk—e.g. with indifference curves such as I_1—will not engage in gambling in any form, either as bookie or as gambler.

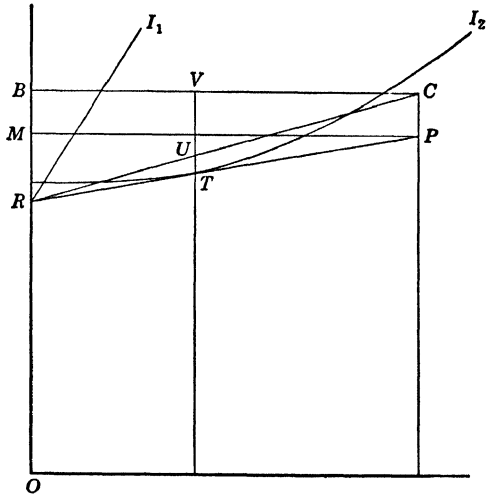


FIG. 56. The theory of insurance.

The theory of insurance

These considerations can also be applied to the theory of insurance. Suppose, for instance, that we have a house, valued at \$10,000, which has a 1 in 100 chance of burning down during the year. If it does not burn down we have \$10,000: if it does we have nothing. Without any

[4] Using the notation of footnote 3, the bookie's expected value is

$$v = a + \frac{n(y - x)}{y}$$

His range is $r = nx$, and the equation of his opportunity line is

$$v = a + r\left(\frac{1}{x} - \frac{1}{y}\right)$$

insurance, then, the range of possible values of our capital is 0–10,000, or 10,000: the expected value, assuming that the house either burns completely or not at all, is 10,000 ($\frac{99}{100}$), or 9900. Suppose now that we can insure at a rate of \$1 per \$100. If we insure \$1000 worth at a cost of \$10, the range of capital values is 10,000–10, or 9990, if the house does not burn down, and 990 if it does: the range has been reduced from 10,000 to 9000. The expected value is still 9990. The result of insurance at a rate equal to the expected chance is thus to reduce the range without changing the expected value. It moves the individual from P along PM towards M in Fig. 56, P representing his position without insurance. If the cost of insurance is greater than the expected risk—as we might reasonably suppose, if insurance companies are to live—the opportunity line will have a positive slope, such as PR, indicating that the diminution of risk has to be paid for by some diminution in the expected value.[5]

Suppose, then, that PR is the opportunity line, P representing no insurance and R full insurance. If the individual has a strong distaste for risk and has indifference curves such as I_1 which are steeper than the opportunity line at all points where they cut it, he will insure fully, i.e. will move right to the full insurance point R where risk is entirely eliminated. If he has less distaste for risk and has an indifference curve such as I_2 which touches the opportunity line at T, he will not insure fully, but will proceed down the opportunity line only to the point T. It is not difficult to construct from the figure the amount spent on in-

[5] Suppose the value of the insured object is A; the chance of its destruction, $1/x$. Then suppose that a sum a is paid in insurance at a premium rate y so that if loss occurs an amount ay is paid. Then if the accident does not happen the capital value is $A - a$; if it does happen the value is $ay - a$. The range therefore is $r = A - ay$. The expected value is then

$$v = (A - a)\frac{x - 1}{x} + a(y - 1)\frac{1}{x}$$

Eliminating a, we have the equation of the opportunity line:

$$v = A\frac{(x - 1)}{x} + (A - r)\left(\frac{1}{x} - \frac{1}{y}\right)$$

$$= A\left(1 - \frac{1}{y}\right) + r\left(\frac{1}{y} - \frac{1}{x}\right)$$

If the insurance is made at the exact actuarial cost, $x = y$ and the opportunity line is

$$v = A\left(1 - \frac{1}{y}\right)$$

The expected value is constant and equal to the expected value of the capital sum itself.

surance at the equilibrium point T. Draw the horizontal line BC so that OB is equal to the capital value, A. Drop TV to BC to cut the line RC in U. Then the amount spent on insurance,[6] i.e. the insurance premium, is UV.

In a similar fashion the opportunity line for an insurance company can be drawn. Suppose, for instance, that it writes insurance amounting to ay for a premium a. If the accident or insured event does not happen, the company is richer by a: if it does happen it is poorer by $a(y - 1)$. It has introduced a range into its capital of ay. If, however, the actuarial value of the risk is less than the premium, the expected value of the company's capital will be greater.[7] The insurance company will therefore have an opportunity line very much like the gambler's line MP in Fig. 55. This not only applies where the company insures just a single item, but applies no matter how many similar items it insures.[8]

[6] The amount spent on insurance, $a = (A - r)/y$. We have $BC = A$, $BV = r$, so that $VC = A - r$.

$$RB = \frac{A}{y}, \quad \text{so that} \quad UV = \frac{A - r}{y}$$

which is the insurance premium.

[7] If K is the company's capital, the expected value is

$$v = (K + a)\frac{x - 1}{x} + (K + a - ay)\frac{1}{x}$$

$$= K + a - \frac{ay}{x}$$

The range, $r = ay$: hence the equation of the company's opportunity line is

$$v = K + r\left(\frac{1}{y} - \frac{1}{x}\right)$$

[8] Suppose that the insurer insures n items for a premium a at a rate y, the probability that he will have to pay off on any item being $1/x$. The following table shows the various possibilities, his initial capital being K.

Items on Which He: Pays off	Does not pay off	Expected Value of Capital	Probability
0	n	$K + na$	$\frac{(x-1)^n}{x^n}$
1	$n - 1$	$K + na - ay$	$\frac{n(x-1)^{n-1}}{x^n}$
2	$n - 2$	$K + na - 2ay$	$\frac{n(n-1)}{1\cdot 2}\cdot\frac{(x-1)^{n-2}}{x^n}$
k	$n - k$	$K + na - kay$	$\frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdots k}\cdot\frac{(x-1)^{n-k}}{x^2}$
n	0	$K + na - nay$	$\frac{1}{x^n}$

[Footnote cont. on p. 124]

It may be argued, of course, that the simple range of extreme values is not an adequate measure of the uncertainty or risk of a situation, and that more refined moments of the probability distribution, such as the standard deviation, would be better measures of the degree of risk. Whatever measure of risk we adopt, however, the essential argument is unchanged: there will always be some opportunity function (not, of course, necessarily a straight line) relating the expected value to the various measures of risk, and some indifference system which can be applied to it. We can, of course, extend the analysis to several dimensions of risk, expressing our opportunities and preferences in terms not only of a simple measure of the extent of the risk, but also of measures of its quality, e.g. skewness. This may be actually of some importance in the theory of gambling, where, for instance, a gambler may prefer one of two situations which are identical in expected value and in deviation, but which differ in respect to the "length" of the chances involved (i.e. the skewness of the probability distribution). People are often more willing to take "long chances" than short ones which are otherwise identical. It may be doubted, however, whether many propositions of interest to the economist can come from such refinements, and for the purposes of this work the very simple theory of these pages will suffice to indicate the form of the principles involved, even if it oversimplifies their content.[9]

Given the curves of Figs. 55 and 56 for all individuals, it would not be difficult to construct something like market curves, or demand and supply curves, for gambling or for insurance which would give us the "market odds" for gambling or a market premium rate for insurance. Thus in Fig. 57 we suppose a gambler with a given expectation of the subjective odds. The greater the objective (market) odds in this case, the steeper his opportunity line, which will move from MP_0, where the market and his subjective odds are equal, to positions such as MP_1,

Multiplying each capital sum by its probability and summing, we get, after simplifying the binomial expansions thus obtained,

$$v = K + na - \frac{nay}{x} \tag{1}$$

The range of variation,

$$r = K + na - (K + na - nay) = nay \tag{2}$$

Eliminating na between (1) and (2), we get the equation of the opportunity line:

$$v = K + r\left(\frac{1}{y} - \frac{1}{x}\right) \tag{3}$$

The opportunity line is, therefore, independent of the number of like cases which the insurer insures.

[9] Cf. H. Makower and J. Marschak, *Economica*, N.S., vol. 5, p. 261 (August 1938).

MP_2. His point of equilibrium will trace out a locus, $P_0P_1P_2$, from which we could derive his "demand curve" for bets, indicating how much he will bet at various levels of odds. Generally, as we see from the figure, a rise in the market odds will induce him to bet more, assuming that this does not affect his own expectations. The possibility of backward-sloping demand curves, however, should be taken under consideration: if the odds are very good in the gambler's estimation he may actually be willing to bet less. Such an outcome, however, is unlikely in this case; it could happen only if the marginal rate of risk substitution increased with increase in the money stock. In fact the marginal rate of risk substitution is more likely to diminish with increase in money stock—the more money we have, the more inclined we are to have a little fling, and the smaller the increase in expected value which would be necessary to compensate for a given increase in risk.

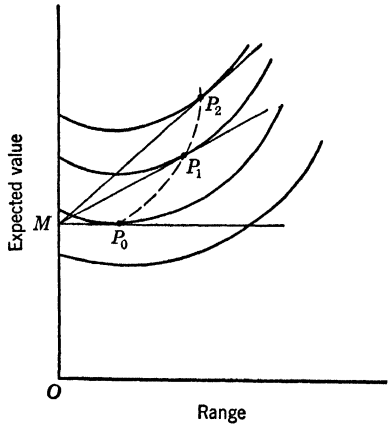


FIG. 57. The demand for gambling.

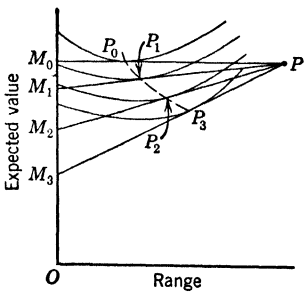


FIG. 58. The demand for insurance.

Similarly we could derive "supply curves" for the bookmakers, showing how much they would be willing to take in bets at each level of odds. The market rate of odds is that at which the volume of bets offered by the gamblers is equal to the volume which the bookies are willing to take. We do not have to assume, of course, that there is a sharp division between gamblers and bookies, though in practice such may often be the case. For each individual we can theoretically derive a market curve showing how much he will bet, or accept in bets, at various levels of odds. The determination of the odds in betting in a "competitive market" is thus exactly analogous to the determination of price in a competitive market.

Similarly, in the insurance market, as in Fig. 58, we can suppose that an increase in the premium rate, other expectations being constant, will

result in a steepening of the opportunity line through successive positions such as PM_0, PM_1, PM_2. The equilibrium point follows a locus $P_0P_1P_2$, etc., indicating that the higher the premium rate, the less insurance will be taken out. For each individual, then, demand and supply curves for insurance can be derived, and the market premium rate, at which the quantity of insurance demanded is equal to the quantity offered, can be derived from the summated market demand and supply curves.

The principle of increasing risk

The apparatus of this chapter can be applied to the exposition of Kalecki's "principle of increasing risk." [10] The principle states that as a firm expands by the aid of borrowed capital the chance of loss of its own capital increases. Suppose, for instance, that there is a chance that a firm might lose 10% of its total assets. If it had a net worth of $100,000 and borrowed no capital, it might find—if it was unlucky—that its total capital was reduced to $90,000. If, however, it borrowed $400,000, so that its total assets were $500,000, and the loss of 10% of its assets occurred, it would lose $50,000, representing half of its net worth. If it borrowed $900,000 and lost 10% of its total assets of $1,000,000, its net worth would be wiped out altogether. It is clear, therefore, that with a given chance of losing a given proportion of total assets there will be some point at which the firm will cease borrowing in order to expand its operations simply because of the increasing risk of loss of its own capital.

The principle can be illustrated by means of the simple method of this chapter. Increased borrowing increases the "range" of possible values of the net worth: it also may increase its expected value. We can, therefore, postulate an opportunity line such as MR on Fig. 59; this will be touched by an indifference curve at the point of equilibrium, beyond which the increase in the risk, as measured by the range, more than outweighs the increase in the expected value. It is not hard to show that in this case also the opportunity line is straight, its slope depending on the relative chance of loss or gain.

Suppose that the initial capital or net worth of the firm is A and that it borrows an amount B, so that its total assets are $A + B$. Suppose that there are only two possibilities, one of loss, which will reduce the total assets to $(A + B)y$, where $y < 1$, and another of gain, which will increase the total assets to $(A + B)x$, where $x > 1$. Suppose the chance of the favorable event is z, the chance of the unfavorable event there-

[10] M. Kalecki, The principle of increasing risk, *Economica*, N.S., vol. IV, p. 440 (November 1937).

fore being $1 - z$. Then the range of possible values of the net worth, r, is given by:

$$r = (A + B)(x - y) \tag{1}$$

for the net worth in the favorable event is $(A + B)x - B$ and in the unfavorable event is $(A + B)y - B$. The expected value, v, is given by

$$v = \{(A + B)x - B\}z + \{(A + B)y - B\}(1 - z) =$$

$$(A + B)\{xz + y(1 - z)\} - B \tag{2}$$

Eliminating B between these two equations, we get the equation of the opportunity line:

$$v = A + \left(z + \frac{y - 1}{x - y}\right)r \tag{3}$$

From (1) we see that when $r = 0$, $B = -A$. That is to say, in order to avoid any risk whatever, it would be necessary to lend the whole capital of the enterprise, assuming, of course, that lending can be done without risk. If we assume that the possible loss and the possible gain are equal the equation of the opportunity line can be simplified. Thus, if we write $x = 1 + t$ and $y = 1 - t$, the equation (3) becomes

$$v = A + r(z - \tfrac{1}{2}) \tag{4}$$

Under these circumstances the chance of gain must be greater than the chance of loss ($z > \frac{1}{2}$) if the opportunity line is to have a positive slope; the greater the chance of gain, the greater the slope of the opportunity line, and the more, therefore, is likely to be borrowed. If the chance of gain and the chance of loss are equal ($z = \frac{1}{2}$), the opportunity line will be horizontal: borrowing merely increases the risk without increasing the expected value. Under these circumstances, as we have seen, there will be no enterprise whatever unless the entrepreneur is positively risk-loving, i.e. unless his indifference curves have an initial negative slope. If the chance of loss is greater than the chance of gain, of course, the slope of the opportunity line will be negative, and the entrepreneur would have to be very risk-loving in order to engage in the enterprise. It should be observed that the entrepreneur may have to be risk-loving to engage in the enterprise at all, not merely to borrow. Just to engage in the enterprise involves some risk, i.e. involves moving down the opportunity line, and unless either the upward slope of the opportunity line counteracts the entrepreneur's disinclination to take risk, or the entrepreneur actually likes risk, he will hold his capital in riskless form and will not engage in business.

Application to the theory of interest

This type of analysis also throws light on the effect of a change in the rate of interest on the scale of the enterprise. In the above exposition we assumed that interest was zero. A positive interest can, however, easily be taken into account. Suppose that the amount of interest paid or received on a loan B is Bi. It should be observed that i is not the annual rate of interest unless the period of the loan is one year. Then the net worth will be $(A + B)x - B(1 + i)$ in the favorable event, and $(A + B)y - B(1 + i)$ in the unfavorable event. It is evident that the range is not affected by the introduction of interest, for the interest payment Bi must be deducted from both extremes of the range. The expected value, however, becomes

$$v = (A + B)\{(x - y)z + y\} - B(1 + i) \tag{5}$$

and the equation of the opportunity line is

$$v = r\left(z + \frac{y - 1 - i}{x - y}\right) + A(1 + i) \tag{6}$$

The point of intersection of the opportunity lines (3) and (6) is obtained by solving the two equations. This point is

$$r = A(x - y), \quad v = A\{xz - y(z - 1)\}$$

This is also the point at which the amount borrowed is zero, and the enterprise is conducted solely with its own capital, as may be verified by putting $B = 0$ in equations (1) and (2) or (5). That is to say, the effect of interest is to swivel the opportunity line upon the no-borrowing point P_0 in Fig. 59, a rise in the rate of interest making the line flatter than before. Thus if MP_0R is the opportunity line at zero interest, where $OM = A$, the total capital of the enterprise, $M_1P_0R_1$, is the line at a rate of interest r, where, if t is the period of the enterprise,

$$OM_1 = A\left(1 + \frac{r}{t}\right)$$

The enterprise will move to the point R at zero interest where the opportunity line touches an indifference curve. A rise in the rate of interest, making the opportunity line $M_1P_0R_1$, moves the equilibrium point left to R_1. It is easy to show that this represents a decline in the amount of borrowing and in the scale of the enterprise. Draw MHK horizontally to cut P_0N_0 in H and RS in K. Draw OH to meet SK

(produced) in B_0. Then, putting $B = 0$ in equation (1), p. 127, we see that $MH = A(x - y)$: also $MK = (A + B)(x - y)$, whence $HK = B(x - y)$. But $OM = A$, so by the properties of the similar triangles OMH and B_0KH, $KB_0 = B$, the amount borrowed. At the higher rate of interest a similar construction shows that K_1B_1 is the amount borrowed. At still higher rates of interest capital will be lent. Thus from Fig. 59 a demand-supply curve for loans could be constructed, showing how much the individual will borrow or lend at each rate of interest. If now the indifference curves are vertically parallel, a rise in the rate of

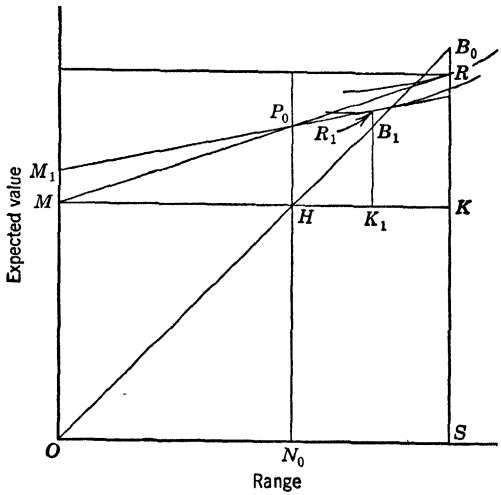


FIG. 59. The principle of increasing risk.

interest must move the equilibrium point further to the left, e.g. from R to R_1, and must therefore lower borrowings or raise lendings. A backward-sloping demand curve for loans, with more borrowing at high rates than at low, is possible if the marginal rate of substitution (the slope of the indifference curves) rises with an increase in expected value, risk being constant. This, however, is very unlikely as it would imply that with increased wealth the aversion to risk also increased.

By summing all the individual demand-supply curves for loans we could obtain a market demand-supply curve, assuming a perfect market, and from this could derive the market rate of interest at the point where the quantity of loans demanded equalled the quantity offered. The assumptions of such an analysis, however, are very far from reality: the loan market is extremely imperfect, and, of course, the analysis of risk

which we give here is quite crude. Nevertheless it throws some light on the phenomena of the real world.

Imperfection in the loan market

Imperfection in the loan market facing the individual firm will reflect itself in discontinuities or curvature in the opportunity line. Thus we may frequently find that the conditions of lending are different from the conditions of borrowing—that, for instance, a bank can usually borrow at a lower rate of interest than that at which it can lend, whereas a manufacturing firm or an individual will frequently find that the rate of interest at which it can borrow is higher than the rate at which it can lend. In either case there is a "kink" in the opportunity line at the no-borrowing no-lending point P_0.

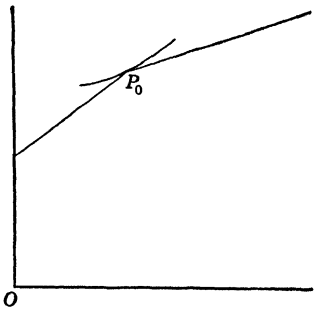


FIG. 60. Imperfection in the loan market.

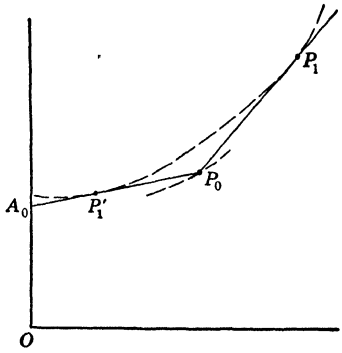


FIG. 61. Imperfection in the loan market.

This situation is illustrated in Figs. 60 and 61. In Fig. 60 the firm can lend only at a lower rate than that at which it has to borrow. Under these circumstances there is the possibility of a rather stable equilibrium at the point P_0: the firm will employ its own capital but will be averse to either lending or borrowing. In the converse circumstances of Fig. 61, where the firm can borrow at a lower rate than it can lend, the point P_0 would represent a minimum utility rather than a maximum utility position, even if the indifference curve touched at that point. In this case the firm will not be content to work with its own capital: it will be either a borrower or a lender, depending on its risk preference. If it abhors risk—as, for instance, a banker is likely to do—it will have very steep indifference curves and will move right to the point A_0. If it is more favorable to risk it may be a borrower. An interesting possibility is that shown in the figure where an indifference curve touches both

arms of the opportunity line at P_1' and at P_1. This is a multiple solution: it will be equally satisfactory to the firm to be either a borrower or a lender, provided that the borrowing or lending is on a large enough scale! This situation, however, is not very likely: almost the only firms which have this privilege are banks, and bankers are characteristically risk-haters, so that they are likely to remain in the lending position. The theory of a bank, however, is more complex than can be treated by means of the apparatus of this chapter.

An imperfect market for loans may take the form of subjecting the firm to higher interest rates the more it borrows. This would make the

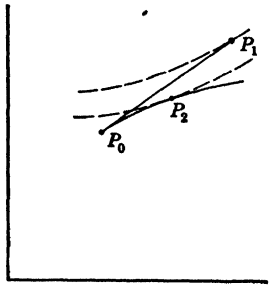


FIG. 62. Imperfect loan market.

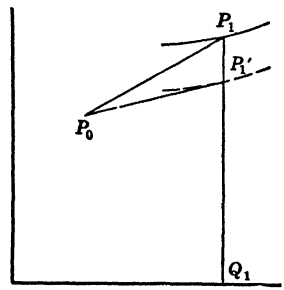


FIG. 63. Capital rationing.

opportunity line P_0P_2 bend back as in Fig. 62: if it has any result it is likely to lead to less borrowing and smaller scales of enterprise. An extreme case, and one of great practical importance, is capital rationing, in which the amount which the firm can borrow is limited by the unwillingness of the borrower to go beyond a certain amount no matter what inducement is offered. In terms of our opportunity line this means that for amounts beyond the rationed sum the rate of interest is infinite, and the opportunity line turns vertically downward. In terms of Fig. 63, if P_1 represents the extreme position which the rationed amount of borrowing will allow the firm to reach, the opportunity line is $P_0P_1Q_1$. Here it is clear that the point P_1 is likely to be a fairly stable position of equilibrium. As in the rationing of commodities, a change in the "price"—in this case the rate of interest—provided that the rationing continues to be effective in the sense that more would be bought (or borrowed) if there were no rationing at the existing price (or rate of interest), does not affect the amount taken (borrowed). Thus in Fig. 63 a rise in the rate of interest will shift the opportunity line from P_0P_1 to P_0P_1', the expected value will decline as more interest is paid, but the range will

not change and the amount borrowed will not change.[11] It does not follow, of course, that a change in the rate of interest is wholly without effect under capital rationing. A sufficient rise in the rate of interest may, of course, make rationing ineffective and reduce the borrowings below the rationed quantity; in Fig. 63, if the line P_0P_1' is touched by an indifference curve at P_1', any further rise in the rate of interest is likely to reduce borrowings below the rationed amount. Even if the rationing is still effective, however, a rise or fall in the rate of interest by permitting a fall or a rise in the profits of the firm may affect both the composition of its assets and its rate of growth.

The flexibility of assets

Another important impact of risk and uncertainty on the structure and size of the firm is on the flexibility of its assets and program. This is a quantity which does not lend itself particularly well to exact quantification and hence is not very amenable to mathematical analysis; nevertheless it is clearly an important property of an asset structure. There are certain assets, e.g. specialized machines or buildings, which can be used only in certain ways for the production of specific commodities. There are other assets which can be used in many ways, particularly, of course, money, but also certain buildings which can be used for many purposes, raw materials which can be made up into several different kinds of commodities, and so on. It is as impossible to quantify this character of specificity or flexibility as it is to quantify the related character of liquidity: indeed, flexibility is to production what liquidity is to exchange. It is, however, possible to define asset preferences, either by an indifference-curve system or by preferred ratios, as we have seen, and we can roughly measure the flexibility of an enterprise by the proportion of total assets which are held in flexible form, just as we can measure liquidity roughly by the liquidity ratio.

An increase in uncertainty is likely to increase both liquidity and flexibility preferences, and to move the enterprise into a more liquid and flexible asset structure, even at the cost of a sacrifice in expected profits. This fact may be of great importance in interpreting the dynamic, and especially the cyclical, behavior of firms. Any general increase in un-

[11] From equation (1), p. 127, we see that

$$B = \frac{r}{x - y} - A$$

That is to say, the amount borrowed is a function only of the range and not of the expected value: lines of equal borrowings in these figures would be vertical straight lines.

certainty, e.g. a change in the climate of business opinion such as characterizes a turning point of a boom, is likely to result in a decline in preference for finished products and for specialized capital and a rise in preference for raw materials and money.

Risk versus uncertainty

A word should perhaps be added in regard to Knight's famous distinction between "risk" and "uncertainty." [12] Risk is insurable: uncertainty is not. Risk, that is to say, can be expressed in terms of a given probability: in uncertainty not only the value of our assets is in doubt, but we cannot even state the probability of the various possible values. This distinction arises because some events, although uncertain in themselves, are members of a large family or universe of like events. A probability is simply the proportion of times a given event occurs in a large universe of cases. It is the existence of a universe of like cases which makes insurance possible: thus from the compilation of the age at death of large numbers of people we can derive life tables which give the probability of death of any single individual. Similar probabilities can be derived for such things as fires, automobile accidents, embezzlements, and weather. There are other events, however, which are also uncertain, but which are not members of large groups of similar events, and for which, therefore, no probability or universe-frequency tables can be derived. In Knight's view the profits of enterprise are such events, and this fact gives profits their peculiar character.

In some cases unquestionably the non-insurability of profits arises from the above cause: as, for instance, in highly specialized or new enterprises where there is not sufficient experience of other enterprises of like character to justify the construction of probability tables. There are other cases, however, where there are a large number of individuals: the probability of bankruptcy of retail stores of various kinds, for instance, could be calculated with fair degree of success. One suspects that in such industries it is another element which prevents the application of insurance to profits. Insurance is possible only if the *fact* of insurance does not affect too greatly the probability ratings of the thing insured. Life insurance may promote suicide and fire insurance arson, but fortunately these effects are small enough to be handled. There seems to be so close a connection, however, between the hope of profits, and even more, perhaps, the fear of losses, and the activity of the entrepreneur that to insure profits would be likely to diminish entrepreneurial activity and to make losses much more likely. In such an event the

[12] F. H. Knight, *Risk, Uncertainty, and Profit* (New York, 1921), p. 46.

premium for the insurance would have to be so high that it would not be worth paying.

There is some possibility of measuring uncertainty in terms of the "probability of the probabilities" of events. Such refinements, however, do not seem to lead to useful propositions, at least at this stage of knowledge, and they rest on such dubious philosophical and empirical foundations that we may well leave them with a bare mention. Too great a refinement in theory which has not even any value in the interpretation of aggregates of behavior is better avoided.[13]

[13] This chapter was written before G. L. S. Shackle's *Economics of Expectations* came to my attention. It may be observed that the concept of the "focus outcome" which is there developed is admirably adapted to the use of the "range of values" as a measure of uncertainty.

Chapter 8

The Theory of Consumption

The prime mover of production is consumption, and it is clear that any adequate account of the economic system must contain some consumption theory. The theory of the household and of the firm, and indeed the whole structure of economics, have been weakened by a general failure to appreciate the nature of consumption. For this reason the present theory of the household is especially unsatisfactory. Nevertheless it is by no means easy to formulate a worthy substitute. The theory of consumption turns out to be, rather surprisingly, one of the most difficult parts of economics, where it seems to be particularly hard to formulate rational principles and establish propositions from them.

The principle of minimizing consumption

We must first distinguish clearly between the consumption and the utilization or "enjoyment" of capital. Consumption in its literal sense means destruction: if we ask destruction of what, the answer is destruction of capital, i.e. of real assets. When we eat food, burn fuel, and wear out clothes we have a smaller stock of these things than we had before the act of consumption. Economists have frequently written as if consumption was the desideratum, the end product of all economic activity. Such, however, is not the case. It is true that there are some commodities which must be consumed in the utilization, such as food and fuel. This, however, is a technical accident. For most commodities consumption is merely incidental to their use and, far from being a desideratum, is to be avoided as much as possible. We want to live in houses, and we derive our satisfactions from their existence, not from their wearing out or depreciation. If our houses did not depreciate: if roofs never leaked, if paint never weathered, if beams never sagged and floors never rotted, we should have less consumption of houses, but we would clearly be much better housed. Similarly we want to wear clothes rather than to wear them out, to use furniture rather than to use it up, and so on down the whole list of commodities. If we had unabradable fabrics, unbreakable china, and indestructible furniture we

should again have less consumption, but we would be all the wealthier for this fact.

The principle that economy in consumption is always desirable applies even to one-use-goods such as fuel and food which cannot be utilized without being consumed. We do not want to consume fuel except as a means to getting heat, light, or power: the more of these things we can get from a unit of fuel the better off we are, and not a little of the technical advance of the past hundred years has been in the economy of fuel. We burn coal in a house furnace only because the temperature of the house is a depreciating asset in winter, which has to be replaced by the consumption of fuel. In the case of food the principle of minimizing consumption may seem to run into difficulties. We can imagine unconsumable houses, furniture, clothing, automobiles: we might even imagine an unconsumable fuel—indeed, for all practical purposes, water power is such. It is perhaps stretching the imagination too far to conceive of an unconsumable steak. Even with food, however, the principle of minimizing consumption has meaning. It is only the technical difficulties of economizing consumption in this case which make it difficult to conceive of the process. Clearly, however, an invention which enabled us to satisfy our hunger and to provide fuel and maintenance for the body with a smaller consumption of food would leave us better off than before. Indeed, small changes in this direction have actually taken place: with the development of the science of nutrition and the balanced diets which it advocates it has become possible to maintain health on an appreciably smaller intake of food than was possible with the unbalanced diets of an earlier generation.

It may be objected that sometimes satisfaction is indeed derived from the very act of consumption itself, and cannot be disassociated from it. The food pill of popular scientific fiction is no adequate substitute for steak and french fries, even though it might satisfy all bodily needs just as well. Even here, however, the oil cruse of the fortunate widow [1] would represent a substantial technical improvement. The principle of the minimization of consumption is not invalidated merely because there are technical cases in which consumption cannot be reduced without reducing satisfaction. One cannot escape the feeling also that satisfaction in consumption itself verges on the sadistic and morbid. The sadist, the pyromaniac, and the glutton are alike in that they obtain satisfaction from the destruction or disintegration of those orderly arrangements of matter which constitute capital and which are the result of productive activity. Such satisfactions, however, can hardly be regarded as healthy.

[1] II Kings 4.

The principle of minimum consumption may then be stated as follows: of the various types of assets which satisfy a given want, we will prefer that which is consumed (or depreciated) at the slowest rate in terms of the value unit. This does not necessarily mean, of course, that the asset with the longest length of life is always to be preferred. The rate of depreciation or consumption of an asset is a product of two factors: the value of the asset itself and the proportionate rate of value decline. It is this latter quantity alone which is related to the length of life of the asset. Thus suppose that we have an asset—say, a bed—which has a value of $200 and which depreciates at a rate of 5% of its value each year: its rate of value consumption is $10 per annum. We might have another bed just as good to sleep in or look at which was valued at $100 and which depreciated at a rate of 10% per annum: this, in spite of its smaller durability, because of its low cost would have consumability equal to that of the costlier, more durable bed.

Modifications of minimum consumption

It may be necessary at this point to make a certain modification of the principle of minimum consumption. Of two assets, each serving the same need, and having equal rates of consumption, we may prefer the

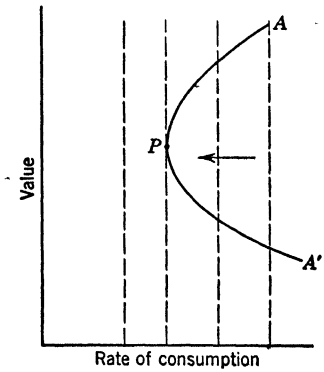


FIG. 64. The principle of minimum consumption.

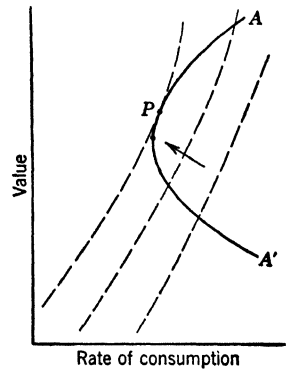


FIG. 65. The principle of minimum consumption.

more valuable, more durable asset because of certain prestige values. This modification, however, can easily be taken care of by means of the usual indifference-curve analysis. Thus in Figs. 64 and 65 we plot indifference curves (dotted lines) on a field, showing combinations of value and rates of consumption for various assets filling a given need. In Fig. 64 we suppose that the value of the asset itself has no effect on prefer-

ences: the indifference curves are vertical straight lines, the surface of which they are contours rising towards the value axis, indicating that an asset with a smaller rate of consumption is preferred to one with a larger. In Fig. 65 we suppose that there is a certain value preference: the indifference curves have a positive slope, indicating that the individual is willing to suffer a certain increase in the consumability of his asset in order to obtain an asset with larger value. Thus, moving in a vertical direction, we cut successively higher indifference curves, indicating that of two assets of equal consumability the individual will prefer that with the higher value.

The opportunity line in this case is likely to have the shape of the solid line APA' in the figures. Assets with very high values and assets with very low values alike may have high consumability, the former because the high value of the expensive asset more than compensates for any increase in its durability, the latter because the low durability of the cheap asset more than compensates for its low value. There will be some point such as P in Fig. 64 at which the consumability of the asset is a minimum. Each point on this opportunity line represents a different type or quality of a general kind of asset: thus if the asset is automobiles, the point A might represent a high-priced, durable car, the point P a medium-priced, less durable car, the point A' a low-priced, non-durable (perhaps second-hand) car. The opportunity line is, of course, more likely to be a series of discontinuous points rather than a continuous curve, but we shall not fall into any serious error by assuming a continuous variation of value and consumption. In the case of Fig. 64 the point of minimum consumability is the preferred point. In Fig. 65, however, the preferred point is P, which is at a higher value and a higher consumability than the point of minimum consumability, owing to the positive slope of the indifference curves. Figure 65 is likely to be characteristic of the rich, who prefer high-value assets even at the cost of greater consumability. A reverse situation is also possible (not shown in the figures) in which the indifference curves slope negatively, indicating that the individual is willing to pay a certain cost in terms of consumability in order to get a low-value asset. This may well be the situation with the poor, whose economic behavior is dominated by the inadequacy—or even negative magnitude—of their net worth, and who therefore have to sacrifice economy of consumption in order to economize assets.

This phenomenon is clearly associated with the whole borrowing-lending position of the individual, and particularly with the element of capital rationing. If an individual can borrow or lend freely in a perfect loan market the only limitation on his *ability* to minimize consumability will be the payment of interest. Because of this we may expect the in-

difference curves in Fig. 64, in the absence of any positive preference for high or low values, to have a slight negative slope: high-value assets have a slight disadvantage because of the interest charge. The equilibrium point then will lie somewhat below the point of minimum consumability: the greater the rate of interest, the more pronounced the slope of the indifference curves and the greater the divergence from the point of minimum consumability. With imperfect capital markets or capital rationing the curvature of the indifference curves will be more pronounced, and the divergence from the point of minimum consumability will be all the greater. The principle of minimum consumability is, of course, perfectly general and applies to all organisms, to firms just as much as to households. If there are a number of different machines, for instance, capable of performing the same service, that machine will be selected which has the least annual rate of depreciation, subject to the modifications outlined above.

Production, consumption, expenditure, and receipts

The next step in the theory of the single economic organism is to outline the relationships among production, consumption, expenditure, and receipts. These relationships are quite general and apply to firms just as much as to households. The consumption concept we have already defined as the destruction of capital. We may if we like distinguish between planned consumption, such as the eating of food, the wearing out of shoes, and the burning of fuel, and accidental consumption, such as the burning down of a house, the smashup of a car, the tearing of clothes. It may be doubted, however, whether this distinction is of great importance, especially in the large. Once we admit that it is not in the consumption but in the utilization of capital that satisfaction lies, the attempt to distinguish between planned and accidental consumption becomes less necessary. *All* consumption—with the morbid exceptions noted above—is undesirable. Whether a house burns down, blows down, or merely wears out, all forms of consumption are equally incidental to its use.

The consumption concept we conceive of originally as a physical concept, i.e. it is a list of so many loaves of bread, so many suits of clothes, etc., consumed in a given period of time. As such, of course, it is heterogeneous and non-aggregative: in order to aggregate consumption we must reduce its various heterogeneous items to a common *numeraire* by valuing each at a price, just as in the case of the aggregation of capital. The value aggregate of consumption may be given the name *outgo*, for, as we shall see, it is strictly analogous to income.

Corresponding to the consumption concept we have, of course, the concept of production, which is the creation of real assets. Production, like consumption, is an aggregate of heterogeneous items; when these are valued at a price the value aggregate is *money income*. The income and outgo concepts, which are essentially value aggregates of additions to and subtractions from the total stock of assets, must be distinguished clearly from the receipts and expenditure concepts. Receipts consist of the additions to liquid assets or money. Expenditures consist of the subtractions from liquid assets or money. It is hardly any exaggeration to say that the failure to distinguish clearly between receipts and income on the one hand, and between expenditure and consumption on the other, has been the source of most of the confusion in economics, and, especially in macroeconomics, in the past generation. The confusion has arisen because frequently income and receipts on the one hand, and outgo and expenditure on the other, are both involved in the same transaction. Thus the payment of wages seems to be both income and receipts. Nevertheless in principle the two must be separated. A wage payment is not income: it represents simply an exchange of one form of asset—a wage claim on the employer—for another kind of asset, money, or perhaps bank deposits, which are a claim on a bank. The income is earned as the wage claim on the employer is created by the daily sale of labor. Even if no wages are paid during the day when a man works, income is still earned, for he increases his total assets by the amount of the wage claim on the employer. Similarly a farmer earns income as the value of his crop grows: the moment of sale when he obtains receipts is merely an exchange of one form of asset—the crop—for another form of asset—money.

In "outgoings" the distinction between expenditure and outgo is even clearer. When a durable good is purchased it is evident that no consumption is involved: all that happens is an asset transfer or exchange, money diminishing and goods increasing in the balance sheet of the purchaser. Consumption comes later when the good depreciates or is destroyed. Even for non-durable goods the distinction still holds: the purchase of an ice cream is not the same thing as its consumption. In the case of services the expenditure and the consumption may be simultaneous. When psychic capital is taken into consideration, however, it may be doubted whether there are any really non-durable goods. Even the things usually labeled as services, such as movies, in fact produce psychic capital with a limited rate of depreciation. We go to the movies in order to produce the mental state of just having gone to the movies. This mental state is the commodity which we purchase with the price of

admission: it is a commodity which depreciates like every other commodity. For some individuals it depreciates rapidly, so that it has to be replaced in a week or less: in other individuals the rate of depreciation is slower, so that they do not have to go to another movie for a month or even several months. Oddly enough, the mental state of having gone to a bad movie may depreciate at a slower rate than the state of having gone to a good one! There are interesting problems of the impact of disappointment and of appetite whetting involved here, which are to be neglected at peril by anyone who undertakes to study the dynamics of demand.

The concept of saving

Income minus outgo is the addition to total assets, which may properly be called "saving." There is a certain ambiguity in this concept, arising from a similar difficulty in the concepts of income and outgo. In physical terms there is no ambiguity: if, for instance, a man produces 100 lb of potatoes and consumes 90 lb he has clearly added 10 lb to his stock, if in consumption we include, not merely eating, but also spoiling. Similarly if we always value the heterogeneous items of production and consumption at constant prices there is no ambiguity: if I produce $10,000 worth of assets and consume $9000 worth then clearly I have "saved," i.e. added to my stock, $1000 worth. The ambiguity arises when the prices at which stocks, and additions to or subtractions from stocks, are valued change in the accounting period. Suppose, for instance, by way of illustration, that there is only one commodity, wheat, and that the individual starts with 500 bushels, produces 1000 bushels and consumes 900 bushels, adding 100 bushels to his stock, which makes it 600 bushels. If the price of wheat is $2 per bushel his income is $2000, his outgo is $1800, and he adds $200 to the value of his stock, which rises from $1000 to $1200; there is no ambiguity. Suppose now, however, that the price of wheat rises on the last day of the accounting period to $3 per bushel. At the beginning of the period he has 500 bushels valued at $2 per bushel, worth $1000. At the end of the period he has 600 bushels, worth $3 per bushel, or $1800. The value of his assets has increased not by $200, but by $800, although the value of his production is still only $2000 and of his consumption $1800. This "saving" of $800 may be considered to consist of two parts: the value of the goods added to his stock, i.e. the value of his real saving at the old price, which is $200 as before, plus the increase in the value of his assets due to a change in their price, which is $600—each of his 600 bushels on the last day of the period having increased in value by $1. In this case, therefore, instead

of simply writing

Value of saving (S)

$$= \text{Value of production } (P) - \text{Value of consumption } (C) \quad (1)$$

we shall have to write

Value of saving, or Increase in net worth (S)

$$= \text{Value of production } (P) + \text{Increase in value of assets due to price changes } (P') - \text{Value of consumption } (C) \quad (2)$$

If, now, we wish to define income (Y) as $P + P'$, the equation can again be written in the simple form,

$$S = Y - C \quad (3)$$

For many purposes, however, it is important to distinguish between that part of income which represents gross addition to physical assets, and that part which is generated by the revaluation of physical assets, as these do not possess the same significance. It is tempting to regard only P as "true" income, and to regard P' as in some sense spurious—income which is thrown up, as it were, by our accounting conventions. There are, in fact, certain attempts being made to revise accounting conventions in order to minimize P'; the last-in-first-out method of inventory accounting, by which, in effect, inventory is valued at approximately the same price level both at the beginning and at the end of the accounting period, is a case in point. In the valuation of durable fixed equipment and plant, also, accountants have been very loath to take account of changing price levels and have tended to perform the valuation at some fairly constant "standard" level of prices. It is something of an oversimplification, however, to regard all income due to the rise in prices of existing assets (and, of course, negative income due to a fall in price) as spurious. The man who owns real assets with rising prices in an inflationary period is clearly better off than the man who owns money or contractual claims (debts and mortgages) fixed in money terms: there is a real shift in the distribution of wealth as a result of any revaluation of different kinds of assets, and it would seem not unreasonable to say that those who have benefited from such redistribution have received income at the expense of those who have suffered from the redistribution. The United States income tax law recognizes that these "capital gains" are income of a sort, though of a peculiar kind, by taxing them at a different rate from ordinary income, but, of course, not all income of this

kind shows up as capital gains. Unrealized capital gains are not subject to tax; this is somewhat illogical in principle, though probably justified in practice by the difficulties of accounting on an accrual basis. Certain profits and losses also which have their origin in a revaluation of inventory or other capital appear as ordinary income.

This difficulty actually is a very deep-seated one and strikes at the root of much conventional accounting and legal practice. It is a general accounting and legal convention that the realized is the real: that is to say, an asset value is uncertain and doubtful until it is "realized" by exchange into money, and that therefore an increase in liquid assets is the most certain measure of income or profits. The famous formula of Marx, M-C-M'—money converted into commodity, which is reconverted into more money than before—is the standard formula of enterprise according to the accountant. All this assumes, however, that the purchasing power of money is approximately constant, and that while other asset values fluctuate relative to each other the general "value of money" remains about the same. In times of general inflation or deflation this assumption, of course, breaks down, and there is no greater certainty in the general purchasing power of money than there is of any other asset. Hence there is no particular reason to suppose, in the absence of any stability in the value of money, that money values are an accurate measure of income or of profits: it may very well be that "corn" or labor, as Adam Smith suggested, or some more sophisticated index of a group of commodities, would be a better "measure of value" than money. Thus in an inflation it is an unsafe thing to hold money; even the most dubious real asset may be preferable.

Hoarding and saving

The net increase in an individual's liquid assets (money) may be called "hoarding" to distinguish this quantity from "saving" which we have defined as the increase in total assets. As receipts represent an increase in money stock, and expenditures represent a decrease, we can write:

$$\text{Hoarding } (H) = \text{Receipts } (R) - \text{Expenditures } (E) \qquad (4)$$

Hoarding in no sense has to be equal to saving: it is quite possible for the two quantities even to have different signs. Thus an individual who buys a house is likely to find that in that year he has saved but has dishoarded: his total assets have increased but his liquid assets have diminished. Similarly an individual who sells property during a year may find that he has dissaved, his outgo having exceeded his income: yet he may well have hoarded, i.e. have increased his liquid assets. If N represents the total change in the value of an individual's non-liquid

assets we have

$$S = N + H \tag{5}$$

That is, the change in total assets must be equal to the change in liquid plus the change in non-liquid assets. It follows that

$$N = S - H = P - C - R + E \tag{6}$$

These relationships can be generalized for any number of kinds of assets. Suppose that the individual has a number of different kinds of assets, A, B, C, etc. Let V_a, V_b, etc., be the value of the change in the quantity of the asset held during the accounting period ("year"). Then the stock of each asset is augmented by production and by purchase, which is the exchange of some other asset for it. The stock is diminished by consumption and by sale, which is the exchange of the asset for some other. Thus, if P_a is the production, C_a the value of its consumption, E_a the value of the amount purchased, and R_a the value of the amount sold, we have:

$$V_a = P_a - C_a + E_a - R_a \tag{7}$$

In the case of money we assume that there is no production or consumption of the asset, unless our individual is a counterfeiter! The "purchase" of money constitutes a receipt, and its "sale" constitutes an expenditure. For money, therefore, equation (7) reduces to (4). If we sum all such equations as (7) for all assets we find that the E's and R's will cancel out, according to the principle that in exchange equal values are exchanged. We are then left with equation (1), assuming that the assets do not change in relative value.

A theory of the consumer

Let us now see what is the problem which a theory of a consumer—who may be a firm as well as a household—must solve. During a "day" a number of things can happen to him. In the first place some of his assets will be consumed. Part of this consumption may simply be a function of the kinds and quantities of assets which he holds at the beginning of the day: his furniture, house, clothing, equipment, etc., will depreciate during the day at a certain rate whatever his other activities. The simplest theory of consumption would be to assume that all consumption was of this nature, in which case it would not be affected by the plans or activity of the consumer in the present, but would simply be a result of his previous accumulations of stocks of assets. If he holds highly consumable assets his consumption will be large: if he holds durable or non-consumable assets it will be small. This picture, although quite useful as a first approximation, is not altogether satisfactory be-

cause there is a certain amount of consumption which apparently depends on the degree of use to which the individual puts his assets. His car may be consumed a little more, i.e. may decline more in value, if he drives 300 miles during the day than if it sits at home in the garage. We cannot assume, however, that there is any very simple relationship between use and consumption. For many assets there seems to be an amount of use which minimizes consumption: if a car sits in the garage all day long it may actually depreciate faster than if it is driven a few miles. The same may even be true of tires. On the other hand there are some assets—e.g. gasoline—for which the minimum consumption comes at the point of zero use. It is clear that we cannot employ any principle of minimum consumption in this case: we may well prefer to use our car well beyond the point of minimum consumption, and there are conceivable situations in which we might wish to use an asset in a degree below the point of minimum consumption. The marginal principle can clearly be applied to this problem: we will push the use of each asset to the point where we judge the marginal gain (utility) involved by the consumption to be just worth the marginal cost (disutility) of the consumption itself. The marginal cost (disutility) of consumption is itself composed of two elements: consumption is disagreeable because it involves the loss of assets; it is disagreeable because if these assets are to be replaced there must be production, which itself may be disagreeable. If we assume that assets are always to be replaced, then every act of consumption will involve an act of production. If exchange is costless this means that in deciding whether or not to consume a dollar's worth of assets the gain resulting from the use of assets which involves their consumption must be balanced against the sacrifice involved in earning the dollar of income which is necessary to replace the dollar of outgo. If this sacrifice is constant, then, of course, the marginal utility of a dollar's worth of consumption should be equal in all possible lines of the use of assets. This is the famous equimarginal principle, and we can, of course, if we wish, restate it in terms of the less objectionable marginal rates of substitution.

It is evident, however, that the problem has more dimensions than the usual theory allows. The assumption which is implicit in the older theory, that the consumer wants to have a stable asset structure, is by no means necessarily valid. A consumer may wish to diminish his stocks of a particular asset by consuming it without replacement, or he may wish to increase his stocks of it by replacing more of it than he consumes. The problem is further complicated by the fact that particular stocks of assets can be augmented or diminished in two ways: either through production and consumption or through purchase and sale. A

theory of consumption by itself is not sufficient to give us a theory of consumer's demand. Nor is a theory of production or income sufficient to give us a theory of the supply of factors of production. A theory of demand is a theory of purchases: in particular, it is a theory of the determinants of the demand function, relating purchases to price. Purchases, however, are not the same thing as consumption, nor will purchases be equal to consumption except by accident, or unless (i) asset stocks are held constant, and (ii) production of the asset in question by the purchaser himself is impossible. The usual theory of demand assumes implicitly that consumption and purchases are identical, or at least equal, and that, therefore, the individual rushes out to replace by purchase (not by production) the assets which he has consumed.

There are several questions, therefore, for the theory of consumption to answer. The first is, Why does an individual consume the amounts of various kinds of assets that he does? The second is, How much of the consumed asset does he want to replace? The third is, How does he want to replace it: by production or by exchange, and if by exchange, what other assets does he wish to give for it? The fourth question is, What other assets does he wish to acquire, either by production or by exchange? In order to answer these questions we need to know not merely what determines the individual's pattern of consumption, but also what determines his asset structure. If we are given a consumption pattern and an asset structure we can then picture the dynamic process of the individual's economic parameters through time. His consumption pattern continually wastes his assets: if he did nothing but consume, his asset structure would not only diminish in total quantity or value but would also change its composition rapidly; non-durables would disappear before durables—food would vanish from the larder long before his clothes wore out or his house fell down.

The fact of consumption, therefore, forces him to do two things: to produce, i.e. earn income which adds to his *total* assets, and to exchange, i.e. change the *form* of his assets. If he did nothing but consume and earn (produce), he might be able to replace the *total value* of his consumption, or even more than replace it if he were saving, so that the total value or quantity index of his assets increased. The composition of his assets, however, would be likely to change, because the things which he adds to his assets by his production or earning are not necessarily the things which he consumes. Earnings are usually, though by no means always, in the form of money: consumption is usually of various real assets. If the consumer did nothing but consume and earn, then, the total value of his assets might not change, but their composition would change radically, such things as food and clothing diminish-

ing while liquid assets increased. In order to rectify this situation the consumuer is forced to exchange: he will usually have to exchange his product for money, and exchange the money for the various assets which have been consumed. Thus the farmer whose product is in the form of wheat sells the wheat and with the money so acquired buys food, clothing, etc., of various kinds to replace what he consumes and perhaps also to add to his total stock. The worker whose product is a wage claim likewise exchanges this wage claim for money when he gets his paycheck, and replaces his consumption by exchanging the money for other assets.

It is clear that we have here a complex system of interrelated variables. It would not be difficult to postulate an equilibrium of such a system in terms of utility or preference-function analysis, based on the propositions that in any day the individual produces of various assets that amount which he thinks is worth producing, consumes of each asset that amount which he thinks is worth consuming, and exchanges those amounts which he thinks are worth exchanging. A mere formal system of equations, however, does not tell us a great deal: it is the character of the functions that matters. This character is complex and may vary considerably from individual to individual. The patterns of the consumer are in fact the patterns of living and require complex psychological and sociological study. It may well be that variables which fall under the general field of sociology—e.g. Veblen's conspicuous consumption, emulation of the neighbors, conventional patterns associated with certain occupations or certain "positions" in society—play a more important role in determining the consumer's economic behavior than any strictly "economic" considerations.

Effects of price on the consumption and stocks

What we need for economics, however, is mainly some kind of insight into the impact of certain economic variables such as prices and money stocks on the whole complex of consumer behavior. The economist is familiar with the theory of the impact of prices on consumption. A shift in the relative price structure of different assets will normally discourage the consumption of higher-priced goods and encourage the consumption of lower-priced goods. There may be exceptions to this rule with "inferior goods" because of a strong "income effect" of the price change: these goods are consumed more at lower incomes than at high, and hence a relative rise in their price, by reducing the real income of the consumer, may actually force him to consume more.[2] Such cases,

[2] See Boulding, *Economic Analysis*, Rev. Ed., p. 756.

however, are rare. The impact of price changes on consumption is also complicated by the existence of complementary and competitive relationships among goods. Where many goods cooperate to serve the same ends—e.g. automobiles, oil, gas, roads, gas stations—a fall in the price of one of them is likely to lead to an increase in consumption of the whole group, subject to the possibilities of substitution. Goods that are competitive with the commodity that has suffered a price fall are likely to be consumed less as the lower-price good is substituted for them.

All this is familiar ground to economists. What has been neglected, however, is the impact of price changes on the desire of consumers to build up stocks. Purchases, as we have seen from equation (7), are equal to accumulations plus consumption less production and less sales. Even assuming that production and sales are zero, we see that the willingness to purchase depends not merely on the willingness to consume, but also on the willingness to accumulate. In the case of durable consumer's goods the willingness to accumulate may be a much more important factor determining consumer's demand than the willingness to consume. What, then, will be the impact of price changes on the willingness to accumulate (or decumulate)? It is evident that we cannot even discuss this question without taking dynamic considerations into account. The willingness to accumulate depends not merely on the present price structure: it depends also on the individual's expectations in regard to future price structures. Generally speaking, we will be more willing to accumulate goods when we expect their prices to rise than when we expect their prices to fall. The impact of a price change, then, depends on the impact of the change itself on the expectation of further changes. If the elasticity of expectations, to use Hick's useful concept,[3] is positive, then a fall in price, by inducing expectations of a further fall, may induce decumulation of the good and hence may actually lower purchases. Similarly, a rise in price, by stimulating expectations of a further rise, may actually raise the purchases of a good, contrary to the general "law of demand." It is true, of course, that these impacts are likely to be temporary. If a price rises to a permanently higher level, expectations of a further rise will soon cease, and there may be decumulation of stocks previously accumulated. The impact of a permanent rise in price on purchases and on consumption may therefore be somewhat as illustrated in Fig. 66. There is a gradual decline in consumption as the consumption pattern adjusts itself, assuming that the commodity has substitutes in consumption and the income effect is not great. Purchases may rise at first, as the rise in price stimulated accumulation: eventually, however,

[3] Hicks, *Value and Capital*, 1st Ed. (Oxford University Press), p. 205.

purchases will fall to the level of consumption as the stock held reaches a position of equilibrium. There may be a period of decumulation (*B*) when purchases fall below consumption following the period of accumulation (*A*). In Fig. 66 the vertical distance between the purchases and the consumption lines indicates the rate of accumulation when purchases exceed consumption, of decumulation when consumption exceeds purchases.

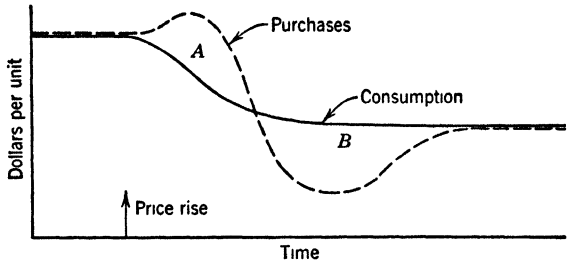


FIG. 66. Effect of price rise on purchases and consumption.

Complementarity and competitiveness of stocks and consumption

The problem of the long-run effect of changes in the price structure on the desired structure of assets, i.e. on the willingness to hold various quantities of the different kinds of assets, presents certain difficulties. Perhaps the simplest assumptions which we can make are first, that the individual wishes to distribute the total value of his assets in certain fixed proportions—say, 10% in the form of money, 20% in the form of clothing, 25% in the form of automobiles, and so on—and, second, that he wishes to hold the total value of all his assets at a constant level. If this were so, the asset-price elasticity would be unity: as the price of an asset decreased the quantity held would increase in the same proportion, so that the total money value of the asset remained constant. If, however, a fall in the price of an asset, as may well be the case, diminished the total value of all assets that the individual wished to hold, the asset-price elasticity would be numerically less than unity: that is to say, a fall in the price of an asset would increase the amount held less than proportionately, so that the total value of the asset declined. If a fall in price has the effect of diminishing the preference for the asset, i.e. if it diminishes the preferred ratio of the value of the asset in question to the total value of assets, the asset-price elasticity will also be diminished. It is more likely, however, that a fall in the price of an asset will increase its preference ratio: in order to obtain the increased consumption which the lower price is likely to encourage, more of the asset must be held, un-

less it is possible to increase its consumption simply by using it more intensively. This effect will increase the numerical value of the asset-price elasticity and may be sufficient to make this value more than unity, in which case a fall in the price of the asset will result in an increased stock of it.[4]

There may be relationships of complementarity and substitution in regard to the holding as well as in regard to the consumption of assets which may complicate the picture still further. These relationships, as we have seen earlier, can be described through the cross relationships of the asset-preference ratios. If an increase in the preference for one asset increases the preference for another, the assets are complementary: in the reverse situation the assets are competitive. Normally we should

[4] If T is the total value of assets held by the individual, and r_1 is the preferred ratio of the value of asset 1 to the total value of all assets, p_1 being the price and q_1 the stock of this asset, we have, by definition,

$$p_1 q_1 = r_1 T \tag{1}$$

Differentiating with respect to p_1, we have

$$p_1 \frac{dq_1}{dp_1} + q_1 = r_1 \frac{dT}{dp_1} + T \frac{dr_1}{dp_1} \tag{2}$$

Dividing by q_1, we have the asset-price elasticity, e_1:

$$e_1 = \frac{p_1}{q_1} \frac{dq_1}{dp_1} = -\left(1 - \frac{r_1}{q_1} \frac{dT}{dp_1} - \frac{T_1}{q_1} \frac{dr_1}{dp_1}\right) \tag{3}$$

It follows that, if T and r_1 are invariant with respect to p_1, $e_1 = -1$. According as the values of $\frac{dT}{dp_1}$ and $\frac{dr_1}{dp_1}$ are positive or negative, so will the numerical value of e_1 be diminished or increased. $\frac{dT}{dp_1}$ is more likely to be positive, i.e. a change in price is likely to change the desired value total of assets in the same direction, for even if we do not change the physical quantities of our assets a rise in price raises and a fall in prices lowers their money value. $\frac{dr_1}{dp_1}$ on the other hand is more likely to be negative, as a rise in price encourages and a fall in price discourages the holding of a particular asset. Whether the asset-price elasticity is greater or less than -1 depends on the relative strength of these opposing factors. The first factor is somewhat analogous to the "income effect" and the second to the "substitution effect" of the Hicksian value theory.

We observe that the asset-price elasticity is almost certain to be negative: for e_1 to be positive we would have to have a very large effect of price on the preferred total asset value, or a large inverse effect on the asset preference. Returning now to Fig. 66, this means that with a price rise there will be decumulation of the asset: the area B, which represents decumulation, is therefore likely to be larger than the area A, which represents accumulation.

expect asset complementarity or competitiveness to correspond fairly closely to consumption complementarity or competitiveness, because of the generally positive relation between asset stocks and asset consumption. However, cases are conceivable in which, for instance, complementarity in consumption arises mainly through differences in the intensity of use of assets, so that we might have two commodities which were complementary in consumption but competitive, or at least less complementary, in regard to stocks held. Thus a fall in the price of eggs may increase the consumption of frying pans by causing us to use them more intensively for frying the increased number of fried eggs, but it is not likely to affect the stock of frying pans. As the average stock of eggs is likely to have increased, the relative stock of frying pans, even if the absolute stock is unchanged, may have diminished as a result of the fall in the price of eggs, indicating a slightly competitive relationship between the *stocks* of eggs and frying pans.

The household as entrepreneur

One further problem of interest in the theory of consumption may be mentioned, if only because of its importance for the theory of the household and for monetary theory. This is the problem of the relation of receipts and expenditures, or their correlatives, sales and purchases, to income and outgo, or to production and consumption. We have seen that there is no necessary equality between receipts and income, or between expenditures and outgo. A consumer who produces a lot of his own consumables at home, e.g. a subsistence farm household, is likely to have a total of income or outgo well in excess of receipts or expenditure, for a substantial part of his income consists of the value of food raised, furniture made, buildings built, cloth spun by himself and his household which never gets into the market economy at all but which nevertheless has an impact on the market through its effect on the demands and supplies of the household. Similarly a part of his outgo consists of the value of the consumption of his own products which accrue to him through his own production and not through purchase and expenditure. On the other hand there are households which have a total of receipts or expenditures well in excess of the total of income or outgo: households, for instance, which produce very little of their own consumables, but sell all their produce for money, and which furthermore do a good deal of trading back and forth in the secondhand market, buying and selling automobiles, furniture, etc.: we should perhaps also include those households which buy and sell securities. The extent of the "want ads" in the newspapers is some indication of how households may expand their receipts-expenditure totals beyond their income-outgo

totals. A difficult question here is when a household ceases to be a household and becomes a business: the "trading" activities of a household have more of the character of a business than the productive or consumptive activities, and it is tempting to try to define a "pure household" as one in which receipts equal income and expenditures equal consumption. Such a definition, however, would be too restrictive to be particularly useful: it is more important to recognize the various activities of households rather than to try to set up a narrow "ideal type." The excess of receipts-expenditures over income-outgo is a good measure of what may be called the "entrepreneurial" character of the household, with the self-subsistent farmstead and Robinson Crusoe at one end of the scale and the highly specialized, urbanized, market-conscious modern household at the other.

The entrepreneurial character of a household seems to be determined by two broad principles. The first is that of comparative advantage, which determines the extent to which specialized production for exchange in the market is worth while. No household, of course, carries specialization to an extreme. Most men shave themselves, dress themselves, and do innumerable small productive chores around the home. Domestic production by wives and mothers is even more significant and constitutes a substantial part of the national product. In general it is clear that as between the two possible methods of acquiring assets—direct production on the one hand, and indirect production through producing some specialized product and exchanging it for the desired asset on the other—we will choose that which is "cheaper" in terms of disutility: this is simply the principle which I have elsewhere called the "principle of the best alternative." [5]

Adam Smith's pregnant dictum, that the degree of specialization (division of labor) depends on the extent of the market, clearly applies to the individual household. Robinson Crusoe, because he has no market available whatever, is forced to be completely self-sufficient: he has income and outgo, or at least production and consumption, but no receipts or expenditure. The market opportunities open to the primitive peasant household are meager; hence there is a great deal of production for internal consumption, and the proportion of receipts-expenditure to income-outgo is likely to be low. In the modern urban household, where the market opportunities are extensive, the domestic production may be mainly done by hired servants while the wife as well as the husband sells her labor or other income sources elsewhere for money receipts, and may even do a good deal of buying and selling of household goods and other assets.

[5] Boulding, *Economic Analysis*, Rev. Ed., pp. 28 and 45.

The "extent of the market" is not the only principle which must be invoked to explain the "entrepreneurial" character of a household. Two households may be faced with the same market situation, and yet one may be highly specialized and entrepreneurial, while the other is less specialized with a much smaller proportion of receipts-expenditure to income-outgo. This difference arises from the entrepreneurial character of the household itself or of the individuals who direct it. Some people like trading: others do not. The "supply of enterprise" or of "trading" in an economy is a real factor in the system, although extremely difficult to quantify satisfactorily. Again we find, of course, that the entrepreneurial householder tends to become a specialized marketer or middleman. We see this process going on in agriculture, where, for instance, a farmer who is good at buying and selling and enjoys making a good bargain tends to develop the marketing side of his business until eventually he may give up his farm altogether and become a specialized dealer. The farmer who does not have this "trading" capacity or preference is likely to specialize all the more in pure production and to pass on the trading function to dealers or to cooperative marketing associations.

The extent to which the total volume of trading in a society exceeds the volume of production and consumption depends in large degree on the extent to which the members of the society have the "trading itch." In general, of course, the stronger the desire to trade—or, to use another famous phrase of Adam Smith, the "propensity to truck"—the more trading there will be and the less the cost to society of a "unit" of trading. It is when there is a general unwillingness to perform the trading function that dealers' margins and commissions are high. On the other hand there seems to be an optimum propensity to truck from the point of view of social welfare. At times the propensity to truck becomes a speculative fever, in which the normal activities of production and exchange are neglected in the excitement of rapid turnover on a rising market. The Dutch bulb speculation, the South Sea bubble, the bull market in the New York Stock Exchange in 1927–1929, and the Florida land boom of the nineteen twenties are but a few of the examples of these excesses of trading fever. At such times it is difficult to picture the household as the somewhat passive consumer seen in traditional economics!

What is happening in economics in this respect is that it is becoming increasingly difficult to distinguish clearly between a theory of the firm and of production on the one hand and a theory of the household and of consumption on the other. The same concepts, the same problems concern both, in various degrees, and what we are moving towards is a gen-

eral theory of an economic organism and of the interactions of these organisms. It must be confessed, however, that the theory of consumption, whether in the household or the firm, is probably the least satisfactory part of theory at present, and it may well be that more case study is necessary before further fruitful generalizations—e.g. about saving and consumption habits—can be formed.[6]

[6] The studies of the Survey Research Center of the University of Michigan are an important first step in this direction: see, for instance, the *1949 Survey of Consumer Finances* (Federal Reserve Bulletin, 1949 and 1950).

Chapter 9

The Equilibrium of Production and Consumption

Our next task is to bring together the processes of production and consumption and to consider their interaction, as we have previously considered the interaction of marketers. The general lines of this theory are familiar to economists, and we shall do no more than outline.

The theory of normal price

The equilibrium of production and consumption is usually considered to be brought about through the pricing system. For any commodity or asset type, both production and consumption are regarded as functions of the price of the commodity, the production-price function being the normal supply curve, usually regarded as positive in slope, and the consumption-price function being the normal demand curve, which usually has a negative slope. The normal price is that which creates an equilibrium of production and consumption. If we assume that there is no change in the stocks of the commodity then the normal price is that which *equates* production and consumption. There is no need to make so restrictive an assumption, however. If there is a certain rate of growth of stocks which is regarded as "normal," then the normal price can be defined as that at which production is equal to consumption plus the normal growth of stocks.

The usual theory assumes, furthermore, that this equilibrium is stable if a price above normal produces an excess production, i.e. an excess of production above consumption greater than that which will yield a normal change in stocks, so that stocks of the commodity are increasing faster than usual. Similarly if the price is below normal, consumption will be in excess, so that stocks will increase less rapidly than usual or will decline. The assumption here is that excess production brings forces into play which will lower the price. The force usually assumed is the increase of stocks, which increase, as we have seen earlier, in the absence of any changes in stocks of money or of asset preferences, will lower the

price at which people are willing to hold the stock in existence.[1] Similarly excess consumption is assumed to bring forces into play which will raise the price, mainly through its effect in bringing about a decline in stocks.

Output reactions versus price reactions

It is important at this point to re-examine these conditions, particularly with a view to their generality, for although there is no doubt that the price system does operate in the way outlined above for many commodities, this type of equilibrium may not be universal. The assumption of the above analysis is that the price-production and price-consumption functions are stable, and that the main impact of excess production or consumption is upon price. There are circumstances, however, in which this may not be the case. A situation in which there is excess production of a commodity, and hence unwanted or unplanned accumulation of inventory, may result not in an adjustment of price but in a direct adjustment of production or, less likely, of consumption.

Consider, for instance, the situation of a manufacturer of some rather bulky household equipment, such as washing machines or stoves. His continued production depends on his ability to sell the product, i.e. to transfer his assets back from final product to money, at a rate approximately equal to his rate of output. If there is a sudden decline in demand for his product he will find that he has unplanned inventories on his hands: his asset structure shifts, his money stocks declining as he continues to pay out production costs, and his stocks of finished goods increasing. In a relatively short time—perhaps only a matter of days—this situation will create a severe crisis, demanding action of some kind. The crisis may impinge on the firm in several different ways. Storage facilities for the finished product, especially if it is bulky, may be used up and other facilities may not easily be available. The firm finds itself with washing machines in the corridors and in the offices, and begins to suffer from acute industrial constipation! Even if storage facilities for finished product are easily available, the firm may run into difficulties because of the diminution in its money stocks which the accumulation of finished goods involves. Workers cannot be paid in electric stoves, and if the conversational clichés of business men are any indication of the economic neuroses to which they are subject, "meeting a payroll" is evidently an ever-potential crisis. If the firm has ample credit facilities it may, of course, meet the situation by borrowing to finance the inventory accumulation without running out of cash. A situation where

[1] From the market equation, $p_a = Mr_a/Ar_m$, we see that an increase in the stock A, other factors constant, will lower p.

unwanted inventory is accumulating, however, is not one in which banks or other lending agencies are likely to feel particularly tender towards the firm's borrowing. Much depends, of course, on whether in the judgment of the business community the crisis is temporary or permanent. Banks are generally entirely willing to extend credit to a firm in a temporarily illiquid position. If, however, the inventory accumulation is taken to be a sign of a general downturn in business, lenders will be as unwilling to lend as buyers seem to be unwilling to buy the product, and the crisis of the firm may not be resolvable by borrowing.

Under these circumstances the only way in which the firm suffering from a sudden loss of demand can correct its dangerously illiquid position is to cut back production, fire some of its workers, and stop buying raw materials until the gradual sale of the unwanted stockpile of finished goods relieves the pressure on warehouse space and replenishes the depleted money stock. Here we clearly have a situation in which excess production and the consequent abnormal accumulation of stocks of goods result not in a lowering of the price of goods in order to encourage their sale, but in a cutback in their production: the production adjusts itself to the price rather than the price to the production. This is likely to be the case with highly durable commodities produced under conditions of considerable imperfection in the market, so that pricing becomes a matter of long-run policy to the firm and cannot be used as a mechanism for adjusting production to consumption in the short run. If the firm finds that there is a permanent decline in demand, i.e. in the rate of purchase at the existing price, it may experiment with lower prices, but this is likely to take place only after a substantial period of undercapacity production of the commodity in question.

At the other extreme we have highly perishable commodities produced under competitive conditions with approximately perfect markets, or at least with markets which show an appreciable elasticity of demand for the sales of the individual firm. Fresh fruit and vegetables are good examples. Here the traditional price mechanism operates even in short periods: if the price is such as to create excess production on to any particular market, with consequent piling up of unwanted stocks, the stocks exert a very powerful influence on the price, which falls until there is a sufficient temporary excess consumption to absorb the stocks—unless, of course, the surplus is extreme, in which case the sellers may prefer to throw stocks away rather than try to sell them.

Excess consumption, leading to an abnormal decline in inventory, may likewise provoke two quite different kinds of reaction. The price of a commodity may be so far below its short-run equilibrium that acute shortages become apparent, and yet its price may show little tendency

to rise, either because of sheer inertia, or because the sellers regard price policy as a long-run matter not to be disturbed for temporary changes in demand. Certain automobiles after World War II provide an excellent example of this phenomenon. The inventory pipe line was empty, the magnificent salesrooms stood unadorned by the elegant new models, there were long waiting lists for new cars, and it was almost impossible to obtain one with no more "pull" than the purchase money. Nevertheless in the face of this excess demand the price of new automobiles, which, if the second-hand market revealed anything, could have been hundreds of dollars greater, rose only very slowly, presumably because the established manufacturers did not wish to risk a loss of consumer goodwill by seeming to take advantage of a tight situation. Hence instead of the excess demand creating an upward pressure on price sufficient to restore a free equilibrium in the market, consumer purchases were adjusted to the output of cars by informal rationing on the part of dealers. In the case of our perishable fruits again, however, a "shortage," i.e. a price below normal at which stocks move very rapidly off the market, produces a situation in which there is a strong upward pressure on the price until the high price itself performs the rationing function, chokes off purchases, and adjusts purchases to the short supply.

The reaction to a situation of excess demand (consumption) or excess supply (production) may also, of course, lie somewhere between these two extremes. An unwanted accumulation of stocks may produce *both* a cutback in production and a fall in price: an abnormal decumulation of stocks may produce both a cutback in consumption through rationing, formal or informal, and a rise in price. Generally speaking, the more flexible are prices, the less will abnormal accumulations or decumulations produce production cutbacks or shortages. Price declines and output declines represent alternative ways of dealing with the crisis of abnormal accumulations: price advances and consumer-purchase rationing represent alternative ways of dealing with the crisis of abnormal decumulations.

The equilibrium of wages

The special problem of equilibrium wages is worthy of consideration in connection with the above principles. Labor cannot be stored at all, except in the very slight measure that a rest today may enable a man to work harder tomorrow. Hence the divergence of a wage—say, of carpenters—from a "normal" level cannot be reflected at all in accumulations or decumulations of labor. Nevertheless the concept of an equilibrium wage has meaning, even if the significance of divergence from the equilibrium is complex. There is some wage above which the labor

market exhibits one group of phenomena which can be described collectively as a "labor surplus" and below which the labor market exhibits another group of phenomena of a converse kind to the first which can be described as a "labor shortage." The first is characterized by perfectly elastic supply curves of labor to the employer—he puts out a "help wanted" notice and gets applicants immediately: the second is characterized by less than perfectly elastic supply curves of labor to the employer—a mere "help wanted" sign produces no results; the employer must either incur a "buying cost" in order to attract labor, or must raise wages in some form if he needs more help. It is clear that a labor shortage is much more likely to produce a "crisis" situation for an employer than a labor surplus: consequently there is likely to be a considerable degree of asymmetry between the consequences of labor shortages and surpluses. A labor shortage almost always produces a marked upward pressure on wages. Firms find themselves with diminishing inventories, unfilled orders, clamoring customers, and feel an enormous pressure to increase outputs and hire more labor; but in a labor-shortage market, once the slim resources of overtime are exhausted, there is hardly any way to do this except by attracting workers away from other employers by offering higher wages or more attractive working conditions. In a labor-surplus market on the other hand there is not a great deal of pressure to lower wages. Lowering wages is in itself a sociologically disagreeable thing to have to do and creates personal tensions between employer and employed which most employers dislike. The crisis which the labor-surplus situation engenders in the firm is more likely to be solved by cutbacks and layoffs than by wage reduction: similarly the crisis in which a labor-surplus situation places an unemployed worker is more likely to be met by withdrawals from savings going into other occupations, dependence on relatives, and going on public relief than by offering to take the jobs of the already employed by offering to work for less. This is not to say, of course, that a labor surplus creates no pressure on wages, but the pressure is much less in a surplus than in a shortage situation.

General equilibrium

The implications of the above analysis for the theory of general equilibrium are of great importance and bring us to the verge of the domain of macroeconomics—the study of income, employment, and other aggregates of the system. The simplest statement of the theory of the general equilibrium is to postulate that the excess demand or supply of each commodity is a function of all prices, and to impose the further condition on this system that the excess demand or supply of each com-

modity (including, of course, labor, securities, etc.,—anything which is priceable) should be zero. We then have a system of n equations and n unknowns, and if this system has a unique real solution, the problem is theoretically solved. In practice the equilibrium is supposed to be approached by a series of *tâtonnements*: whenever a commodity appears in an excess demand the price of that commodity rises; whenever it appears in excess supply the price falls. These price changes will, of course, create excess demands or supplies elsewhere in the system, depending on the nature of the functions relating excess demand-supply and price. Generally speaking, a rise in a price will increase excess demand for commodities complementary with the commodity that has risen, and will increase excess supply for commodities competitive with it. These excess demands and supplies will in turn create price changes, which will echo and re-echo all through the system of related commodities until the turbulence subsides in the stillness of equilibrium—unless, of course, as is likely, some new disturbance sets off a new train of consequences.

Even accepting the underlying assumption that excess demands and supplies actually produce an effect on prices, we should, of course, distinguish between the general equilibrium of *exchange*, which consists of that set of prices at which people in general are willing to hold the existing stocks of all assets, and the general equilibrium of production, which is that set of prices at which the stocks of all commodities grow or decline by "normal" amounts. This distinction, however, presents no real difficulty: it corresponds roughly to the distinction between short- and long-run equilibrium. What is much more damaging to the usefulness of the general equilibrium concept is the assumption which we have questioned above, that excess demands or supplies do in fact affect prices. If they do not in certain cases, then the general equilibrium concept may be actually misleading; it does not guide us to the dynamic changes in the system, but by emphasizing *weak* causal relationships may actually divert attention from the real problems. The failure of the Marshallian economics to deal adequately with the problem of unemployment may well stem from this weakness: there is no sense in postulating an equilibrium concept unless a divergence from equilibrium brings appreciable forces into play to restore the equilibrium.

A non-price equilibrium

As an extreme exercise in a non-price equilibrium system it is instructive to discuss the adjustments of the economic system to a completely arbitrary system of prices—say, a system of fixed prices imposed by law. Such a system is one in which adjustments to changes in tastes and techniques can clearly be made, yet by definition these adjustments cannot

operate through the price system and must operate through other variables. In such a system there will, of course, be excess demands and supplies, unwanted accumulations and decumulations of assets. Excess demands will simply result in rationing of some kind. If no formal rationing is imposed by authority, there will be informal rationing in the form of inconvenience, time and trouble costs of acquiring the commodity, dealer rationing to favored customers, and the like. Part of these informal rationing methods consist of arbitrary favors to fortunate customers: part operate something like a "trouble price," in that those who are willing to pay the highest cost in terms of time and trouble taken get the commodity. Whatever the method, however, the result is a shift in demand curves until at the fixed legal price demand is just sufficient to take care of supply, consumption just equals production, and there is no abnormal decumulation of stock beyond the level (which may be lower than under free pricing) necessary to impose a sufficient degree of rationing. Excess demand may also have an effect in increasing the production of a commodity. With commodities in excess demand there will be no selling problem: the sales curve to the individual producer will be perfectly elastic, and he will be able to sell as much as he likes at the fixed price. With such an unlimited market the energies of producers may well go into cost reduction rather than into sales promotion, so that there may be a faster rate of technical progress in those commodities for which the fixed price yields an excess demand; this progress, by pushing the supply curve to the right, will diminish the excess demand and may eventually do away with it.

There will be other commodities which are in excess supply at the fixed price. The initial impact will be the development of surpluses of these commodities: these surpluses cannot, however, produce any impact on the price. The result will be a cutback in production if producers have to hold these surpluses, as would be the case beyond that level of stocks which specialized marketers are willing to hold at the fixed price. This cutback will be found even in processes where it does not normally occur, as in agriculture. The wheat farmer who finds himself with a year's crop of wheat in his barn at sowing time is likely to take a holiday! Production, therefore, will be cut back to the point where it can just take care of the consumption at the fixed price plus any normal accumulations.

There may also, however, be greater emphasis on sales promotion. Because of the excess supply producers are likely to find themselves with highly imperfect markets, even though they cannot vary the price: that is to say, they will be able to sell less than they want to produce, will be unable to meet this situation by cutting prices, and hence may find that

sales promotion is a fruitful way of disposing of their surpluses. This sales promotion will, of course, increase purchases at the fixed price, i.e. increase demands, and will therefore lessen the excess supply. Moreover, the fact that the attention of producers is diverted to sales promotion may mean that in these commodities less attention will be paid to technical progress, which will lag behind, so that the costs of these commodities rise relative to those which were in excess demand. This will also have the result of diminishing the excess supply.

It is evident, therefore, that even in a system of absolutely rigid prices forces will come into play on the side of costs, technical change, selling and buying promotion, and rationing which will tend to do away with the excess demands and supplies and which will eventually so alter the fundamental determinants of the equilibrium price system that the arbitrary fixed price system itself *becomes* the equilibrium system! It is clear that the equilibrium system is not, dynamically speaking, determinate: the course of the equilibrium itself depends on the way in which excess demands and supplies exert their influence. It is true that the system abhors excess demands and supplies, and that these are a potent symptom of disequilibrium. It is not true that the movement towards equilibrium must be through the relative price structure. As we have seen from our "strong case," the movement to equilibrium can take place without any change in the relative price structure. Normally, of course, this is not so, as almost all prices are somewhat responsive to excess demands and supplies, and many prices are extremely responsive to them. The dynamic course of the system, however—i.e. whether the adjustments to excess demands and supplies take the form of price adjustments, or whether they take the form of the various non-price adjustments mentioned above—depends entirely on the relative adjustability of the various parameters. Where prices are adjustable and the other factors non-adjustable, naturally the adjustment will take place through the price system: where it is the other factors which are more easily adjustable, these factors, rather than price, will adjust.

The case for non-price adjustments

The implications of these considerations for welfare and for economic policy are important, and though we cannot deal with them fully at this point in the argument, some mention should be given. There has been a fairly general presumption among economists that price adjustments are desirable and non-price adjustments are undesirable. In large part this proposition seems to have arisen from the structure of economic thought itself: from the fact that economics has concentrated on price adjustments almost to the exclusion of non-price adjustments. In fact

there seems to be very little justification for such a conclusion. Indeed, in so far as non-price adjustments consist in technical progress there seems to be every reason to prefer them to price adjustments. Technical progress, in part at least, is a response to a "crisis" situation; i.e. it may result from the attempt to find a solution to some problem imposed by the price system. Not all technical progress is of this character, of course: some arises from pure inventiveness, some from spontaneous discovery, and so on; but there seems to be an ingrained conservatism in human nature sufficient at least to give some plausibility to the proposition that nobody changes his established ways of doing things unless a crisis of some sort forces him to take action. In so far, therefore, as technical progress is a substitute for price adjustments there is everything to be said for preventing price adjustments and so diverting the adjustment process into more socially useful channels.

On the other hand, of course, there are many non-price adjustments which are socially harmful. Sales promotion beyond a certain point—a point probably well below the present level of such activity—is not only economically wasteful of resources, but has undesirable sociological and psychological by-products in the shape of the disintegration of character, the development of the sly arts of ingratiation and deception, and the vulgarization of periodicals and the radio. The most harmful non-price adjustment, however, is unquestionably unemployment, i.e. production at less than the capacity of the system. We shall take up this problem in much greater detail in the second part of the book: nevertheless it will be worth while now to glance at certain possible mechanisms of micro-economic adjustment which may result in general underemployment of resources.

Adjustments to change: "Ricardian" unemployment

It will be useful first to describe the "ideal" process of adjustment in a general equilibrium system with highly flexible prices. Let us suppose, for instance, that such a system is in equilibrium, the production of each commodity being just equal to its consumption plus normal accumulation at the prevailing price structure. Suppose, now, that this equilibrium is disturbed, say, by a shift in tastes: suppose, for instance, that a religious revival turns large numbers of people into total abstainers from liquor and increases the support of missions. The first impact of this change is, of course, that money which previously flowed to the distillers in exchange for liquor now flows to the mission boards. The distillers find their cash balances declining, and their stocks of finished goods piling up. Mission boards, on the other hand, find their cash balances increasing as the increased flow of contributions more than

outweighs their current expenditures. The persistence of such a situation will force action of some kind. If prices are flexible, liquor prices come down, thus, alas, encouraging the consumption of liquor among the unconverted. At these low prices marginal distillers will go out of business and production of liquor will decline, until the diminished production meets the expanded consumption, equilibrium is restored, and the drain on the cash balances of distillers ceases. Mission boards, on the other hand, finding themselves so abundantly blessed with resources, will hire more missionaries, more staff members, and will branch out into larger enterprises until the expansion matches their increased receipts and their cash balances also cease to expand. Equilibrium is now restored; there will, of course, have been innumerable other readjustments. We do not have to assume that the displaced distillery workers become missionaries. There will be a long chain of adjustments throughout the labor market: missionaries will be drawn from carpenters, teachers, etc.; others will step into the gaps left by these departures for the mission field, still others will step into the gaps created by these second movements, and so on all through the system until the distillery workers also become soft-drink workers, teamsters, farmers, etc. In all this process there is nothing to suggest the development of *general* unemployment: the picture is one of temporary labor surpluses in distilling and temporary labor shortages in other fields, with a subsequent rearrangement of resources between the surplus and the shortage areas.

This process of resource transfer is not necessarily dependent on price and wage adjustment for its completion. Even in a system with completely rigid prices and wages a transfer of resources could take place. The expanding receipts of the mission boards would enable them to employ more missionaries, probably without raising their salaries. The inability of the distilleries to sell their present output of liquor would force them to shut down and to discharge some of their employees even if there were no price or wage adjustments. Similarly all the various intermediate adjustments could perfectly well take place by the mere transfer of resources, without any change in the price-wage structure. Indeed this is precisely the model of the Ricardian economics, if we except food production. Ricardo assumed what is not in most cases very far from the truth—that industrial production takes place under conditions of constant cost. Under these circumstances there is no reason to suppose that a shift in tastes or in demand will affect the price structure: it will merely effect a redistribution of resources.

How, then, can general unemployment arise? It can clearly arise in the first place because of differences in the tempo of adjustment in various occupations. A sudden shift in demand from liquor to missions

may occasion a very rapid decline in employment in distilling, for the distillers will be faced with the immediate crisis of unsold stocks and depleted cash balances. The expansion in employment in the mission field, however, may be slow, and for a long time the main result of the change may be the accumulation of cash balances by the mission boards. It takes time to find and to train missionaries, to develop new fields, and so on: hence there may be a long period during which total employment is below normal because of the different pace of the adjustment to shortage and to surplus situations. This too is perfectly good Ricardian economics: he recognized clearly the effects of "sudden changes in the channels of trade." [2]

"Malthusian" unemployment

The type of unemployment mentioned above is, however, temporary, no matter how long it may endure, in the sense that it arises out of the time lags of the system. There may on the other hand be unemployment of a more fundamental and permanent kind: this might be described as "Malthusian," as opposed to "Ricardian," unemployment, Malthus having given the first systematic description of the phenomenon in his *Principles*.[3] This is the unemployment which results from the general unwillingness or inability of the system to consume and to accumulate as much as it can produce. We shall investigate this phenomenon in much greater detail in Part 2. We can see here, however, how a chain of microeconomic reactions might create general unemployment through underconsumption and underinvestment. We have not yet considered the reaction of the diminution of employment and output in the distilling industry on the consumption of the people affected. It is evident that the crisis in the distilling industry will not stop with the laying off of men and the shutting down of plants. The men laid off, as well as the employers, will find their economic situation suddenly worsened by a sharp decline in income and in receipts. They too will find themselves both dishoarding and dissaving: spending more than they receive and consuming more than their income. The natural reaction to this crisis is a curtailment of both consumption and expenditure in an attempt to diminish this dishoarding and dissaving; it may be easier to deal with a dishoarding-dissaving crisis by cutting down expenditure and consumption than by trying to expand receipts or income. The curtailment of consumption and expenditure means, however, that receipts and sales of other commodities are diminished. The out-of-work distillery workers do not buy the clothes, the furniture,

[2] Ricardo, *Principles of Political Economy* (Everyman Library), Chapter 19.
[3] Malthus, *Principles of Political Economy*, Book II.

the radios which otherwise they would have bought: consequently the manufacturers of these things, who have perhaps produced them in expectation of sales to the distillery workers, also find themselves with unsold stocks and depleted cash balances; their reaction to this situation may also be to lay off men and curtail output, which will again have consequences on other parts of the system. So we get a chain reaction, unemployment leading to a decline in consumption, which leads to further distortions of firms' asset structures and to further unemployment. This reaction may easily push the system down to an underemployment equilibrium from which the later development of the expanding occupations may not be sufficient to rescue it. Wherever the immediate reaction to dislocation is a cutback of output, a rapidly changing system is always likely to be slipping into underemployment.

The ecological framework

In order to complete this outline of microeconomics we should return to the concept of an ecological system outlined in the first chapter—from which we seem to have departed a long way! Can we fit the theory of the firm, of the consumer, and of the general equilibrium of the pricing system into an ecological framework? The economic system seems to resolve itself into certain highly complementary groups of populations which we can roughly designate as "industries," and within each industry we can postulate some fairly simple "population relationships" among various kinds of assets, firms, and individuals. There is an important difference, however, between natural and social ecological systems, in that in natural ecology the population functions—i.e. the functions relating the equilibrium size of one population to the sizes of others —are fairly stable and are quite outside the wills or control of the participating individuals. In social ecology, however, we have the factor of "demand" determining the relationship between, for instance, the population of automobiles and that of households. It is important, however, to realize that the population functions of Chapter 1 are themselves the result of complex patterns of functional relationship and may be subject to change. In natural ecological systems change comes about usually as a result of certain "exogenous" events—the incursion of man who cuts down the forest, changes in climate, volcanic eruptions, or the advent of new species through the mutation of genes—rather than through changes in the population functions themselves. In social ecology the presence of the factor of conscious choice gives to the population functions an instability which they do not seem to possess in the field and forest.

In every ecological system there is something akin to the circulation of assets. In natural ecologies we have, for instance, chemical cycles

such as the nitrogen cycle or the oxygen cycle, the waste products of some organisms forming the nutriment for others. The chemical composition of a tree is something like a balance sheet: its respiration and nutrition processes are something like exchange; the circulation of the essential chemical elements by these exchange processes through the various organisms is a process somewhat akin to the circulation of money. We must beware, however, of analogy: it is one thing to perceive similarity in the various systems of the universe, such is the greatest task of the philosopher. It is quite another thing to argue by analogy from one system to another. The student of each system must investigate its own laws, and argument by analogy is a sure indication of a weak science. In so far as all definition involves the concept of a population of the thing defined, all empirical, quantitative sciences are based on an ecological system of some sort. The superstructure of every subject, however, it must erect itself.

Part 2

Macroeconomics

Chapter 10

The Nature and Significance of Economic Aggregates

Justification of macroeconomics

Macroeconomics is the study of the nature, relationships, and behavior of aggregates and averages of economic quantities. Microeconomics studies the nature and behavior of individual economic quantities and organisms: individual prices, outputs, firms, households, the wage of union carpenters in Seattle, the price of wheat on the Chicago grain exchange, the output of automobiles from a given plant, and so on. Since the aggregates and averages which are studied in macroeconomics are merely aggregates and averages of these individual economic quantities, the question may well be raised, Why bother with macroeconomics at all? If we knew the behavior of all the microeconomic variables and their relationships one with another we could discover all the macroeconomic relationships by the simple process of summation or averaging. All that we need to know about the economic system is already implicit in microeconomics: moreover, the particular prices, wages, outputs, etc., which are studied there are the true realities of economic life; aggregates and averages at best are merely mathematical constructions and are not "real" in the sense in which their individual components are. Why, then, burden ourselves with dangerous and inevitably inaccurate generalizations in regard to aggregates, when we could be content with both exactness and completeness at the microeconomic level?

The question is worth the asking, and it has several answers. We may first of all defend our *interest* in aggregates and averages—such concepts as the general level of prices or wages, the national income—from a severely practical point of view. Any discussion of economic policy must be couched primarily in terms of economic aggregates and averages. There is good reason for this. By economic policy we mean the economic impact of government on the economic system. Government, however, must operate in the mass. The very concept of law implies some condition which applies to all individuals of a certain class, i.e.

conforming to a certain definition. It is not merely that government does not have the technical ability to regulate all individual prices, wages, rents, firms, households, etc., separately. In any society which has a responsible government—a government of laws and not of men, to use the familiar cliché—it is basic to the very concept of government that it should operate not on individuals as such but on individuals in their capacity as members of a certain defined group. This is theoretically true even if the individual is the sole member of the group: when, for instance, a President is impeached he is impeached not as an individual but as a member of the group known as Presidents—which, however, has only one member! It is only arbitrary and tyrannical government that acts on individuals as individuals, and the whole development of law and constitutionality has been in the direction of preventing this type of governmental action.[1]

As governmental policy and behavior, then, must operate on large groups and classes of economic variables, in order to throw light on the problems of government economic policy it is necessary to have a theory of the behavior of economic aggregates.

It is not necessary, of course, to justify a theoretical interest in aggregates by a reference to its practical utility. Any science must be interested in attempting by various devices to picture its universe as a whole. Description of a large and complex universe of facts like the economic system is impossible in terms of individual items: we must generalize and aggregate if we are to get any idea of what the system is like and how it functions as a whole. As a matter of fact, even microeconomics is not disinterested in aggregates; in so far as it attempts to generalize it tries to construct models of particular economic behavior which are characteristic of large groups of similar items. The theory of the firm is not a list of the behavior patterns of General Motors and the Podunk Ice Company: it is an attempt to construct a model which will be of value in interpreting the behavior of large groups of firms. Even the most primitive form of science—taxonomy, or the classification of a universe of objects—involves some aggregation. The botanist is interested in his flower not as an individual but as a representative of a group of similar plants.

The main justification for macroeconomics, however, is that the character and behavior of aggregates or populations of things cannot be

[1] It must not be thought, of course, that the collective or aggregative character of governmental action is a *sufficient* justification for it, even though it is necessary. It is quite possible to have government operating on individuals as members of groups, e.g. in the institution of slavery or of genocide, in a manner which is repulsive to our sense of justice.

obtained simply by generalizing from the character and behavior of the individual components. The forest as an object of study, even though it is merely an aggregation of trees, does not exhibit the characteristics and behavior of individual trees. A tree germinates, grows, dies, and falls. A forest in equilibrium may go on for ever with exactly the same composition in regard to the age and character of the trees that compose it: as individual trees grow, mature, and die new individuals spring up to replace them and the forest remains the same. It is almost impossible to burn an isolated tree: forests are subject to fires. An isolated tree will not affect the climate in which it grows: a forest will. Many more examples from the natural world could be given, but perhaps the most striking examples of the differences between group and individual characteristics come from social, and especially from economic, life. In hardly any other science are the dangers of generalizing from individual experience so great as they are in economics, and perhaps the greatest source of error in economic thought has arisen because of the difficulty which the mind seems to have in conceiving of a closed system as a whole and because of the consequent tendency to apply propositions derived from individual experience to society or to the economy as a whole.

What may be called the "macroeconomic paradoxes"—the propositions which are true when applied to a single individual but which are untrue when applied to the economic system as a whole—are by now familiar to most economists, and we need only outline them here. Perhaps the most important is the proposition that although an individual can accumulate cash balances (money stock) by the process of "hoarding" (allowing his receipts to exceed his expenditures) a society can increase the aggregate of its money stocks only by the creation of money, and cannot increase its money stocks by the process of individual hoarding. A moment's serious reflection is usually sufficient to convince a student of the truth of this proposition, yet it is astonishing how universal is the failure to recognize it among the present generation of men of affairs, and even among the last generation of economists. In a society where the total quantity of money is constant individuals can increase their personal money stocks only at the expense of the money stocks of others. If an individual "hoards" by spending less than he receives, it follows—if the aggregate stock of money held by all individuals is constant—that the hoarding of one individual forces others, *in toto*, to spend more than they receive and hence to dishoard. The speeches of bank presidents during World War II provided many good examples of failure to understand this very elementary proposition. These speeches frequently fell into two parts. In the first part the speaker pointed with pride to the wonderful record of "saving" on the part of

the public during the war, meaning by saving, of course, the increase in bank deposits, i.e. in the money stocks of the public. In the second part the speaker viewed with alarm the immense deficit in the government's budget and suggested remedies for reducing it, without, apparently, being aware of any connection between the budget deficit which he deplored and the public's "saving" which he lauded. In fact, of course, if there is a cash deficit in the accounts of government, i.e. if government is receiving less money than it pays out, the public cannot help finding itself with a cash surplus, i.e. it will find itself receiving more than it pays out; and if there is no other creation of money in the system the cash deficit of government will exactly equal the increase in the money stock of the public, no matter what decisions individuals make about "hoarding." That is to say, the aggregate hoarding of society is not a result of decisions to hoard, but of decisions to create or destroy money. This process will be examined in greater detail later and need not be further explained here.

Another important macroeconomic paradox is that of saving and investment. This again is a problem to which we will return, and need but mention here. An individual "saves" by the process of consuming less than he produces, thus increasing his net worth. The aggregate of saving, however, is not determined by the sum of decisions to save, but by the sum of decisions to accumulate assets, i.e. to invest. Hence individual decisions to save may result not in saving, i.e. not in an increase in net worths, but in a decline in output and income.

There is yet a third paradox, that of profits and wages, which is not so familiar as the other two and yet is perhaps equally important. We shall show later that the distribution of the national income between profits and wages, or rather between labor income and non-labor income, is not determined directly by the wage bargain or by the productive efficiency of management, but by a combination of other factors, the most important of which are decisions of management to invest, i.e. accumulate real assets, and the complex of decisions of the whole society about liquidity preferences.

It is these paradoxes, more than any other factor, which justify the separate study of the system as a whole, not merely as an inventory or list of particular items, but as a complex of aggregates.

Dangers of aggregative thinking

Although there is ample justification for the study of economic aggregates, it must not be thought that aggregative thinking is without dangers. It is, in fact, peculiarly subject to error, and it may be well at this point therefore to indicate some of the pitfalls. The greatest danger

we have already mentioned: that of generalizing to the system as a whole from particular individual experience, and hence failing to understand the macroeconomic paradoxes. Another danger, perhaps equally great, is that of thinking in terms of aggregates which are too heterogeneous to be workable. It must always be emphasized that aggregates are not reality, but a picture or approximation of reality. The real things, in economics as in any other subject, are the individual components of the aggregates. If our aggregate (or average) is composed of very heterogeneous individuals it will not be meaningful. The significance of any aggregate depends on the homogeneity of the individuals that compose it. Adding six apples to seven apples makes a quite meaningful aggregate of thirteen apples. Adding six apples to seven oranges still makes a fairly meaningful aggregate of thirteen fruit—i.e. apples and oranges have enough in common to give their aggregation some meaning. Adding six apples to seven thumbtacks makes thirteen small objects: an aggregate which may again have some sort of meaning, e.g. for an artist, but which certainly is losing meaning rapidly. Six apples added to seven skyscrapers makes an aggregate which could hardly have meaning to anybody except perhaps the surrealists.

The significance of averages depends in addition on the frequency distribution of the things averaged. An average is a ratio of aggregates: the average weight of Americans, for instance, would be found by aggregating the total weight of all Americans and dividing by their total number. All the considerations regarding homogeneity, therefore, which we have applied to aggregates apply likewise to averages. An average, however, to be interesting must be derived from a frequency distribution which is at least unimodal and which does not have too great a dispersion. Averages are interesting only as they are *representative* of the individuals averaged. The average weight of Americans is a fairly interesting figure because we may expect a great many individuals to weigh around the average, and have a pretty fair idea of the chance of finding individuals who weigh considerably above or below the average. It is interesting because in a sense the average approximates to our idea of a "norm," though we must, of course, beware of applying ethical standards of normality too readily to statistical averages. In the case of weight, however, we are reasonably justified in supposing that an individual who weighs about average for his age and height weighs "about right," and that individuals who are much heavier or lighter than the average are, in a normative sense, "overweight" or "underweight."

If, however, we took the average weight of Americans and of their automobiles the result—say, about 1000 lb—would not be an interesting figure in the least, not so much because the aggregates from which the

average is derived are meaningless as because the average does not represent any individuals whatever. There is nothing wrong with aggregating weight: we can add the 180 lb of a man to the 3000 lb of an automobile and get a meaningful aggregate of 3180 lb, an aggregate which may, for instance, be quite meaningful to the operator of an overloaded ferry! The aggregation of numbers of automobiles to numbers of people does not make very much sense, for automobiles and people are not usually sufficiently homogeneous in character to be an interesting aggregate: nevertheless as an aggregate it is not complete nonsense. What makes nonsense out of the average is that the frequency distribution of the population which is being averaged is bimodal, with one mode at the modal weight of Americans and the other mode at the modal weight of their automobiles. Hence the average—no matter what kind of mean it is—will lie between the two modes and will not represent the weight of any individual whatever! A 1000-lb man would be a monstrosity; a 1000-lb automobile at least a curiosity!

These considerations are of great importance in the construction of economic aggregates and averages, for there is real danger in the use of aggregates which are too heterogeneous or averages which are too unrepresentative to be sufficiently meaningful. The difficulty arises because we cannot be too finicky about economic aggregates and averages: we cannot insist, for instance, on perfect homogeneity of our aggregates, because many of the things in which we are interested are in fact aggregates of rather heterogeneous individuals. An excellent example of such heterogeneity is to be found in the concept of an aggregate of real capital. Real capital consists of a list or inventory of a fantastically diverse collection of objects—so many elephants, so many pianos, so many toothbrushes, so many china dogs, so many turret lathes, and so on. It is clear that simple aggregation in this case is nonsensical; there is no point in adding fifteen elephants to twenty-five pinches of snuff: the result is not forty anythings, not even forty sneezes! Nevertheless the aggregate of real capital is something which interests us. We want to know whether real capital is increasing or decreasing, whether it is larger or smaller in one place or time than in another, and we cannot be fobbed off by the answer that the question is statistical nonsense.

Our sense that the aggregate of real capital is meaningful is derived from a conviction that the various items of real capital can be *valued* in terms one of another; i.e. that we can postulate certain ratios of equivalence, or prices, which will enable us to convert all these heterogeneous items into a homogeneous sum. The commonest method of valuation is, of course, in terms of money: if we value elephants at $5000 apiece and pinches of snuff at 1 cent apiece, then we can say with considerable con-

fidence that fifteen elephants and twenty-five pinches of snuff are "worth" $75,000.25. No matter how heterogeneous an aggregate, if each item can be expressed in terms of its equivalence or "value" in some common denominator or *numeraire* then the value of the aggregate can be summed. It is not necessary, of course, to sum the aggregate of capital in terms of money. As long as we have a series of value ratios, enabling us to express all the various items in terms of a single homogeneous variable, the heterogeneous aggregate can be summed. Thus suppose we knew that an elephant was equivalent to a ton and a pinch of snuff to an ounce of wheat; we could reduce inventory of elephants and snuff to a common total in terms of wheat.

Measurement of the aggregate capital or income

It is because we have a certain faith that the value or equivalent ratios of different kinds of capital are fairly stable that we have confidence in the concept of a total stock of capital as a single magnitude. It should be noted that these value ratios do not have to be exchange values, i.e. ratios at which the assets can actually be transformed one into another through exchange. It is true that the value ratios are likely to be based on exchange values: nevertheless there are frequently occasions when people feel that the existing structure of exchange values is "abnormal," in which case they are clearly referring to some past, or to some average, of past relationships as the "true" value system. After World War II, for instance, most people felt that the prices of houses and secondhand automobiles were "out of line," i.e. were above their "normal values" in the general structure of relative value relations. Accountants frequently assign somewhat arbitrary values to fixed capital and equipment in an attempt to avoid taking account of merely temporary fluctuations in exchange relationships. Just as our common experience of buying and selling builds up in our minds a certain conception of the "value of money" or "what a dollar is worth," so in our common experience we find that the structure of relative values is moderately stable and we get a certain feeling for what things are "really" worth in terms of each other. It must be emphasized that this is a purely subjective feeling, which may change, of course, as our experience changes. Most people in middle life after World War II had a certain feeling that prices were high: that butter at a dollar, small houses at $10,000, Chevrolets at $1400, and so on were not what they were brought up on—which is true! Young people who had never known any lower price level, of course, would not have such a feeling: to them 50 cent butter and $800 Chevrolets would seem cheap. Similarly those whose economic frame of reference was set in the thirties might have a feeling that food-

stuffs after World War II were high relative to other things—as, compared with prices in the thirties, they were; to those who knew no other system of relative values, however, the postwar state of affairs would seem quite normal.

It is clear that an unequivocal measure of the physical stock of capital can be given only if the value equivalents of all the various items remain constant. Only if a typewriter is always equivalent to two suits can we say whether the total stock of capital has increased, decreased, or re-

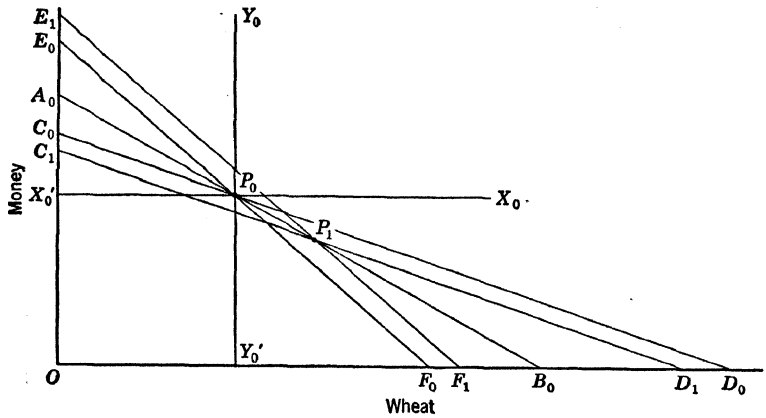


FIG. 67. The aggregation of capital.

mained the same when we add x typewriters and take away y suits from the total stock. If the relative values of commodities changed rapidly and irregularly we would have no sense that the total stock of physical capital was a measurable aggregate. Because in fact the relative values of commodities, whether measured by their actual exchange ratios or by some subjective "normal" value ratio, do change in the course of time the total stock of capital cannot be measured exactly: we can, however, postulate certain extreme values for its variation.

The asset-valuation analysis of Part 1 throws a good deal of light on this whole problem. We will return to the simple case of two assets—call them "wheat" and "money." In Fig. 67 the stock of wheat is measured horizontally, the stock of money vertically. P_0 represents an initial stock of money and wheat: P_1 a new combination of stocks. The question is, Does P_1 represent a larger or a smaller stock of capital than P_0? If P_1 were to lie within the field bounded by the horizontal line P_0X_0 and the vertical line P_0Y_0 there would be no question, as P_1 would represent more of both assets than P_0, and unless one or more of the

assets is a discommodity an increase in both assets clearly represents an increase in the total. There is likewise no problem in comparing the combination P_0 with a combination in the quadrant bounded by P_0X_0' and P_0Y_0', where there is less of both assets than before.

The problem arises in comparing the combination P_0 with a combination either in the quadrant $X_0P_0Y_0'$, where there is more wheat but less money than before, or in the quadrant $Y_0P_0X_0'$, where there is more money but less wheat than before. In these quadrants the change from P_0 to P_1 may represent an increase, a decrease, or no change in the total value of assets, depending on the value equivalents or prices employed in obtaining a common-value denominator. Thus if the price of wheat for valuation purposes is equal to the slope of the line P_0P_1, which cuts the money axis at A_0 and the goods axis at B_0, the shift in asset composition represents no change in the total: the money value of both the combination P_0 and the combination P_1 is OA_0; the wheat value of both is OB_0. If now the valuation price of wheat is less than this—say, equal to the slopes of C_0D_0 or C_1D_1—the movement from position P_0 to P_1 represents a decline in the index of total capital, whether we measure it in money values, which decline from OC_0 to OC_1, or in wheat values, which decline from OD_0 to OD_1. If, however, the valuation price is greater—say, equal to the slopes of E_1F_1 or E_0F_0—the movement from P_0 to P_1 represents an increase in the index of total capital, money values increasing from OE_0 to OE_1, wheat values from OF_0 to OF_1. It is clear that the principle of this figure can be extended to any number of dimensions: no matter how many different items of capital we have there are some systems of relative valuation prices which will yield no change in the index of total capital as we move from one combination to another. If we are given all valuation prices except one, a level of that one price can be found which will make the values of two combinations equal.[2]

[2] Suppose that we have quantities A_0, B_0, C_0, etc., of the different kinds of capital to begin with, and A_1, B_1, C_1, etc., in the new position. Let the valuation price of B in terms of A be b, of C in terms of A be c, and so on: that is to say, one unit of B is regarded as equivalent to b units of A, and so on. Then the total A-value of the first position is $V_0 = A_0 + B_0b + C_0c + \cdots$, and the total A-value of the second position is $V_1 = A_1 + B_1b + C_1c + \cdots$. If all the valuation prices b, c, etc, are given except one—say, b—a value of b can be found which will equate V_0 and V_1:

$$b = \frac{A_0 - A_1 + c(C_0 - C_1) + \cdots}{B_1 - B_0}$$

The value of b that equates the two values may not, of course, be positive, in which case it would not be a valid solution unless the asset it priced was a discommodity. There is always likely to be a wide range of positive relative valuation prices, however, at which the values of any two combinations of asset quantities are equal.

In general, in comparing the values of two different combinations of assets, we will be given two extreme sets of valuation prices—one at which the first combination is actually valued; the other at which the second combination is actually valued. Thus, in Fig. 67, suppose that the price of wheat in the combination P_0 is equal to the slope of E_0F_0 (or E_1F_1) and that the price in the combination P_1 is equal to the slope of C_0D_0 (or C_1D_1). Then OE_1 is the value of combination P_1 at the P_0-values, and OC_0 is the value of combination P_0 at the P_1-values. If we are constructing an index of the change in capital between P_0 and P_1, with, say, $P_0 = 100$, then at P_0-values the index of the value of P_1 would be $\frac{OE_1}{OE_0}100$, and at P_1-values the index of the value of P_1 would be $\frac{OC_1}{OC_0}100$. In the case of the figure we see that the index rises when P_0-values are used and falls when P_1-values are used: this is not necessarily true, of course, and it would be perfectly possible for both indices to rise or fall. These two indices may be said to denote the extreme statistical range of the measure of change in total capital, for there is clearly no significance in going to any more extreme values, possible as these may be.

It is possible to carry the problem a little further, conceptually at least, with the aid of the indifference-curve analysis. If we can assume that there is a stable indifference-curve system which is common to the two asset combinations (a fairly large assumption!) then we can define the *larger* capital combination as the *preferable* one, i.e. the one that stands on the highest indifference curve. Thus, in Fig. 68*a*, P_1 would be a "larger" capital stock than P_0; in Fig. 68*b* it would be smaller, as it stands on a lower indifference curve. If we knew the indifference curves there are several ways in which the capital change from P_0 to P_1 might be measured. We might measure it by the "money equivalent" or the "wheat equivalent." OM_0 is the money equivalent of the combination P_0—that is to say, it is that amount of money combined with no wheat which would be just as satisfactory to the individual as the combination of money and wheat represented by P_0. Similarly OM_1 is the money equivalent of P_1, and the ratio OM_1 to OM_0 is one possible measure of the increase or decrease in total capital from P_0 to P_0. Similarly we might measure it in terms of the change in the wheat equivalent, OW_1/OW_0. The wheat equivalent will always have the same sign as the money equivalent, as indifference curves cannot cross one another, but it will not, in general, have the same magnitude. This method can clearly be generalized to more than two asset types: it is not,

however, particularly valuable empirically, and when there are a large number of assets the single-asset equivalent of a large group of assets may be virtually infinite, the whole significance of the asset group lying in their combinations.

Another possible measure would be to value each asset quantity at its marginal rate of substitution with the *numeraire* or measure of value. In terms of the figure this would involve running tangents to the indifference curves at P_0 and P_1 to meet the money axis in S_0 and S_1; then

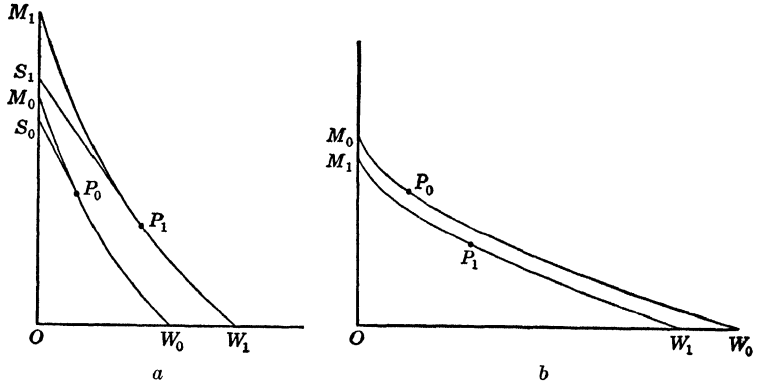


FIG. 68. Asset aggregation by indifference curves.

the ratio OS_1/OS_0 would be the money measure of the capital change: running these same tangents to the wheat axis would give the corresponding wheat measure.

There are many objections to all these procedures. The root of the whole difficulty lies in the fact that what we are trying to measure is utility in an objective, cardinal sense, which is a problem we have not yet succeeded in solving and one which is probably insoluble. What we really want to know is the "relative height" of the two indifference contours, and a perfectly possible, though not very satisfying, answer to the question is that the question itself is an illegitimate one! There is a particular objection to the method of valuation at the marginal rate of substitution—that if the curvature of the indifference curves is considerable it may easily produce a nonsensical result, giving a lower value to a combination on a higher indifference curve. The most fundamental objection of all, however, is that in measuring the total capital of a society we have no way of summing the various individual indifference curves of the persons who comprise the society. Suppose individual A prefers combination P_1 to P_0, whereas individual B prefers P_0 to P_1: there seems to be no way of resolving these differences or of adding them.

In order to construct "social indifference curves" we should have to break both the commandments of indifference-curve analysis: we would have to assume a cardinally measurable utility and would further have to assume the possibility of interpersonal comparisons of utility, i.e. that the utility unit was the same for all individuals. While, therefore, the indifference-curve analysis at least throws some light on the nature of the problem of the measurement of heterogeneous stocks of capital, it does not take us very far towards the solution of the problem: we are left with the proposition that the measurement of the stock of capital is possible in completely unequivocal terms only if the valuation-price structure is stable; if there are fluctuations in the valuation-price structure we must be content with a range of possible values, and the most we can say is that this range is usually small enough so that the concept of a change in total assets has enough significance to be interesting!

The problems involved in the measurement of a total of assets are repeated exactly in the problem of the measurement of a total of "real income" or "real outgo." Just as the total of assets is a heterogeneous inventory of different items, so the gross additions to this total, which constitute real income, or the gross subtractions from it, which constitute real outgo, are also heterogeneous inventories of different items, and if they are to be reduced to a common measure each item must be multiplied by a valuation price. Unless the valuation-price system is constant we again have an ambiguity in the concept of a total of production, i.e. real income, or consumption. There is an added difficulty in the case of aggregating income and consumption—that the valuation prices may themselves change during the accounting period. There seems to be no solution to this problem except by taking accounting periods short enough so that we can assume no change in valuation prices *during* the period and a change only on the last moment of the period. There is thus no escape from the fundamental ambiguity of the total capital concept, as some writers seem to have supposed, in attempting to reduce capital to some equivalent of income. Income is simply a differential of capital, i.e. additions to or subtractions from it, and all the difficulties of the capital concept are equally inherent in the income concept.

Price-level indices

Similar difficulties apply to the concept and measurement of a general or average price level of an aggregate of goods, whether we conceive of the price level as a characteristic of a stock of capital or of a flow of income. Just as we can measure changes in a stock of heterogeneous goods unequivocally only if the valuation prices remain constant, so we can

measure changes in the price level of an aggregate of goods only if the composition of the aggregate remains constant. If we wish to compare the price levels of two different aggregates of goods with different prices, we can obtain only a range of possible values, the extremes of which are set by measuring the change in the total value, under the impact of the change in prices, first of one aggregate of goods and then of the other.

Suppose that we have two aggregates of goods, each with its price: the aggregate P_0, consisting of q_a of A priced at p_a, q_b of B priced at p_b, and so on, and the aggregate P_1, consisting of q_a' of A, priced at p_a', q_b' of B priced at p_b', and so on. From these figures we can derive four aggregate value quantities:

$$\begin{aligned} V_0 &= \Sigma p_a q_a \\ V_1 &= \Sigma p_a' q_a' \\ W_1 &= \Sigma p_a q_a' \\ W_0 &= \Sigma p_a' q_a \end{aligned} \tag{1}$$

V_0 and V_1 are "real" value quantities: V_0 is the total value of the aggregate P_0, each item valued at its own price; similarly V_1 is the total value of the aggregate P_1. W_0 and W_1 are imaginary or hypothetical values: W_0 is what the total value of the P_0 aggregate of goods would be, valued at the prices of the P_1 aggregate; W_1 is what the total value of the P_1 aggregate would be, valued at the P_0 prices. There is an assumption present here that all goods represented in both aggregates have prices represented in both aggregates. It is not necessary to assume that all goods are present in both aggregates. Thus we do not need to assume, for instance, that B exists in both aggregates: if $q_b = 0$ we can still aggregate it! We must assume, however, that we can postulate a price even for goods which do not exist in one or the other of the aggregates, in order to calculate W_1 or W_0. This may present a real problem in attempting to compare price levels of widely different aggregates of goods, e.g. the price levels of widely different ages or cultures.

From these four value aggregates the extreme levels of both quantity and price indices can be immediately derived. We have the two quantity indices, expressing the quantity aggregate of P_1 as a proportion of P_0 ($P_0 = 1$),

$$Q_0 = \frac{W_1}{V_0}, \qquad Q_1 = \frac{V_1}{W_0} \tag{2}$$

Q_0 is the ratio of the quantity aggregate P_1 to P_0, valued at the prices of

P_0; Q_1 is the ratio of the quantity aggregate P_1 to P_0, valued at the prices of P_1.

Similarly we have the price indices:

$$\pi_0 = \frac{W_0}{V_0}, \qquad \pi_1 = \frac{V_1}{W_1} \tag{3}$$

π_0 is the ratio of the value of the P_0 aggregate of goods measured at the prices of P_1 to the value of the same aggregate measured at the prices of P_0. π_1 is the ratio of the value of the P_1 aggregate of goods measured at the prices of p_1 to the value of the same aggregate measured at the prices of P_0. π_0, that is to say, measures the relative change of the weighted arithmetic mean of the prices between the two situations, weighted by the quantities of P_0: π_1 is a similar change in the weighted arithmetic mean of the price change, weighted by the prices of P_1.

As there is no way of telling which of the two extreme positions of either the quantity or price index is "right"—or whether any intermediate position is right—it does not matter very much which index we select. If we use the Q_0 quantity index, however, we should use the π_1 price index, and if we use the Q_1 quantity index we should use the π_0 price index. The reason for this is that price and quantity indices should be selected which have the property that the *product* of these indices should equal the index of change in total value at own prices. We see from the formulae that

$$Q_0\pi_1 = Q_1\pi_0 = \frac{V_1}{V_0} \tag{4}$$

With the indices selected in this way the general quantity- and price-level measures will obey the rule which all individual prices and quantities follow—that the total value of any quantity is equal to the quantity multiplied by its price.

There are, of course, a great many other possible indices of price and of quantity besides the simple weighted arithmetic averages used here. There is, for instance, the geometric mean, or even more complex means such as Fisher's so-called "ideal" index number, which is the geometric mean of the two extreme geometric means. Some of these other means or aggregates have convenient statistical properties. None, however, has any more economic significance than the simple arithmetic means used here, and none can escape the fundamental conceptual ambiguity of an aggregate of heterogeneous items. There seems to be very little point, therefore, in using them.

In the two-asset case these relationships can be visualized very simply in our usual diagram. In Fig. 69 we again suppose two assets, money and wheat. The price of money is, of course, equal to 1. P_0 and P_1 are two money-wheat combinations. The price of wheat in the P_0 aggregate is equal to the slope of P_0V_0, and in the P_1 aggregate is equal to the slope of P_1V_1. Then $OV_0 = V_0$, the value of the P_0 combination at P_0 prices, and $OV_1 = V_1$, the value of the P_1 combination at P_1 prices. If, now, we draw P_0W_0 parallel to P_1V_1 to meet the money axis in W_0, $OW_0 = W_0$, the value of the P_0 combination at P_1 prices. Similarly, drawing P_1W_1 parallel to P_0V_0, we get $OW_1 = W_1$, the value of the P_1 combination at P_0 prices. With prices constant, the change in quantity can be measured by the movement either from V_0 to W_1, or from W_0 to V_1. With quantities constant the change in prices can be measured by the movement either from V_0 to W_0, or from W_1 to V_1. The value movement from V_0 to V_1 can be divided into a price and a quantity movement either at W_0 or at W_1: if at W_0 the price movement accounts for the movement from V_0 to W_0 (π_0); the quantity movement, from W_0 to V_1 (Q_1). If we divide the value movement at W_1, V_0W_1 (Q_0) is the quantity change, W_1V_1 (π_1) the price change.

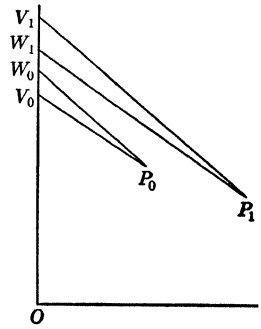


FIG 69. Price and quantity levels.

The indifference-curve analysis has been used also in an attempt to give greater precision to the concept of a change in the price level, the change being measured by the compensating payment which just is sufficient to compensate for the price changes. This definition, however, suffers from the fatal defect of all indifference-curve analysis, from the macroeconomic point of view, that it can be said to have precision only from the point of view of a single individual. Hence a given objective change in the price structure may represent as many different price-level changes as there are individuals, and it is hence impossible, even conceptually, to devise a measure of the price level which will avoid this ambiguity unless we are prepared to allow interpersonal comparisons of utility. This ambiguity is not merely a statistical difficulty: it is embedded in the very nature of the concept of a price level or of any other attempt to value a heterogeneous mass of items in terms of a single parameter.

Suppose, for instance, that we wish to compare the price level in Denver and in New York City. For a single individual this amounts to

the question, Will a dollar buy more in New York City or in Denver? The answer to this question, however, depends on what the individual wants to buy. Some things which can be bought in New York, such as theater tickets, are not available in Denver. Some things which are available in Denver, such as rides to the mountains, are not available in New York. For an individual the question really resolves itself into the question whether he would rather live in Denver or in New York City in similar positions on the same salary. Clearly this is a question which might be answered differently even by the same individual at different levels of salary: one might very well prefer to live in Denver on $3000 a year, and in New York on $20,000. It is certainly a question which might be answered differently by two different individuals. It is true that for a single individual, at a given level of salary, under given working and living conditions, the difference in price level between New York and Denver might be expressed accurately by means of a compensating payment. Thus, if the individual would just as soon have a salary of $5000 in Denver as $6000 in New York, then we could say that for him, at this level of income, the price level in New York was 120 as compared with 100 in Denver. With the same set of prices in the stores another individual might well just as soon have $6000 in Denver as $5000 in New York: for him the price level in New York would be $83\frac{1}{3}$ as compared with Denver's 100. Theoretically, of course, it would be possible to aggregate these various price levels by taking all the individuals who have some sort of option of living in New York or Denver, finding out for each what would be equivalent incomes in the two places, and aggregating these incomes; then if for, say, 100 individuals we found that $500,000 in Denver was just equivalent to $550,000 in New York, we might say that the price level in New York was 110 to Denver's 100. But this process has very little to do with the usual statistical processes of comparison of price levels, and the result obtained from the "compensating payment" method might not even lie within the range of the price level as determined by the weighted average of goods found on the market in the two cities. In determining individual decisions a great many "free goods" inevitably enter the picture which cannot by their very nature get into any statistical index: the excitement of New York and the fresh air of Denver are unknown to the world of economic statistics, but they may play a major part in the decisions of an individual in choosing a place of residence.

The fallacy of aggregation

One further danger in macroeconomic thought must be emphasized. It is frequently necessary to treat economic aggregates and averages as

simple homogeneous variables for purposes of mathematical analysis. Thus we frequently find the economist writing, "Let Y be the national income" or "Let W be the level of wages." Within such innocent-looking mathematical symbols there lurks a world of danger. One of the most fundamental postulates of mathematical analysis, at least on the simple level of Newtonian calculus, is that the variables of the analysis are homogeneous quantities which have no structure that is significant for the problem in hand. In engineering, physical, and astronomical problems this frequently is a fair approximation to the truth. Thus if we write, "Let D be the distance from the earth to the sun," we have a fair degree of confidence that this distance has no significant structure, i.e. that every mile, as far as its property of extension is concerned, is exactly like every other mile. The engineer on the other hand who writes, "Let W be the weight of the beam," may get into some difficulties if he neglects the distribution of weight in the various parts of the beam, but he at least has the satisfaction of knowing that every gram, once he has located it, behaves in the same way as every other gram, for his purposes at least.

The economist, however, who writes, "Let Y be the national income," is running into much greater danger: he is assuming implicitly that for the purposes of the problem he has in hand the *structure* of the national income may be neglected and the only significant thing is its aggregate size. For some purposes this may approximately be true; for most problems, however, the structure of the national income may be as important as, if not more important than, its aggregate size. No amount of mathematical analysis with the variable Y will reveal this fact, for mathematical analysis cannot discover any more than was already implicit in the problem, and unless we break down Y explicitly into some of its components mere mathematical manipulation will never break it down for us. The inadequacy of the predictions of the macroeconomists in regard to the immediate postwar period, when most of them predicted a severe slump during the transition period, seems to have been the result of relying too much on assumed functional relations, such as the consumption function, between heterogeneous aggregates without realizing the importance of the composition of these aggregates.

The "art" of macroeconomic analysis consists largely in the breakdown of the great aggregates of the system into smaller aggregates which are not too numerous to handle but which are small enough so that the inevitable heterogeneity is insufficient to upset the results of the analysis. The smaller the aggregates with which we conduct our analysis, the less likely are we to find that the neglect of their internal structure leads us seriously astray. On the other hand, the smaller our basic aggregates,

the more of them we shall have and the more complex our analysis of the aggregates themselves: we can handle four or five aggregates with ease, but ten or twenty are quite hard to juggle all together. Thus for most purposes the "average wage" or the aggregate wage bill is an aggregate too heterogeneous to be of much use: the structure of relative wages may be a much more potent factor in determining the course of economic events than the average level of wages. If on the other hand we define a hundred different wage groups, such as the average wages of carpenters, plumbers, and so on, our system will have so many variables that it will be difficult to treat either analytically or statistically.

The "fallacy of aggregation," i.e. of working with aggregates with too significantly heterogeneous a structure, has been a common one in the history of social thought. One particularly interesting example of the fallacy is the Marxian system, which, in spite of its many valuable insights, loses most of its validity and predictive power because of its assumption that heterogeneous groups like "the working class" are much more homogeneous than they really are. In fact the Marxian "working class" is a pure fiction: an aggregation of even less internal consistency and homogeneity than the "nation." What we have instead is a heterogeneous inventory of carpenters, ditch diggers, teachers, and the like, the various groups of which have a good deal of internal homogeneity but very little in common with each other. Hence the Marxian interpretation of the labor movement fails completely to recognize the significance of the intergroup or interindustrial struggle as opposed to the "class struggle." Workers and employers in a particular industry—say, leather—have a much stronger common bond and common interest than, say, workers in leather and rubber or employers in leather and rubber. Hence the "class struggle" in industry is frequently merely shadow boxing, and the real struggles lie between the industrial groups—between competing employer groups on the one hand or rival unions on the other.

Another interesting example of the fallacy of aggregation has been the theory of international trade, which has usually failed to recognize the economic heterogeneity of political states and the homogeneity of international with domestic trade. The international trade theorists have tended to visualize nations as individuals trading around a table, instead of as vast aggregates of individuals engaged in an intricate web of individual intercourse. There are, of course, aggregates of trade which have enough unity to serve as objects of discourse, but they seldom coincide with national boundaries. The whole problem of the proper treatment of spatial and political aggregates is of great interest, and much work needs to be done on it; the topic, however, has been excluded from the scope of this book.

Chapter 11

The Structure of Macroeconomic Aggregates: I. Capital and Income

If we are to analyze the aggregates of the economic system without falling into the fallacy of aggregation we must first look into the structure and composition of these aggregates to see, if we can, what subclassifications will be useful. From these subclassifications we shall be able to derive the fundamental macroeconomic identities which are the basis for any system of macroeconomic analysis.

The age structure of capital

Consider first the aggregate of physical capital. It consists, as we have seen, of a vast heterogeneous inventory of diverse items. The problem of the "structure" of capital is essentially that of the subclassification of these diverse items into meaningful groups. It is evident that hypothetically there might be many such subclassifications, each one useful for a different purpose, and that these classifications might cut across one another. There is, for instance, the possibility of classifying according to the *purpose* for which the item of physical capital is used. The textbook distinction between producer's and consumer's goods is based on such a principle, though it is doubtful whether such a distinction is very significant. Another possible classification is according to some characteristic of the owner: it might be useful, for instance, to distinguish between goods owned by persons of different income levels, or with different liquidity preferences. Yet other classifications may be based on some physical properties of the capital items themselves. Probably the most important classification of this kind is by the *age distribution* of the capital stock. The members of any particular item of the capital stock can be regarded as a "population." When an item is produced it is "born" into its population, and when it is consumed—i.e. ceases to conform to the definition by which the population is de-

fined—it "dies." The time elapsed between its birth and the present is its "age." It may be observed that any definition defines a population of the things defined, and that when anything conforms to the definition it is "born" and when it ceases to conform it "dies."

The age distribution of physical capital may be a very important element in the explanation of economic cycles and the course of production and consumption. By the age distribution is usually meant the proportion of the total population in each age group. Thus we may have 5% from 0 to 1 year old, 10% from 1 to 2 years old, 12% from 2 to 3 years old, and so on. The mathematical study of these age distributions and their relations to the history of the population is one of the principal subjects of actuarial science and is very complex. It is not difficult, however, to outline the essentials of the relationships involved. The first principle is the identity that the increase in any population in a given period of time is equal to the total number of "births," i.e. gross additions to the population, less the total number of "deaths," gross subtractions from the population. This proposition, simple as it is, is perhaps the most fundamental of all science, and we shall find it illuminating in economics time and time again.[1] This principle may perhaps be called the "accumulation principle." It can be applied immediately in the derivation of a second important law of population: that the number in any given age group diminishes in a given period of time by the number of deaths which have taken place among the individuals of that age group. The individuals who were born in 1950 will be between the ages of 0 and 1 on January 1, 1951, between the ages of 1 and 2 on January 1, 1952, and so on. If we regard all individuals born in 1950 as a population it is clear that no further "births" can take place into this particular population after 1950 has expired, as by definition only those born in 1950 can belong to this particular population. The number in this population on January 1, 1951, will be equal to the number born in 1950 less the number of deaths: this may be called the "net births" into the population. The number in this population on January 1, 1952, will be equal to the number on January 1, 1951, less the number of deaths during the year 1951, and so on. The proportion of a given number of net births that die in each succeeding "year" may be called their "survival distribution."

It is also convenient to have a term which describes the absolute number of individual items in each age group—the "age grouping." Sim-

[1] This is essentially the same as the principle which I have elsewhere, rather facetiously, called the "bathtub theorem"—the addition to the population of drops of water in a bathtub being equal to the number of drops that have run into it less the number that have run out of it. See my *Economics of Peace* (Prentice-Hall, 1945), p. 127.

ilarly we need a term—the "survival grouping"—which describes the absolute number which survive into each successive year out of a certain absolute number of net births. Thus suppose that a million automobiles are produced in 1950, and that 950,000 survive into 1951, 900,000 into 1952, 800,000 into 1953, and so on down to the last survivor in the museum of the year 2000: the series 1,000,000, 950,000, 900,000, 800,000, etc., would be the survival grouping. The survival distribution would be 1, 0.95, 0.90, 0.80, etc., indicating the *proportion* of the output of 1950 that survived into successive years. Similarly in the year 1950 we might have 1,000,000 cars between 0 and 1 year old, 900,000 between 1 and 2 years old, 800,000 between 2 and 3 years old, etc.: this would constitute the age grouping. If out of the total population of cars 10% were between 0 and 1, 9% between 1 and 2, and so on, this series would be the age distribution.

The history of the numbers and age distribution of any population can be predicted from the following data: (i) the survival distribution, or, if this is not invariant, the functional relationships which will give us the survival distribution of each "class," i.e. each year's number of net births; (ii) the existing age grouping; (iii) the number of net births per annum, or, if this is not invariant, the functional relationships which will give us the number of net births per annum. Again, the mathematical solution of these functions may be very difficult, but the principle involved is not difficult to follow. Suppose, to shorten the exposition, that we have a population of four age groups. Let its initial age grouping be $a_0b_0c_0d_0$, i.e. a number a_0 in the first age group, b_0 in the second, c_0 in the third, d_0 in the fourth. The total initial population $P_0 = a_0 + b_0 + c_0 + d_0$. Let the subscripts 1, 2, etc., refer to succeeding years: then the course of the population can be described by a table such as Table 1. B_0, B_1, etc., are the net births in the years 0, 1, etc. Then the horizontal series give the age groupings; the vertical

TABLE 1

MOVEMENT OF A POPULATION THROUGH TIME

		B_0	a_0	b_0	c_0	d_0
	B_1	a_1	b_1	c_1	d_1	
B_2	a_2	b_2	c_2	d_2		
B_3	a_3	b_3	c_3	d_3		
B_4	a_4	b_4	c_4	d_4		
. . . . .						

series give the survival groupings. Of the net births B_0 in year 0, a_1 survive into year 1, b_2 into year 2, c_3 into year 3, d_4 into year 4, and none into year 5. Of the B_1 net births in year 1 a_2 survive into year 2, b_3 into year 3, and so on. Starting from any given age grouping, if we can find the net births of each year, say, as a function of the total population, and if we know the survival grouping from each class of net births, we can chart the whole future course of the population in such a table. Depending on the assumptions we make regarding the various functional relations, we may get stable equilibrium populations, damped or expanding oscillations, and so on: the whole theory of these relationships is very complex and need not detain us here, as it is sufficient by a few examples to illustrate the principles involved.

If the net births and the survival distribution are constant, the population, no matter what its initial position, very soon (i.e. within the lifetime of the longest-lived individual) reaches an equilibrium position at which the total population and the age grouping remain constant from year to year. If B is the constant number of births per annum, of which Bs_1, Bs_2, Bs_3, etc., survive to be 1, 2, 3, etc., years old, the course of the population, starting from the position a_0, b_0, c_0, d_0, will be as in Table 2. It is clear that by the year 4 the population has reached an

TABLE 2

MOVEMENT OF AN EQUILIBRIUM POPULATION

								Year
				B	a_0	b_0	c_0	d_0	0
			B	Bs_1	b_1	c_1	d_1		1
		B	Bs_1	Bs_2	c_2	d_2			2
	B	Bs_1	Bs_2	Bs_3	d_3				3
B	Bs_1	Bs_2	Bs_3	Bs_4					4
Bs_1	Bs_2	Bs_3	Bs_4						5

equilibrium. The total population is constant, given by

$$P = B(s_1 + s_2 + s_3 + s_4) \tag{1}$$

The age grouping is not only constant from year to year, but is identical with the survival grouping, which, as the number of births is constant from year to year, is also constant. The rate of deaths D—i.e. the annual

number of deaths—is equal to the rate of births, B:

$$D = B\{(1 - s_1) + (s_1 - s_2) + (s_2 - s_3) + (s_3 - s_4) + s_4\} = B \quad (2)$$

This condition has to hold, of course, if the population is constant. It can be shown also that the ratio of the population to the rate of births is constant and equal to the average length of life of the individual. Suppose that of the births B, $B(1 - s_1)$ die at age 0, $B(s_1 - s_2)$ at age 1, and so on. Then the average length of life, L, is given by:

$$L = \frac{B\{(1 - s_1)0 + (s_1 - s_2)1 + (s_2 - s_3)2 + (s_3 - s_4)3 + (s_4)4\}}{B}$$

$$= s_1 + s_2 + s_3 + s_4 = \frac{P}{B} \quad (3)$$

This formula may not be quite accurate where the age-group intervals are finite, as actually the average age at death of those who die between 0 and 1 is not 0 but some figure between 0 and 1, and so for the other figures. This inaccuracy disappears, however, in the limit when the age-group interval becomes zero and the age grouping becomes a continuous function.

The Austrian theory of capital

Equation (3) has a good deal of historic interest for economic thought because it is the assumption on which the so-called "Austrian" theory of capital is based. The Austrian theory, which actually goes back to Adam Smith's distinction between productive and unproductive labor and Ricardo's concept of capital as embodied labor (later put to such un-Ricardian uses by Marx!), conceives of capital as a population of "value" rather than a population of goods. Goods, that is to say, are regarded as "embodied value." Whenever value is added to them a new value item is "born," and whenever value is subtracted from them by consumption a value item "dies." The average length of life of these "embodied values," i.e. the average interval of time which elapses between their "birth," or embodiment in a commodity, and their "death," or release as a satisfaction, is the famous "average period of production." If such a population of values were an equilibrium population we see that the average period of production would be equal to the ratio of the total value of capital to the annual amount of value embodied or disembodied, which is income or outgo—the rate of births or deaths (these being equal) in the capital population. If the capital "population of value" is not an equilibrium population, of course, it is no longer true that the ratio between the value of the capital stock and the value of income or outgo is

equal to the average period of production. Even in this case, however, the average period of production or, more accurately, the time structure of capital is perhaps the most important single factor determining the capital-income ratio.

There is a great deal of truth in the Austrian viewpoint on capital, even though the concept of the average period of production has not proved to be a particularly refined instrument of analysis and the applications of the theory have been, on the whole, disappointing. The attacks on the theory, made especially by J. B. Clark, F. H. Knight, and their followers, have not been particularly fruitful and are based on a view of capital which seems to be even less realistic than the one criticized.

According to the Clark-Knight view capital is a homogeneous fund of automatically replaceable values which, by definition, cannot possibly have a time structure. In every act of labor (or other factor of production) values are immediately created, which may be immediately realized if the producer wishes through exchange, which is itself only a further form of production, so that there is no time interval between the creation and the realization of values. The weakness of this view, oddly enough, seems to be the same weakness that makes the Austrian theory itself somewhat unrealistic—the view that "values" in some way have an existence independent of the physical commodities in which they are "embodied." The very word "embodied" is, as a matter of fact, highly misleading: it derives from a rather crude embodied-labor theory of value which is often credited to Ricardo, although it is very doubtful whether Ricardo himself held it. The creation of an item of physical capital, such as a bushel of wheat or a house, does not necessarily involve the creation of a "value" which, as it were, resides in the wheat or the house as a soul in the body. Valuation is an essentially "present" process, arising out of the opinions, beliefs, and sentiments of the owners of physical capital (including money) operating on the stocks of the various forms of physical capital actually in existence at the moment of valuation. Neither the past history of capital nor even its future prospects can affect its present value except in so far as these things affect the state of mind of its owners. It is true, of course, that both past history and future prospects do affect the present state of mind of owners, but this connection is by no means a stable one, and valuation always proceeds through the intermediary of the present state of mind, or "asset preferences," of owners.

There is hardly any more subtle or corrupting fallacy in economics than that of misplaced concreteness as applied to values, the view that every good goes through its life with a birth certificate in the form of a

price tag. What persists, what has a time structure, is the physical commodity—the bushels, the tons, the houses, the machines. Valuation is a *process*, essentially a mental process, though it may express itself in physical acts of exchange applied to physical commodities. A value is not a "thing" like a bushel of wheat; it is a part description of an act or process. It is assuming too much, therefore, to think of capital with the Austrians as a "population of values": values are not identifiable individuals in a population, or cannot be so regarded without stretching the concept too far from reality.

On the other hand it is assuming far too much to suppose with Clark and Knight that capital is a homogeneous fund of values; in fact capital is a heterogeneous collection of physical objects *which are valued*, and are revalued continuously. All the principles of population analysis, however, apply with rigor to any defined aggregation of physical objects, which certainly are born, exist for a finite length of time, have an "age," and die. And in economic theory we neglect at our peril the fact of an age structure of physical capital, because of the empirical fact that the consumption or revaluation of physical capital in most cases is closely related to its age. For most items of physical capital, automobiles, refrigerators, all machines, clothing, most furniture, china, and so on, it would be possible to construct "life tables" much as we do for the human population. There are, of course, certain differences between the human population and the various populations of goods. The characteristics of individual human beings probably depend more closely on their age than do the characteristics of physical capital. It would be impossible for an individual having the characteristics of a seventy-year-old man to be "born": every human being that is born is a baby. It would be *possible* to reproduce the characteristics of a ten-year-old automboile in a new one straight from some very odd factory—but it is highly unlikely that this would happen! In a few isolated cases where certain desirable characteristics appear with age, e.g. in wines and antique furniture, there may be some reproduction of the characteristics of age in a new piece: quantitatively, however, such reproduction can be neglected, and for all practical purposes we can treat a population of goods almost exactly as if it were a human population. And unquestionably, if the population of automobiles were in equilibrium, the ratio of the total stock of automobiles to the annual production and consumption would be equal to the average length of life, i.e. to a figure closely allied to the Austrian "period of production." So far the Austrians are right. However, the most interesting aspects of the population character of capital is what happens in disequilibrium.

Populations of capital in disequilibrium

The "progressive" population analysis outlined above can handle the problem of disequilibrium in a way that the Austrian theory cannot. We can see immediately, for instance, how a disturbance in the age distribution of a population not only perpetuates itself from year to year—the distortion gradually moving up through the age distribution as time goes on—but also gives rise to cyclical movements in consumption and therefore in production. The general mathematical theory of these changes is quite difficult, and I do not propose to expound it, but the principles are simple and can be understood without much mathematical training. It is clear that the annual consumption of any particular item of capital—say, automobiles—depends mainly on the age grouping, i.e. on the numbers in each age group. If we could calculate a stable "life table," so that we knew exactly what proportion of those in each age group would "die," i.e. be smashed up or scrapped; then if we knew the numbers of cars in each age group we should be able to calculate exactly how many would disappear in a given year. In fact, of course, the life table is not quite stable enough for exact prediction; during World War II, for instance, automobiles somehow managed to last much longer, because of gasoline rationing and a conservation program, than otherwise might have been expected. However, for automobiles the life table is certainly stable enough to predict consumption within fairly narrow limits. The life table for automobiles follows the usual pattern of human life tables: the older the car, the greater its chance of being scrapped during the year. Hence a given population of automobiles will have smaller consumption if it is concentrated in the "young" age groups than if it is concentrated in the "old." If therefore there is a distortion in the age grouping it starts, of course, in the younger age groups.

The war provided an excellent example: very few automobiles were produced in the years 1943, 1944, 1945, and 1946. By 1947, therefore, there was a marked distortion in the age grouping; there were very few "young" automobiles, almost the whole automobile population being over the age of 5 years. Consequently the mortality could be expected to be well above normal, and the demand for replacement likewise was well above normal. As a result there was very large production of automobiles from 1947 to 1950: this is likely to continue in the fifties. By, say, 1955, the situation will have completely reversed itself. Nearly all the automobiles will be between 0 and 8 years old: there will be practically no automobiles in the 8–12-year age group, which normally has a fairly high mortality rate. Consequently by that time we may confidently expect that the *replacement* demand for automobiles will fall

pretty sharply, and unless there is a corresponding expansion in the total numbers of automobiles we may expect a sharp drop in automobile output, which might well lead to a general crisis.

The situation is made worse by the fact that any severe distortion in the age distribution of capital is likely to perpetuate itself into the lives of successive "generations" of goods—fortunately on a somewhat diminishing scale. Thus if because of the peculiar age grouping of automobiles there is a decline in output in the years 1955–1959, this will also set up a new distortion in the age grouping, probable less intense than the initial 1943–1946 distortion, but nevertheless noticeable. It may take three or four "generations" of automobiles before the distortion is smoothed out. This is sometimes called the "shadow effect."

In the case of human (and all other biological populations) there is a further important source of perpetuation of distortions in the age distribution because of the fact that the rate of births as well as the rate of deaths depends on the age grouping. The birth rate of a biological population depends in large measure on what proportion of the population is of childbearing age. We may have two equal human populations with the same birth rates in each age group; but if one has a larger number of people in the years from 20 to 40 than the other, it will have a larger birth rate and a larger annual rate of births. Hence distortions in human-population age distributions due, say, to the decline in birth rates during a war, may well produce "shadows" in the next generation when the potential parents are in smaller numbers than they otherwise would be, and hence the birth rate falls because a smaller proportion of the population is of childbearing age. A population of children and old people will have great difficulty in reproducing itself, at least until the children mature. The course of the German population since 1914 is a tragic example of this phenomenon. The "birth shadow" is, of course, much less likely to occur in populations of physical capital, as the "birth rate" of, say, automobiles depends on a much more complex mass of variables than the age distribution of automobiles themselves. There are no automobiles of childbearing age, and although the history of automobile design gives quite an uncomfortable amount of support to the facetious thesis of Samuel Butler that human beings are merely the reproductive organs of machinery, the new Ford, in spite of appearances to the contrary, is not the result of biological union between a Buick and a Studebaker, and its birth rate in no way depends on the number of cars in the lusty prime of life. Nevertheless the annual rate of birth of physical capital, i.e. its output, is very closely related to its consumption, which, as we have seen, is itself closely related to the age distribution. There may even be a certain small amount of genuine "birth-shadow"

effect in the case of physical capital. Automobiles themselves do not beget automobiles, but machine tools and skilled workers do beget them, and there may be cyclical movements in production equipment, both mechanical and human, which will influence the cycle of the output of automobiles.

The major distortions in age distribution of capital, which give rise to fluctuations in consumption and output, seem mainly to be the result of the economic disturbances occasioned by war. To a certain extent in fact the major economic fluctuations in Western society at least can be regarded as a "peak-load" problem, with the "peak" at war. Just as an electric power company must have a certain amount of unused capacity at 3 A.M. in order to take care of the peak load at 6 P.M., because of the technical fact that it cannot transform its equipment into any other use in a few hours, so the unused capacity of the thirties may be regarded in part as the unused war potential of a time of peace. Such a view, however, is oversimplified, though it represents an important contributing factor to the longer cyclical movements in Western capitalism: certainly the price cycle is mainly a war cycle with peaks in the major wars, and at least part of the so-called Kondratieff long cycle in output may in fact be a result of the cyclical nature of political disorganization rather than of anything which might specifically be called economic.

The equilibrium rate of growth of capital

There are, however, other factors in the structure of capital which might set up distortions in age groups, the most important of which is the rate of growth of capital itself. When capital is growing at a rate more rapid than a certain minimum it is impossible to avoid distortions in age groupings. This is seen most clearly in the most extreme life-table situation—that of the "one-hoss shay." Suppose that we had a population of capital goods of this nature in equilibrium, with every item lasting for n years and then suddenly disintegrating. If there were B births and deaths into the population in each year, each age group up to the nth would contain B members, and the $n + 1$th and subsequent age groups would contain nothing, all the n-year-olds dying off in the nth year. If now there is an increase in demand for the item, so that the population grows, there must be more than B births in each year. If the whole increase in demand is satisfied in one year there will be an output of, say, $B + b$ items in that year. The next year, however, only B items disappear—the n-year-olds—so that it is necessary only to produce B items in order to maintain the larger population. This will be true for n years. In the nth year, however, the $B + b$ items that were made in the year of large output suddenly reach the allotted span and have to be

replaced by another $B + b$ items. Here it is evident that a perpetual cycle is set up: output will be $B + b$ in every nth year and B in all the intervening years.

Even in this extreme case it is possible for growth to take place without setting up distortions in the age distribution if the growth takes place at exactly the right pace. If we wish to increase the total population by an amount p we must add p/n to the population in each year, for n years, the number of births in each year being $B + (p/n)$ indefinitely. At the end of n years there will be $(P + p)/n$ individuals in each annual age group, the total population being $P + p$ (P being the original population), and the annual number of births and deaths $B + (p/n)$. The following population tables may help to visualize these processes more clearly. A population of four age groups is assumed for the sake of simplicity.

In Table 3 the population is in equilibrium at a level P in year 1; in year 2 there is a rise in demand, so that the population rises to $P + p$ by year 3 through an increased number of births in year 2. In year 3,

TABLE 3

CYCLICAL DISTORTIONS IN A "ONE-HOSS SHAY" POPULATION

Year	Births											Total Population
1	$\frac{1}{4}P$							$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	P
2	$\frac{1}{4}P + p$						$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$		P
3	$\frac{1}{4}P$					$\frac{1}{4}P + p$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$			$P + p$
4	$\frac{1}{4}P$				$\frac{1}{4}P$	$\frac{1}{4}P + p$	$\frac{1}{4}P$	$\frac{1}{4}P$				$P + p$
5	$\frac{1}{4}P$			$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P + p$	$\frac{1}{4}P$					$P + p$
6	$\frac{1}{4}P + p$		$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P + p$						$P + p$
7	$\frac{1}{4}P$	$\frac{1}{4}P + p$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$							$P + p$

however, only $\frac{1}{4}P$—the fourth age group—dies off so that we only need $\frac{1}{4}P$ to replace it. There is, however, as a result of the large number of births in year 2 a distortion in the age distribution resulting in an unusually large age group of those born in that year. This age group gradually becomes older, passing through the population until it finally dies and has to be replaced in year 6, resulting in a large number of births in that year also, which likewise set up an unusually large age group. With a constant total population, then, it is possible to have a cyclical movement in births and deaths, i.e. in output and consumption in the case of populations of goods, if a distortion in age distribution is set up.

In Table 4 the increase of population p is accomplished by a $\frac{1}{4}P$ increase in each year: this results in an equilibrium population again by year 6.

TABLE 4

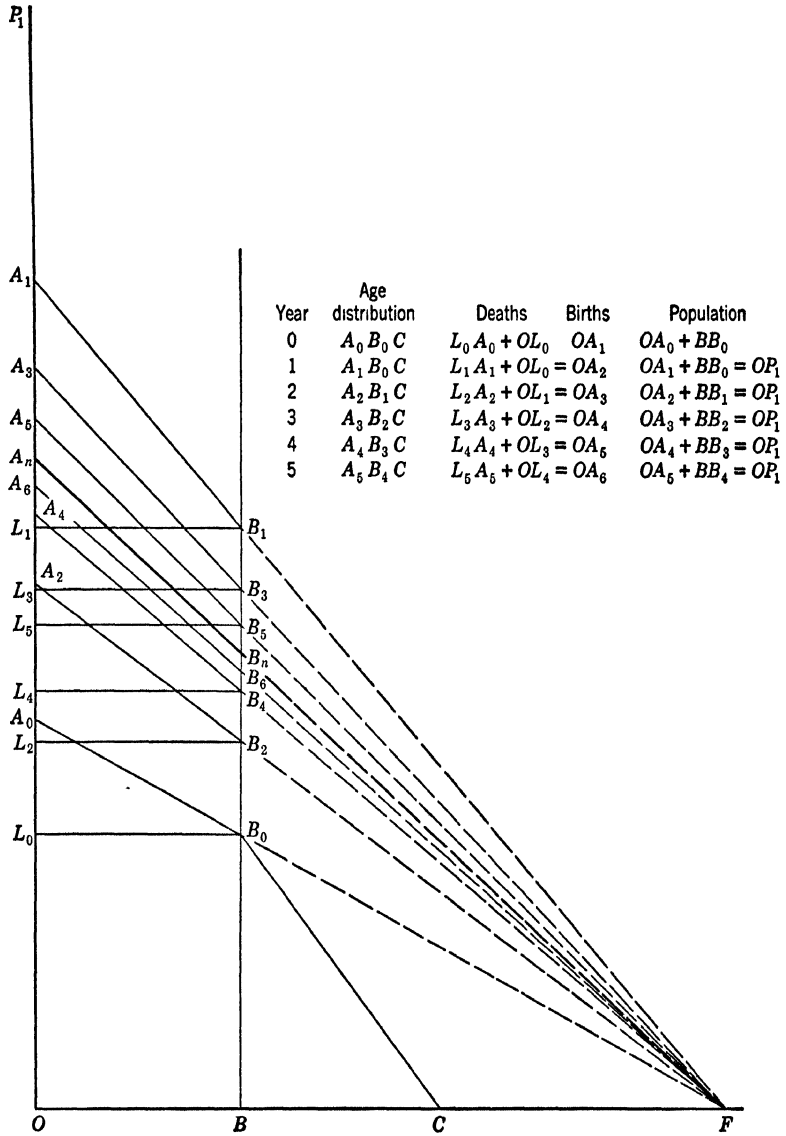
INCREASE IN POPULATION WITHOUT DISTORTIONS

Year	Births										Total Population
1	$\frac{1}{4}P$						$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	P
2	$\frac{1}{4}(P+p)$					$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$		P
3	$\frac{1}{4}(P+p)$				$\frac{1}{4}(P+p)$	$\frac{1}{4}P$	$\frac{1}{4}P$	$\frac{1}{4}P$			$P+\frac{1}{4}p$
4	$\frac{1}{4}(P+p)$			$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}P$	$\frac{1}{4}P$				$P+\frac{1}{2}p$
5	$\frac{1}{4}(P+p)$		$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}P$					$P+\frac{3}{4}p$
6	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$						$P+p$
7	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$	$\frac{1}{4}(P+p)$						$P+p$

It is interesting to observe that a population of this kind cannot be increased at a slower rate than p/n if it is desired to raise the whole population by an amount p, for if, say, the population is increased at a rate $k(p/n)$ $(k < 1)$ for n years an equilibrium population will be established at the level $P + kp$, which is less than the desired population: every year the age group born n years before will die and simply be replaced, and there will be no growth of population.

The "one-hoss shay" population is, of course, a very extreme case which is never likely to be found in practice. Usually there are a number of deaths in each age group in each year. Where this occurs the cyclical movement introduced by too rapid a growth in a population is damped—usually fairly rapidly—and the population recovers an approximate equilibrium. This is because the number of deaths, and therefore of replacement births, which takes place in each year is a function not merely of the last age grouping but of all of them.

The algebraic or analytic treatment of this problem is rather complicated, and the geometrical method can be used with advantage. In Fig. 70, numbers of individuals are plotted vertically. Three age groups are postulated, A, B, and C, the last having zero members. The numbers in group A are measured along OA_1, in group B along BB_1. Suppose, now, that there is an equilibrium population with OA_0 in the first and OB_0 in the second age group. The deaths in the first age group amount to L_0A_0 $(OL_0 = BB_0)$, in the second age group to BB_0, total deaths $(OL_0 + L_0A_0)$ being equal to total births (OA_0). Now suppose that it is desired to increase the population to OP_1, and this is done in a



Year	Age distribution	Deaths	Births	Population
0	A_0B_0C	$L_0A_0 + OL_0$	OA_1	$OA_0 + BB_0$
1	A_1B_0C	$L_1A_1 + OL_0 = OA_2$		$OA_1 + BB_0 = OP_1$
2	A_2B_1C	$L_2A_2 + OL_1 = OA_3$		$OA_2 + BB_1 = OP_1$
3	A_3B_2C	$L_3A_3 + OL_2 = OA_4$		$OA_3 + BB_2 = OP_1$
4	A_4B_3C	$L_4A_4 + OL_3 = OA_5$		$OA_4 + BB_3 = OP_1$
5	A_5B_4C	$L_5A_5 + OL_4 = OA_6$		$OA_5 + BB_4 = OP_1$

FIG. 70. Population oscillations.

single year, the population being held constant thereafter, and suppose, furthermore, that the "life table," i.e. the proportion who die in each age group in each year, is invariant. In year 1 there will be OA_1 in group A ($A_1P_1 = OL_0$), BB_0 in group B. In order to find the number of deaths from the OA_1 members in group A we project A_0B_0 to meet OB in F, and then draw A_1F, cutting BB_1 at B_1. Then BB_1 shows how many of the number OA_1 survive into the next age group, and L_1A_1 is the number of deaths from the age group A. The number of deaths from age group B is still BB_0 ($= OL_0$), so the total number of deaths is $OL_0 + L_1A_1$, or OA_2. To maintain the population constant, then, there must be OA_2 births, so that the next year this will be the number in the first age group: the number in the second age group is now OB_1, surviving from OA_1. Similarly, drawing A_2F to cut OB_1 in B_2, we get the number of survivors, BB_2 ($= OL_2$), from OA_2 into the second group and the number of deaths from the first group L_2A_2. The number of deaths from the second group that year is BB_1 ($= OL_1$), so that the total number of births and deaths is $L_2A_2 + OL_1$, which is equal to OA_3. We can repeat this process indefinitely, getting successive values for the annual births and deaths, OA_1, OA_2, OA_3, OA_4, etc. It will be seen from the figure that these form a damped cyclical movement, fluctuating with diminishing amplitude around the equilibrium value OA_n. If we draw A_nF to cut OB_1 in B_n, then $BB_n = A_nP_1$. The constant life table or survival distribution is given by OB/OF. This is the proportion of the first age group that dies: BF/OF is the proportion that survives into the second age group. This method is clearly applicable no matter how many age groups we have, and it gives the same general result of a damped cyclical oscillation following a rapid growth. In order to prevent this oscillation in the present example we need only increase the annual rate of births to OA_n and keep it there. In two years the population will arrive at its desired value with an equilibrium age distribution. The method is clearly applicable to any number of age groups, though naturally the geometry becomes more complicated as the number increases.

Income structures

The structure of the aggregates of income or production, of outgo or consumption, and of accumulation or investment—the three great aggregates of flow quantities—may be just as important in the interpretation of economic dynamics as the structure of the stock of capital. It is clearly not merely the aggregate total of production, consumption, and accumulation that matters: it matters *what* is produced, consumed, and accumulated, i.e. of what goods these aggregates are composed. An

aggregate output which consists mainly of highly consumable goods will have very different economic effects from an output which consists mainly of durables; an output which consists largely of war goods will have a different effect from one composed of peace goods. Output may also be classified roughly into wage goods and non-wage goods, wage goods being those commodities which are purchased by workers. In the short run it is probably the distribution of output among the various kinds of goods that determines the real distribution of income, especially if prices are flexible. Thus real wages, in the short run, depend mainly on the output of wage goods: if money wages rise without any change in the output of wage goods, workers can increase their purchases only at the expense of business inventories. The degree to which this is possible is strictly limited, and any marked decline in inventories will result in a rise in the price of wage goods sufficient to prevent further depletion. This means, however, that the rise in money wages is completely absorbed by the rise in the prices of wage goods, and real wages will be exactly where they were before the rise in money wages. Under certain circumstances, therefore, real wages may be quite independent of money wages or may even move in the opposite direction. We merely mention this problem now by way of illustrating the importance of breaking down aggregates of output; we shall return to it in more detail later.

Price- and wage-level structures

In a similar way it may be important for many problems to break down price levels into smaller aggregates. The movement of the general price level itself may be in part a function of the structure of relative prices. An important classification of prices may be that according to the elasticity of demand for the commodity priced, for a change in price of commodities with inelastic demands may have quite different effects on the system from a change in price of commodities with elastic demands. Classification according to elasticities is particularly important in constructing supply or demand functions relating aggregates of output, sales, purchases, or consumption to price levels of these aggregates. If these aggregates consist of commodities with widely different individual demand or supply elasticities the aggregate supply or demand curve may be almost meaningless, for consistent with a given price level of the aggregate there may be a wide range of possible values of the quantity level, depending on the *composition* of the price level. Thus suppose that we have an aggregate consisting of commodity A, which has an elastic demand, and commodity B, which has an inelastic demand. A rise in the price of A coupled with a fall in the price of B may leave the

price *level* unchanged: it will, however, diminish the purchases of A very substantially and increase the purchases of B very little, because of the differing elasticities of demand, and hence the index of quantities purchased may decline even though the price level remains unchanged. Exactly similar considerations apply to supply. It is possible to draw aggregate supply and demand curves, therefore, only where the commodities aggregated have approximately the same elasticities of demand and supply. For this reason aggregate demand and supply curves which include very large groups of commodities, and even more, aggregate demand and supply curves which attempt to cover the whole of output are very unreliable and dangerous weapons of analysis.

The considerations which apply to the structure of prices also apply to the structure of wages, a wage, of course, being merely a special case of a price. The structure of relative wages may easily be a more important factor in the economy than the absolute or average level of wages as a whole. Again the most significant aggregates of labor quantities and wage levels are those of occupations in which the elasticities of demand or supply are approximately equal, and it may be important for purposes of analysis to classify different kinds of labor according to their elasticities of demand or supply. Failure to appreciate this point has led to some errors in the theory of employment. For any particular occupation, or of any group of occupations with about the same demand elasticities, the wage and the amount of employment offered are likely to be inversely related; i.e. a lower wage, demand being stable, will produce a larger volume of employment. It is not true, however, as has sometimes been supposed, that we can generalize from this proposition and assume that a general fall in money wages will increase the volume of employment. It is not merely that there may be adverse dynamic effects resulting from the adverse effect on business expectations which a *fall* in wages might produce, but even without these dynamic effects there is nothing to guarantee that a general fall in money wages, without change in the relative structure, will affect employment one way or another. The demand function for any particular kind of labor depends on the wages of other kinds of labor. Hence it is quite illegitimate to aggregate the demands of various kinds of labor, for each demand curve is drawn on the assumption that the wages of other kinds of labor do not change. It may well be that the principal wage factor affecting employment is the relative structure of wages, as, say, between high-elasticity and low-elasticity occupations, in so far as the wage factor is important at all. A rise in wages in the inelastic-demand occupations and a fall in wages in the elastic-demand occupations, even keeping the general wages level constant, might well produce a favorable effect on employment. The primary effect of the rise in the inelastic-demand occupations will

be a small fall in employment there, the primary effect in the elastic-demand occupations may be a large rise there, the sum of the primary effects will therefore be a rise. The secondary effects, through the increased income in both types of occupation, may be favorable to both.

Price-flexibility structures

From the point of view of the theory of prices and wages it may be necessary to classify commodities and occupations also according to the state of the market in which these prices or wages are determined, i.e. according to whether the market is monopolistic or competitive. Prices in competitive markets are apt to be much more changeable than prices in monopolistic markets, mainly because in competitive markets the price policy of buyers and sellers is short-run policy, i.e. is the most obvious and rapid reaction to any crisis situation, such as an inability to sell or to buy in the quantities desired. In monopolistic markets on the other hand price policy is long-run policy, mainly because there are other reactions possible to a selling or buying crisis, such as increased sales or purchase effort, cutbacks in production, and shifts in consumption. In the theory of the business cycle the differing changeability of prices is an important interpretive factor. Thus, in a period of decline in total receipts or in total income, a change in the relative structure of prices may be more important than the change in the absolute level. Thus it is a familiar fact that in a period of low income (low value of output) the decline in the total dollar *value* of the outputs of various industries is approximately the same. In some industries, however, such as agriculture, the decline in the value of output is achieved almost wholly through a fall in its price, the quantity of output remaining fairly stable. In other industries the decline in the value of output may be achieved almost entirely through a change in the output, prices being approximately stable. In other industries again the adjustment in the value of output is achieved partly by an adjustment in output, partly by an adjustment in price. The identity

$$\text{Value of output} = \text{Price of output} \times \text{Quantity of output}$$

clearly exhibits these various relationships.

It is evident that if there are going to be fluctuations in the value of output it is highly desirable that they should take the form of price fluctuations, which are merely inconvenient and lead perhaps to a certain redistribution of wealth, by no means all of which is unjust, rather than the form of output fluctuations with their consequent underemployment and general poverty. We need inquire further, therefore, into the causes of price flexibility.

Chapter 12

The Structure of Macroeconomic Aggregates: II. The Theory of Payments

Problems in the structure of aggregates are by no means confined to those aggregates which are heterogeneous. Even the homogeneous aggregates, such as payments or present money values, may have a structure, and in fact the structure of payments may perhaps be the most important feature of the whole macroeconomic system. A payment is a transfer of money from one person or account to another, usually, though not invariably, in return for some other asset. A payment therefore may be classified both according to the persons or accounts between which it passes, and also according to the type of asset for which it is made.

The payments table

The classification according to person or account may best be made by means of a "payments table" such as Table 1. We suppose here

TABLE 1

THE PAYMENTS TABLE

	A	B	C	D	Total Expenditure	Balance of Payments
A		a_b	a_c	a_d	E_a	$R_a - E_a$
B	b_a		b_c	b_d	E_b	$R_b - E_b$
C	c_a	c_b		c_d	E_c	$R_c - E_c$
D	d_a	d_b	d_c		E_d	$R_d - E_d$
Total receipts	R_a	R_b	R_c	R_d	T	$T - T = 0$

that in a closed society there are four accounts, A, B, C, and D, into and out of which payments can be made. The analysis can clearly be extended to any number of accounts, but four are sufficient for purposes

of illustration. Then we indicate the four accounts both vertically and horizontally in a chessboard diagram as in the table. In each cell of the diagram we write the payment *from* the account indicated at the left of the row in which the figure stands, *to* the account indicated at the head of the column in which it stands. Thus a_b is the payment from A to B, b_a the payment from B to A. The sum of each row is the total expenditure from the account at the row's head. Thus we have the total expediture from account A:

$$E_a = a_b + a_c + a_d \tag{1a}$$

and similarly for the other accounts:

$$E_b = b_a + b_c + b_d \tag{1b}$$

Etc.

The total receipts into each account is the sum of its *column*. Thus in Table 1 we have the total receipts of account A:

$$R_a = b_a + c_a + d_a \tag{2a}$$

similarly we have

$$R_b = a_b + c_b + d_b \tag{2b}$$

Etc.

Each figure in the cells of the table enters once as a receipt and once as an expenditure, for every payment is a receipt to the account that gets it and an expenditure to the account that pays it out. Hence the grand total of receipts is equal to the grand total of expenditures, T, each consisting of all the payment figures, but each added up in a different way.

The balance of payments of each account, or the amount of hoarding (or dishoarding if negative) performed by each account, is its receipts less its expenditures. That is,

$$H_a = R_a - E_a \tag{3a}$$

$$H_b = R_b - E_b \tag{3b}$$

Etc.

It follows that the algebraic sum of all balances of payments, i.e. of all amounts hoarded and dishoarded, is equal to zero:

$$\Sigma H_a = \Sigma R_a - \Sigma E_a = 0 \tag{4}$$

In the foregoing it has been assumed that there has been no creation or destruction of money in any of the accounts of the system. If there is creation or destruction of money the payments identities must be some-

what modified. A creation of money represents an addition to the money stock of the person or account which has created the money, which is not represented by any expenditure elsewhere in the system. Similarly the destruction of money represents a diminution of the money stock which is not represented by any receipt. If M_a is the net creation of money in account A, therefore, and H_a is the increase in A's money stock, we have:

$$H_a = M_a + R_a - E_a \tag{5a}$$

whence

$$\Sigma H_a = \Sigma M_a \tag{5b}$$

That is to say, the algebraic sum of all hoardings or balances of payments (increases in money stocks) is equal to the net amount of money created. If the amount of money destroyed exceeds the amount created, net hoarding will, of course, be negative. This may be called the "balance-of-payments identity."

The aggregation of payments

The next proposition is that the balance-of-payments identity is not affected by the aggregation of accounts. Suppose that in Table 1 the accounts of A and B on the one hand and C and D on the other are amalgamated, so that only the external payments of each group are shown, as in Table 2. For each group, as for the individual, it follows

TABLE 2

THE PAYMENTS TABLE: AMALGAMATION OF ACCOUNTS

	$A + B$	$C + D$
$A + B$		$a_c + a_d + b_c + b_d$
$C + D$	$c_a + c_b + d_a + d_b$	

that the excess of external receipts over external expenditures is equal to the net hoarding, or increase in the group's money stock, in the absence of any creation or destruction of money, or is equal to the excess of external receipts over external expenditures plus the net creation of money in the various accounts of the group. We have, from Table 1,

For A:

$$H_a = M_a + b_a + c_a + d_a - a_b - a_c - a_d$$

For B:

$$H_b = M_b + a_b + c_b + d_b - b_a - b_c - b_d$$

Summing these two identities, we have for the group $A + B$:

$$H_a + H_b = M_a + M_b + (c_a + c_b + d_a + d_b) - (a_c + a_d + b_c + b_d) \tag{6}$$

This process applies no matter how many accounts we aggregate, as in summing the balance-of-payments identities of all the individual accounts in a group the internal payments always cancel out, leaving us with the increase in the group's money stock (its hoarding) equal to its net creation of money plus its balance of payments (excess of external receipts over external expenditure. This proposition is important, as it means that the properties of the payments table apply to groups just as to individuals. It is particularly important in the theory of international balances of payments: the balance-of-payments identity applies to the group of accounts called a "nation" just as it does to any other group of accounts. In the case of international balances the problem is, of course, complicated by the fact that the various payments and balances are reckoned in different currencies, but the principle remains.

In order for the balance of payments to be in complete equilibrium it is, of course, necessary for the receipts of each individual account to be equal to its expenditures; otherwise some accounts will be gaining money at the expense of others. That is to say, we must have

$$R_a = E_a, \qquad R_b = E_b, \qquad \text{etc.} \tag{7}$$

Payments determined by partial velocities

What then, determines the various items of the payments table? Clearly we cannot expect to have a complete theory of its determination at this stage, for it is mutually determined with all the elements of the economic system. Nevertheless a simple assumption will enable us to separate out in some degree the problem of the payments structure from the other elements in economic life. Let us construct a simple model on the assumption that the expenditure in the next accounting period out of any account towards any other account is a fixed proportion of the total money stock in the account. Thus we suppose that if m_a, m_b, m_c, etc., is the money stock in the accounts A, B, C, etc., then we have:

$$\begin{aligned} &a_b = m_a v_{ab}, \qquad a_c = m_a v_{ac}, \qquad a_d = m_a v_{ad}, \qquad \text{etc.} \\ &b_a = m_b v_{ba}, \qquad b_c = m_b v_{bc}, \qquad b_d = m_b v_{bd}, \qquad \text{etc.} \\ &\text{Etc.} \end{aligned} \tag{8}$$

If the v's are given we can determine an equilibrium position of the whole system of payments consistent with the given v's, for if there are n accounts we have $n(n-1)$ equations of type (8), $(n-1)$ independent equations of type (7), and one equation (9):

$$m_a + m_b + m_c \cdots = M \tag{9}$$

where M is the total quantity of money. We assume here that no money is created or destroyed; if it is, the equations (7) must be modified to read

$$R_a = E_a + M_a, \text{ etc.}$$

where M_a is the amount of money created by account A, and M_a, M_b, etc., are given.

Although there are n equations of type (7), only $(n-1)$ of them are independent, because if the condition of equality between receipts and expenditures is fulfilled for all but one of the accounts it must also be fulfilled for the other one, the total of receipts being equal to the total of expenditures. That is, if R is the total receipts and E the total expenditures, given that

$$R_a = E_a, \qquad R_b = E_b, \cdots, \qquad R_{n-1} = E_{n-1}$$

we can deduce that

$$R_n = R - R_a - R_b - \cdots - R_{n-1} = E - E_a - E_b - \cdots - E_{n-1} = E_n$$

This gives us $\{n(n-1) + (n-1) + 1\}$, or n^2 equations in all. We also have n^2 unknowns: $n(n-1)$ payments and n money stocks of the various accounts. The v's, which we may call the *partial velocities* as they are analogous to velocities of circulation, and M we assume to be given. As all the equations are linear a unique solution is possible.

It will be instructive to examine the working of such a system in some simple cases. Let us take first a system with only two accounts, A and B. The payments table will be like Table 3.

TABLE 3

	A	B
A		a_b
B	b_a	

$R_a = b_a$, $E_a = a_b$, so that in equilibrium we have

$$a_b = b_a \tag{10}$$

If v_{ab} and v_{ba} are the corresponding partial velocities, and m_a and m_b are the equilibrium money stocks of the two accounts,

$$a_b = m_a v_{ab} \tag{11a}$$

$$b_a = m_b v_{ba} \tag{11b}$$

We also have

$$m_a + m_b = M, \text{ the total money stock of the society} \tag{12}$$

With two accounts we have $2^2 = 4$ equations (10, 11*a*, 11*b*, and 12), which can easily be solved to give:

$$m_a = \frac{Mv_{ba}}{v_{ab} + v_{ba}} \qquad m_b = \frac{Mv_{ab}}{v_{ab} + v_{ba}} \tag{13}$$

$$a_b = b_a = \frac{Mv_{ab}v_{ba}}{v_{ab} + v_{ba}} \tag{14}$$

The total of receipts or expenditures in such a system, T, is given by

$$T = a_b + b_a = M\frac{2v_{ab}v_{ba}}{v_{ab} + v_{ba}} = MV \tag{15}$$

V is the transactions velocity of circulation of money in the whole system, i.e. the ratio of the total volume of payments to the total stock of money. We observe that V is the harmonic mean of the partial velocities, v_{ab} and v_{ba}. We can also express the same set of facts in terms of the reciprocal of the velocities, k_{ab} and k_{ba}, which in this case represent the *period of money turnover* in the two accounts, i.e. the fraction of the accounting period in which receipts could replenish the money stock were there no expenditures, or expenditures deplete it were there no receipts. By definition

$$k_{ab} = \frac{1}{v_{ab}} \quad \text{and} \quad k_{ba} = \frac{1}{v_{ba}}$$

If k is the average period of money turnover for the whole society, k is the arithmetic average of the individual periods, i.e.

$$k = \tfrac{1}{2}(k_{ab} + k_{ba})$$

The significance of the velocities of circulation (or their reciprocals, the periods of money turnover) is often overlooked, and it will be worth while to illustrate the concept further. The *individual's* velocity of circulation, e.g. v_a, is the sum of all his partial expenditure velocities. Thus,

$$v_a = v_{ab} + v_{ac} + \cdots + v_{an}$$

The reciprocal of the individual velocity,

$$k_a = \frac{1}{v_a}$$

or the period of money turnover, is a measure of the individual's "de-

mand for money" and has sometimes been given this name—somewhat inappropriately, as a single parameter should not properly be called a "demand," this term being reserved usually for a functional relationship. The period of money turnover may quite properly be called a measure of the "*preference* for money," especially for what may be called "payments balances," though we should beware of attempting to identify any *particular* part of an individual's money stock as his payments balance. An individual who wishes to hold a money stock equal to six months' receipts clearly has in some sense a higher preference for money than one who is content with a money stock equal to three months' receipts. The individual's velocity of circulation is, of course, inversely related to his money preference: an individual who holds a money stock equal to six months' receipts has a velocity of circulation of twice a year; one who holds three months' receipts has a velocity of circulation of four times a year. If receipts and expenditures are not equal, i.e. if the individual is accumulating or decumulating money, it may be necessary to distinguish between the expenditures velocity and the receipts velocity, one being the ratio of the total receipts, the other of total expenditure, to average money stock. In balance-of-payments equilibrium, of course, the two concepts become equal.

The individual's velocity of circulation, or period of money turnover, is an index of his state of mind in regard to money. It may change with changing circumstances. Generally speaking, an increase in uncertainty, leading to a desire for greater flexibility in assets, expectations of falling prices leading to a preference for money as opposed to other assets, and the expectation of greater irregularity in payments, will all lead to an increase in the individual's money preference, which is expressed as an increase in his period of money turnover or, what is the same thing, a decline in his individual velocity of circulation. The average period of money turnover, or average (payments) velocity of circulation, for a whole society is a weighted average of the individual periods of turnover or velocities of circulation of all its constituent members. As such it is an index of the state of mind of the whole society—or, to be more accurate, as society does not have a mind to have a state of—a summary or aggregate of the states of mind of all the various individuals in the society; any change in the attitudes of individuals towards money will be reflected in a change in the average velocity V or its reciprocal K.

The paradox of payments

Here we run into another of the macroeconomic paradoxes. An individual adjusts his money preference by changing his money stocks, keeping his total of expenditures and receipts approximately constant.

If there is no creation or destruction of money in the society, however, a general change in money preference can take effect only through its impact on the total of payments. If the total money stock of a society is constant, an increase in money preference, reflected in declining V or increasing K, will diminish the total of payments until it bears the appropriate proportion to the constant money stock. The *individual* seeks to increase his individual k's by diminishing his expenditure and so increasing his money stock by having a positive balance of payments. The result of his action, however, is to diminish the receipts of those to whom he makes payments and to create unfavorable balances of payments among them: the money that he gets into his pocket must come from the pockets of others, if the total in all pockets is constant. If these individuals who have suffered a decline in money balances accept this fact, thus in effect increasing their own individual velocities of circulation to offset the decline in velocity on the part of the first individual, there need be no change in total payments: the result of the shift is simply a transfer of money from one account to another. In terms of our simple two-account society, if individual A increases his preference for money he will seek to expand his money stock by cutting down his expenditure, a_b. His expenditure, however, is B's receipts, so as a result of his cut in expenditure B finds his receipts cut. The favorable balance of payments which A experiences creates an equal, unfavorable balance for B: A increases his money stock; B finds his diminished by an equal amount. If now B diminishes his preference for money, so that he raises v_{ba} by as much as A has diminished v_{ab}, there will be no change in total payments; the effect of the shift in preference is simply to redistribute the money stock between A and B, A taking more, B less, as shown by equations (13). If, however, B is not willing to suffer a decline in his money preference, and wishes to hold k_{ba} or v_{ba} constant, the result of A's action is a decline in the total volume of payments. B will cut down his expenditure in response to the fall in his receipts, which will give A a smaller volume of receipts than before; A may cut his expenditure again in his endeavor to accumulate money; B will suffer another loss of receipts, and will cut his expenditure, which will further lower A's receipts. This circular process will go on until the equations (13) and (14) are once more satisfied. If v_{ab} has declined without a change in v_{ba}, the total volume of payments T will decline, and there will still be some redistribution of the money stock away from B and towards A until B finds that the ratio of his diminished money stock to his diminished receipts is what it was before.

In general, the average payments velocity of circulation of a whole society is the weighted *arithmetic* average of the individual velocities of

circulation, weighted by the money stock of each individual. Thus, if v_a, v_b, etc., are the total individual velocities of the individuals A, B, etc., we have by definition, if R_a, R_b, etc., are their total receipts (assumed equal to expenditures),

$$R_a = E_a = m_a v_a, \qquad R_b = E_b = m_b v_b, \qquad \text{etc.} \tag{16}$$

Then, if M is the total stock of money in the society, V its average velocity, we have the total receipts of the society,

$$R = MV = \Sigma m_a v_a \tag{17}$$

whence,

$$V = \frac{\Sigma m_a v_a}{\Sigma m_a} \tag{18}$$

In the special case of the two-account society we can substitute the values of m_a and m_b from equations (13) and obtain, as in (15),

$$V = \frac{2 v_a v_b}{v_a + v_b}$$

The significance of the partial velocities, v_{ab}, v_{ac}, etc., should also be noted. These are parameters which express the demand, or more accurately, the preference of one individual for the product or asset of another, assuming that payments are made in return for assets of some kind received. Thus, if individual B sells shoes in return for the payments which he receives, the parameter v_{ab} is in some sense a measure of A's demand or preference for B's shoes, in the sense at least that if A's preference for B's shoes increases, v_{ab} is also likely to increase; i.e. if the amount of money which A possesses does not change, A will spend more for shoes than before. It will be observed, of course, that this parameter does not take any account of the *price* of shoes and hence is not, of course, a perfectly satisfactory expression of demand. If there is a rise in the price of shoes then A may increase his expenditures for shoes, i.e. a_b may increase or decrease, and consequently v_{ab} will increase or decrease, as we are not assuming any change in A's money stock, depending on whether his demand for shoes has an elasticity less than or greater than one. However, at this stage of the analysis we are neglecting price changes, and considering the partial velocities as independent parameters expressive of the preferences of individuals. The functional variability of the partial velocities can be introduced at a later stage of analysis.

Dynamics of the payments system

The dynamics of payments systems presents some interesting and important problems, for it is in the study of the movements of a payments

system that we shall find an important clue to the problem of the business cycle. From a theoretical viewpoint, unfortunately, we have an embarrassment of possible riches. There are a great many possible dynamic systems which may be constructed, all of which may have some significance for reality but none of which individually is very realistic. Consequently the elaboration of mathematical systems can in this case easily turn out to be a waste of time, for any mathematical dynamic system must be based ultimately on the assumption of certain patterns of stability in the "reaction functions," i.e. in the way in which the present situation of an individual affects his behavior in the next "accounting period." It is necessary to assume some kind of stable reaction functions in order to be able to derive the systems of difference equations, i.e. equations relating successive values of a given variable, which are the foundations of any dynamic system. The difficulty in this case is not merely that there seems to be no *a priori* ground for selection among a number of possible assumptions in regard to reaction functions, but that there is no guarantee that the reaction functions themselves in anything like a simple form are stable.

It may, however, be illuminating to consider some relatively simple dynamic systems and to trace out the consequence of certain reaction patterns, as almost all these patterns *may* be found in practice.

Perhaps the simplest assumption is that the partial velocities remain stable in each accounting period, the accounting period being assumed to be the "period of reaction." Any change in the money stocks held by an individual at the beginning of a period produces a proportionate change in all his expenditures. As the mathematical analysis of the payments table is quite complex we shall content ourselves for the purposes of this chapter with arithmetical illustrations. We shall use a three-account payments table, as most of the important properties of larger tables are already found in this. In Table 4*a*, then, we show a payments

TABLE 4*a*

	A	*B*	*C*	*E*	*M*	*dM*
A		15	30	45	300	0
B	20		15	35	250	0
C	25	20		45	500	0
R	45	35	45	$T = 125$	1050	0

table of three accounts, A, B, and C, which is in balance-of-payments equilibrium. Column E shows the total expenditure, M the money stock (assumed) of each individual, dM the change in the money stock

into the next period. The partial velocities are $v_{ab} = 0.05$, $v_{ac} = 0.1$, $v_{ba} = 0.08$, $v_{bc} = 0.06$, $v_{ca} = 0.05$, $v_{ab} = 0.04$. These are assumed to be constant, so that in each period A spends 5% of his money stock towards B and 10% of his money stock towards C, and so on for the others.

Now we suppose that \$300 of new money is given to A, and we wish to trace out the effects in the system. These effects are shown in Tables 4*b*, *c*, and *d*. The tables are worked out to the nearest digit. The process

TABLE 4*b*

	A	*B*	*C*	*E*	*M*	*dM*
A		30	60	90	600	−45
B	20		15	35	250	+15
C	25	20		45	500	+30
R	45	50	75	*T* = 170	1350	0

TABLE 4*c*

	A	*B*	*C*	*E*	*M*	*dM*
A		28	55	83	555	−36
B	21		16	37	265	12
C	26	21		47	530	24
R	47	49	71	*T* = 167	1350	0

TABLE 4*d*

	A	*B*	*C*	*E*	*M*	*dM*
A		26	52	78	519	−29
B	21		16	37	277	11
C	28	22		50	554	18
R	49	48	68	*T* = 165	1350	0

of working is as follows. The payments in each table are found by multiplying the money stocks M of each individual by the proper partial velocities. Thus, in Table 4*b*, $a_b = v_{ab}m_a = (0.05)(600) = 30$. The dM column in each table is obtained by subtracting the E from the corresponding R: thus, in Table 4*b*, $dM_a = R_a - E_a = 45 - 90 = -45$. Then the dM figure is added (or subtracted if negative) from the M figure in the same row to get the M of the next table. Thus dM_a in Table 4*c* is equal to $M_a + dM_a$ of Table 4*b* ($= 600 - 45 = 555$). Then from the M's of Table 4*d* the payments, the R's, E's, and dM's, are calculated as before.

It will be observed that in this case the new money is gradually distributed to the other members of the system, and A's money stock and expenditures fall while B's and C's rise. The total of payments rises sharply at first, but as the money is distributed to B and C the total falls. This is not a general property, but depends on the fact that in the example B and C have lower individual velocities (0.14 and 0.09, respectively) than A (0.15). The total velocity being a weighted average, weighted by the money stocks, as A's weight declines and B's and C's weights increase, the general velocity will fall, and as the total of the money stock is constant after the initial increase the total of payments will fall with it. If the table had been constructed giving B and C higher individual velocities than A the total volume of payments would first have fallen as A had the larger money stock, and then would have risen again as the money was redistributed. If all the individual velocities are equal the payments total after the initial shock will be constant, as the redistribution of the money stock does not then affect the general velocity, V. In the present example the system approaches the equilibrium given by Table 4e (approximately, to the nearest digit) in a steady

TABLE 4e

	A	B	C	E	M	dM
A		19	39	58	386	0
B	26		19	45	321	0
C	32	26		58	643	0
R	58	45	58	$T = 161$	1350	0

but diminishing progression. The general velocity of circulation (0.119) is the same in both equilibrium positions, in Table 4a (125/1050) as in 4e (161/1350). The reason is that in the solution of equations (7), (8), and (9) the equilibrium money stocks of the various individuals, m_a, m_b, etc., appear as given *proportions* of the total money stock M, the proportions, of course, depending on the v's. As in this case the v's have not changed, the proportions of the money stocks have not changed, and hence the relative weights of the weighted average of the individual velocities, which is the general velocity, has not changed. This means that the final result of an increase in the quantity of money, *no matter to whom the money is given*, is an increase in the total volume of payments equal to the increase in money multiplied by the velocity of circulation, provided that the particular velocities do not change. That is to say, if the velocity of circulation is seven times a year, each increase of $100 in the money stock will eventually increase the total volume of payments by $700. When we say that the velocity of circulation is seven times a year

we mean in effect that on the average every dollar creates $7 of payments during the year.[1]

The "surge" phenomenon

It cannot always be assumed, however, that the progress towards the equilibrium position is as smooth as in Tables 4*a*–4*d*. The smoothness of the progress depends on the *v*'s being small. If they are of the order of magnitude of 1, i.e. if the reaction period, or accounting period, is about equal to the period of turnover of money, violent fluctuations both in balances of payments and even in payment totals may occur in the attempt to reach an equilibrium. This is illustrated in Table 5, where the same relative values of the partial velocities are given as in Table 4, but the absolute values are ten times as great. The individuals in each reaction period spend an amount about equal to or exceeding their total money stock. We find now that the money stock, instead of gradually disseminating itself through the system as in Table 4, surges violently from *A*, as *A* disgorges *more* than the addition to its stock, towards *B* and *C*, from whom it surges with equal or even greater violence back towards *A*. In the present example these surges actually increase in intensity until *A* is drained of money completely.

The assumptions which give rise to fluctuations such as in Table 5 must be made clear, with a view to assessing the likelihood of this alarm-

TABLE 5*a*

	A	*B*	*C*	*E*	*M*	*dM*
A		15	30	45	30	30
B	20		15	35	25	0
C	25	20		45	50	0
R	45	35	45	*T* = 125	105	30

TABLE 5*b*

	A	*B*	*C*	*E*	*M*	*dM*
A		30	60	90	60	−45
B	20		15	35	25	15
C	25	20		45	50	30
R	45	50	75	*T* = 170	135	0

[1] We have the total volume of payments, $E = MV$. It follows that

$$\Delta E = M\,\Delta V + V\,\Delta M + \Delta V\,\Delta M$$

If V is unaffected by the change in M, as will be the case where the particular velocities are unchanged, then $\Delta V = 0$, and

$$\frac{\Delta E}{\Delta M} = V$$

TABLE 5c

	A	B	C	E	M	dM
A		7	15	22	15	50
B	32		24	56	40	−17
C	40	32		72	80	−33
R	72	39	39	T = 150	135	0

TABLE 5d

	A	B	C	E	M	dM
A		32	65	97	65	−55
B	18		14	32	23	19
C	24	19		43	47	36
R	42	51	79	T = 172	135	0

TABLE 5e

	A	B	C	E	M	dM
A		5	10	15	10	60
B	33		25	58	42	−20
C	42	33		75	83	−40
R	75	38	35	T = 148	135	0

TABLE 5f

	A	B	C	E	M	dM
A		35	70	105	70	−66
B	18		13	31	22	21
C	21	17		38	43	45
R	39	52	83	T = 174	135	0

TABLE 5g

	A	B	C	E	M	dM
A		2	4	6	4	72
B	34		26	60	43	−23
C	44	35		79	88	−49
R	78	37	30	T = 145	135	0

ing situation arising. It is assumed first that decisions as to expenditure are made at the beginning of each planning or accounting period, and that these decisions are carried through without regard to what happens during the planning period. Then at the end of the period the results are assessed and new plans are made for the next period. For results such as in Table 5 it is necessary to assume that the planning period is long enough

so that individual stocks of money can be almost entirely depleted before any change in policy is made. Clearly this is not often likely to be the situation with individuals. In the case of international balances, however, the liquidity position of the members of a nation in regard to purchasing power over foreign assets is apt to be hidden from them by the existence of a national monetary and banking system. In spite of the fact, therefore, that the payments system is a perfectly general system covering all individuals and any arbitrary group we like to construct, the fact of national sovereignty and independent national monetary systems may create a real problem of adjustment in international balances which does not exist to anything like the same extent with domestic balances.

The mathematical treatment of the dynamics of the payments table is complicated severely by the fact that although a "planning period" unquestionably exists, so that we cannot assume that the adjustments to changes in balances are simultaneous, the planning period itself is likely to differ from individual to individual and, what is worse, does not even remain constant with a single individual. The period of adjustment is likely to be a function of the need for adjustment. Rather than assuming a fixed planning period, it would probably be more realistic to assume that there is a certain "range of inertia" within which it is not worth troubling to make adjustments, but that when the money stock, or the partial velocities, or whatever other index we wish to take as signifying a need for adjustment transgresses the limits of this range, then adjustments are made, no matter how long or how short the time period involved.. In such a case whether the "surge" phenomenon is observed or not depends on the extent of the range of inertia: the smaller this range the less likely is the system to set up "surges" of money back and forth among the various accounts. If, however, adjustments are not made until they have to be drastic to deal with the situation then we may easily get "surges" which will create cyclic movements in balance-of-payments structures or even in the payments total.

Systematic variation in velocities

A further complication in the adjustment pattern is that movements in payments totals are likely to set up changes in velocities; i.e. the velocities themselves are functions of changes in the payments total. The velocities are in part an expression of the willingness to hold real assets as against money: if there is a decline in liquidity preference, with a given money stock people will be more willing to make expenditures (i.e. get rid of money for other assets), and the partial velocities will rise. If this rise is general there need be no great disturbance of the balance-of-pay-

ments structure: the money stocks are a "widow's cruse" which cannot be consumed, and the more people spend, the more they find—perhaps to their astonishment—that they have to spend. This is shown in Table 6, where, starting even from the dangerous and unstable situation of Table 5*a*, we see that a general rise in velocities will raise both receipts and expenditures without disturbing the balance-of-payments structure.

TABLE 6*a*

	A	*B*	*C*	*E*	*M*	*dM*
A		15	30	45	30	0
B	20		15	35	25	0
C	25	20		45	50	0
R	45	35	45	$T = 125$	105	0

TABLE 6*b*

	A	*B*	*C*	*E*	*M*	*dM*
A		20	35	55	30	0
B	25		20	45	25	0
C	30	25		55	50	0
R	55	45	55	$T = 155$	105	0

It is interesting to observe that one general change in velocities which will leave the balance-of-payments equilibrium unchanged is an equal absolute change in all particular expenditures, as in Table 6*b*. Other changes are also possible, of course, which will leave the balance-of-payments equilibrium undisturbed, such as an equal *proportional* change in the velocities. This is shown in Table 6*c*, where the velocities are assumed to have doubled as compared with 6*a*.[2]

TABLE 6*c*

	A	*B*	*C*	*E*	*M*	*dM*
A		30	60	90	30	0
B	40		30	70	25	0
C	50	40		90	50	0
R	90	70	90	$T = 250$	105	0

[2] If v_{ab}', v_{ac}', etc., represent the change in the partial velocities, then the balance-of-payments equilibrium will remain undisturbed as long as:

$$m_a(v_{ab}' + v_{ac}') = m_b v_{ba}' + m_c v_{ca}'$$

$$m_b(v_{ba}' + v_{bc}') = m_a v_{ab}' + m_c v_{cb}'$$

$$m_c(v_{ca}' + v_{cb}') = m_a v_{ac}' + m_b v_{bc}')$$

If now payments have been rising, and because of that are expected to go on rising, prices are also likely to rise and may be expected to go on rising. Under these circumstances there will very likely be a rise in velocities, as people will be anxious to get rid of their money, which is losing purchasing power, and will wish to hold more of their assets in the form of goods. This rise in velocities, however, will itself create a further rise in the payments total, which may in turn raise prices still further. It is easy under these circumstances to set up an "expectational cycle" in the total of payments. We have only to assume that the change in velocities with rising payments loses force as the payments total rises, so that when the payments total is "high" the rise in velocities ceases or even reverses itself to set the opposite force in motion—a fall in payments, leading to a fall in velocities, which leads to a further fall in payments, and so on until payments are so low that velocities cannot fall any further and start to rise again. There remains a problem, however, of how such expectional cycles get started, and why they are not damped off sharply, as one would normally expect them to be.

The principle of asymmetrical reaction

A partial answer to this question may perhaps be found in another interesting possible property of the payments system: the fact that reactions to an increase and to a decrease in an individual's money stock may not be symmetrical. This may be particularly true where the reactions are to unexpected changes in the money stock. If, for instance, an individual finds that he has an unexpectedly *favorable* balance of payments, with a consequent increase in his money stock (due perhaps to a sudden rise in demand for his product), will his reaction be the same as that of an individual who is confronted with an unexpectedly *unfavorable* balance of payments with a consequent decline in his money stock? It is true, of course, that a favorable balance is likely to create an increase in an individual's expenditures, just as an unfavorable balance is likely to create a decrease. When there is a positive balance with expansion of money stocks, however, there is nothing particularly urgent about this pressure, and an unexpected positive balance of payments may not produce much in the way of an expansion of expenditure for quite some time. An unexpected negative balance, however, may produce a much sharper pressure on the individual. It is one thing to lie back while one's money stock is accumulating and wonder for a while how one is going to spend it without actually changing the established patterns of expenditure. It is quite another thing to suffer a sharp negative balance of payments with its consequent decline in one's money stock: there are sharply defined limits on the downward side in regard to the money

stock with which one feels comfortable; there are no such sharp limits on the upward side. It may well be, then, that the dynamic reactions to positive balances of payments are weaker than the reactions to negative balances. *Where this is the case any pronounced disturbance of the payments structure is likely to have a net deflationary effect on the payments total.*

A disturbance of the payments structure means that some accounts are accumulating money (have a positive balance) while other accounts have a negative balance and are decumulating money. If the reactions to these events are symmetrical there need be no change in the payments total: those accounts which are gaining money will increase their expenditures and those which are losing money will decrease their expenditures, but the total of receipts will not change if the increase in expenditures in some accounts just equals the decrease in expenditures in others. If, however, the accounts which are gaining money do not increase their expenditures, or only increase their expenditures a little, whereas the accounts that are losing money cut their expenditures sharply, the net result will be a decline in the total of expenditures and therefore, as they are the same thing, of total receipts.

The generalized algebraic or analytic description of these processes is so complex and would involve so many arbitrary assumptions that we may again best illustrate the types of process involved by means of an arithmetical example. Table 7*a* shows a three-account society in a con-

TABLE 7*a*

	A	*B*	*C*	*E*	*M*	*dM*
A		11	6	17	10	0
B	3		10	13	10	0
C	14	2		16	10	0
R	17	13	16	$T = 46$	30	0

dition of payments equilibrium. We suppose, to begin with, that each account has a money stock of 10 units. Column *dM* shows the change in the money stock of each ($= R - E$); *M* shows the money stock itself. As the system is in equilibrium, receipts equal expenditure and *dM* is zero for all three accounts. Now let us suppose that these three accounts represent three industries, and that there is a sharp decline in the purchases of the product of industry *A* by the other accounts, without any corresponding expansion in their other expenditures. In the next accounting period the situation will be as in Table 7*b*.

We have supposed simply that *B* and *C* have reduced their expenditures towards *A* from 3 to 1 and from 14 to 9, without any other changes. This might happen, for instance, if *A* produced a durable product the

total stock of which had been expanding for a while and then suddenly ceased to expand. The automobile industry in 1929 was a case in point. The result is to create a profound disturbance in the balance-of-payments structure. A finds himself with a negative balance of 7, reducing his

TABLE 7b

	A	B	C	E	M	dM
A		11	6	17	3	−7
B	1		10	11	12	+2
C	9	2		11	15	+5
R	10	13	16	$T = 39$	30	0

money stock to 3: B and C find themselves with positive balances of 2 and 5, increasing their money stocks to 12 and 15. The total money stock we assume constant; the disequilibrium in the balance of payments results in its redistribution. As a result of the decline in expeditures we find the total volume of payments has shrunk from 46 to 39.

Such a situation cannot, of course, endure: if it continued for another accounting period A would not only be completely denuded of money but would run into debt to the amount of 4 units. The situation clearly calls for drastic action on the part of A; it does not necessarily call with the same immediacy for action on the part of B and C, who may be simply content to allow their money stocks to increase. In Table 7c, then,

TABLE 7c

	A	B	C	E	M	dM
A		6	4	10	3	0
B	1		10	11	9	−3
C	9	2		11	18	+3
R	10	8	14	$T = 32$	30	0

we see that A has adjusted his situation by making drastic cuts in his expenditures, thus "balancing his budget" and stopping the decline in his money stock.

This reduces the total payments still further to 32. A has balanced his budget, but now we find that B is dishoarding while C is hoarding still further. Meanwhile A's money stocks are still very low, and C has accumulated most of the money in the system! In the next period, then, suppose that A tries to increase his money stock by limiting his expenditure still further, to 8 units, and that B, finding that he has a dangerously negative balance of payments, also reduces his expenditure from

11 to 10, but that C, rejoicing in his swollen bank balances, expands his expenditure from 11 to 14. The expansion in C's expenditure now offsets the decline in A's and B's, and the total volume of payments remains constant at 32. Some of C's swollen balances return to A; B, however, is still losing cash. This is shown in Table 7*d*.

TABLE 7*d*

	A	B	C	E	M	dM
A		5	3	8	6	+3
B	1		9	10	8	−1
C	10	4		14	16	−2
R	11	9	12	$T = 32$	30	0

A and B now find their money stocks bearing about the same proportion to their payments as before the disturbance. C, however, finds his money stock proportionally much higher than it used to be. It may well be, then, that C will eventually adjust to this situation by diminishing his money stock, which will involve his expanding his expenditures. A, being about satisfied with his money stock, will expand his expenditures to cut down his positive balance of payments. The next period then may see something like Table 7*e*. We observe that the total of payments is

TABLE 7*e*

	A	B	C	E	M	dM
A		6	4	10	9	+3
B	1		9	10	10	+2
C	12	6		18	11	−5
R	13	12	13	$T = 38$	30	0

rising again: it is now 38. A, in spite of expanding his expenditures, still finds himself with a favorable balance, because the expansion of C's expenditures yields him an increase in his receipts. C may now feel that he has about the right amount of money, so he cuts down his expenditures again to check his negative balance. B and A expend a little more to check their positive balances. So we get a payments equilibrium established as in Table 7*f*.

It will be observed that the balance-of-payments equilibrium has been restored at a level of payments below the initial level (46) of Table 7*a*. This is not, of course, a necessary result, and we can easily devise examples in which it is not so. For the condition of Table 7*f* to be really stable we should have to assume that the individual velocities of circula-

tion of the individual marketers had changed. In Table 7*a* these velocities are: *A*, $\frac{17}{10}$, or 1.7; *B*, $\frac{13}{10}$, or 1.3; *C*, $\frac{16}{10}$, or 1.6. In Table 7*b* the corresponding velocities are *A*, $\frac{12}{9}$, or 1.33; *B*, $\frac{12}{10}$, or 1.2; *C*, $\frac{15}{11}$, or 1.36. The average velocity in Table 7*a* is $\frac{46}{30}$, or 1.53. In Table 7*f* it is $\frac{39}{30}$, or 1.3. Table 7*f* therefore cannot represent a true equilibrium if the "normal" average velocity of circulation has not changed, i.e. if the dynamic changes which have been reflected in the tables are regarded as "temporary." It is, of course, quite possible for "normal" velocities to change; there seems to be some evidence, for instance, that the general level of velocities has exhibited a secular decline over the past two or three decades. It is difficult, however, to judge such a movement from such a short experience, as to some extent the low velocities may merely reflect the consequences of a highly disturbed dynamic situation.

TABLE 7*f*

	A	*B*	*C*	*E*	*M*	*dM*
A		7	5	12	9	0
B	2		10	12	10	0
C	10	5		15	11	0
R	12	12	15	$T = 39$	30	0

The payments multiplier

We could go on constructing possible models of a payments system almost *ad infinitum:* enough has been said, however, to indicate its general properties and use in analysis. It remains only to comment on a concept which has sometimes been applied to the payments system with not altogether happy results—the multiplier. The general concept of the multiplier as applied to payments is the ratio of the total expansion in payments which results from a single increase in expenditures on the part of one account. The increased expenditure appears as increased receipts in the accounts of its recipients; in consequence they increase their expenditures, which again appear as receipts in other accounts, and so the process goes on until the whole volume of expenditure and receipts is expanded by a multiple of the original increase in expenditure. The same process applies as a "divider," of course, in the case of a single decrease in expenditure. The assumption is usually made that the whole increase in payments takes place within the accounting period, and that every increase in expenditure is a definite proportion k (assumed to be less than 1) of an increase in receipts. Thus, if k is 0.8 an increase in receipts of \$10 results in an increase in expenditure of \$8. Then, if e is the initial increase in expenditure, other accounts find receipts increased by e and

of this increase spend ke: this again finds itself as receipts and $k \cdot ke$, or k^2e, is spent. So the process goes on, the final increase in expenditure, me, being given by

$$me = e + ke + k^2e + k^3e + \cdots \text{ to infinity}$$

$$= e\left(\frac{1}{1-k}\right)$$

$m\left(= \frac{1}{1-k}\right)$ is the multiplier. Following the Keynesian terminology, k might be called the "marginal propensity to spend." It must be distinguished from the concept of the marginal propensity to consume—a name by which it is sometimes miscalled—a concept which is of much more significance and which will be developed later.

The weakness of the multiplier or propensity-to-spend concept as applied to the payments system is that it must assume that payments are divided into two categories, one (which we may call spontaneous expenditures) which are independent of the volume of receipts, and the other (which we may call induced expenditures) which depend on receipts.[3] The multiplier approach is also quite incapable of analyzing the "network" character of payments, the structure of velocities, or the effects of redistribution of the money stock.

It also confuses two quite different sets of phenomena: the increase (or decrease) in the payments total which is a result of an increase (or decrease) in the total money stock, and the change in the payments total which may be a result of a change in partial or individual velocities. In the case of an increase in the money stock, as we have seen (p. 218), if the general velocity of circulation is itself unaffected by changes in the money stock there is a "multiplier" which is the velocity of circulation itself. The velocity of circulation, that is to say, is the ratio of pay-

[3] Let total expenditures, E, consist of spontaneous expenditures, E_0, and induced expenditures, E_r. Then:

$$E = E_0 + E_r$$

$$\frac{dE}{dE_0} = 1 + \frac{dE_r}{dE_0} = 1 + \frac{dE_r}{dE} \cdot \frac{dE}{dE_0}$$

$\frac{dE}{dE_0} = m$, the multiplier: it is the change in total payments per unit change in spontaneous expenditures. $\frac{dE_r}{dE} = k$, the marginal propensity to spend: it is the proportion of any increase in total payments (receipts) which results in induced expenditure. We have, therefore, as above, $m = \frac{1}{1-k}$.

ments to money stock, and therefore, if this velocity is constant, is also the ratio of an increase in payments to the increase in the money stock which gave rise to it. If the velocity is not constant but is a function of the money stock, as in short periods it may be, then the "multiplier" is not the "average" but the "marginal" velocity.[4]

A change in the payments total may also take place because of a change in velocities. The effects of this change are more complex, as it almost always involves a redistribution of the money stock among the accounts of the payments system, and hence changes the *weights* of the partial or individual velocities which have not changed. The results of a change in a single partial velocity may be a rather complex change in the general velocity, V. Nevertheless as the total volume of payments is a function only of the total quantity of money and the partial velocities—the equilibrium distribution of the money stock being determined by the partial velocities themselves—we can postulate a concept of a "partial-velocity multiplier,"—i.e. the change in the total volume of payments which results from a unit change in a partial velocity. This in turn may be analyzed into the change in the general velocity brought about by a unit change in the partial velocity, multiplied by the total money stock, as the change in the total volume of payments is equal to the change in the general velocity multiplied by the money stock, assuming that the money stock itself is not a function of the velocity.[5]

[4] We have $E = MV$; therefore

$$\text{The marginal velocity} = \frac{dE}{dM} = M\frac{dV}{dM} + V$$

[5] As before, $E = MV$, and

$$\frac{dE}{dv_{hk}} = \frac{dV}{dv_{hk}} \cdot \frac{dE}{dV} = \frac{dV}{dv_{hk}}\left(M + V\frac{dM}{dV}\right)$$

If M is constant, of course, $dM/dV = 0$. With a flexible banking system it is quite possible for dM/dV to be positive: i.e. a rise in velocities, inducing a rise in payments, may encourage the expansion of bank loans and deposits. Where this is the case, of course, the "velocity multiplier" is all the more increased.

Chapter 13

The Theory of Output and Employment

A given total of payments is compatible with any volume of physical output and employment: a theory of payments, therefore, is not a theory of income or of employment. The next step, consequently, is to develop a theory of employment, income, and output in general which can be related to the theory of payments.

A simplified Keynesian model

We will develop first a very simple model based on the Keynesian system, though actually it is even simpler than the Keynesian model. Perhaps the most fundamental identity of macroeconomics is "the accumulation identity" (see p. 190) between accumulation on the one hand and the excess of production over consumption on the other. For any single item of capital, K, or for that matter of anything whatever, production (p_K) is the gross addition to the total stock, consumption (c_K) is the gross subtraction from the total stock, and so accumulation (a_K)—the net addition to the total stock—must be equal to production minus consumption. That is,

$$a_K = p_K - c_K \tag{1}$$

If now we construct indices of accumulation, production, and consumption, using the same valuation weights, w_K, in each case, the identity (1) is also true of the indices of physical production (P) (output), consumption (C), and accumulation (A), for

$$A = \Sigma a_K w_K = \Sigma p_K w_K - \Sigma c_K w_K = P - C \tag{2}$$

We can also put the identity the other way round:

$$P = A + C \tag{3}$$

In this form the identity is, of course, strictly *ex post*. That is to say, looking back over any given period, we can say with certainty that the total output was in fact equal to the total accumulation plus the total

consumption, simply because any output of a period which has not been consumed during the period must be in existence somewhere, i.e. must have been accumulated, during the period. This identity is based fundamentally on the proposition that any stock of definable items is a population: the growth in the number of items in any period (accumulation) must be equal to the number of "births"—additions to the defined population, or production of items—minus the number of "deaths"—subtractions from the defined population, or consumption of items.[1] An item is "born" when it begins to conform to the definition which defines the population, and "dies" when it ceases to conform to this definition. The change which carries any physical aggregation of matter over the boundary line of a population definition may, of course, be quite small. The birth of a baby does not represent a very drastic change in its physical substance, yet it represents a transition from one population (unborn foetuses) to another (human beings); this is a "death" to one population and a "birth" to another. Thus every birth into one population is a death out of another: this is the fundamental principle of conservation. In economics, however, we are considering only part of the universe—that part which has value, even more narrowly perhaps, that part especially which has monetary value. The principle of conservation does not, therefore, apply to capital, which can grow or decline at the expense of, or to the increase of, the rest of the universe.

If now we ask why anyone indulges in the activity of production, i.e. adding to the stock of goods (valuable things), the answer must be, in order to replace what has been consumed in whole or in part, and to increase the existing stock. The simplest assumption that we can make in regard to production (as we have seen in Chapter 2) is the *principle of homeostasis*—that what people wish to do is to preserve their existing asset structure. On this assumption any consumption will immediately call forth enough, and just enough, production to replace it. Consumption in this case is the sole and sufficient cause of production, and if we know what determines consumption that is all the production theory we need. This assumption is, of course, too simple to be of much use. However, we can make it a good deal more useful by assuming that people plan to accumulate or decumulate certain quantities of assets of different kinds in a given "planning period." If we know what is the planned or desired accumulation of any asset, $\bar{a}$, and we know what is the expected consumption, $\bar{c}$, then the planned production must be equal to the desired accumulation plus the expected consumption:

$$\bar{p} = \bar{a} + \bar{c} \tag{4}$$

[1] This is the "bathtub theorem" again. See p. 190.

All that is now needed to complete a simple theory of production is the assumption that planned accumulation and expected consumption are themselves functions of realized production of the period immediately past, and the further assumption that planned production is realized in the next period. Given these conditions, we can immediately derive a difference equation in production. If p_0, p_1 are the production in successive planning periods, and if $\bar{c}_1 = C(p_0)$ and $\bar{a}_1 = A(p_0)$, then

$$p_1 = \bar{a}_1 + \bar{c}_1 = C(p_0) + A(p_0) \tag{5}$$

If these functions are *stable* we have

$$p_2 = C(p_1) + A(p_1)$$

$$p_3 = C(p_2) + A(p_2)$$

and so on.

Clearly, if we have these functions and if we know the output of any one period, we can deduce the outputs of all other periods, both past and future. The equilibrium output p is given by the condition

$$p = C(p) + A(p) \tag{6}$$

That is, the equilibrium output is that at which the expected consumption plus the planned accumulation, which together may be called the "total planned absorption," is equal to the output itself. At any level of output at which the total planned absorption is less than the output, the output must decline, as it cannot be absorbed. At a level of output at which the total planned absorption is greater than the output, either the output must rise, or some change must take place in the functions.

These principles are illustrated in Fig. 71. Along the horizontal axis we plot the output of "year" n; along the vertical axis the planned accumulation, expected consumption, and planned output in year $n + 1$ which correspond to each level of output in year n. Then the curve $\bar{C}$ is the expected-consumption curve for the year n, showing how much consumption will be expected in year $n + 1$ at various levels of output in year n. We assume that expectations are realized because the functions are assumed to be stable and known; i.e. it is known how much consumption will follow a given level of output. The curve marked $\bar{C} + \bar{A}$ is the expected-consumption-plus-planned-accumulation, or total-planned-absorption, curve. The vertical distance between the curve $\bar{C}$ and the curve $\bar{C} + \bar{A}$ is the planned accumulation at various levels of output. In the figure it has been assumed that at low levels of output accumulation will be negative, being zero where the two curves cross and positive at higher levels. This is a plausible assumption but is not, of course, neces-

sary for the argument. The curve $\bar{C} + \bar{A}$ is the planned-output curve, i.e. the curve which shows the output planned—and realized—for year $n + 1$ at each level of output of year n. This is the curve corresponding to equation (5).

The equilibrium output is $NK = ON$, where K is the point of intersection of the planned-absorption curve with the line making a 45° angle with the axes, OK. At this point $p_n = p_{n+1}$, and the output con-

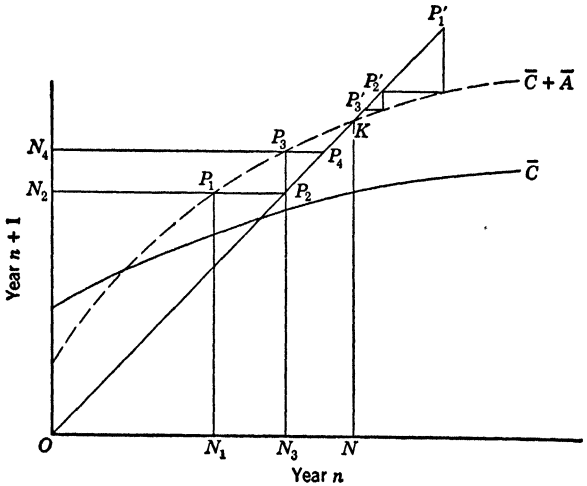


FIG. 71. Theory of employment.

tinues at a stable level. That this is a stable equilibrium on the given assumptions is shown by assuming an initial output less than the equilibrium, say, ON_1 in year 0. At this level of output consumption and planned accumulation will amount to N_1P_1; hence in the year 1 an output equal to $N_1P_1 = N_2P_2$ will be produced. Similarly the next year the output will be $N_3P_3 = N_4P_4$, and so on up to the equilibrium point K. Similarly, if we start with an output above the equilibrium point, output will decline, following the points $P_1'P_2'P_3'$, etc., back to the point K. The stability of the equilibrium depends on the planned-output curve cutting the 45° line from above, indicating that an output smaller than the equilibrium results in a total planned absorption greater than the output, and an output greater than equilibrium results in a total planned absorption smaller than output. If the reverse is the case, with the planned-output line cutting the 45° line from above, the equilibrium is unstable; if output declines below equilibrium, accumulation and consumption will fall still further and will drag output still further down.

There is nothing in this system so far to indicate that the equilibrium level of output is identical with "capacity" or "optimum" output. It may well be that at the optimum level of output the system is unwilling to absorb the total output, in consumption and in planned or willing accumulation, and under these circumstances the output is bound to decline below the optimum, and there will be unemployment of resources, in the sense that the level of employment is less than the optimum. If the consumption and accumulation functions happen to be "right," of course, the system may give us optimum output. It may be, on the other hand, that at the optimum output the system may wish to consume and accumulate more than it can produce. This may lead for a time to a state of hyperemployment, in which output is above that which would be socially desirable, e.g. from the point of view of conservation, in the long run. There is some point, however, which represents the absolute capacity of a system above which output cannot possibly go. If at this point the system still insists on trying to accumulate and consume more than it produces it will, of course, fail: accumulation will be less than expected, there may even be decumulation, and there is likely to be inflation of prices which will check consumption. The object of the various disturbances which will take place is, of course, to adjust the consumption and accumulation functions to the point where at "full employment" output is exactly absorbed by consumption and willing accumulation. If these disturbances fail the result may be hyperinflation.

Microeconomic assumptions in the Keynesian model

In order to examine the validity of the assumptions which underlie this system, it is necessary to return to the theory of the firm and of the individual economic organism. Perhaps the most dubious assumption of the model is that there is a stable consumption function relating output, i.e. real income, to consumption. We have seen that consumption may be more a function of the total size and the composition of the capital stock than of output. Even if consumption is quite independent of output, however, the system would still be useful, especially if the level of consumption were fairly stable. Thus, in Fig. 71, we could perfectly well assume that the consumption curve $\bar{C}$ was horizontal, indicating that the level of consumption was given independently of output, and the analysis would still hold. Even if the level of planned accumulation also were independent of output, so that the planned-output curve was horizontal too, the analysis would still hold formally; i.e. the equilibrium level of output would be equal to the otherwise-determined sum of consumption and accumulation.

The theory of the firm which underlies this analysis seems to involve the assumption that the production plans of each firm for each planning period depend on the realized real national income (output) of the previous period. Stated in this way, of course, the assumption seems highly unrealistic. Nevertheless it is likely to approximate to reality if certain other assumptions are made. Thus suppose we assume that any unplanned accumulation of inventory makes a firm cut down its output, and any unplanned decumulation of inventory makes a firm increase its output; we have a state of affairs which will approximate the conditions of our model, even without any conscious forecasting of consumption on the part of individual firms. In any period the actual rate of accumulation must be equal to output (production) minus consumption. If accumulation is proceeding at a rate faster than capitalists are willing to accept, individual capitalists will find themselves making "unplanned" accumulations—for instance, will find themselves with unsold inventories of product. We have seen that a very common reaction to this crisis is to cut down output in the next planning period in order to conserve money stocks and deplete inventories. The equilibrium output then will be that at which on the whole the actual rate of accumulation is such that it neither forces capitalists to contract nor encourages them to expand output. If output is below equilibrium, of course, consumption will absorb so much of the total output that capitalists will find themselves with depleted inventories and expanded money stocks, which will encourage an expansion of output.

In this model—and the defect is somewhat characteristic of the Keynesian analysis—there is a sharp neglect of possible price reactions. We have seen that the contraction of output is not the only possible reaction of a firm to bloated inventories, nor an expansion of output the only possible response to depleted inventories. Price reactions are also possible: if a firm finds itself with larger stocks of a commodity than it wishes to hold, its reaction may be to speed up the rate of transformation of these stocks into money by lowering the price at which it is willing to sell. Similarly a situation of depleted stocks may be met by a rise in price, thus lowering the rate at which the stocks are being depleted by purchase. The effects of general changes in the price structure on consumption and on the desire to accumulate are complex. In so far as they exist the usefulness of our simple model is threatened, for the consumption-output and investment-output functions will not be stable, and the reaction to a disequilibrium situation will not be a shift in output only but also a shift in the functions which determine the equilibrium output: i.e. a shift of the equilibrium towards the actual rather than of the actual towards the equilibrium output. Where this happens, however,

the interpretive value of the equilibrium concept is largely destroyed, no matter how much its formal validity may be retained.

The investment-savings identity

Before going on to discuss these problems, it may be well to discuss some further aspects of the simple Keynesian model. There are first some difficulties which should be cleared up concerning the relationship between the "real" concepts of accumulation, output or production, and consumption and the monetary concepts of investment, income, and outgo. Income and outgo, we have already seen, are the money value of production and consumption, respectively, valued at present prices. There is a problem in connection with this valuation when the prices of the constituent commodities of the "real" aggregates change during the accounting period; we must assume to avoid theoretical confusion that all prices remain the same during the accounting period and change only on the dates which separate successive accounting periods. This assumption does not raise any serious practical difficulties: we can always make our accounting period short enough so that the assumption holds, and in statistical practice we shall not go far astray, as long as the movement of prices is not too violent, if we value income and outgo at some average of prices. For the savings-equals-investment identity to be true, however, it is necessary that we value income, outgo, and investment at the same set of prices. If, then, we define investment as the total money value of accumulation during a period, we have:

$$\text{Investment} = \text{Income} - \text{Outgo} = \text{Saving}$$

The concept of the difference between income and outgo is probably the concept which best deserves the ambiguous term "saving." For any individual it represents the increase in net worth, and for society as a whole it represents the total increase in net worths, which is the value of the increase in capital. We have already discussed some of the difficulties inherent in these concepts when there are increases in net worth not as a result of additions to real capital at constant prices but as a result of price increases which raise the book value of existing capital.

The multipliers

In the Keynesian system the assumption is frequently made that investment or, better, planned accumulation is an independently determined variable of the simple system. We can then postulate an "investment multiplier," m—the change in income which results from a unit change in investment—which is in turn dependent on the "marginal propensity to consume," c, i.e. the increase in consumption which

results from a unit increase in income. These two ratios are related according to the formula: [2]

$$m = \frac{1}{1 - c} \tag{7}$$

The similarity with the payments multiplier formula on p. 227 should be noticed. The two formulae, however, are conceptually quite distinct and must not be confused: they pertain to two quite different though, of course, related processes in the economic system. A geometrical proof is given in Fig. 72. The line $C + A$ is the original "absorption curve"

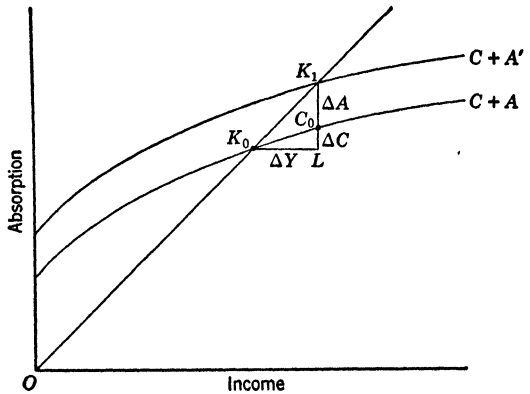


FIG. 72. The multiplier.

before the increase in investment. An increase in investment, assuming that investment is an independent variable, raises this curve vertically to the position $C + A'$ by an amount equal to the increase in investment, ΔA. The old position of equilibrium is K_0, the new K_1, these being at the point of intersection of the 45° line with the old and the new absorption curves, respectively. K_0L is the increase in income, ΔY, LC_0 is the increase in consumption as a result of the increased income, ΔC. C_0K_1 is the increase in investment, ΔA. Then, if K_0C_0 is a straight line, i.e. if the marginal propensity to consume is constant over the relevant range

[2] If Y is income, C is consumption, and A is investment,

$$Y = A + C$$

Hence

$$m = \frac{dY}{dA} = 1 + \frac{dC}{dA} = 1 + \frac{dC}{dY} \cdot \frac{dY}{dA} = 1 + cm$$

i.e.

$$m = \frac{1}{1 - c}$$

of income, we have:

$$m = \frac{\Delta Y}{\Delta A} = \frac{K_0 L}{C_0 K_1} = \frac{K_1 L}{C_0 K_1} = \frac{LC_0 + C_0 K_1}{C_0 K_1} = 1 + \frac{\Delta C}{\Delta A} = 1 + \frac{\Delta C}{\Delta Y} \cdot \frac{\Delta Y}{\Delta A}$$

$$= 1 + cm$$

whence

$$m = \frac{1}{1 - c}$$

The investment multiplier, it must be emphasized, has nothing directly to do with money flows, with payments, with expenditures and receipts: it is essentially a "real" concept even though it may be expressed in money-value terms.

Other multiplier concepts can, of course, be constructed, and may be of some interest. In general, if c is the "marginal propensity to absorb," i.e. the slope of the absorption curve $C + A$, assumed constant in the relevant range, and any change raises the absorption curve by a fixed amount, the effect on income will be equal to the change itself multiplied by a multiplier given by the above formula. Thus instead of a change in investment or accumulation there may be a spontaneous change in the consumption function, i.e. a vertical shift in the whole consumption function. For every dollar increase in consumption at each level of income, there will be an increase of m dollars in income. Similarly we can have a "government multiplier." This is the number of dollars increase in income which results from a unit increase in government absorption of the product, government absorption being the net output which government purchases.

Government absorption and deficits

In Fig. 73 the curve PP_1P_2 is the "private-absorption curve": it shows how much product is willingly absorbed, either in consumption or in voluntary accumulation, by the private sector of the economy. Total absorption is, of course, equal to private absorption plus government absorption. If, then, we are given the private-absorption function we can find immediately how much government absorption is necessary to yield any desired level of output and employment.

In the absence of any government absorption the level of output will be ON_1 ($= N_1P_1$), where OP_1 is the 45° line cutting the private-absorption curve in P_1. Suppose that ON_2 is the desired or optimum ("full employment") level of output. At this output private absorption is only N_2P_2, which is short of the output itself (N_2G_2) by an amount P_2G_2. The amount of government absorption, then, which will yield optimum output is P_2G_2, and at that level of government absorption the

total-absorption curve will be GG_2. The assumption is made here, of course, that government absorption does not affect the position of the private-absorption curve. The conditions under which this assumption is likely to hold are examined in Chapter 16.

It should be noted that the significant quantity here is not the budget deficit—that merely affects the total of payments—but the government absorption, i.e. the goods and services which are purchased by government from private firms and households and not resold. This absorption may, of course, be expressed in value terms, but it is the physical absorption which is significant. That is not to say, of course, that the payments effects of governmental action have no significance; they operate, however, by changing the private-absorption function itself. Thus a budget deficit, by increasing the quantity of money in the hands of the public, and therefore increasing the total volume of payments, may very well have the effect of raising the consumption function, and hence raising the private-absorption function, and consequently will enable a society to reach a given level of output with a smaller amount of government absorption. These effects, however, are indirect, and on the whole are likely to be dynamic in character; i.e. they depend for their efficacy on certain "lags," e.g. in prices, or on certain reactions to dynamic change. The closer we get to the "long run" the more is the economics of Adam Smith as opposed to that of Keynes likely to hold: the quantity of money, the volume of payments, and the price level may come into line at *any* level, their absolute level being a quite insignificant quantity. In the short run, of course, the *changes* in the payments-prices structure may be of great importance; even then, however, their main importance lies in the effects they have on real consumption, real accumulation, and real output.

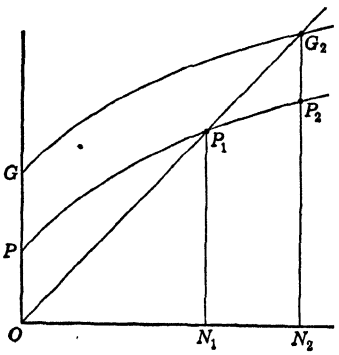


FIG. 73. Government absorption.

The Keynesian theory of interest

The Keynesian system, as set forth in the *General Theory*[3] and as reduced to symbolic form by, e.g. Meade,[4] did not stop short at the very

[3] J. M. Keynes, *The General Theory of Employment, Interest, and Money* (Harcourt, Brace & Co., New York).

[4] J. E. Meade, A simplified model of Mr. Keynes' system, *Review of Economic Studies* (February 1937).

simple model outlined above, in which the volume of planned investment is regarded either as one of the given factors in the system, or as determined simply by the previous or existing level of output. Keynes was much interested in the problem of the determinants of the volume of investment, thinking as he did that this was the principal direct determinant of income, output, and employment. On the whole his thought in this matter stemmed from Wicksell, who regarded the volume of investment as in the main determined by the divergence between a "natural" rate of interest and the "market" rate. In the Keynesian system Wicksell's "natural" rate of interest (never very satisfactorily defined by him) becomes the famous "marginal efficiency of capital," which is the *expected* marginal rate of profit on investment in real capital. The rate of investment Keynes then regarded as a function mainly of the difference between the marginal efficiency of capital and the rate of interest, by the "rate of interest" meaning, of course, the whole complex of credit facilities.

We shall examine the impact of the credit system in a later chapter: meanwhile it is sufficient to observe that Keynes apparently regarded the marginal efficiency of capital as "given" in each situation by the state of mind of the capitalist class (the significant marginal efficiency of capital being the *ex ante* concept, what capitalists *think* will be earned by new investment, not the *ex post* concept, what investment actually earned, or even what the investment actually will earn), and that hence he regarded the rate of interest, or the ease or stringency of credit, as the main *controlling* factor in the employment situation. A rise in the rate of interest, the marginal efficiency of capital being given, would then diminish, and a fall would increase, the volume of investment, and therefore of income. Hence investment is now regarded as a function not only of income (output) but also of the rate of interest.

Similarly consumption might also be considered as a function of income and of the rate of interest: however, it is doubtful whether the rate of interest has any perceptible effect on consumption. Then the critical question becomes, of course, What determines the rate of interest? The Keynesian answer to this is found in the liquidity preference theory, which, going back to the mercantilists, sees the main determinant of the rate of interest in the quantity of money (liquid assets). In terms of our previous analysis, the rate of interest is an inverse function of the price of securities, and given the preferred liquidity and security ratios, a rise in the quantity of money will raise the price of securities and will therefore lower the rate of interest. The implicit assumption behind this theory is, of course, that the future payments from securities do not change. If an increase in the quantity of money raises the prices of commodities, over the course of time as old bonds are paid off and new bonds

written the higher price of bonds will be offset by higher payments, and there is no particular reason to suppose that an increase in the quantity of money would diminish the rate of interest. This is Adam Smith's famous reply to the mercantilists, in essence, and in the long run he again is likely to be right. That is to say, the liquidity preference theory of interest is essentially a short-run theory which depends for its validity on two assumptions: first that the principal impact of an increase in the quantity of money is felt in the security markets—especially in the bond market—and is not felt in commodity markets; and, second, that there is a sufficient volume of long- and medium-term bonds so that a general rise in the prices of bonds, which does not affect the contractual payments from these bonds, has a perceptible effect on the rate of interest. Both these assumptions—and especially the first—constitute rather special cases which may hold in some circumstances but which are by no means generally true.

Here, as in many other macroeconomic propositions, we may fall foul of the "aggregation" danger. The creation of new money may have very different *dynamic* effects, depending on how it is created and into whose hands it first comes, even though it may be true that once the new money is disseminated through all the accounts of the system by the payments process it produces an approximately even effect on all markets. In the initial stages, however, the impact of new money depends very much on its distribution. If the new money comes mainly into the hands of operators in the security markets, it will, of course, have its first impact on security prices, and the Keynesian condition is most likely to be fulfilled. If, however, the new money comes into the hands of people who operate mainly in commodity markets it may express itself chiefly in a rise in wholesale commodity prices. This will, of course, raise the marginal efficiency of capital invested in commodities, which from the Keynesian point of view may be just as good as a fall in the rate of interest, but which is certainly not the same thing as a fall in the rate of interest. If the new money comes mainly into the hands of wage earners, e.g. through tax reductions in lower brackets, the effects may lie chiefly in an increase in consumption. It will be observed that by and large the effects of the creation of new money are likely to be favorable to employment, especially in a period of unemployment, but the mechanism by which this takes place may be very different from the one postulated by Keynes.

The relation of payments and output

A question of considerable difficulty but of great importance in the theory of employment is that of the exact relationship between the pay-

ments system and the volume of output and employment, and particularly between the total of payments and the total volume of output and employment. Payments (with the exception of pure transfer payments, or gifts) are always made *for* something. That is, corresponding to the payments table of the previous chapter we could postulate a table showing what non-liquid assets had passed from each individual to every other during the "year." The values of these "purchase-sales" [5] would, of course, be equal to the payment made for them, on the assumption that in exchange equal values are exchanged. A purchase, however, does not have to be consumed by the purchaser in the accounting year in which it is made; hence there is no reason for the total volume of purchases to be equal to the volume of consumption. Similarly there is no necessity for a sale, i.e. a commodity sold, to have been produced in the accounting year: it may have been produced long previously and may be merely changing hands. There is no reason to suppose, therefore, that the total volume of sales will be very directly related to the volume of production.

We could quite easily conceive of a system in which there was no production or consumption, no income or outgo whatever, but in which a set of immortal assets perpetually shifted ownership and in which, therefore, there would be a purchase-sale system and a corresponding payments system. Over short periods, in fact, the securities market approximates to such a system: there is very little production or consumption of securities, i.e. very little added to the total by new issues or subtracted by withdrawals, yet there may be a great volume of transactions and a great volume of payments. It is evident that normally in the system as a whole the volume of transactions and of payments will exceed, perhaps by many times, the volume of output or of income. The problem of what determines the relation between the volume of payments and the volume of output is an important, if difficult, one, important if only for the reason that government can control the total volume of payments probably more easily than it can control the total volume of output and employment. It is certainly a mistake to regard the problem of the control of the total volume of payments as synonymous with that of controlling the volume of output. It would be perfectly possible for payments to rise while output fell if, for instance, there was a sharp increase

[5] There seems to be no substitute for the clumsy expression "purchase-sale" to indicate the transfer of an asset (which may, of course, be a commodity, a security, or a service) for which a payment is made. It is important to emphasize that every such transfer is *both* a purchase and a sale—a purchase to the account that buys the object, a sale to the one that sells it.

in the price level, or if there was an increase in the transactions-output ratio, i.e. more transactions per unit of output.

From the point of view of the production-consumption-employment system the main significance of exchange is that it shifts assets into the hands of those who are more willing consumers or accumulators. If there were no exchange it is a commonplace of economics that each individual would have to produce what he consumed himself. The specialization which exchange permits, however, is significant not merely because it improves the *efficiency* of production and hence permits a greater potential or capacity volume of output from given resources. It is significant also because it permits a greater "absorption" of output, either as willing consumption or as willing accumulation. If there is a breakdown in the exchange system (which is in part what constitutes a depression) we find not only a regression in specialization, as, for instance, a return to the self-subsistent farm, but we find also a lessened willingness to consume or to accumulate. That is to say, a deflationary movement in the payments system—occasioned, for instance, by a disequilibrium in the balance of payments as described on pp. 222–226—may easily lead to a downward shift in the output-consumption or output-accumulation functions, and hence in the equilibrium level as well as the capacity level of output.

The prime link between the payments-exchange system and the income-output system is the price system. If the physical volume of exchanges is constant, a rise in the prices at which the exchanges are made will raise the volume of payments. This is seen clearly in Fisher's equation of exchange, where if E is the total volume of payments,

$$E = MV = PT$$

A rise in P, T being constant, must raise E in the same proportion. We see clearly therefore that even if E were constant, fluctuations in opposite directions could take place in P and T.

A ratio W can be postulated, so that $W = T/Q$, where Q is the volume of output: this ratio has sometimes been called the "velocity of circulation of goods," which is something of a misnomer, as T includes exchanges of securities. The Fisher equation can then be written:

$$E = MV = PQW$$

The ratio W is capable of some independent fluctuations: however, it is a very heterogeneous parameter and must be used with caution.

Chapter 14

A Macroeconomic Theory of Distribution

One of the most important problems in economics, both from the point of view of its central place in any theoretical structure and also from the point of view of its significance for public policy, is the distribution of the social income. By this we really mean the structure or composition of the social product from the point of view of the claims upon it or the "real" ownership of it. We have seen that income consists essentially in the gross addition of value to the total stock of capital, which reflects itself in a gross addition to net worths. The question naturally arises, Who gets these additions, i.e. in whose balance sheets, in whose accounts, do these gross increases appear? Any act of production results, in itself, in an increase in the total of net worths: the essential problem of distribution is how this increase is distributed among the net worths associated with different individuals, different classes of people, or different categories of assets.

The microeconomic theory of distribution

The problem of the distribution of income among different people or classes of people is known as "personal" distribution. Economists have usually carried the problem further back to what is known as "functional" distribution, i.e. distribution according to the kind of asset or "factor of production" owned. Each factor is thought of as "producing" income which accrues into the account of the owner of the factor. The personal income of each individual is then determined by the amount of factors which he owns, and the income per unit of each factor. In order to include labor income in this picture we must, of course, regard a man's body and mind as one of his assets. The problem is complicated further by the fact that the income per unit of factor owned depends on the intensity of utilization of the factor, and that frequently the distribution problem has been discussed in terms of income per unit of utilization rather than in terms of income per unit of factor. A "wage per hour" is

income per unit of utilization: wage per year is more likely to represent income per unit of factor. Thus we have:

$$\text{Income} = \text{Quantity of factor owned} \times \text{Income per factor unit}$$
$$= \text{Quantity of factor owned} \times \text{Coefficient of utilization} \times \text{Income per utilization unit}$$

In the case of labor income a man, say, owns one body and works it 2000 hours per year (this is the utilization coefficient) at an hourly wage of $2 per hour: his income from his body is $4000 per annum. Or suppose that a man has 10 houses, rents each for 9 months in the year at $50 per month net: his annual income from houses is $10 \times 9 \times 50$, or $4500.

Because of the fact that the personal income of an individual can be derived from these three elements—his property (including his own person), the degree of utilization of property, and the price of the unit services of property—the problem of distribution has generally been treated as a microeconomic problem. The distribution of property is taken as given by historical factors such as inheritance and personal saving: the degree of utilization of property can be treated as a problem in the theory of the individual economic organism by means of indifference curves,[1] and the problem of the price of services (e.g. wages per hour, rents per unit of property) is regarded as a special case of the theory of particular prices. Thus the marginal-productivity theory is essentially of this microeconomic nature, in spite of the attempts by J. B. Clark and others to give it a macroeconomic significance. Its real significance is that it gives us a useful first-approximation theory of the demand of firms for input—i.e. it enables us to derive, for instance, a function relating the money wage and the amount of employment offered for any particular type of labor.[2] Then by relating this demand function to the conditions of supply and the nature of the market (e.g. competitive or monopolistic) an economic equilibrium money wage can be determined.

This microeconomic theory, however, throws very little light on a number of important problems. Because demand and supply curves cannot be aggregated for the whole economy, it tells us very little about the impact of a *general* wage change, for instance, on employment and output; although it makes a good deal of sense to draw a demand curve for carpenters' labor in Chicago, it makes very little sense to draw a demand curve for labor as a whole, simply because the general volume of

[1] Boulding, *Economic Analysis*, p. 742.

[2] Even at this level, as we have seen (pp. 107–111), the marginal-productivity theory is valid only if we can neglect the asset preferences of the firm, as it is based on the assumption of maximizing profit without regard to the composition of assets.

employment may not be a particularly significant function of the average money wages. The microeconomic theory also throws very little light on the question of money wages versus real wages: it can discuss "real" wages only in the insignificant sense of the purchasing power of a particular kind of labor in terms of its own particular product, a quantity which is not of great interest to the worker, who generally spends only a very small proportion of his wage on his own product. It is not only how many shoes he can buy with an hour's labor that interests the shoe worker, but also how many loaves of bread, bottles of milk, and so on. Hence the microeconomic theory, because it assumes that particular wage changes take place in an environment where all other things are equal, cannot throw much light on the question whether a general rise in money wages will result in a rise in real wages and a shift in distribution towards wages away from profits.

Need for a macroeconomic theory

It is clear that the microeconomic theory of distribution, which is excellent, for instance, in dealing with problems of relative wages and relative profits,[3] is not adequate to deal with problems involving the general levels of wages and profits and the determinants of the broad aggregates of the distributional structure. We need, therefore, a macroeconomic theory of distribution. It is interesting to observe that the classical economists, including Marx, were aware of the necessity for some such theory and made some attempts to provide it. The wage-fund theory of Smith and Ricardo—and even in some measure of Böhm-Bawerk—and the surplus-value theory of Marx are attempts to find in essentially macroeconomic terms an answer to the broad question, What determines the distribution of the product into labor's share and capital's share? Of these Ricardo's is undoubtedly the most satisfactory, rent being determined by the degree of pressure of population on the available land, wages being determined by a like pressure of population on the wage fund, and profits being essentially a residual. In the Marxian theory (the significance of which is often lost because so much of it is essentially macroeconomic theory couched in microeconomic terms) wages are determined at the subsistence level, apparently by the greater bargaining power of employers, and hence "surplus value" (i.e. non-labor income) is the residual, i.e. the difference between the total output and the total wage bill. None of these theories, however, is very satisfactory: the wage-fund theory rests on a fundamental confusion between "stocks" and "flows," and not even Böhm-Bawerk's attempt to rescue it was really successful. The Marxian wage theory is so clearly unrealistic in

[3] See, for instance, Adam Smith's classic chapters on the subject, which have hardly been bettered (Book 1, Chapters 8 and 9).

the modern economy, a characteristic feature of which is a remarkable stability in the proportion of the total product which goes to labor, that it cannot be considered a serious theoretical contribution. Nevertheless the problem remains as to what are the determinants of the distribution of the total product between labor and non-labor income. The distribution of non-labor income between contractual income (interest and rent) and residual income (profits) at any one time is historically determined by the nature and extent of the contractual obligations, and does not present many serious theoretical problems. The distribution of total income between labor income (wages) and non-labor income, which we may call "gross profits" (the Marxian "surplus value"), is a theoretical problem of first magnitude.

The balance-sheet identities

We shall approach it by the method of comparative balance sheets that we have found useful earlier, treating income as essentially a gross increase in net worth. We shall assume that there are two kinds of economic organisms, businesses and households. Government will be neglected for the moment: it can formally be regarded as a rather peculiar kind of household! We shall assume that there are three kinds of assets: goods, Q (real assets), securities or debts, K, and money, M.

We shall assume that in the balance sheet of an individual business the items may be classified into seven types, as listed in Table 1.

TABLE 1

BALANCE SHEET OF A BUSINESS

Assets		Liabilities	
Money	m_b	Debts to businesses	k_b'
Goods	q_b	Debts to households	k_h'
Debts from other businesses	k_b	Net worth	g_b
Debts from households	k_h		

By reason of the definition of net worth, the total of assets must be equal to the total of liabilities: hence we have the fundamental "balance-sheet identity":

$$m_b + q_b + k_b + k_h \equiv k_b' + k_h' + g_b \tag{1}$$

If now we sum the balance sheets of all businesses we get a collective balance-sheet identity,

$$M_b + Q_b + K_b + K_h \equiv K_b' + K_h' + G_b \tag{2}$$

M_b ($= \Sigma m_b$) is the total money stock held by all businesses. Q_b ($= \Sigma q_b$)

is the *value* of the total stock of real capital, or goods, held by all businesses: it includes inventory of finished goods, goods in process, and raw material: buildings and equipment, land, etc.; it may also include certain non-material goods such as patent rights and copyrights. The test for inclusion of an item in Q is that it should appear in only one balance sheet: if it appears as an asset in one balance sheet and as a liability in another (or others) the item is properly classified under one of the K items.

K_b is the total amount of debts from businesses to businesses; K_b' is the total amount of debts to businesses from businesses. It is evident that K_b' and K_b are the same quantity, for every interbusiness debt will appear twice in business balance sheets: once as an asset and once as a liability. We have therefore

$$K_b' \equiv K_b \tag{2a}$$

K_h is the total debts due to businesses from households. It consists mainly of credit of various kinds extended to households by businesses, such as book credit and installment credit. It is mostly short run in nature.

K_h' is the total debts due from businesses to households. This consists mainly of the contractual obligations (bonds) of businesses held by households. The total net worth of business (G_b) is also represented by stocks which are held by households. Stocks held by other businesses we shall regard as included in K_b and K_b', as simply part of interbusiness debt.

We can then combine identities (2) and (2*a*) and reformulate as follows to obtain what may be called the *business-net-worth identity:*

$$G_b \equiv M_b + Q_b + (K_h - K_h') \tag{3}$$

That is to say, the total net worth of businesses is equal to the money holdings of business, plus the value of the goods (real asset) holdings of business, plus the net household indebtedness to business.

A similar balance-sheet identity can be constructed for households. The items in the balance sheet of any household may be classified as follows:

Assets		Liabilities	
Money	m_h	Debts from household to businesses	k_{hb}'
Goods (value of)	q_h	Debts to other households	k_{hh}'
Debts from businesses to household	k_{bh}	Net worth of household	g_h
Net worth of businesses owned by household	g_{bh}		
Debts from other households	k_{hh}		

Again we have a balance-sheet identity, as follows:

$$m_h + q_h + k_{bh} + g_{bh} + k_{hh} \equiv k_{hb}' + k_{hh}' + g_h \tag{4}$$

We can sum these balance sheets also, and so obtain for all households, writing $M_h = \Sigma m_h$, $Q_h = \Sigma q_h$, and so on,

$$M_h + Q_h + K_{bh} + G_{bh} + K_{hh} \equiv K_{hb}' + K_{hh}' + G_h \tag{5}$$

We know that the total of debts from households to households (K_{hh}) is equal to the total of debts to households from households (K_{hh}'). We also know that the total of debts from households to businesses (K_{hb}') is equal to the total of debts to businesses from households (K_h), as these are the same thing. Similarly we have $K_{bh} \equiv K_h'$. We have also $G_{bh} = G_b$, as all the net worth of businesses must ultimately be allocated to households. We can rewrite the identity (5) then as follows:

$$G_h \equiv M_h + Q_h - (K_h - K_h') + G_b \tag{6}$$

Combining identities (6) and (3), we have:

$$G_h \equiv M_h + Q_h + M_b + Q_b \tag{7}$$

That is to say, the total net worth of households is equal to the total quantity of money in the system ($M_h + M_b$) plus the total value of the stock of real assets (goods) in the system ($Q_h + Q_b$). This result expresses the fact that the values of all *titles* to property cancel out when we add up all balance sheets, and we are left with the identity that the net worth of households is equal to the value of all "real" property, including money. This reflects the fact that the "net worth" of businesses is in a sense an accounting fiction, for businesses as such (being "imaginary persons") cannot be the ultimate owners of property—the final equity in all property must be held by persons or households. There is an interesting problem in how to treat certain non-profit organizations such as churches, colleges, orphanages, and charitable foundations, as the net worth of these bodies is not represented in the balance sheets of any households, even though households may benefit from their property and activities. Either such organizations may be regarded themselves as households or some imaginary apportionment of their net worths may be made among their beneficiaries.

The business-savings identity

All the above identities can be differentiated and still hold good if the symbols represent not the stocks of assets or liabilities but the *changes* in these quantities in any given time period. Thus from equation (3)

we can write:

$$dG_b \equiv dM_b + dQ_b + dK_h - dK_h' \tag{8}$$

where the d in front of the symbol signifies the change in the quantity so modified in a given time period. All the quantities of the identity (8) represent important concepts. The quantity dG_b is the change in the net worths of all businesses during the period under review (say, a year). It may be called *business savings* if positive, business dissavings if negative. It represents that part of total profits which has not been distributed in dividends. The quantity dM_b is the increase in the total money holdings of businesses; dQ_b is the increase in the *value* of the goods (real capital) holdings of businesses; dK_h is the total increase in business credit extended to households; dK_h' is the increase in business indebtedness to households. As it stands, of course, this "business-savings" identity is a truism like all the others. It is a significant truism, however, because on the whole the various items in it are determined with some degree of independence of each other. Hence we get another "macroeconomic paradox": that business savings are for the most part not determined by the decisions of businesses to save, but are a result of a complex of other decisions which determine the volume of business savings independently of the decisions of individual businesses to save, i.e. to increase their net worths.

If we accept the fact—which will be elaborated later—that the total of business savings is independently determined, then business *decisions* to save, as reflected in dividend policy, determine not business savings but the level of profits itself. This can be seen clearly in a further identity: if V is the total profits of business, this must be equal to business savings plus business distributions in interest and dividends. Total profit is the gross increase in the net worth of businesses as a result of their operations. Out of this gross increase interest and dividends are paid, representing a diminution in net worth. What remains after interest and dividends have been subtracted from gross profits is business savings. If, therefore, D is total business distributions in interest and dividends, and V is total profits of business, we have

$$V \equiv dG_b + D \tag{9}$$

As interest payments are in large measure determined in any given year by past contractual obligations, the total of business distributions is mainly governed by decisions as to dividends declared. If we can assume that dG_b is determined independently of business distributions—an assumption which is not quite true, as we shall see later, but which is true enough to serve as a first approximation—then business distributions

are a "widow's cruse and Danaïd jar" [4]: the more business distributes in interest and dividends, the more will be returned to it in profits to distribute; the less it so distributes, the less there will be to distribute. Distributions, as it were, run through the system in a rapid reflux back into the pockets of business again.

Suppose, for instance, that business savings, as given by the various items of equation (8) which determine the net increase in business assets, i.e. in business net worth, are 10 billion. If businesses expect a profit of 30 billion and decide to save 12 billion they will distribute only 18 billion; hence actual profits will turn out to be 28 billion instead of 30, and in spite of the decision to save 12 billion businesses will succeed in saving only 10 billion. Similarly, if under the same circumstances businesses decide to save only 8 billion, and hence distribute 22 billion, the total of profits will be 32 billion and again business, in spite of its decision to save 8 billion, will find itself saving 10 billion.

This seems like a strange macroeconomic world that we are inhabiting, especially for those accustomed to thinking in terms of the experience of a single enterprise and generalizing from that experience. A single firm will certainly not find that an increase in its distributions automatically brings a corresponding increase in profits! An increase in distributions by one firm, however, swells the profits of others; an increase in distributions of all firms swells the profits of all.

The profits and wages identities

Bringing equations (8) and (9) together, we can now derive an identity for total profits:

$$V \equiv dQ_b + dM_b + dK_h - dK_h' + D \tag{10}$$

The next step is to derive a similar identity for wages. This can also be approached by developing first an identity for household saving by differentiating equation (7):

$$dG_h \equiv dQ_h + dQ_b + dM_h + dM_b \tag{11}$$

For the present we are assuming no creation or destruction of money, so that the increase in the money holdings of business must be equal to the decline in the money holdings of households, as the total money holdings of businesses and households together are constant: i.e. $dM_b + dM_h = dM = 0$, where dM is the change in the total money stock. Equation (11) therefore becomes:

$$dG_h \equiv dQ_b + dQ_h \tag{12}$$

[4] Cf. J. M. Keynes, *Treatise on Money* (Macmillan, London), Book 1, p. 139.

That is, the total savings of households—which, of course, are equal to the total savings of the whole economy, as business savings are all ultimately owned by households—must be equal to the increase in the value of total real capital, including that held by households as well as that held by businesses. This is the change in the net worth of households. Household savings can also be approached from another aspect, however: they are equal to household income less household consumption. Household income, Y_h, is the gross additions to the net worths of households. It consists of wages earned—which are not necessarily equal to wages paid (W); from the point of view of the total assets of the household a wage claim is just as good as money—plus dividends paid (D) plus business savings (dG_b) which in theory, at least, accrue to households. If household consumption is C_h we have:

$$dG_h \equiv W + D + dG_b - C_h \equiv dQ_h + dQ_b \tag{13}$$

Writing in the value of dG_b from equation (8) and of dG_h from equation (12), we have:

$$W \equiv C_h + dQ_h - (dM_b + dK_h - dK_h' + D) \tag{14}$$

It will be observed that the item $dM_b + dK_h - dK_h' + D$ ($= T$) appears in both equations (14) and (10). It may be called the "transfer item." We can now bring together the equations for wages and for gross profits and show that between them they exhaust the total product:

$$W \equiv C_h + dQ_h - T \tag{15a}$$

$$V \equiv dQ_b + T \tag{15b}$$

$$\text{Total net product} \equiv P_n \equiv W + V \equiv C_h + dQ_h + dQ_b \tag{16}$$

Equation (16) may be compared with equation (3) on p. 229: P_n, the total net product, is equal to P, the gross product, minus an item of "business consumption," or depreciation of business capital, C_b. Total consumption therefore is $C_b + C_h = C$; total accumulation, $A = dQ_h + dQ_b$, i.e. household accumulation plus business accumulation. Equation (16) can therefore be derived directly from equation (3), p. 229, and we see that the whole *net* product is distributed. The significance of the gross-product concept will be examined elsewhere: for the moment we may content ourselves with observing that the net product is all that is available for distribution, as depreciation does not represent "income." The net product is the concept usually given the name of "national income," though there is no reason to confine the concept to national boundaries, and it can be applied either to any group of people or to the whole world economy. We observe that the *composition* of the net

product can be expressed in either of two ways: as household consumption (C_h) plus total accumulation ($dQ_h + dQ_b$), or as household "absorption" ($C_h + dQ_h$) plus business accumulation (dQ_b). Household absorption is the net total of goods and services which households take from businesses. It represents the gross increase in household holdings of goods, and must therefore be equal to household consumption plus household accumulation, for whatever a household does not consume it must accumulate. If I purchase six shirts during a year and wear out four, I must have increased my stock of shirts by two. Household absorption is what the Keynesians have called "consumption": in spite of the general use of the term it appears to be such a misnomer that it seems better to reserve the word "consumption" for its more exact sense of asset destruction.

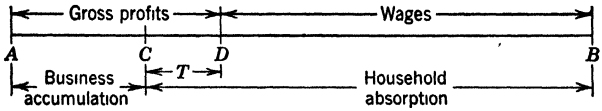


FIG. 74. The distributional identities.

We now see from equations (15) that the *distribution* of the product between wages and gross profits is determined by two elements: the *composition* of the product absorption on the one hand as between business investment and household absorption, and a transfer factor which we add to business absorption to get the total of profits, and subtract from the total of household absorption to get the total of wages. Figure 74 illustrates this proposition graphically: AB is total income, divided into AC, business accumulations; CB, household absorption; and AD, gross profits, and DB, wages. Then CD is the transfer factor, T. As they stand, of course, these identities are merely truisms. Whether they are useful truisms depends on whether the components are useful as "parameters of behavior," i.e. whether the components vary in a fairly regular fashion with certain changes in human behavior. In particular it is important to examine the composition of the transfer factor, T, to see whether it is likely to have any stability, and whether it is likely to be determined in some degree independently of the business-accumulation and household-absorption factors.

The transfer factor: the balance-of-payments item

We see from equations (10) and (14) that the transfer factor can be broken down into four parts:

$$T = dM_b + dK_h - dK_h' + D \tag{17}$$

Each of these four components is a highly significant economic quantity. The item dM_b, the increase in the money balances of businesses (which will, of course, be negative if there is a decrease in business money stocks) is a pure balance-of-payments item: it is also equal to the decline in the money stocks of households. It represents, that is to say, the favorable balance of payments of businesses or, what is the same thing, the unfavorable balance of payments of households. If the total of money payments from business to households is b_h, and from households to businesses is h_b, then

$$dM_b = h_b - b_h \tag{18a}$$

and

$$dM_h = b_h - h_b = -dM_b \tag{18b}$$

For the moment we are neglecting changes in the total money stock of society, which we assume to be constant; as we shall see, relaxing this assumption does not make much difference to the argument. We can then think of society's money stock as a shifting cargo, which shifts away from household pockets into business pockets when households are spending towards businesses more freely than businesses are spending towards households, and shifts back again from business pockets into household pockets when businesses are spending more freely than households. Clearly the longer period of time which we take, the smaller, relative to the other quantities of the system, will this balance of payments tend to be. A period of positive dM_b must inevitably be followed by a period of negative dM_b; the balance of payments is a shifting thing, and no one part of society can continue to have a positive—or the other part a negative—balance of payments indefinitely, or the whole money stock of a society would be concentrated in the one part of it.

The determinants of this net transfer of money from households to businesses are the relative liquidity preferences of households and businesses. If both are attempting to disgorge money with an equal degree of intensity, obviously neither will succeed: the result will be an increase in the total volume of payments, but no change in the balance of payments. Similarly if both attempt to hoard money with equal intensity neither will succeed; there will be a decline in total volume of payments, but no change in the balance of payments. If, however, households exhibit a stronger desire to hoard, or a weaker desire to dishoard, than businesses, there will be a shift in money stocks from businesses to households: households will find themselves with a generally favorable balance of payments and businesses will find themselves with a generally unfavorable balance.

The "velocity of circulation" parameter is a convenient shorthand method of expressing the relative liquidity preferences, as we have seen

in Chapter 11. What we have in effect is an amalgamated two-account payments system, as in the following table:

	B	H	dM
B		b_h	$h_b - b_h$
H	h_b		$b_h - h_b$

If M is the total quantity of money, m_h and m_b are the quantities of money held by households and businesses, respectively, and v_h and v_b are the partial velocities of households and of businesses, then:

$$\begin{aligned} M &= m_a + m_b \\ b_h &= v_b m_b \\ h_b &= v_h m_h \\ h_b &= b_h \quad \text{(in equilibrium)} \end{aligned} \tag{19}$$

Solving these equations, we obtain (cf. p. 211)

$$m_b = \frac{v_h}{v_b + v_h} M, \qquad m_h = \frac{v_b}{v_b + v_h} M, \qquad h_b = b_h = \frac{v_b v_h}{v_b + v_h} M \tag{20a}$$

Now let us suppose that there is a shift in the partial velocities to v_b' and v_h' between dates 0 and 1. There will be a redistribution of the money stock, resulting in m_b' in the hands of businesses, m_h' in the hands of households. We have, however,

$$m_b' = \frac{v_h'}{v_b' + v_h'} M, \qquad m_h' = \frac{v_b'}{v_b' + v_h'} M \tag{20b}$$

Comparing equations 20*a* and 20*b*, we get:

$$dm_b = m_b' - m_b = \frac{v_h' v_b - v_b' v_h}{(v_b + v_h)(v_b' + v_h')} = m_h - m_h' = -dm_h \tag{21}$$

dm_b is, of course, the positive balance of payments of business, equal to the negative balance of payments of households, between the two dates. We see immediately from equation (21) that this will be positive, zero, or negative, as

$$\frac{v_b'}{v_b} \gtreqless \frac{v_h'}{v_h}$$

i.e. if the partial velocities change in the same proportion, there will be no redistribution of the money stock and a zero balance of payments; if, however, the partial velocity of business rises relatively more, or falls

relatively less, than the partial velocities of households, the balance of payments of business will be negative, of households will be positive. The partial velocity expresses the desire to get rid of money, and if the desire of business to get rid of money rises more than that of households businesses will "win" in the tug of war and will succeed in pushing money into household balances. If households increase their velocities more than businesses, however, the money cargo will be pushed into business balances out of household accounts.

The transfer factor: debt and distributions items

The second item in the transfer-factor equation (17), dK_h, is the increase in household indebtedness to business, i.e. in credit extended by businesses to households. The main component in this item is installment credit: however, ordinary book credit and charge accounts form a not inconsiderable part. It is clear then that extending credit to households has a directly favorable effect on gross profits, and shifts the basic distribution pattern from labor to non-labor income. This is an effect which I do not believe has been recognized clearly before; nevertheless the strong prejudice in favor of consumer credit on the part of the business community may reflect a subconscious appreciation of its favorable effect on profits. There does not seem to be any corresponding distaste for consumer credit in the labor groups, mainly because we have here a clear case of a "macroeconomic paradox," where the effect of a certain pattern of behavior on a group is not equal to the sum of individual experiences. Thus it is very often to the interest of the individual worker to avail himself of consumer-credit facilities which enable him to command in his domestic enterprise a greater volume of assets than his net worth would otherwise allow him to have. Yet the general direct effect is to reduce the proportion of the total product which accrues to labor.

The third item in the transfer factor, $-dK_h'$, likewise presents another example of a macroeconomic paradox. The figure dK_h' is the increase in business indebtedness to households, and represents for the most part the volume of securities (bonds) sold to households during the period. This represents a negative item in the transfer factor; hence the greater the volume of securities sold to households, the more the distribution pattern is shifted towards wages and away from gross profits. Hence the sale of securities adversely affects the profits in expectation of which the securities are offered. Just as the individual interest and the group interest may diverge when a worker uses consumer credit, so the purchase of securities (usually by the owner class), which may be of real advantage to the individual purchaser, may be detrimental to the

owner class as a whole. This is not to say, of course, that the sale of securities to the public is fatal to profits, which are mainly determined by the volume of business investment: it merely operates a direct force tending to reduce gross profits and expand wages.

The fourth and last item in the transfer factor, the volume of business distributions, D, is perhaps the most tantalizing and paradoxical of all. We see that the direct effect of business distributions is to increase gross profits by the same amount: this is the "widow's cruse" effect. We know far too little, however, about the determinants and the effect of business distributions. In so far as business distributions consist of contractual interest payments there is no great problem, for in any one period these can be taken as approximately given, though there is a real long-run problem, to be discussed later, regarding the historical determinants of the volume of contractual debt. The short-run problem, however, is mainly that of the determinants of dividend distributions on shares, and of what comes almost to the same thing, withdrawals for household use from private businesses and partnerships. Here we almost have to leave the solid ground of economic maximization for as we have seen earlier it is extremely difficult to obtain any rational theory of consumption either in the household or in businesses, and for the corporation, of course, dividend distribution represents consumption of assets, particularly of liquid assets. The variety of policies in this regard is enormous. On the one hand we have the almost automatic declaration of a fixed rate of dividend (as in A. T. and T.), in which case dividends become conventionally almost the same as interest payments and the shares approximate in character to bonds. At the other extreme we have highly erratic dividend distributions which seem to obey no particular principle beyond a certain broad tendency to pay out more in times of high profits. An expanding corporation, manned by ambitious directors, may be very hesitant to "waste good money" on a crowd of invisible and useless stockholders. The same phenomenon even attends cooperative societies, where a corresponding ambition and acquisitiveness frequently manifest themselves in an unwillingness to declare rebates. On the other hand, a business which has reached its peak size and which does not seem to have many opportunities for expansion may be inclined to pay out almost all its profits in dividends to avoid the piling up of unwanted assets. The attitude of directors towards dividends may also depend on the extent to which they are themselves large stockholders. It is clear that in such a complex situation simple models of behavior are at least to be regarded with some distrust.

The effect of dividend distributions on the recipient stockholders is almost as much a mystery as the subtle forces which motivate the

policy of directors. If the market value of a stock reflects accurately the net worth of its corporation (which is, of course, by no means always the case) it is always possible for an individual stockholder to "declare his own dividends" by selling a portion of his stock. Thus, if a firm has 10,000 shares of stock and a net worth of $1,000,000, we might reasonably expect the market value of the share to approximate its "book value" of $100. If the firm earned $200,000 in profits, so that its net worth rose to $1,200,000, the book value of the stock would rise to $120, and if the market value rose accordingly the individual stockholder, by selling off 20% of his holdings at the end of the year, would still have the same value of stock as at the beginning of the year and would have realized a cash "dividend" of 20%. Of course, if all stockholders did the same, the market value of the stock would fall, reflecting a decline in the *liquidity* of the profits shared by its owner. This may be one factor forcing corporations to declare regular dividends. It is clear then that dividend policy does not directly affect the *income* of the shareholders; apart from a "liquidity discount" income accrues to a shareholder if his corporation makes profits, whether dividends are declared or not. In the one case the income is expressed as an increase in the value of his shares, in the other as an increase in his cash balance. We should not expect dividend policy, therefore, to affect the consumption patterns, except in so far as these are affected by the liquidity position or preferences of the shareholder.

In considering short-run fluctuations we cannot afford to neglect the possibility of secondary interactions among the four items in the transfer factor. The most important of these is likely to be the reaction of changes in debt or in dividends on the business-household balance of payments. Thus an increase in dividend payments may in part be offset by the fact that some of the money which is distributed simply remains in household balances without being spent. This means in effect that dividend distributions may produce temporary reactions at least on the partial velocities of households and businesses. There may be a rise in the partial velocities of business as well as a fall in the partial velocities of households: businesses do not react immediately to the decline in money stocks, and households do not react immediately to their increase. This will result in a decline in dM_b, as the money stock shifts out of business into household balances, which will in part at least offset the increase in D. The longer the period of time we take, however, the less important will this effect be. The longer the period, the larger D will be, for distributions go on all the time; the net balance of payments, however, exhibits no secular trend, as money merely shifts back and forth between the various accounts. There may be similar short-run

impact effects of changes in the volume of business-household credit. A rise in consumer credit may be reinforced in part by the fact that it enables households to get along with smaller money stocks, and there may be a shift of money away from household and into business accounts (a positive dM_b), which will augment the profit-raising effects of the credit increase. Similarly the sale of securities to households may result in a shift in money stocks from households into business balances, in so far as securities are to some extent a substitute for money and households feel themselves able to get along with smaller money stocks as a consequence of their increased holdings of securities. This will serve to offset the distributional shift from profits to wages. All these effects are likely to be more important, the shorter the time which we take, as the longer the period, the smaller the balance-of-payments item is likely to be relative to the other three.

A factor which does not get into the equations directly but which may have some importance in affecting the business-household balance of payments is the development of interbusiness credit facilities and other devices, such as mergers, which enable businesses to get along with a smaller money stock in proportion to their receipts or expenditures. This may lead to a rise in the partial velocities of businesses relative to those of households, leading to a negative balance of payments of business, which will be adverse to profits.

Distributional equilibrium: model I

With the identities of this chapter as a foundation it is possible to construct a large number of "models" of macroeconomic equilibrium and even of dynamic change, each depending, of course, on the assumptions which are made about the nature of the functional relationships among the various items of our identities. All that is necessary to do to construct such a model is to link the distributional pattern in some way to the determinants of employment and output, i.e. to the various components of "absorption." An instructive, if not too realistic, model can be constructed on the assumption that there is a definite relationship between the *proportional* distribution of the product (between wages and gross profits) and the volume of consumption and investment—or, if we like to make the more usual breakdown, between household and business absorption. We may suppose, for instance, that the greater the proportion of output which goes to wages and the less to gross profit, the greater will be the volume of household absorption and the less will be the volume of business absorption. Summing the household-absorption and the business-absorption functions gives us the total-absorption function, and the condition that total absorption shall be equal to out-

put gives us the equilibrium output and distribution. In this model the determining factor is T, the transfer factor.[5]

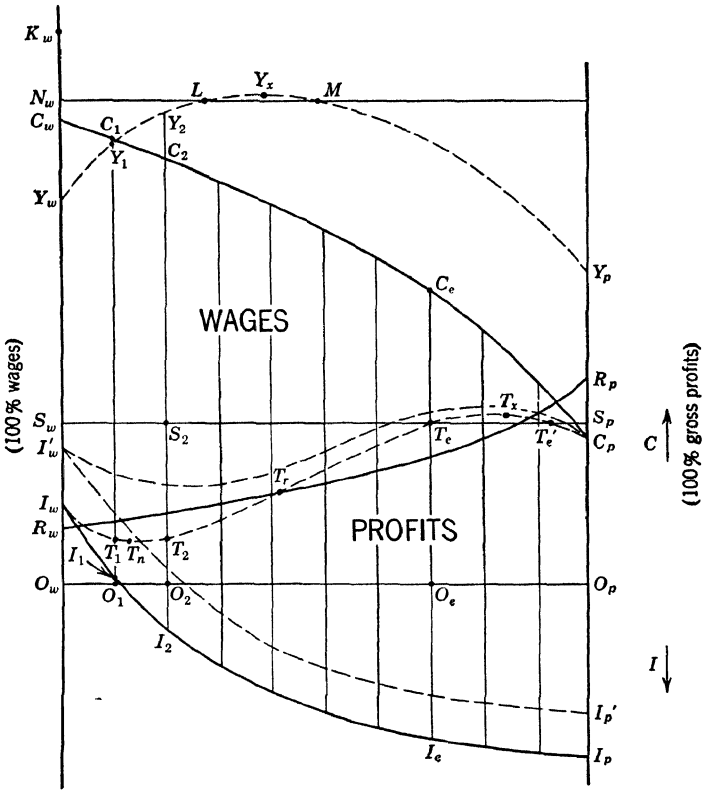


FIG. 75. Distributional equilibrium: model I.

A geometrical solution for this model is shown in Fig. 75. Along the base O_wO_p we measure the proportionate distribution between wages and gross profits; at O_w 100% of the product goes to wages, and at O_p

[5] Let C be the household absorption, I the business absorption, Y the total output, W the total wage income, P the total profit income, and T the transfer factor. Let w be the proportion of total output going to wages and p the proportion going to profits. Assuming T to be given, we have seven unknowns: C, I, Y, P, W, p, and w. We also have seven independent equations: the household-absorption function, $C = F_c(w)$; the business-absorption function, $I = F_i(p)$; the identities of composition, $Y = C + I$, $Y = W + P$, and $W = C - T$ (or, what is the same thing, $P = I + T$); and the definitions $w = W/Y$, and $p = P/Y$. The equations are therefore capable of a solution.

100% of the product goes to gross profits. The household-absorption or "consumption" function is C_wC_p; the business-absorption or "investment" function is I_wI_p. We assume in this figure that household absorption rises as the proportion of wages increases, but at a diminishing rate. We also assume that business absorption rises at a decreasing rate as the proportion of profits increases. Household absorption is measured upwards from the base line O_wO_p, and business absorption is measured downwards from the same line. Thus at a proportionate distribution O_2, representing 20% to profits and 80% to wages, O_2C_2 is the household absorption, O_2I_2 is the business absorption, and I_2C_2 $(= O_2C_2 + O_2I_2)$ is the total absorption. If we draw $O_2Y_2 = I_2C_2$, Y_2 is on the total-absorption curve, $Y_wY_2Y_p$. In the figure the total-absorption curve exhibits a maximum value of Y at Y_x. At this point the slopes of the household and business-absorption curves are equal; i.e. the rate of decline of household absorption is equal to the rate of increase of business absorption as the ratio of profits increases. This maximum is not necessarily exhibited within the range O_wO_p, though if the business- and household-absorption functions exhibit "diminishing returns" throughout there must be a maximum total absorption somewhere; the maximum may, however, lie to the right of O_p (wages negative) or to the left of O_w (profits negative).

The line I_1C_1 is now divided at T_1 so that

$$I_1T_1 = \frac{O_wO_1}{O_wO_p} I_1C_1 \quad \text{and} \quad T_1C_1 = \frac{O_1O_p}{O_wO_p} I_1C_1$$

In the figure O_wO_1 is one-tenth of O_wO_p, and represents a distribution with 10% going to gross profits and 90% to wages. I_1T_1 is 10% of I_1C_1, i.e. 10% of total absorption, which in equilibrium would be equal to total output; I_1T_1 therefore is total profits, and similarly T_1C_1 is total wages, at the distribution O_1. Likewise at distribution O_2, representing 20% profits and 80% wages, we divide I_2C_2 at T_2, where

$$I_2T_2 = \tfrac{1}{5}I_2C_2, \quad \text{and} \quad T_2C_2 = \tfrac{4}{5}I_2C_2$$

I_2T_2 then gives us total profits and T_2C_2 total wages at the distribution O_2. The locus of T_1 and T_2 is the *transfer-factor line* $I_wT_2C_p$: this divides the total output into wages (the upper half) and profit (the lower half) at each level of distribution. At any distribution, for instance, O_2, O_2T_2 is the equilibrium transfer factor, T $(= P - I$, i.e. $I_2T_2 - O_2I_2$, or what is the same quantity, $C - W$, i.e. $O_2C_2 - T_2C_2)$. In the figure T exhibits a minimum at T_n and a maximum at T_x: these need not, of course, be exhibited within the range O_wO_p.

If now the actual transfer factor is given, and is equal to O_wS_w, the equilibrium of the whole system is given at the point T_e, where the horizontal line from S_w cuts the transfer-factor curve from above. The proportionate distribution is O_e: business absorption is O_eI_e; household absorption is O_eC_e; the total output is I_eC_e; total profits are I_eT_e; total wages are T_eC_e. The dynamics and stability of this equilibrium are as follows. Consider the position O_2; the actual transfer factor O_2S_2 is larger than that which would yield a distribution given by the point O_2. Profits therefore will rise at the expense of wages; i.e. the distribution will move from O_2 to the right, towards O_e. Similarly if the actual transfer factor (O_wS_w) is below the equilibrium, as at T_x, the distribution will move in favor of wages, i.e. to the left, again towards O_e. It is evident that T_e represents a point of stable equilibrium. The point T_e', however, where the line S_wS_p cuts the transfer-factor curve from below, is an unstable equilibrium: a slight movement to the left of T_e' will generate a further swing to the left until the stable equilibrium T_e is reached; a slight disturbance to the right of T_e' will generate a swing to the right, carrying the system to O_p or even beyond, were that possible.

It is instructive now to relax the assumption that the transfer factor, T, is given, and to consider T as a variable of the system, dependent on the proportionate distribution. This means introducing another variable, T, into the equations of footnote 5, plus another equation, $T = F_t(p)$. In Fig. 75 this is the curve R_wR_p, which we may call the R-curve; in the figure this cuts the transfer-factor curve, or T-curve, at T_r, which is the point of equilibrium.

The R-curve is likely to have properties similar to those drawn. We recall (p. 252) that the transfer factor, $T = dM_b + dK_h - dK_h' + D$. The first two items are not likely to be related to proportionate distribution in any very clear way; though the immediate effects of a shift towards wages might be to increase household balances and perhaps to diminish consumer credit (or perhaps to increase it!), these effects are not likely to be important, especially in longer-run equilibria. The third item, dK_h', is likely to be closely correlated with the volume of investment: in the extreme case where all investment is financed by the sale of securities to households we should have $dK_h' = I$. The fourth item, D, is likely to be correlated with profits: the higher profits, the higher will distributions be. If all profits are immediately distributed we have $P = D$. In the extreme case where $dK_h' = I$ and $D = P$, and we can assume $dM_b + dK_h$ to be zero, we have for the equation of the R-curve $T = P - I$; this is identical with the T-curve, and the solution is indeterminate—there is a neutral equilibrium in the whole range. Normally, however, we should expect that at very low levels of profit

there will be little or no sales of new securities—as there is no investment to finance—and distributions will be confined to contractual rents and interest payments, equal, at zero profits, to O_wR_w. Once gross profits are in excess of contractual distributions, however, both D and dK_h' are likely to rise. As, with T positive, gross profits are larger than business absorption, D is likely to rise faster than dK_h'; hence the R-curve is likely to slope upwards. The alternative possibility, however, is by no means out of the question if dividend policy is conventional and not much dependent on profits, and if there is a strong tendency to finance investment by the issue of new securities.

If the slopes of the T-curve and the R-curve are similar the equilibrium, if it exists, will be highly insecure and "shiftable"; i.e. a slight change in the underlying functions will cause a large shift in the equilibrium position or will even destroy the possibilities of equilibrium altogether. Thus if the R-curve lies wholly above the T-curve there will be no equilibrium, but a constant tendency to increase the profit proportion until the model breaks down. Similarly if the R-curve lies wholly below the T-curve there will be a similar constant tendency for the wages proportion to increase. The possibility of a shift from a stable equilibrium to an apparently chaotic state of perpetual disequilibrium revealed by the model has some basis in experience. Thus a relatively small decline in the willingness to invest, which would be represented in Fig. 75 by a shift of the curve I_wI_p in an upward direction, signifying in this case that at each distributional level less will be invested than before, would shift the T-curve upwards, and especially if the R-curve is "anchored" at R_w by a relatively fixed rent-and-interest charge this shift might easily bring the T-curve wholly above the R-curve and precipitate a situation of acute disequilibrium characterized by a constant increase in the wage proportion and the disappearance of profits. It may well be that something of this nature happened in 1929, and that the distributional shift towards wages and away from profits that followed, with its accompanying unemployment, could be interpreted as an irresistible flight towards wages, caused by a shift in the investment curve, say, to the position $I_w'I_p'$ in Fig. 75 and the consequent shift in the T-curve to a position $I_w'C_p$, wholly above the R-curve. A similar decline in the consumption function could possibly create the reverse situation, bringing the T-curve wholly below the R-curve and creating a cumulating decline in wages and increase in profits. That such an event has never happened in modern history is perhaps a tribute to the stability of the consumption function by comparison with the instability of the investment function. In this model it may be observed that a

rise in the consumption function will always shift distribution towards wages, and a rise in the investment function will shift it towards profits.

This model is of interest in that it shows clearly that a shift in the distributional proportions which is due to a change in the transfer factor may have either a favorable or an unfavorable effect on output and employment, depending on which side of the maximum of the total-absorption curve the system stands. Thus, if the total-absorption curve exhibits a maximum at Y_x in Fig. 75, then to the left of this point a rise in the profits proportion and a fall in the wages proportion will have a favorable effect on output and employment. To the right of Y_x, however, the reverse is the case: a rise in the profits and a fall in the wages proportions have an unfavorable effect on output. Whether a shift in the distributional proportion towards profits and away from wages has a favorable or unfavorable effect on output depends, of course, on the relative slopes of the consumption and investment curves: if the fall in wages diminishes consumption by less than the rise in profits increases investment, then a fall in the wages proportion will be favorable to employment and output. If, however, a fall in wages discourages consumption more than the resultant rise in profits encourages investment, a fall in the wages proportion is unfavorable to employment.

There is nothing in this model, of course, to indicate either that the maximum output at Y_x is the equilibrium output, or that the maximum or any other output is the *optimum*. It could well be, for instance, that the socially most desirable level of output was, say, O_wK_w in Fig. 75 above the maximum point on the total-absorption curve. In that case no amount of distributional manipulation could give us "full employment" unless there was a change in the investment or consumption functions. On the other hand the socially optimum output might be below the maximum level Y_x curve, such as at ON_w. Then there may be two distributional systems which will yield the optimum output, represented by the points L and M. There is no reason to suppose, of course, that the equilibrium distribution will be either at L or at M.

An interesting problem arises as output approaches "capacity." In the model as set up there are no apparent limits on output short of those set by the ability of the system to absorb it: if consumption and investment increase, output increases with them. This is clearly unrealistic at high levels of output, although it may be a quite adequate description of the state of affairs at low levels where there is a large volume of unemployed resources. As the system approaches "full" employment, i.e. capacity output, a limitation on the expansion of output appears not on

the absorption side but on the production side. This may reflect itself in several ways. If the price system is flexible there will be price inflation beyond a certain level of output, which will have some sort of effect on the consumption and investment functions themselves. Even if the price system is inflexible, e.g. under price control, as output rises "shortages" will appear which will have a direct effect, especially on the consumption function. We must therefore postulate some level of output which we may call the "inflation threshold," above which the production limitations begin to be felt in modifying the consumption and investment functions. This inflation threshold may not, of course, be either the distributional equilibrium output as defined by the above model, or the "ideal" output from the point of view of society. If it comes below the ideal output, as it may well do, society is faced with a serious problem, in that the attempt to get full employment with the existing consumption and investment patterns will inevitably result in inflation, whether actual or suppressed.

Another valuable insight which the above model provides is that the effect of a distributional change on employment depends on the *causes* of such a change. We have seen that a distributional change may come about either as a result of a change in the transfer factor (or R-curve) or as a result of a change in the consumption or investment functions. Governmental intervention to affect distribution, as we shall see, may operate through either channel. The effects of a given change in the distributional ratio, however, may be very different, depending on how it has been brought about. As we have seen, a change in the distributional ratio, e.g. a shift towards wages, brought about by a change in the transfer factor may either raise or lower employment, depending on the relative slopes of the consumption and investment curves. A change in the distributional pattern which is a result of a rise in either the consumption or the investment function will almost always have a favorable impact on employment (assuming, of course, that the system is not at capacity), and a change which is due to a fall in these functions will almost always have an unfavorable effect on employment. Consequently a fall in the wage proportion which is due to a rise in the investment function will have a favorable effect on output and employment; a fall in wages which is due to a fall in the consumption function will have an unfavorable effect.

Distributional equilibrium: model II

Many models such as the above can be constructed, depending on the type of functional relationships which we assume. It will be sufficient perhaps to outline one further model, based on the assumption—which

may be more realistic than that of the first model—that consumption is a function not of proportional distribution but of the total absolute level of income (output), and that investment is a function of the absolute level of profits. This model is illustrated in Fig. 76. Output is measured along the horizontal axis, and the various components of output are measured vertically. The consumption function is given by the curve CC', showing how much will be consumed at each level of output (income). If the transfer factor is independent of output, equal to CW, the

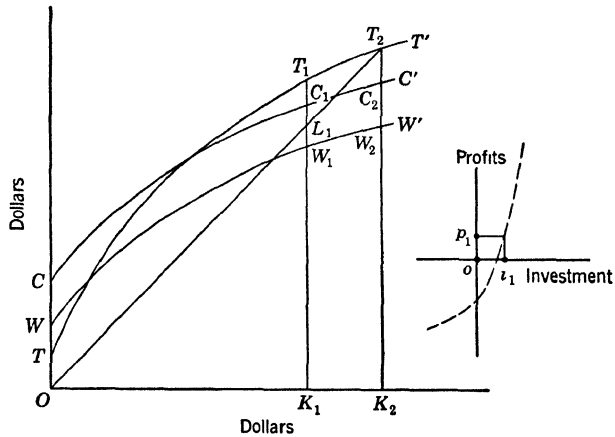


FIG. 76. Distributional equilibrium: model II.

wages curve, WW', lies below the CC'-curve by the amount of the transfer factor. This shows the absolute wage bill at each level of output, when output is at its equilibrium level. At any output OK_1, K_1C_1 is the amount consumed, K_1W_1 is the wage bill. If the 45° line OT_2 is drawn to cut K_1C_1 in L_1, then actual output is K_1L_1 (= OK_1 by construction), and profits will be W_1L_1, actual investment being C_1L_1. (As drawn this is negative, indicating actual disinvestment.) We may now postulate an investment function relating profits to planned investment, as shown in the small figure. Thus at a level of profits equal to L_1W_1 (= op_1 on small figure) the planned investment will be C_1T_1 (= oi_1 in small figure). K_1T_1 is then the planned absorption at the level of output OK_1. By means of a similar construction we can trace out the total-(planned) absorption curve TT'. In the figure we assume that investment is positive even at zero profits, but the method does not necessitate this assumption. The final equilibrium is then at the point T_2, where the 45° line cuts the total-absorption curve. Equilibrium output is

OK_2 (= K_2T_2). Investment is C_2T_2, consumption is K_2C_2, the wage total is K_2W_2, and the profit total is W_2T_2.[6]

The effect of changes in the various components of the model is illustrated in Fig. 77. The curves WW', CC', TT' have the same meaning as in Fig. 76. Now we suppose an increase in the transfer factor, resulting in a shift from wages to profits. This will give us a new wage curve, W_1W_1', lying below the old curve WW'. The transfer factor

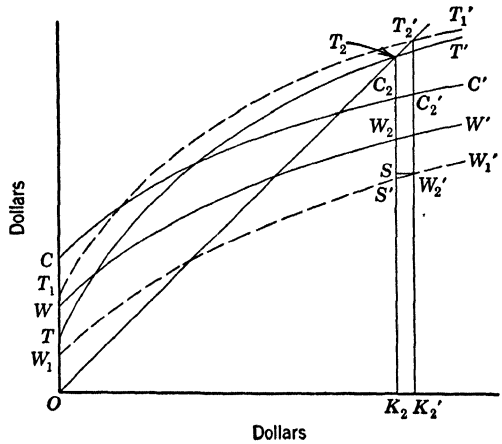


FIG. 77. Dynamics of model II: change in transfer factor.

(assumed independent of output) rises from CW to CW_1. Profits at each level of output increase, and as a result, deriving a new total-absorption curve T_1T_1' from the small figure of Fig. 76, we find that the new total-absorption curve lies wholly above the old one. The reason is that we have assumed the consumption curve to be independent of the *distribution* of income—i.e. we have assumed consumption to be dependent only on the size of income, not on its composition. The equilibrium point moves from T_2 to T_2': output rises from OK_2 to OK_2', total profits rise from W_2T_2 to $W_2'T_2'$, and total wages fall from K_2W_2 to $K_2'W_2'$. The fall in wages is somewhat less than the change in the transfer factor; if we drop $W_2'S$ perpendicular to K_2W_2, and if W_1W_1' cuts K_2W_2 in S', then the change in the total wage bill is SW_2, which is smaller than

[6] The algebraic expression of this model is as follows: using the symbols of footnote 5, we have five unknowns: C, I, Y, W, and P. There are two independent equations: $C = F_c(Y)$ (the consumption function); $I = F_i(P)$ (the investment function); and three independent identities, $Y = C + I = W + P$ (the income identities), and $W = C - T$ or $P = I + T$ (the transfer identity).

the change in the transfer factor, $S'W_2$. Similarly it can be observed that profits rise by more than the transfer factor. The effect of the shift from profits to wages can therefore be analyzed into a "transfer effect"—the rise in the transfer factor simply being added to total profits and subtracted from total wages—plus an "income effect"—the rise in income due to the shift being divided between wages and profits, mitigating the fall in wages and augmenting the rise in profits. The proportion

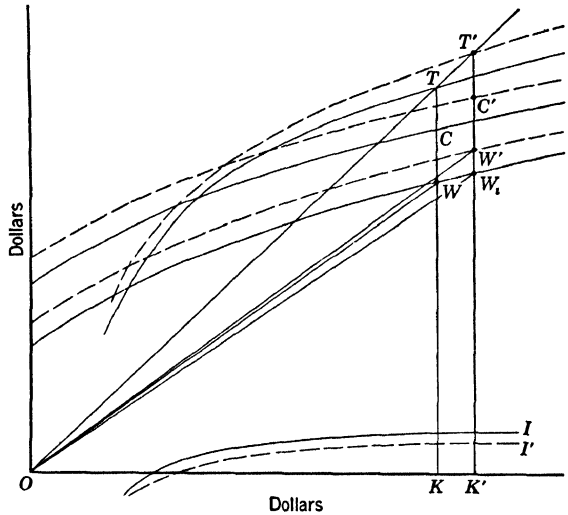


FIG. 78. Dynamics of model II: change in absorption functions.

of this division depends on the relative slopes of the consumption and investment functions. The steeper the consumption function, the greater the amount of the increased income which will accrue to wages, for the greater will SS' be.

It will be observed that on the assumptions of this model a shift in distribution towards profits and away from wages always increases output (unless, of course, capacity has been reached). This is a consequence of the assumption that consumption depends only on the size of income and not on its distribution. If consumption is lessened by a shift from wages to profits, as may well be the case, this operates to counteract the output-increasing effects of a shift from wages to profits: if the effect is large enough it will more than offset the output-increasing effects, and may produce net output-decreasing effects, as in the first model.

The effects of a change in the consumption function itself are shown in Fig. 78. The solid curves are the consumption, wage, and total-

absorption curves as before, with equilibrium output at OK ($= KT$), total wages at KW, total profits at WT, consumption at KC, investment at CT. Now we suppose a rise in the consumption curve to the dotted position at C'. The transfer factor is assumed to be unchanged, so that the wage curve rises by the same amount as the consumption curve to the dotted position at W'. This means that at each hypothetical level of output there has been a shift towards wages and away from profits: this occurs because the distribution of output depends on its composition. Hence a rise in the consumption curve, without any change in the investment-profits function, shifts distribution towards wages and away from profits, as we see from the formula $W = C - T$. The investment-output curve, therefore, falls somewhat, from I to I', and the total-absorption curve T' is lower relative to C' than T is relative to C. If the fall in the investment-output function is not too great, the total-absorption curve T' will still lie above the curve T, and the result of the rise in the consumption function will be a rise in output, as in the figure. If, however, investment is extremely sensitive to profits, so that there is a sharp fall in the investment-output curve, the rise in the consumption function could actually lead to a decline in output, if the decline in investment exceeded the rise in consumption at the equilibrium output.

A similar analysis can be applied to the problem of the effects of a spontaneous rise in the investment-profits function. In this case there is no change in the C- or W-curves, but the T-curve rises from, say, T to T'. Here there is no doubt about there being an expansion of output. The bulk of the increased output, however, will appear as profits: there will be only a small expansion in total wages, from KW to $K'W_1$. The expansion in profits, however, is substantial, from WT to W_1T'. Most of the expansion in output goes into profits. It is interesting to note that in this model a rise in output which is caused by a rise in the investment function is almost certain to bring about a decline in the *average* wage. If the equilibrium is to be stable the T-curve must cut the 45° line from above, i.e. must have a slope of less than 45°. The slope of the W-curve is even less than that of the T-curve, so it must also be under 45°. If output is assumed to measure the volume of employment, which will approximately be the case, the average wage will be proportional to tan WOK (KW/OK) in the initial equilibrium. After the rise in the investment function the average wage is proportional to tan W_1OK, which is smaller than tan WOK. If the rise in output is due to a rise in the consumption curve, however, the average wage may rise, e.g. from tan WOK to tan $W'OK$. This does not have to be the case, however, if investment is highly responsive to changes in profits,

and hence there is a large expansion both in investment and in consumption.

The dynamics of all these models is likely to be complicated by the dynamic instability of the transfer factor, and even of the consumption and investment functions themselves. They are therefore to be taken as illustrative rather than definitive pictures of the economic system, and yield us perhaps more insight than knowledge. They serve to illustrate, however, the complexities of the real world.

Chapter 15

Money, Debt, and Interest

Changes in the quantity of money

In the analysis of the previous chapter it was assumed for the sake of simplicity that the total quantity of money in the system was fixed. We must now remove this assumption, and consider the effect of changes in the quantity of money. These changes can take place in three ways. The first is the creation or destruction of money by ordinary businesses or households. This can virtually be neglected in our society. There is a certain amount of money destroyed by fire, etc. There may also be a certain amount of successful counterfeiting, and in a society which has commodity money such as gold the production of the money commodity counts as the creation of money by private businesses or households. In the western world, however, these items are small. The two other sources of money creation or destruction are government and the banking system.

Government increases the amount of money in private hands whenever it has a "cash deficit," i.e. whenever it takes in from the private sector less money than it pays out. Similarly the amount of money in private hands decreases whenever the government has a cash surplus. The amount of the increase or decrease is exactly equal to the cash deficit or surplus. Thus if government in a given period takes in 100 million dollars and pays out 90 million, there must as a direct consequence be an increase of 10 million in the money stocks held by private businesses and households. This is, of course, simply a special case of the balance-of-payments identity—that the positive (or negative) balance of one part of a bisected closed system must be equal to the negative (or positive) balance of the other part. It derives special significance, however, from the fact that government is virtually unique in its capacity to create money, without creating offsetting liabilities, and hence is in a sense possessed of an indefinitely large potential stock of money. A negative balance of payments for government, then, i.e. a net outflow of money from government to the rest of the system, is of much less immediate

consequence to government than it would be to a private party without the power to create money.

The creation or destruction of liquid assets in the form of bank deposits by the banking system has some features like government creation of money, but there are also important differences. When a bank makes a loan it at the same time creates a deposit—a deposit being simply a debt (liability) of the bank, payable on the demand of the owner of the deposit, to whom, of course, it represents a liquid asset. In the instant of the transaction the banker increases his assets by the value of the borrower's note, and increases his liabilities by an equal deposit. There is no change in the banker's net worth in the moment of the transaction, merely an equal addition to both assets and liabilities. As time goes on, however, the value of the asset (the note) grows towards maturity, whereas the value of the liability (assuming for the moment that the deposit is not withdrawn, or if transferred is merely transferred to other depositors in the same bank) does not change; the banker's net worth therefore grows, this being his profit-making process. For the borrower likewise, there is no change in immediate net worth: there is merely an equal increase in his assets (the deposit) and his liabilities (the note). If he merely retained the deposit his net worth would decrease as interest on the note accrued, and the banker's profit would simply be the borrower's loss. However, the borrower does not usually rest content with holding his deposit and making a gift of interest to the bank. He exchanges the deposit for other assets, and by the manipulation of these assets through exchange and production he hopes to increase the value of his total assets at a rate faster than that at which his liability to the bank is increasing, i.e. to make a higher rate of profit than he has to pay in interest. The question now arises, in macroeconomic terms, does the profit which a particular individual makes—if any—represent a net addition to the total of profit, or does it in part result in a deduction from the profits of other entrepreneurs? Some insight into this question can be given with the aid of the analysis of the preceding chapter.

Effects of bank credit on profits

We see from equation (10), p. 250, that the creation of bank deposits by a loan to a business has no direct immediate effect on the profits of the business community, as it merely results in an equal increase in asset and in liability totals. If we regard the banks as part of the business community, then the direct effect of a note-deposit creation is to increase the volume of interbusiness debts. As these, however, cancel out when we add up all business balance sheets, they are not in the profits identity, and their volume has no *direct* effect on profits. Even if we

regard banks as something outside the business community, say, as a kind of "household," the result of a loan-deposit creation is an increase in the quantity of "money" (bank deposits) held by businesses and an equal increase in business indebtedness to "households." In terms of the profit identity (p. 250) dM_b increases, but is offset because dK_h increases by an equal amount so that there is no change in total profits.

An expansion (or contraction) of bank loans may, however, have substantial indirect effects on profits. In the first place an expansion of bank loans is likely to augment the real capital in the hands of businesses, as the liquid assets created by the loans are spent to build factories, machines, stocks of raw materials or other inventories, and so on. In terms of the profit identity, dQ_b has increased. The increase in physical capital may, however, in part be offset by a shift of the newly created bank deposits into the hands of households. If we include banks in the business community a shift in the total of deposits out of business into household accounts is equivalent to an increase in dK_h', which diminishes total profits. If we exclude banks from the business community a shift in deposits into household accounts is reflected in a decline in dM_b, which also has the effect of decreasing profits. An expansion of bank loans by making businesses generally more liquid may also have the effect of encouraging the extension of credit by businesses to households, e.g. in installment credit. This, as we have seen, has a favorable effect on profits. There may also be a reduction in securities sold by businesses to households: if businesses can borrow from banks there will be less need to raise money by the sale of new securities to households. This reduction also has a favorable effect on profits, for as we have seen the sale of securities to households has a direct effect in diminishing profits; because of this effect the negative item dK_h' may be smaller than it otherwise would be.

We thus see that, although an expansion (or contraction) of bank credit has no direct effect on profits, it may have a substantial indirect effect through the impact of the resulting changes in the *structure* of balance sheets on the various items in the profits identity. The effect on employment likewise is a complex, indirect matter. In so far as the expansion of bank credit permits an increase in the physical capital of business which otherwise would not have taken place, it will tend to increase employment and output, provided that unused resources are available. In so far as there is a transfer of bank deposits into the hands of households, resulting in an increase in household-held deposits, there may be an expansion in household purchases from businesses, and if unemployed resources are available this will not result in a decline in business inventories, but rather in an increased production to replace

the more rapidly depleting stocks of goods. This consumption effect will in that case also have an employment-increasing effect.

All these complex effects can be illustrated by means of a simple diagram such as that used in the preceding chapter. Thus, in Fig. 79, AB represents the increase in physical capital held by businesses (dQ_b), BC is the increase in household indebtedness to businesses (dK_h), CH is the increase in business indebtedness to households (dK_h'), HE is the increase in business money stocks (dM_b), EF is business distributions (D), and BG is household purchases from businesses. Then BF is the transfer factor, AF is total profits, FG is total wages, AG is total income. Possible effects of an increase in bank credit are shown in the

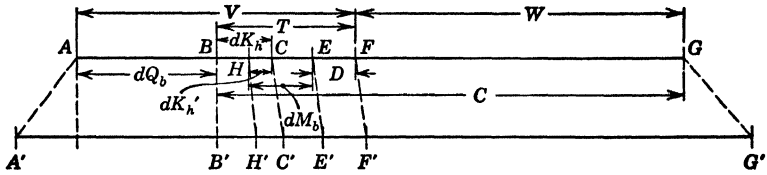


FIG. 79. Distributional effects of banking system.

lower line, $A'G'$. AB expands to $A'B'$, BG to $B'G'$, resulting in an increase in total income (output) to $A'G'$. What happens to the transfer factor BF depends, of course, on the relative strengths of the impact on the various components. Favoring an expanded transfer factor are possible increases in consumer credit, possible decreases in the sale of securities to households by business, and also possible expanded dividend distributions as a result of a more liquid position on the part of business. Favoring a contracted transfer factor is a possible loss of money from business to household balances. On the whole it would seem that the expansive factors are likely to predominate, and the transfer factor may therefore expand, say, from BF to $B'F'$. Unless the expansion in household purchases is greater than the expansion in the transfer factor the absolute amount going to wages will decrease. If, however, the expansion in household purchases (from BG to $B'G'$) is greater than the expansion of the transfer factor (from BF to $B'F'$), there will be a net absolute expansion of wages from FG to $F'G'$. The likelihood of an absolute expansion of gross profits (from AF to $A'F'$) is considerably greater, both because of the probable expansion of the transfer factor and also because an increase in bank credit is likely to have a more direct and immediate effect in increasing business investment than it is in increasing household purchases. It is difficult to make any very significant propositions in regard to the relative shares; there seems to be a certain

predisposition in favor of an increase in the relative share of profits at the expense of wages, but circumstances are quite conceivable in which the reverse would be true.

Effects of non-bank credit

The expansion or contraction of bank credit is, of course, only a special case of the general expansion or contraction of credit. Its peculiarity lies in that an expansion of bank credit by means of advances or loans results in a double creation of credit: a credit asset and a debt liability are written into the balance sheets both of the banker and of the borrower, and one of these (the bank deposit) is a liquid asset, capable of easy transfer to other accounts. When a bank buys an already existing security, or when a private person lends money on a note or mortgage, there is only one debt-credit item created in the balance sheets. Thus, if a man borrows $5000 from a bank, in effect he first "issues a security" (his note or promise to pay $5000) and "sells" this to the bank in return for another $5000 "security" (the bank deposit) which the bank has issued. The immediate result of the transaction is that the bank now has a $5000 asset (the note) which it did not have before, and which is listed in its balance sheet under loans or advances, and has a $5000 liability which it did not have before (the borrower's deposit); likewise the borrower has a $5000 liability (his note) and a $5000 asset (the deposit). If now a bank buys an existing bond for $5000, it adds $5000 to its assets (listed under investments) and adds $5000 to its liabilities in the shape of the deposit created for the purchase of the bond. The former owner of the bond, however, has merely made an exchange, or asset transfer: nothing is added to his liabilities, nothing to the *total* of his assets, but the bonds item in his balance sheet is diminished by $5000 and the bank-deposits item is increased by $5000. Similarly, when a private person or business borrows $5000 from another, the borrower issues a "security" (his note or promise to pay) which he sells to the lender: the lender diminishes his cash by $5000 and increases his notes receivable by $5000; the borrower increases his cash by $5000 and increases his liabilities by the $5000 of notes payable. The sale of goods on credit has somewhat similar effects on the balance sheets; in effect the goods are purchased by the creation of an informal kind of security known as accounts payable. When $5000 worth of goods is sold on credit the seller merely registers an asset transfer in his balance sheet, $5000 being deducted from the goods item and $5000 being added to the accounts-payable item. The buyer registers an equal addition to both assets and liabilities, $5000 worth of goods being added to his assets, and $5000 in accounts payable added to his liabilities.

The effects on employment and distribution of the creation of forms of debt (credit) other than bank loans can also be analyzed by means of the transfer-factor apparatus. Thus the purchase of existing securities (bonds) by banks increases the liquidity of non-bank firms and households at the expense of bank liquidity. If this represents an actual diminution of liquidity preference on the part of banks, i.e. if the purchase of the security does not merely mean that bank loans are diminished by an equal amount, there is a net increase in the liquidity of non-bank firms and households. If this increased liquidity leads to an expansion of business accumulations and household purchases there will be an increase in the national income. The effect on distribution depends, as before, first on the composition of this increase—increased business accumulation going to profits, household purchases to wages—and also on the effects on the transfer factor. The extension of credit among businesses does not have any direct effects on national income or its distribution, as it merely results in an increase in interbusiness debts which cancel out in the addition of balance sheets. It may, however, have considerable indirect effects. It may, for instance, permit a larger volume of business accumulation than otherwise might have been the case, as it results in a shift in the control of resources into more venturesome hands. This will increase both national income and profits. It may also have an effect on relative business-household liquidities, for an easing in interbusiness credit means that businesses have to hold less cash, and hence there may be a balance-of-payments shift out of business into household cash holdings. This diminishes the transfer factor and lessens the favorable impact on profits.

The extension of credit from businesses to households, as we have seen, increases the transfer factor directly, and so has a favorable impact on profits. By permitting households to get along with smaller money stocks it may also have a secondary balance-of-payments effect, money being shifted from household to business holdings. This likewise has a favorable effect on profits. It may be also that easier credit from business to households will result in a net expansion of household purchases. This will increase national income and will raise the proportion that goes to wages.

The increase in business indebtedness to households, e.g. through the sale of securities to households, has somewhat reverse effects. There is again a direct effect on the transfer factor, unfavorable to profits. There are also likely to be indirect effects: in so far as the securities are a substitute for cash in the minds of the householders, there is likely to be a net balance-of-payments shift out of households into business holdings of cash, increasing profits at the expense of wages. In so far as households

regard the securities as substitutes for goods holdings, however, there may be a diminution in household purchases of goods, leading to a decline in national income, but again shifting distribution towards profits. In so far as the sale of securities to households permits and encourages businesses to increase their real assets, however, national income will be increased, with a further shift towards profits and away from wages. In spite of the direct unfavorable effect on the transfer factor, therefore, the sale of securities to households, through its indirect effects, may shift distribution towards profits.

The significance of debt and interest

The above analysis throws a good deal of light on the significance of the debt phenomenon in the economic system, and also on the nature and significance of interest. The principal function of the financial system is to separate the ownership of real capital from its control, that is, to enable people to administer real capital without owning it, and to own it without administering it. When we add up all the balance sheets of a society, as we have seen, the financial assets and liabilities (debts and securities) cancel out, and we are left with the total of real capital on the one side and the total of net worths on the other. The tangle of financial relationships and financial assets that lie between real capital and its correspondent net worth (ownership) has the effect of organizing real capital into complexes under unitary control which do not correspond with the complexes of ownership. If there were no financial system, no system of debt, credit, and securities, no financial assets and liabilities, then every balance sheet would consist simply of real assets on one side and net worth on the other: every person would have to administer the real assets which he owned, no matter how incapable he might be at such a task. The financial system has grown up because the *personal ownership* of real assets is a result of a long process of historic accident, through individual saving and inheritance, the structure of which need bear no relationship to the structure of administrative abilities or economic needs. A farm, for instance, may by inheritance come into the possession of an elderly spinster, quite physically incapable of operating it. In the absence of any financial system, however, she would either have to operate it herself or not operate it at all. If, however, she holds a mortgage on a farm the "real" ownership, i.e. the equity in the farm, may still be in her hands, but the operation of the farm is in the hands of an "owner-operator" in whose balance sheet the farm appears as an asset, counterbalanced by the mortgage liability. The operator, even though he has legal possession of the farm and is free to operate it

pretty much as he pleases, does not have equity in the farm; his net worth is not derived from the value of the farm. By contrast the mortgage holder's net worth is derived from the value of the farm, but she takes no responsibility for operating it, apart from the certain residual responsibility which is implicit in the right of foreclosure.

The creation of explicit debts or financial asset-liabilities is not, of course, the only way by which this separation of ownership from administration can be accomplished. Real capital may be *rented;* that is, instead of the owner relinquishing the right of administration for an indefinite period, he relinquishes this right, or certain specified rights, to the renter for a definite period of time. A rental contract, however, is theoretically equivalent to a short-term asset-liability: the renter adds to his assets the value of the use of the rented property for the specified period, and to his liabilities the contractual rent to be paid. The owner of the rented property should, in strict theory, subtract the value of the use of the rented property from his assets, and add the value of the rental contract. We can therefore clearly extend the concept of the financial system to include rental contracts.

Interest and employment

In the period that stretched between Böhm-Bawerk and Keynes economists paid a great deal of attention to the rate of interest as a determining and also a regulative factor in economic changes. The name of Wicksell [1] is particularly associated with this emphasis, and the notion that general economic change, whether of prices or of outputs, is conditioned mainly by the divergence of the market rate of interest from some "natural" rate is important in the writings even of men as divergent as Irving Fisher,[2] Hawtrey,[3] and the Keynes of the *Treatise.*[4] A natural corollary of this theoretical emphasis on the rate of interest as a regulative factor was the emphasis on central banking policy as the key to the regulation of the business cycle. Even as late as the mid-thirties economists were optimistic about the possibility of controlling depressions by means of the open market and rediscount policy of central banks. However, the experience of the great depression was an all-too-clear demonstration of the inadequacy of these devices in the face of a serious depression, and the interest of economists turned away from

[1] Knut Wicksell, *Interest and Prices* (translated by R. F. Kahn, Macmillan, London, 1930).

[2] Irving Fisher, *The Purchasing Power of Money* (Macmillan, New York, 1912.)

[3] R. Hawtrey, *Currency and Credit,* 3rd Ed. (Longmans, Green, London, 1928).

[4] J. M. Keynes, *A Treatise on Money* (Macmillan, London, 1930).

monetary and banking policy to fiscal policy, the revenue-expenditure system of government being rightly regarded as the most powerful mechanism available for controlling payments and income.

With the analytical apparatus now at our disposal we can perhaps give some appraisal of the significance of the "rate of interest" for employment and distribution theory, though it is not possible at this stage to appraise its exact quantitative significance. It must be made clear at the outset that there is no such thing, strictly speaking, as *the* rate of interest. Not only is there an elaborate *structure* of interest rates on various types of securities—short-term and long-term, secured and unsecured, etc.—but also the rate of interest itself is not an objective "price" determined in the market as the price of, say, wheat is determined, but is rather a subjective estimate of the profitability of a security derived from its present price and from its expected returns. There are cases, of course, such as gilt-edged government securities and short-term bank loans in which the returns are so certain that the rate of interest takes on a disguise of objectivity. But as in strict theory all future events are uncertain, so all interest rates, because they depend on assumptions about future events, are likewise uncertain and therefore subjective. The same security at the same price may easily represent different rates of interest for different people, according to their different expectations of the future returns from it.

The situation is further complicated by the imperfections in the market for securities, whether old or new issues. No individual or firm can borrow or lend indefinite amounts at a constant interest rate: that is to say, no one can buy or sell unlimited amounts of securities having certain expectations at constant prices. These imperfections in the market may not be significant for every individual, for they may develop only beyond the point at which he himself ceases voluntarily to expand his borrowings or lendings. But for many, perhaps for most individuals and firms, these imperfections are of great importance, and as Dr. Albert Hart [5] has pointed out, capital rationing may be a determining factor limiting the scale of operations of a firm or household.

In the writings of the Wicksellian school the "rate of interest" has often been used as a symbol of the general state of the capital market or, more exactly, of the loan and securities market. The rate of interest has been regarded as an important determinant of the volume of real investment, and also of the total amount of borrowing and lending. It is important, however, to discuss this problem explicitly in terms of the conditions in the loan and securities markets, for the structure of rates

[5] A. G. Hart, Anticipations, business planning, and the cycle, *Quarterly Journal of Economics*, vol. LI (February 1937), p. 3.

of interest or of security prices forms only a part of these general conditions and cannot without danger of considerable error be taken as a symbol of the whole.

It is important to distinguish at the outset between two sets of conditions: the condition of the market for "old" securities, and the condition of the creation (issue) or withdrawal of securities. The relation of these two problems is similar to the relation of the market-price problem to that of the long-run equilibrium of production and consumption in the case of a commodity. The so-called "liquidity preference" theory of interest associated with the name of Keynes is a theory of the market price of existing securities. It differs in no way from the general "liquidity preference" theory of market price as we have developed it earlier (pp. 53–59). The rate of interest on securities with given expectations is an inverse function of their price: the more we have to give for any given security, the less the rate of interest it will earn for us. Hence, with a given roster of securities, with given expected returns, the problem of the rate of interest and of the price of these securities is identical. Indeed, the so-called "rate of interest" is simply a method of quoting the prices of securities on the assumption of known returns from them, and it is the price of securities, not the rate of interest, which is determined in the actual securities and loan market. Loans, of course, are simply a form of security—the lender issues, and the borrower buys, a "bond" in the form of a promissory note.

The price of securities, like any other price or price level, may be expressed in terms of the market identity,

$$p_s = \frac{Mr_s}{Sr_m}$$

where p_s is the price level of securities, S is an index of the quantity of securities outstanding, M is the quantity of money in the market, r_s is the preferred security ratio, r_m the preferred liquidity ratio. A rise in liquidity preference, expressed as a rise in r_m, and perhaps in a fall in r_s, results in a fall in p_s, which is the same thing as a rise in the rate of interest.

A phenomenon which has been much neglected, both in theoretical and in empirical studies, is the interrelationship between the commodity and security markets. We do not know the extent to which the people who operate in these markets are specialized, operating in one market or in one set of markets only, or the extent to which operators shift from one market to another as they see opportunities. The fact that there have been times when an inflation in the security markets has not been accompanied by an inflation in commodity markets (as in 1927–1929),

and other times when a commodity inflation (as in 1936–1937) has not been accompanied by a securities inflation, indicates that the two sets of markets may be fairly well separated institutionally. Nevertheless there must be some connections between them. This dependency—or lack of it—can be expressed by means of functional relationships among the various preference ratios. Suppose, for instance, that there are only three kinds of assets—commodities, securities, and money, with preference ratios r_c, r_s, and r_m. The various preference ratios are linked by the general relation that their sum must be unity. Within this general relationship, however, a wide variety of internal relationships is possible. Thus a rise in r_m might be offset by an equal fall in r_s; in such case the price change would be confined entirely to the securities market. If, however, a rise in r_m resulted in a fall in both r_c and r_s, any movement of prices due to changes in preferences would be felt in both sets of markets. There is an implicit assumption in much of the Keynesian exposition of the efficacy of low interest rates that the securities and commodity markets are in fact highly independent, so that, for instance, a rise in security prices (i.e. a fall in the rate of interest) can be brought about by pumping money into private balances or by stimulating a change in liquidity preference without any repercussions in the way of rising commodity prices. As we noted in Chapter 13, the interconnectedness of markets does not perhaps affect the general validity of the Keynesian analysis of the usefulness of inflation, for even if an inflation in the securities market does spill over into commodity markets, this will raise the marginal efficiency of capital, if only in money terms, and so will still stimulate investment, at least in the short run.

The long-run theory of interest

We cannot, however, rest content with the liquidity preference theory of interest as such, because it deals essentially only with the market price of existing securities. The problem of "production and consumption" in the securities market is a different, and a more difficult, problem. From a balance-sheet point of view the creation of securities is identical with the process of "getting into debt." It involves in the first instance writing up equal items on both asset and liability sides of the balance sheet. Thus I may write a promissory note for $1000: the note itself is an asset; the promise to pay is a liability. The mere act of creating a security would not, of course, have any significance unless the asset represented by the security was salable, i.e. exchangeable into money and hence into other things. A borrowing operation, however, can be broken down into two fundamentally different processes: one, the creation of a security, i.e. an asset-liability, and the other, the exchange

(sale) of the asset-security for money. The confusion of these two essentially different processes—understandable in view of the fact that they are frequently simultaneous, and that it is the possibility of the second that gives rise to the first—has nevertheless contributed greatly to the obscurity of this whole subject. Lending, by contrast, is a single process of exchange, or asset transfer: the lender does not create any security, he merely exchanges money for the security which the borrower creates. In his balance sheet there is no change in the total of assets or of liabilities in the first instance, merely a transference from the cash item to the "notes receivable" or some other item representing a value of securities.

How, then, do securities come to be created (produced) or destroyed (consumed)? What is it that makes people "get into debt" (issue securities) or "get out of debt" (withdraw securities)? By getting into debt an individual (or corporation) increases the total both of its assets and of its liabilities. The liabilities continually increase in value as interest accrues; hence they will not be incurred unless the individual believes that his asset structure can be manipulated, through exchange and production, to yield a rate of growth of assets higher than the rate of growth of liabilities. The rate of growth of assets is the "average efficiency of capital," in the Keynesian phrase; the rate of growth of liabilities is the rate of interest. Clearly the individual's willingness to increase his liabilities depends on the expectations which he has of being able to increase his assets. In the absence of uncertainty he would presumably increase his liabilities (and assets), i.e. would get into debt, up to the point at which his expected marginal efficiency of capital was just equal to the marginal rate of interest. In the presence of uncertainty, of course, the "principle of increasing risk" comes into play, and the marginal efficiency of capital must be discounted by a risk factor which increases with the amount of borrowing. The amount of borrowing, however, with *given* conditions on the assets side of the picture, will be inversely related to the rate of interest on liabilities. This in turn depends on the price at which given securities can be sold in the securities market.

The link can now be made between the short-run determination of interest rates or security prices by liquidity preference, etc., in the securities market, and the long-run theory of interest which equates it to some kind of rate of return on real capital. If the market rate, as given by the prices of existing securities, is much below the "natural" rate, as determined by the rate of return on asset combinations, it will pay asset holders (entrepreneurs) to create and sell securities in the market in order to increase their asset holdings. The total volume of securities

will increase; hence (by the market formula) in the absence of compensating changes their price will fall, which means that interest rates will rise. This process will go on until there is no further incentive to increase the volume of debt. If on the other hand the price of securities is low, and rates of interest are high, entrepreneurs will find themselves too much indebted, debts will be paid off, the total volume of securities will fall, and this will raise security prices and lower interest rates.

If, therefore, there was a reasonably competitive market in all securities, the concept of a "long-run" rate of interest would make a good deal of sense, and it would bear a close relationship to the long-run rate of return on "real assets," i.e. the "rate of profit." Two things, however, weaken the usefulness of such a long-run concept. One is the very rapid fluctuations in the rate of profit itself and, still more, in the expected rate of profit on new asset combinations. Such fluctuations do not, of course, necessarily invalidate the long-run concept, or even the concept of a disequilibrium between market rates of interest and rates of profit. They do, however, weaken the usefulness of the concept in the interpretation of actual market behavior. A more fundamental weakness, however, arises from the imperfection of the securities market. The market for existing securities of well-established issuers is, of course, very highly organized and is about as close to a perfect market as can be found. The market for the securities of individual issuer who are not well known, however, e.g. for promissory notes and mortgages, is very far from perfect. The buyers of these instruments (banks, loan associations, etc.) operate for the most part in a highly oligopolistic market, exhibiting price leadership and a strongly conventional price structure. There is something sacred about 5%, and rates of interest on bank loans and mortgages strongly resist pressure either up or down. As a result in this particular field the creation of securities is blocked not by a rise in the rate of interest on them but by their inability to find a purchaser at the conventional rate of interest. Similarly, an expansion in the volume of this type of securities may be developed not by any decline in interest rates but by an increased willingness to buy them on the part of banks and loan companies.

The impact of monetary policy

With these considerations in mind we can now review briefly the effect of what is usually called "monetary" policy on employment, output, and distribution. In fact the term "monetary" policy is something of a misnomer, especially when monetary is distinguished from "fiscal" policy, for it is fiscal policy, in modern times at least, that has mainly determined the quantity of money. What is called monetary policy is

in fact "financial" policy, i.e. actions of government which affect the structure and magnitude of financial instruments (debts). The term monetary in this connection is used because the most important element which is subject to control is the financial structure of the banking system, and bank deposits are, of course, regarded by their owners as liquid assets, i.e. as money. Nevertheless we cannot safely discuss the control of bank assets and liabilities except as part of the general system of debt, for a "line of credit," even of ordinary commercial credit, may be equivalent to a liquid asset as far as its effect on behavior is concerned.

The monetary authorities—by which is meant the national treasury and the central banking system—can unquestionably affect the general conditions of the markets in financial instruments, and hence affect the creation and destruction of these instruments. They can do this either by direct regulations and prohibitions, which impose certain restrictions on the structure of individual balance sheets, or by intervention in the market, affecting the structure of private balance sheets through purchases and sales on public account.

The legal reserve requirements of banks and the restrictions on consumer credit are examples of direct regulation. The "bank rate" of the Bank of England, because of the custom of using it to set interest rates on commercial bank loans, was in effect, though not in law, a direct regulation of the terms on which the banks would lend. Open market policy is an example of regulation by public purchase and sale of securities. Just as a monetary authority can regulate the price of gold by buying and selling it at a price which the authority itself has fixed, so the price of securities—especially, of course, of government securities—can be regulated if the monetary authority buys and sells them at a fixed price. At the moment the United States might almost be said to be on a virtual "government bond standard," for the policy of the Federal Reserve System is to peg the price of government bonds by open market operations.

The object, however, of all "monetary" policy is the same: to expand or to contract the total quantity of securities (i.e. credit), or to change its composition in such a way as to serve the ends of social policy. An expansion of credit may reasonably be expected to increase employment in a depression, as we have seen. A contraction of credit, or at least checks on its expansion, may check inflation if that is a problem. Nevertheless the problem cannot be handled in simple, quantitative terms. The key to the great dilemma of employment policy—how to get and keep full employment without inflation—may well lie in the *qualitative* control of the structure of credit. To turn this key, however, is somewhat beyond the scope or power of this book.

Chapter 16

The Economics of Government

Government impinges on the economic system in two ways. There is first the regulative aspects of government, by which it applies its coercive power to prohibit private persons from performing illegal acts of production, consumption, or exchange. Antitrust laws, labor laws, minimum-wage laws, price-control laws, transportation laws, and the like usually fall under this category. The second aspect of government may be called its positive aspect, in which it acts itself like an economic organism, performing acts of exchange, production, and consumption. It is with this second aspect that we shall mostly be concerned.

Positive aspects of government

The positive economic acts of government can be further classified into *exchanges* and *transfers*. Acts of production and consumption by government may also be important because of their long-run effect on the government's ability to exchange, but from the point of view of the direct impact of government on the private sector of the economy the structure and transformations of government-held assets can largely be neglected. A government *exchange* is a transaction in which there is a transfer of one kind of asset from a private party to government, and a reciprocal transfer of another kind of asset from government to the private party. These exchanges may be further classified into government sales, in which liquid assets are taken from private parties and non-liquid assets are given in exchange, and government purchases, in which liquid assets are given to private parties by government in return for non-liquid assets. There may also be barter transactions of government, though these are quantitatively unimportant.

Government exchanges must be sharply distinguished from government *transfers*. An exchange represents a double transfer of assets; a transfer is simply the removal of a single quantity of assets from one account to another. Government transfers are almost always of liquid assets, although there may be some transfers of other kinds of assets, e.g. relief in kind, or some forms of capital levy. We will assume for the

purposes of this discussion that all transfers are of money of some kind. Positive transfers are defined as those in which money passes from government to private ownership, e.g. relief and subsidy payments. Negative transfers are those in which money passes from private to government ownership. Tax payments, of course, are negative transfers, and virtually all negative transfers are tax payments.

One or two difficult cases of definition require attention. Wage and salary payments by government are not transfer payments, but exchanges, even though the asset transferred to the government in exchange for the payment is a somewhat intangible one. We will assume, however, that when labor is hired the worker produces "work" and sells this work to the purchaser, the value of the work (as measured by the wage) being part of total output-value or income. Government payments of interest on the national debt, however, represent a more difficult problem. It certainly seems to be stretching the concept of production very far to suppose that the holders of the national debt, by the mere fact of their ownership, "produce" income which they then "sell" to the government in exchange for the interest payment. Consequently the statisticians of national income usually regard interest payments on the national debt as a pure "transfer payment," i.e. something which affects the distribution but does not affect the magnitude of the national income. If this point of view is held, interest payments on the national debt simply represent "negative taxes": they can be treated simply as a subsidy to the holders of the debt. As such a large proportion of the national debt is held by banks and insurance companies this means, in effect, a subsidy to depositors and policyholders. The extent to which this represents an actual redistribution of income is discussed later.

Effects of government absorption of product

We will consider first the impact of positive government activity on output and employment. The most direct effect of government results from the fact that government is an absorber of the product. In this respect the government can be treated like a peculiar kind of household. We have already analyzed this effect in Chapter 13 (pp. 237–238). If a stable private-absorption function can be postulated there will be a "government-absorption multiplier" dependent on the marginal private "propensity to absorb," and if we are given the private-absorption function it is easy to calculate how much government absorption would be necessary to give any desired level of output, as on p. 237. An increase in government absorption which produced no change in the private-absorption function might be called a "neutral" increase. The multiplier concept is interesting only on the assumption of stable private-

absorption functions; nevertheless the assumption of "neutral" changes in government absorption is highly unlikely. An increase in government absorption may increase or may decrease the private absorption at each income level, i.e. may raise or lower the private-absorption curve. In the first case the effect of government absorption is reinforced, and in the second is offset, by the induced change in the private-absorption function. The case of "complete offset" is interesting, in which an increase in government absorption produces an exactly offsetting decline in private absorption, and hence produces no change in total absorption and therefore no change in output. This might well be called the "Ricardian" case, as it corresponds to Ricardo's general view of the effect of an increase in government activity. On the assumption of continual full employment or, more exactly, a constant level of output, there must always be complete offset, an expansion in government activity being counterbalanced by a decline in private activity.

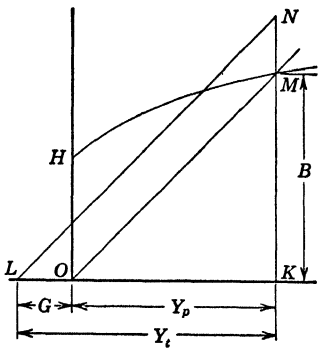


FIG. 80. Effects of government absorption.

It should be observed that complete offset does not necessarily follow from a balanced budget. Even if the value of all government absorption is equal to all net positive transfers, so that there is no direct increase or decrease in private net worths as a result of government activity, and an increase in government absorption results in an equal decrease in private net incomes, it can be shown that an increase in government absorption results in an increase in total absorption. This is to be expected, since if we regard government as a peculiar kind of household an increase in government absorption has much the same impact on the economic system as a spontaneous increase in household absorption, which under conditions of elastic supplies (undercapacity output) will increase output even though the household has a "balanced budget," e.g. finances its increased absorption by the sale of securities.

A simple model to illustrate the above proposition is shown in Fig. 80. Private income (Y_p) is measured from O along OK; government absorption (G) is measured in the reverse direction along OL. Then if Y_t is total income (output), we have

$$Y_t \equiv Y_p + G \tag{1}$$

on the assumption that government operates with a balanced budget

and hence makes no purely fiscal contribution to private incomes. Thus, if $Y_p = OK$ and $G = OL$, $Y_t = LK$. We now postulate a stable private-absorption function, shown by the curve HM, relating total private absorption (B) (consumption and investment by both businesses and households) to *private* income Y_p: i.e.

$$B = \psi Y_p \tag{2}$$

This is more realistic than to assume that private absorption is a function of total income. Private income is, of course, net after taxes have been paid. In output equilibrium we must have

$$Y_t = G + B \tag{3}$$

i.e. the total output must be equal to total willing absorption. It follows that in equilibrium

$$Y_p = Y_t - G = B \tag{4}$$

In Fig. 80, therefore, if we draw the 45° construction line from O, OM, to cut the private-absorption curve in M, $MK = OK$ is the equilibrium level of private income. If a similar line is drawn from L, LN, cutting KM produced at N, $KN = LK$ is total income, which is clearly equal to B (KM) plus G (MN).

Effects of surplus or deficit

This case can now easily be extended to include that in which the government has a deficit or surplus. There are some complex problems involved in the structure of government deficits and surpluses which we will not discuss here: a deficit will be simply defined as a situation in which government takes in less than it pays out, the difference representing an increase in private income. Similarly a surplus is the excess of what government takes in over what it pays out, and hence represents a deduction from private income. A surplus deducts from, and a deficit adds to, private assets. Thus, if F is the government deficit (a surplus would simply be a negative value of F), our identities and equations (1), (3), and (4) must be revised as follows:

$$Y_p \equiv Y_t - G + F \tag{5}$$

in equilibrium

$$Y_t = B + G \tag{6}$$

therefore

$$Y_p = B + F \tag{7}$$

Identity (5) states that private income must be equal to the value of total output, less that part of it which is absorbed by government, plus the government deficit, the government deficit representing additions

to private assets in the form of securities or money. Equation (6) simply states as usual that total income must be absorbed in equilibrium, either by the private sector or by government. Equation (7) is derived from the other two. A graphic representation of this model, again on the assumption of a stable private-absorption function, is shown in Fig. 81. As in Fig. 80 private income is measured to the right from O along OK, government absorption to the left along OL. The vertical axis at

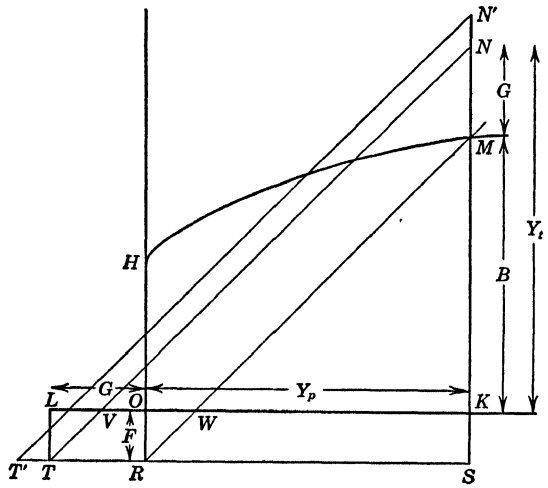


FIG. 81. Effects of government deficit.

O is projected downward to R, so that $OR = F$, the government deficit. HM as before is the private-absorption curve. Then, if a 45° line RM is drawn from the point R, M gives the point of equilibrium, and OK is the equilibrium private income, Y_p, which by construction is equal to MS, or $MK + SK$, i.e. $B + F$: thus equation (7) holds. If now a horizontal line RT is drawn so that $RT = OL = G$, the government absorption, and another 45° line TN is drawn from T to cut KM produced in N and KL in V, $KN = KV$ is the total income, Y_t; we see that $KN = KM + MN$: this is equation (6). Also if RM cuts LK at W, $Y_p = OK = VK - VW + OW = Y_t - G + F$, which is the identity (5).

The effect of changes in government absorption or deficit can easily be seen from these diagrams. Take first an increase in government absorption from RT to RT', without any change in the deficit or in the private-absorption function. Repeating the previous construction in Fig. 81, we see that the effect is simply to increase total income from KN

to KN', leaving private income and private absorption unchanged. That is to say, an increase in government absorption simply increases the total product (income) by an equal amount under these assumptions, if there is no change in the deficit, i.e. if the increase in government absorption is paid for by increasing net taxes (transfers). This, of course, as in all these models, holds true only under conditions of idle and easily available resources.

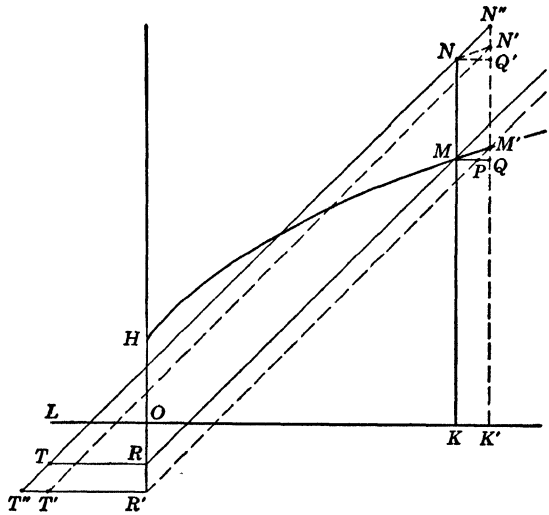


FIG. 82. Effects of change in deficit.

The effects of a change in the deficit, without any change in government absorption, are shown in Fig. 82, where the corresponding symbols have the same meaning as in Fig. 81. We suppose an increase in deficit from OR to OR'. This shifts the equilibrium point from M to M', increases private income from OK to OK', and increases total income (output) from KN to $K'N'$. The increase in private income depends on the marginal private propensity to absorb. If dY_p is the increase in private income, dF the increase in deficit, and k the marginal private propensity to absorb, we have:

$$dY_p = MQ = MP + PQ = RR' + QM' = dF + k\,dY_p$$

whence

$$dY_p = \frac{1}{1-k}\,dF = m_p\,dF \tag{8}$$

The increase in total income is $Q'N'$: this is equal to QM', as $MN = M'N' = G$. If dY_t is the increase in total income we have

$$dY_t = QM' = k\,dY_p = \frac{k}{1-k}\,dF = m_t\,dF \tag{9}$$

m_p may be called the "private-income deficit multiplier" and m_t the "total-income deficit multiplier."

Suppose now that we have an increase in government absorption which is *not* paid for by increased taxation and which results in an equal increase in the deficit. This is represented in Fig. 82 by projecting NT to cut $R'T'$ produced in T''. Then $T'T''$ is the increase in government absorption, equal to the increase in deficit TT' $(= RR')$. The increase in private income is still KK' $(= NQ')$. The increase in total income, however, is now $Q'N''$, which is equal to NQ'; i.e. the increase in total income, dY_t, is equal to the increase in private income, dY_p, and is equal to $\frac{1}{1-k}\,dF$. The expression $\frac{1}{1-k}$ we recognize as the usual multiplier. We see, however, that the multiplier in this sense applies only to the effect on private income of a pure rise in the deficit, or the effect on total income of an equal rise in government absorption and deficit. The pure total-income deficit multiplier, m_t, is considerably smaller.

The "principle of overfinance"

All these models, of course, exhibit a greater degree of unreality the closer the system reaches "full employment" because of the assumption that the private-absorption function is invariant. As full employment is approached there will be increasing disturbances in the price system which will tend to lower the private-absorption curve. Thus there may be price inflation, which will tend to shift real income away from those groups which are "high consumers," such as pensioners and renters, towards groups which are not high consumers; this will reduce the volume of consumption at each level of income. This is the phenomenon sometimes known as "forced saving," though, like any other "saving" concept, forced saving is unsatisfactory as a quantitative tool. There may also be an increasingly unfavorable situation in the loan market, reflected in rising interest rates or in stricter loan rationing. This will decrease the willingness to accumulate. Price inflation, on the other hand, is likely to increase directly the willingness to accumulate, as it makes the holding of real assets unusually profitable. This offsets the effect of price inflation on consumption, and makes price inflation an unreliable method of lowering the private-absorption function.

If the "Ricardian" condition is to obtain, however—i.e. if a rise in government absorption, without change in the deficit, is to produce an equal decline in private absorption—there must be a decline in the private-absorption curve as a result of the rise in government absorption. That is, there must not only be a decline in private income, but also a decline in the amount which the private sector is willing to absorb at each level of private income. Unless this happens, in fact, there cannot

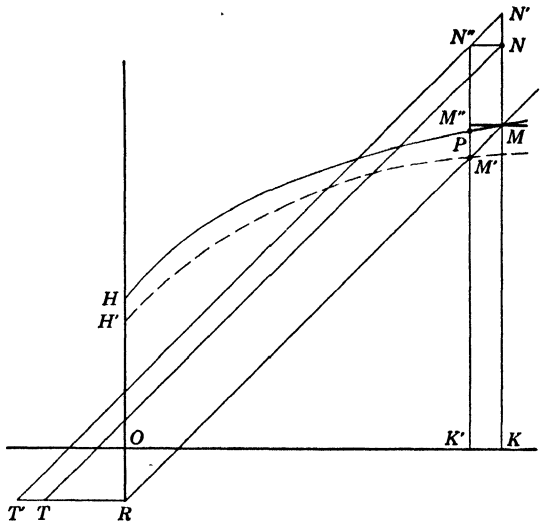


FIG. 83. Government absorption: the "Ricardian" model.

be a reduction in private income as long as the deficit remains unchanged. This proposition is illustrated in Fig. 83. The construction is the same as in Fig. 82. We suppose an increase in government absorption from RT to RT'. In the absence of any change in the private-absorption function total income will rise from KN to KN'. If now the total income is to be unchanged by the increase in government absorption there must be a decline in the private-absorption curve from HM to $H'M'$. In the new equilibrium total income is $K'N''$, equal to what it was before (KN); private income has fallen by an amount KK', equal to TT', the increase in government absorption. Actual private absorption has fallen by an amount equal to $M'M''$, which is also equal to the decline in private income (M_gM''). It will be observed, however, that the decline in the amount of private absorption at each income, i.e. the downward shift in the private-absorption curve itself, PM', is less than

the decline in private income.[1] This downward shift does not have to be uniform: as long as the downward shift in the vicinity of P is sufficient, it does not matter what happens in the rest of the curve.

If an increase in government absorption is to be achieved without any change in total output and without any change in the private-absorption function there must be a decline in the deficit or, what is the same thing, an increase in the surplus. This is shown in Fig. 84, in which the construction is again the same as in Fig. 81. M shows the original position

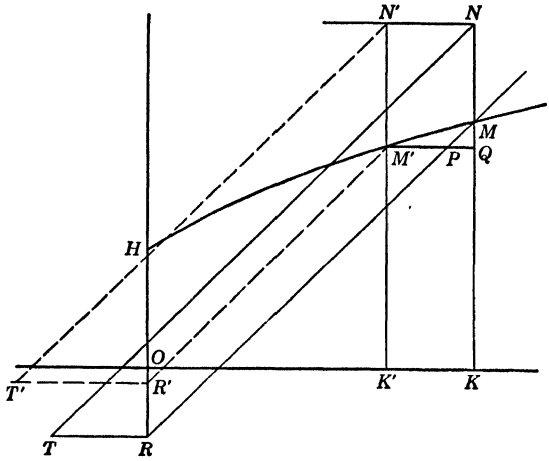


FIG. 84. The principle of overfinance.

of equilibrium, with government absorption at RT, the deficit at OR, private income at OK, and total income at KN. Now we suppose an increase in government absorption to $R'T'$: how much must be the deficit—the private-absorption curve being unchanged—which will give the same total income as before? This is found by moving the line KN to the position $K'N'$, at which $M'N'$ is equal to the new government absorption, $R'T'$. Then, drawing the 45° line from M' to cut OR in R', we get the new deficit, OR'. Private income is now OK', total income, $K'N'$, the same as before.[2]

[1] If dG is the increase in government absorption, k the marginal propensity to absorb between P and M, then $PM' = dG(1 - k)$.

[2] If dG is the change in government absorption, dY_p the change in private income, we have $dG = M'N' - MN = QM = PQ$, and $dY_p = M'Q$. If k is the marginal private propensity to absorb over the range MM' we have

$$k = \frac{QM}{M'Q} = \frac{dG}{dY_p}, \quad \text{or} \quad dY_p = \frac{1}{k}dG$$

[Footnote cont on p. 293]

The conclusion from this analysis is an interesting and important one: that if there is an increase in government absorption at a time of full employment, the only way in which this can be prevented from having an inflationary impact is either to increase taxes by *more* than the value of the increase in government absorption, so that the deficit is reduced or the surplus increased, or to lower the private-absorption curve in some way other than by inflation, e.g. by direct regulation or rationing. An increase in government absorption during full employment will have an inflationary impact even if it is completely financed by increased taxes; it must be "overfinanced" if its effects are to be truly "Ricardian." This may be called the "principle of overfinance."

The principle of overfinance applies also to an increase in private investment at a time of full employment. Even if the investment is financed in the strictly orthodox way of selling securities to households (which is roughly equivalent to having a balanced budget on the part of government) there will still be an inflationary effect unless there is an actual fall in the consumption function.

Government and distribution

The next problem is the impact of positive government action on the system of distribution, and especially on the distribution of the national income between labor income and non-labor income. The analysis of the previous two chapters may have prepared us for some strange conclusions, and indeed it will be found that there is a wide divergence between the microeconomic effects of government activity and the macroeconomic effect. Nowhere are the macroeconomic paradoxes so striking, and so little understood, as in the theory of public finance.

Let us begin by considering a system in which the deficit or surplus, in the sense in which we have been using it, is zero, so that there is no net increase in security or money holdings as a result of direct government activity: i.e. $F = 0$. Let us return first to the fundamental business-savings identity on p. 249:

$$dG_b \equiv dM_b + dQ_b + dK_h - dK_h' \qquad (1)$$

If dF is the change in the deficit, $dF = RR' = M'P$, whence

$$k = \frac{QM}{M'P + PQ} = \frac{dG}{dF + dG}$$

i.e.

$$dF = dG\left(\frac{1-k}{k}\right)$$

The greater the private propensity to absorb, the less decline in deficit (or increase in surplus) will be necessary to compensate for an increase in government absorption.

What modifications must be introduced into this by the inclusion of government as a separate item in the system? As the identity is derived solely from the addition of business balance sheets the question resolves itself into the further question of whether government activity as such directly produces in business balance sheets items not included in the above four. The only clearly distinguishable items produced in business balance sheets by the activity of government (neglecting government sales) are debts, both to and from government. In identity (1) we have assumed government to be a kind of "household," and government debts to and from business may be included in household debts. It is now desirable to make an explicit distinction between debts to and from households proper, which we shall still designate by K_h and K_h', and debts to and from government, which we shall designate K_g and K_g'. The complete balance-sheet identity then reads:

$$G_b \equiv M_b + Q_b + K_h - K_h' + K_g - K_g' \tag{2}$$

and the business-savings identity:

$$dG_b \equiv dM_b + dQ_b + dK_h - dK_h' + dK_g - dK_g' \tag{3}$$

Debts owed by government to businesses (K_g) are, for the most part, government securities owned by businesses, though the item may include some accounts receivable. Debts owed by business to government are not likely to include much in the way of business securities owned by government, but will consist of accounts payable for purchases from government, plus tax liabilities. By accounting convention the tax liabilities include only "current" tax liabilities, i.e. those payable within the fiscal year. This, as we shall see, is a fact of considerable importance.

If there is no government deficit or surplus there will be no change in the total holdings of money and net government debt on the part of both businesses and households. There may, of course, be shifts in the holdings of government securities between household and business accounts, which will produce changes in K_g. These changes, however, are likely to be small, and in so far as government securities are highly liquid assets, good substitutes for money, the shift in government security holdings between household and business accounts can be regarded in the main as part of the general structure of relative liquidity preferences. If this is fairly stable there will be no appreciable change in K_g on that account. Hence with a balanced budget dK_g and dK_g' can in the first approximation be neglected, and identity (2) reduces to identity (1).

The identity for total profits [equation (10) on p. 250] must, however, be modified even on the restricted assumptions of this section, for total profits must now be equal to business savings plus business distributions

plus net business taxes, T_b. Business taxes can be regarded as business distributions to government. Net business taxes consist of the net transfer payments from businesses to government, and are equal to taxes paid out of business accounts (corporation taxes, transfer taxes, payroll taxes, etc.), less subsidies received by business accounts from government. We have, therefore, for total profits, V,

$$V \equiv dG_b + D + T_b \tag{4}$$

The significant figure in this case, however, is not total profits before taxes and subsidies, but profits *after* taxes and subsidies, V'. We have therefore:

$$V' \equiv V - T_b = dG_b + D \tag{5}$$

We emerge therefore with a conclusion at first sight surprising: that, under the conditions we have supposed, the imposition of taxes on business accounts produces no *direct* effects on profits after taxes; the more taxes are laid on business, the more profits there will be to pay them out of, assuming that there is no change in business or household *policies* as a result of the imposition of taxes. It will be recalled that we have assumed a balanced budget; an increase in taxes under these circumstances must, of course, be offset by an equal increase in expenditure. An increase in taxes with no offsetting effect will, of course, reduce the deficit (or increase the surplus) and hence will produce effects on M_b, K_b, etc., which will change business savings.

The truth of the above proposition can be seen very clearly if we consider the effect of a pay-as-you-go income tax. This represents a shift of tax payments from household accounts to business accounts. Its effects on the distribution of income are obviously quite trivial, as it will have very little influence on business behavior. If profits were computed "before taxes" they would show an increase equal to the tax because of the decline in wage payments. In this case the accountant does not bother to compute profits before taxes because the magnitude of the tax does not depend on this figure. Profits "after taxes" are clearly unaffected in any direct way by this change. It is equally true, however, that a shift from household taxes to, say, a corporation income tax would produce no change in profits after taxes if the tax produced no change in policies. Of course, the fact that the corporation income tax varies with the magnitude of the profits-before-tax figure means that it may produce some effect on business policies, because of the way the incidence of the tax affects different firms (what we may call the microeconomic incidence of the tax). These effects on individual policy are likely to be even greater in the case of, say, an excess-profits tax or an un-

distributed-profits tax. But the effect even of these taxes depends on their effects on business or household policies, not on the taxes themselves.[3]

This is not, of course, to deny *all* effect to the *form* of taxation. The corporate profits tax unquestionably has had some effect, for instance, in shifting enterprise into unincorporated forms such as the partnership, though it is of course impossible to say how much of the "income of unincorporated enterprises" represents labor income, as a large proportion of it must, and how much represents true profits. This effect, however, is still an indirect one, operating through the impact of the method of calculating taxes on the policies of individuals, not through the direct effect of the taxes themselves.

The government-activity identities

We can now proceed to obtain an identity for wages which will include government activity specifically. If K_{hg} is the total of government debt held by households, and K_{hg}' is the total of household debts to government, identity (6) on p. 248 may be expanded as follows:

$$G_h \equiv M_h + Q_h - (K_h - K_h') + G_b + K_{hg} - K_{hg}' \qquad (6)$$

That is, the total net worths of households equal the money holdings of households, plus the real capital holdings of households, plus the net

[3] The stability of profits after taxes under a sharp increase in tax collections from corporations is shown very strikingly in the following table.

(Billions of Dollars)

Year	Corporate Profits before Tax	Corporate Tax Liability	Corporate Profits after Tax
1939	6.5	1.5	5.0
1940	9.3	2.9	6.4
1941	17.2	7.8	9.4
1942	21.1	11.7	9.4
1943	25.1	14.4	10.6
1944	24.3	13.5	10.8
1945	19.7	11.2	8.5
1946	23.6	9.6	13.9
1947	31.6	12.5	19.1
1948	34.8	13.6	21.2
1949 *	27.6	10.9	16.7

* Estimates based on incomplete data.

Source: Department of Commerce. Taken from *The Economic Report of the President, January 1950.*

It would certainly seem that the state is simply using the corporation as a tax collector, and a shift in taxes from corporate to private income probably would not make much change in the final distribution.

worths of businesses, plus net business debts to households, plus net government debts to households. Substituting the value of G_b from identity (2), p. 246, we have:

$$G_h \equiv M_b + M_h + Q_b + Q_h + K_g - K_g' + K_{hg} - K_{hg}' \qquad (7)$$

That is, the total net worth of households is equal to the sum of money and real capital in both household and business accounts, plus the net government debt outstanding. Differentiating equation (7), we have:

$$dG_h \equiv dM_b + dM_h + dQ_b + dQ_h + dK_g - dK_g' + dK_{hg} - dK_{hg}' \qquad (8)$$

This gives us an expression for household saving, dG_h.

Household saving, however, must be equal to household income, Y_h, less household consumption, C_h, and less taxes paid out of household accounts, T_h; household income is composed of wages, W, plus business distributions, D, plus business savings, dG_b, which we assume accrue to household accounts. That is,

$$dG_h \equiv W + D + dG_b - C_h - T_h \qquad (9)$$

Combining equations (8) and (9), and substituting the value of dG_b from equation (3), p. 294, we obtain:

$$W \equiv C_h + dQ_h + T_h + (dK_{hg} - dK_{hg}') + dM_h - D - (dK_h - dK_h') \qquad (10)$$

Suppose now that we write

$$dM_h \equiv dM_h' - X \qquad (11)$$

where X is the balance-of-payments factor, i.e. the shift in money stocks from households to businesses as a result of an excess of business receipts from households over household receipts from businesses, and dM_h' is that part of the newly created money which finds its way into household balances. That is, we are regarding the change in the money stocks of households as being in part a result of their balance of payments with businesses, in part a result of the creation of new money. Similarly we can write:

$$dM_b \equiv dM_b' + X \qquad (12)$$

where dM_b' is the new money in business balances, and X is the same balance-of-payments factor.

The transfer factor of Chapter 14 can now be written:

$$T \equiv X + D + (dK_h - dK_h') \qquad (13)$$

The wages identity (10) can then be simplified, substituting from equations (11) and (13), and we have:

$$W \equiv C_h + dQ_h + dM_h' + T_h + (dK_{hg} - dK_{hg}') - T \qquad (14a)$$

Similarly, substituting the value of dG_b from equation (3) into equation (4) and substituting from equations (12) and (13), we have for the profits identity:

$$V \equiv dQ_b + dM_b' + T_b + (dK_g - dK_g') + T \qquad (14b)$$

Summing equations (14*a*) and (14*b*) gives us the national income,

$$P_n \equiv dQ_b + dQ_h + C_h + (T_b + T_h) + (dM_b' + dM_h') + (dK_g + dK_{hg} - dK_g' - dK_{hg}') \qquad (15)$$

Equation (15) identifies the national income with total private investment and consumption, plus total taxes, plus total increase in the money stock, plus the net increase in government debt. The distribution of the national income between wages and gross profits again depends on its composition, modified by the transfer factor. Wages equal household absorption, plus household taxes, plus additions to the money stock accruing to households, plus the net increase in household holdings of government debt, *minus* the transfer factor. Gross profits are equal to business accumulations (investment) plus business taxes, plus the net increase in business holdings of government debt, *plus* the transfer factor.

The paradoxes of taxation

Some apparently paradoxical conclusions seem to flow from these identities. The first is that an increase in taxes increases the national income by an amount equal to the increase in taxes, so that income after taxes is unaffected by the amount of taxes. This is a "widow's cruse" theory of taxation—the more taxes, the more income there is to pay taxes out of: a most delightful doctrine. This conclusion, however, depends on other factors in the identities remaining equal. If the increase in taxes if not offset by a corresponding increase in government expenditure and absorption, there will be an increase in the surplus (or a decline in the deficit) which will be reflected in a decline in money stocks or in government indebtedness, and income will be unchanged. If there is no change in the surplus or deficit, then the increase in taxes reflects an increase in government absorption. If this is not offset by a decline in private absorption there will be an increase in national income, just as any increase in absorption—whether public or private—will increase the national income if it is not offset. In conditions approaching full

employment, of course, an increase in government absorption must produce a decline in private absorption, either through the impact of announcements of increased tax rates on private decisions or through the effects of inflation or direct controls. Even in periods of underemployment, if the announcement of increased tax rates causes households to restrict their absorption of product and businesses to plan for smaller accumulations, there will be a decline in income after taxes.

The identities (14*a*) and (14*b*) can easily be put into the form of "private income" identities, i.e. income after taxes have been paid, or "available income." We divide taxes paid out of household accounts into two parts, T_w, taxes paid out of wage accounts, and T_d, taxes paid out of dividend accounts, so that

$$T_h \equiv T_w + T_d \tag{16}$$

If $W' \equiv W - T_w$ is "available wages," i.e. wages after taxes, and $V' \equiv V - T_b - T_d$ is "available profits," i.e. non-labor income after taxes, then from equations (14*a*) and (14*b*) we have:

$$W' \equiv C_h + Q_h + dM_h' - T + (dK_{gh} - dK_{gh}') + T_d \tag{17a}$$

$$V' \equiv dQ_b + dM_b' + T + (dK_g - dK_g') - T_d \tag{17b}$$

That is to say, an increase in taxes on dividends, other things being equal, will increase available wages at the expense of available non-labor income. This may be the principal effect of increasing progressiveness in income taxation on functional distribution, as dividend incomes are probably—though not necessarily—larger on the average than labor incomes. Other things in this case are not, of course, likely to remain equal, as taxes on dividends may well affect dividend-distribution policies and hence will affect the transfer factor, T. If an increase in dividend taxes reduces dividends, thus reducing the transfer factor, the effect on distribution is thereby reinforced.

Equations (17) can be further simplified if we observe that $dM_h' + dM_b' = dM$, the total increase in the quantity of money, and $dK_{gh} + dK_g - dK_{gh}' - dK_g'$ is the net increase in government indebtedness, dK. The distribution of the increase in money stock between households and businesses depends on the relative liquidity preference, or particular velocities of households and businesses, and is determined by the same forces that determine the distribution of the money stock in general. That is to say, if we allow a period long enough for liquidity preferences to operate, the distribution of "new money" does not depend on who are its original recipients. Even if all the new money is first placed in the hands of businesses it will soon disperse itself through

the economy by means of the payments system and will end up distributed between businesses and households in the same proportion as "old money," provided that there has been no change in relative liquidity preferences in the meantime. The same may be said of government debt: where there is a free market in government securities the distribution of the total volume of government securities between businesses and households will depend on their relative liquidity and security preferences, and this will also be true of the distribution of the holdings of the net increase in government debt. We can therefore write $dM_b' = r\,dM$ and $dM_h' = (1 - r)\,dM$, where r is a ratio determined by the relative liquidity preferences of households and businesses. Similarly we can write $dK_g - dK_g' = s\,dK$, and $dK_{hg} - dK_{hg}' = (1 - s)\,dK$, where s is a ratio determined by the relative security preferences of businesses and households. A rise in business liquidity preference relative to households will raise r; a rise in business security preference relative to households will raise s. Now we also have, if F is the total deficit,

$$dM + dK \equiv F \tag{18}$$

That is to say, the total deficit of government must be equal to the amount of new money, plus the net amount of new government securities. We can, of course, regard money as a peculiar kind of non-interest-bearing government security. The distribution of the deficit between new money and new security issues is a matter of government policy. We can therefore suppose that $dM = aF$ and $dK = (1 - a)F$, where a is a ratio expressing the deficit-finance policy of government: the more the deficit is financed by the issue of currency, the greater a will be. We have then

$$\begin{aligned} dM_b' + dK_g - dK_g' &\equiv r\,dM + s\,dK \\ &\equiv \{(r - s)a + s\}\,dF \equiv b\,dF \end{aligned} \tag{19a}$$

$$\begin{aligned} dM_h' + dK_{hg} - dK_{hg}' &\equiv (1 - r)\,dM + (1 - s)\,dK \\ &\equiv [1 - \{a(r - s) + s\}]\,dF \equiv (1 - b)\,dF \end{aligned} \tag{19b}$$

where $b = a(r - s) + s$ may be called the "business fraction" of the deficit, and $(1 - b)$ the "household fraction." We can then rewrite equations (14):

$$W \equiv C_h + dQ_h + (1 - b)\,dF + T_h - T \tag{20a}$$

$$V \equiv dQ_b + b\,dF + T_b + T \tag{20b}$$

and equations (17):

$$W_a \equiv C_h + dQ_h + (1 - b)\,dF + T_d - T \tag{21a}$$

$$V_a \equiv dQ_b + b\,dF - T_d + T \tag{21b}$$

Just as models can be constructed in the simpler cases, so they can be constructed involving the governmental variables, depending on the various assumptions which may be made regarding the functional relationships among the various quantities. Thus we can suppose that household absorption is a function of available household income, $Y_{ha} =$

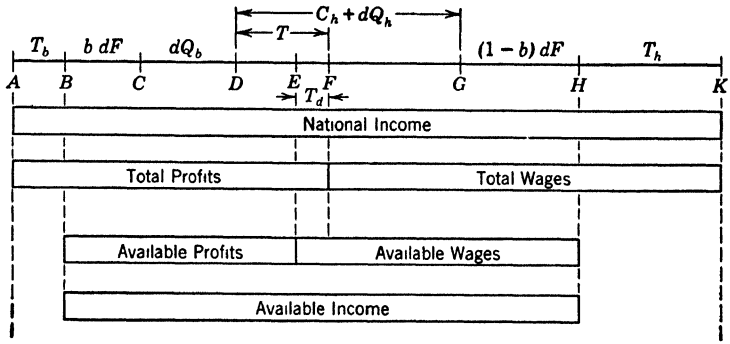


FIG. 85. Government and distribution.

$W_a + D_a$, and that business accumulation is a function of available profits: from these assumptions determinate systems can be obtained, assuming that F, b, etc., are given. It may be doubted, however, whether much is gained by the construction of such models of static equilibrium, as it is extremely doubtful whether any functions can be discovered which are stable enough to be the basis for dynamic systems of difference equations.

Nevertheless the above identities can be used as a useful rough tool of analysis. The diagrammatic interpretation of Fig. 85 may be found helpful. The national (money) income is AK, composed of taxes (AB on businesses, HK on households); the deficit (BC to businesses, GH to households); business accumulations, CD; and household absorptions, DG. DF is the transfer factor, T; EF, taxes on dividends, T_b. Then AF is total profits, FK is total wages, BE is available profits, and EH is available wages. BH is the private (available) national income. It should be observed that total government *absorption* is equal to total taxes plus the total deficit, that is, to $AC + GK$. If there is a reduction in taxes without any change in the government absorption there will be

an equal increase in the total deficit, which will result in corresponding proportionate expansions in *BC* and *GH*.

In the equilibrium of our model all the various quantities must be consistent. If, for instance, business accumulation is a function of available profits, the total of available profits *BE* must be such as to yield a business accumulation of *CD*. If household absorption is a function of available wages, or of total available income, *DG* must be consistent with *EH*, or *BH*, and so on. The extended treatment of these possible models is, however, beyond the scope of this work.

Chapter 17

A Concluding Note

It has not been the principal purpose of this work to develop definitive answers to problems of social policy; it has been the objective rather to develop methods of analysis than applications. Nevertheless it is possible to outline briefly some of the insights into economic problems which the analysis yields.

Economic cybernetics

The most critical problem of economic policy for the Western world is whether the wide fluctuations in output and employment which have characterized its history can be reduced to tolerable dimensions. This is essentially a problem in "cybernetics," to use the word coined by Dr. Wiener.[1] An unrestrained and unrestricted capitalism exhibits complex and irregular fluctuations in its basic aggregates which are symptomatic of a lack of adequate homeostatic mechanisms. What the system clearly needs is a "governor" or "thermostat" of adequate sensitivity. Dr. Wiener himself, like most natural scientists, underestimates the degree of homeostasis, or self-regulating ability, which exists in a free market economy, and rather resents the unwillingness of man to behave like matter. Nevertheless the development of adequate cybernetic mechanisms for a free economy, which will not achieve stability only at the cost of tyranny or stagnation, is a project of the first priority, exceeded in urgency perhaps only by the necessity of developing similar cybernetic mechanisms for the stabilization of peace. These mechanisms clearly must come through the agency of government, which is the only organism that has the power to act like a governor. The mechanism itself, however, is complex, and neither the signals for action nor the actions themselves are easily identified.

The simplest form of cybernetic mechanism is the governor or thermostat, which always brings into play forces to restore the system it is controlling to a constant equilibrium. Thus a governor slows down the

[1] Norbert Wiener, *Cybernetics* (John Wiley & Sons, New York, 1948).

engine that is running too fast and speeds up one that is running too slow. Similarly a thermostat operates to warm up the house when it gets too cold, and cool it down when it gets too hot. All cybernetic mechanisms involve a cycle of some kind; the amplitude of the cycle, however, depends on the sensitivity of the mechanism. If the mechanism is insensitive there may be a considerable cycle. A hand-fired furnace has a thermostat in the shape of the human fireman, who reacts—usually rather slowly—to superoptimum temperatures by going down and checking the draft and to suboptimum temperatures by going down and shovelling coal. Because the reactions of a human fireman are slow, however, his activities are likely to result in a considerable, if irregular, temperature cycle, and the house will most of the time be either too hot or too cold. An automatic fireman is a more specialized brain, which sets the furnace in motion the moment the temperature falls below the optimum, and checks it the moment the temperature rises above the optimum. There is still a temperature cycle, but it is small enough hardly to be noticed. It should be observed that a too-sensitive and too-powerful thermostat will also set up fluctuations of a "surge" nature, the movement towards equilibrium being too violent and carrying the system beyond it in swings which may even be of ever-increasing amplitude.

It is tempting to apply this analogy to the economic system, with its swings of inflation and deflation, and to argue that all that is needed is a "governor" which will operate in an inflationary manner when the system is deflating, and in a deflationary manner when the system is inflating. The old gold standard was, of course, a mechanism of this kind—a deflation, by making gold production unusually profitable, eventually corrected itself by an increase in gold production and the gold stock, and an inflation, by making gold production unprofitable, led to a shrinkage in gold production and in the gold stock. There is some evidence that this mechanism actually operated in the nineteenth century; it operated, however, with a very long time lag, and was subject to all kinds of extraneous irregularities. Various substitutes which have been proposed, such as bimetallism, symmetallism, the composite commodity standard, even the adjustable tax plan, are mainly proposals for increasing the sensitivity of the governing mechanism, and as such, of course, are all to the good, as there seems to be very little danger of getting a too-sensitive "governor."

The analysis of this book indicates, however, that the problem may be much more complex than can be solved by the simple governing mechanism outlined above. It would be fairly easy for government to regulate the total volume of *payments* within reasonable limits by means of its

own payments system, given adequate and rapid information about the total volume of payments. This is the objective of the "adjustable tax plan" described in my *Economics of Peace.*[2] A decline in total payments could be followed automatically by a cut in tax payments, a rise in total payments by a rise in tax payments. To put the matter in another way, changes in payments velocities can be offset by changing the quantity of money in the private sector: if payments velocities rise, this can be offset by diminishing the quantity of money in private hands by running a government cash surplus; if payments velocities fall, this can be offset by running a government cash deficit.

Payments, however, are not output or income, and it would be quite possible to have a stable volume of payments within which output, income, and prices fluctuated violently. Economic cybernetics therefore involves more than a simple governing operation: it is more like the cybernetics of the automatic steersman, in that it involves at least two mechanisms, and perhaps more. It is not enough to stabilize payments through the deficit-surplus: it may be necessary to stabilize output and income through fluctuations in government absorption. This represents a much more intractable political and social problem, as it is more difficult to effect rapid changes in government absorption (except, ironically enough, in time of war) than to effect rapid changes in government payments. Nevertheless the problem must be tackled, as we cannot assume that payments fluctuations can be substituted altogether for absorption fluctuations.

Another possible solution is to make the level of government absorption large enough that it results in a "pressure economy" in which total planned absorption is always pressing on capacity, and then to regulate the economy by the direct control of consumption or of private investment. In a sense this is the Socialist solution, as seen in an extreme form in Russia and in a less extreme form in England. If government absorption is too large, however, the restriction of private activity, both of consumption and of investment, which it entails becomes burdensome and injurious to the general welfare of the economy. This is clearly true of Russia, where the ideology inhibits entrepreneurial activity even of the most obviously socially desirable kind, and it may even be true of England. It may well be, however, that a quite small extension of government absorption might well result in a situation where most of the benefits of private enterprise were retained, without the danger of severe depressions. This is particularly true, of course, if the "government multiplier" is large, as it may be in our society. A government

[2] Boulding, *The Economics of Peace* (New York, 1945), p. 161.

absorption of 20% of the national income, as opposed to about 11–12% in the thirties, might easily give us a resilient "moderate pressure" economy which could be regulated without severe controls.

Secular inflation

The question still remains, however, whether if we avoid the Scylla of secular stagnation we do not fall into the Charybdis of secular inflation. This is a problem which most Keynesians have not adequately faced, though Keynes himself gave it an uneasy recognition. According to Keynes "true inflation" does not begin until there is full employment, but long before this state is reached prices in most of the economy may be rising because of the "bottlenecks" problem—immobility of resources leading to inelastic supplies, and therefore to rising prices instead of rising outputs, in one industry after another as demands increase. In this regard the experience of 1936–1938 is depressing, when a recovery in output began to move off into price-wage inflation at high levels of unemployment, and the attempt to check the price inflation led instead to the sharpest output deflation on record. There are strong long-run forces in our economic institutions which make for secular inflation. It is obviously easier to run an economy with highly organized labor and agricultural groups if the general monetary level is rising slowly, as this provides a setting in which it is relatively easy to satisfy the sociological need of these groups for "victories." A fall in the cost of living for the individual is something that happens to him: a rise in money wages is something that he gets—and for which appropriate credit is usually given! As, therefore, the effect of a change in money wages on employment is so complex that it is almost impossible to predict even its direction, no effective social pressure can be brought to bear in favor of wage-price reductions even in time of unemployment. Hence the political and sociological forces would certainly indicate that a full employment policy would predispose the economy to secular inflation. We have had so little experience of such a phenomenon in the United States, where there has been no secular trend in the price level for well over a century, that we are probably too blind to the problems it may raise, e.g. for insurance, pensions, and social security. There is also a problem as to the extent to which the favorable effects of inflation are due to its unexpectedness. If this factor is important, then inflation would have to be accelerating to be effective—a most unpleasant prospect! On the other hand, one certainly does not wish to come to the conclusion that a full employment policy is possible only with price-wage control, for we have certainly not developed as yet the administrative skill which can make governmental price regulation an adequate substitute for the free market.

The future of capitalism

Another question on which the analysis of this work throws some light is one which concerned the classical economists a good deal, but which has unaccountably dropped out of much modern economics. This is the question of what happens to the distributive shares "in the progress of society." Must profits ultimately disappear as accumulation proceeds, and is there consequently an internal contradiction in capitalism which seals its historic doom? Can we possibly maintain full employment in the stationary state? There are hardly any more important questions for the Western world to answer. The analysis of Chapter 14, in particular, clears up some doubtful points in this connection. It is clear from the business-savings identity (8) on p. 249 that there cannot be any long-run future for business savings. In the long run there cannot be a perpetual increase in business money holdings, or in net household indebtedness to business, and in the stationary state there cannot be any long-run business accumulations. There may be short-run variations in all these items, permitting alternations of periods of business saving with periods of business dissaving, but these must cancel out in the long run. Business saving, therefore, along with all saving, must disappear in the stationary state, and outgo must be equal to income. This does not mean, however, that profits also must disappear. From the profits identity (9) on p. 249 we see that, even though business savings (dG_b) are zero, profits can still be positive as long as they are distributed, according to the "widow's cruse" principle. Hence the distribution schema in the stationary state is quite indeterminate, except as dividend or entrepreneurial consumption policy determines it. There is no need for profits to disappear as long as businesses are willing to distribute them.

Exactly what policy is desirable in regard to such distributions is a difficult question. In general it can be laid down that the proportion of national income going to profits should be at the minimum consistent with efficient operation of a free enterprise system. Where that minimum lies, however, is a difficult problem. A system of free enterprise is conceivable, in a stationary state, in which there is no profit, i.e. in which all income is labor income, and in which each enterprise maintains itself by the process of simple "homeostasis of the balance sheet." The fact of change and uncertainty, however, and the possibility of a retreat of entrepreneurs into liquid assets, mean that a certain average level of profits is necessary if the system is to operate at capacity. Thus the disappearance of profits in 1931 and 1932, due to business disinvestment (a negative dQ_b), price decline, and a cautious dividend policy, led to serious disruptions in the economy and widespread unemployment. The

methods of this work are not adequate to discover what determines the optimum rate of profit. It is encouraging, however, to find that whatever the rate may be, there is no reason why it should not be attained.

The importance of dividend distributions in the determination of profit raises once again the question whether an undistributed-profits tax (i.e. a tax on business savings) should be added to the armory of policy weapons. The administrative difficulties of such a tax are considerable, yet it might prove to be an unusually valuable piece of control apparatus to prevent the disappearance of profits when business investment declines.

As the stationary state approaches, of course, output will come closer and closer to consumption. It may be very important, therefore, to raise the consumption function either by direct encouragement through the tax system or by the expansion of government absorption. There certainly seems to be no reason why the consumption function cannot be adjusted to that level which would yield optimum output, even without any net investment. Hence there seems to be no absolute reason why a suitably modified free enterprise society could not survive its own success in accumulation. To determine whether it can in fact survive would lead us far beyond the limits of economics in the direction of the total dynamics of social ecology, but we can at least show by economics that the task of preserving a free economy is not economically impossible. This note of hesitant optimism may well bring the work to a conclusion.

Index

Absorption, 231
Accumulation principle, 190
Age distribution, 191
Aggregates, in economic policy, 171–172
  measurement of, 177–182
  structure of, 187
Aggregation, dangers in, 240, 174–175
  fallacies of, 186–188
  of payments, 208
Asset flexibility, 132
Asset-price elasticities, 147–150
Asset revaluation, 32
Asset transformation, 30
Asset valuation, 43–45, 65–67, 73
Austrian theory of capital, 193
Averages, significance of, 172

Balance of payments, 208–209, 252–254
Balance sheet, changes in, 29–33
  definition of, 26–29
Balance-sheet identities, 246–248, 294
Bank credit, effects of, 271–274
"Bathtub theorem," 190, 230
Bilateral monopoly, 67–71, 91–94
Böhm-Bawerk, E. von, 245
Boulding, K. E., 53, 81, 104, 105, 147, 152, 190, 244, 305
Business net-worth identity, 247
Business-savings identity, 248, 294
Butler, Samuel, 197

Capital, aggregates of, 178
  as a population, 189–202
Capital gains, 142
Capital rationing, 131, 278
Capitalism, future of, 307
Clark, J. B., 194–195, 244
Compensating payment, 87
Competition, imperfect, 21–24
  of two populations, 7–11
Complementarity, 12, 58, 149–151
Constant cost, 102
Consumer, theory of, 144–154
Consumer credit, 255
Consumption, affected by dividends, 257
  as cause of production, 230
  definition of, 135–136, 139
Contract curve, 69–70
Contract surface, 93
Credit, bank, 271–274
  consumer, 255
  interbusiness, 258
  non-bank, 274–276
Cutbacks in output, 157
Cybernetics, economic, 303–306
Cycles, expectational, 58
  in populations, 24–25, 196–202

Darwin, C., 16
Debt, significance of, 276
Deficit, cash, 270
  effects of, 287–290
Demand, for input, 107–111
  for money, 212
Distribution, government and, 293–302
  macroeconomic theory of, 245–269
  microeconomic theory of, 244
Dividends, theory of, 113–114, 256

Ecology, 6, 166
Ecosystem, theory of, 5–25
Edgeworth, F. Y., 69
Employment, theory of, 229–242, 263
Envelope curve, 98
Equilibrium, general, 159
  mutual-price, 76–79
  non-price, 160–163
  of employment, 232
  of population, 192
  of wages, 158–159

Equilibrium, parasitic, 14–15
shiftability of, 14, 262
Equipreference curve, 90
Exchange, and employment, 242
bilateral, 67–71, 91–94
multilateral, 71–73
Exchange path, 96
Exogenous factors, 4
Expectations, 48, 116
Expenditure-receipts curve, 50–53

Fertility, 17–18
Firm, dynamics of, 156
static theory of, 32–35
Fisher, I., 184, 277

Gain from trade, 88
Gambling, 118–121, 125
Genes, social, 7
Gold standard, 304
Government absorption, 237–238, 285–287
Government activity identities, 296–298
Government and distribution, 293–302
Government bond standard, 283

Hart, A. G., 278
Hawtrey, R., 277
Hicks, J. R., 148
Hoarding, 143, 173–174
Homeostasis, 26–29, 230, 303
Household as entrepreneur, 151
Hyperemployment, 233

Imperfect markets, for loans, 130
for output, 98
Income, 140, 182, 202–205
Indifference curves, 79–84, 180–182
Indifference surface, 83–86
Inflation, secular, 306
Inflation threshold, 264
Insurance, theory of, 121–126
Interest, theory of, 128–132, 238–240, 277–282
International trade, 188
Investment-saving identity, 235

Kalecki, M., 47, 126
Keynes, J. M., 26, 238, 277
Knight, F. H., 133, 194–195

Life cycle, 37
LIFO, 101, 142
Liquidity preference, 240, 279
Loans, effect on balance sheet, 32–33
imperfect market for, 130

Macroeconomics, 3, 171
Makower, H., 117, 124
Malthusian unemployment, 165
Marginal factor cost, 107
Marginal productivity, 244
Marginal propensity to spend, 227
Marginal rate of substitution, 82
Market equation, 54, 71, 77, 279
Market resistances, 60
Marketer, theory of, 39–50
Marschak, J., 117, 124
Marshall, A., 6, 34
Marxian fallacies, 188
Maximizing behavior, 35–38
Meade, J. E., 238
Microeconomics, 3, 171
Models, of distribution, 258–269
of market, 56–59
Monetary policy, 282–283
Money illusion, 82–83
Monopoly, bilateral, 67–71, 91–94
Morgenstern, O., 60
Mortality, 17–18
Multiplier, government, 237
investment, 235–237
payments, 226–228
private-income deficit, 290
total-income deficit, 290
Mutual price, 69, 73, 76–79

Non-price equilibrium, 160–163
Null price, 53

Odds, 119
"One-hoss-shay" populations, 198–200
Opportunity function, 80
Opportunity surface, 85
Outgo, 140

Paradox, macroeconomic, 173
of payments, 212–213
of taxation, 298–302
Parameters of Behavior, 49, 57, 214
Parasitic equilibrium, 14–15

Pay-as-you-go taxes, 295
Payments, aggregation of, 208
  related to output, 240–242
  theory of, 206–228
Payments multiplier, 226–228
Payments table, 206–207
Peak-load, 198
Period, of money turnover, 211
  of production, 194
Population theory, 5–25, 189–202, 230
Preferred asset ratios, 48, 64–65, 88–91, 99, 279
Preferred liquidity ratio, 55
Price, normal, 155
Price discrimination, 61–63
Price flexibility, 205
Price level indices, 182–186
Price level structures, 203
Prices in theory of employment, 234
Principle, of asymmetrical reaction, 222–226
  of homeostasis, 230
  of increasing risk, 126
  of minimizing consumption, 135–139
  of overfinance, 290–293
Production, 95, 111, 140
Profit-making process, 42, 74–75, 97
Profit maximization, 100
Profits identity, 249–250, 295
Purchase-sale, 241

Rate of risk substitution, 120
Ricardian models, 286, 291
"Ricardian" unemployment, 163
Risk, opposed to uncertainty, 133

Sales-purchase curve, 63
Saving, 141
Shackle, G. L. S., 47, 134
Shadow effect, 197
Shiftability of equilibrium, 14, 262
Smith, Adam, 152, 153, 245
Socialist solution, 305
Stationary state, 308
Superparasitism, 20
Supply-demand curve, 50–53
"Surge" phenomenon, 218–220
Survival distribution, 191

Tax, fixed, 103
Taxation, paradoxes of, 298–302
Taxes, effect on producers, 104–107
  effect on profits, 295–296, 308
    on dividends, 299
Thermostat, 304
Transfer factor, 252
Transfers, government, 284
Transformation function, 39, 61, 80

Uncertainty, 45, 118
Undistributed profits tax, 308
Unemployment, 163–166, 229–242
Utility, maximization of, 36

Valuation, as method of aggregation, 176
  of assets, 32, 43–45, 65–67, 73
Valuation coefficient, 40, 45
Veblen, T., 147
Vector lines, 10
Velocities of circulation, of goods, 242
  of money, 254
  partial, 209–210
Volterra, Vito, 7
Von Neumann, J., 60

Wage-fund theory, 245
Wage structures, 204
Wages, equilibrium of, 158–159
  in the balance sheet, 31
Wages identity, 251
Welfare economics, 86–88
Wicksell, K., 277
Widow's cruse, 221, 250, 256, 298, 307

Zero-sum game, 74–75