THE ECONOMICS OF CONTROL

THE ECONOMICS OF CONTROL

THE MACMILLAN COMPANY
NEW YORK • CHICAGO
DALLAS • ATLANTA • SAN FRANCISCO
LONDON • MANILA
IN CANADA
BRETT-MACMILLAN LTD.
GALT, ONTARIO

THE ECONOMICS OF CONTROL

Principles of Welfare Economics

BY ABBA P. LERNER

Associate Professor, the Graduate Faculty of the New School for Social Research

THE MACMILLAN COMPANY

New York

PRINTED IN THE UNITED STATES OF AMERICA

Twelfth Printing, 1960

To Alice,
Lionel and Marion

PREFACE

This study was started about twelve years ago while I was a student at the London School of Economics. During this time it has changed in scope and in direction. Originally it was to be a development of the theory of the price mechanism of a socialist society. The socialist society was conceived of as completely or almost completely collectivist. In the course of the development of my ideas on the subject, while the work was continually being interrupted by other tasks, it gradually became clear to me that the maintenance and further development of the democratic way of life, as it grew under capitalism and was extended by the labor movement within the capitalist society, not only formed a far more essential part of the socialist ideal than the negative "abolition of private property in the instruments of production" but was in much greater need of careful tending. So much is now clear from the history of Russia and Germany. If socialism is to be identified with the belief that the abolition of private property would automatically establish the brotherhood of man—and many socialists did, while some apparently still do, believe this—then socialism must be counted out as false. State control or ownership of the instruments of production where the state itself is not thoroughly democratic is not socialism and is much further removed from socialism than socialism's "opposite," capitalism.

The title *Economics of Control* was proposed in 1932, with the idea that the principles of the price mechanism would also be applicable to nonsocialist but autocratic collectivist societies. The name is perhaps even more appropriate for the present form of the book, now that the stress is taken from collectivism and applied to the idea of conscious recognition of the problems of social organization and the exercise of conscious control over the

economic system. I do not think I ever was guilty of raising collectivism from a *means* of bringing about the socialist ideals to an *end* in itself, but, like many socialists, I tended to overemphasize its importance. The economics of *control* is still contrasted with the economics of *laissez faire*, but control does not necessarily mean collectivism. It suggests the deliberate application of whatever policy will best serve the social interest, without prejudging the issue between collective ownership and administration or some form of private enterprise.

In my original plan I had intended to provide a theoretical solution for each economic problem of a completely collectivized economy and then see to what extent, if at all, and by what means the problem is in fact solved in a capitalist society. But the abandonment of the dogmatically, and therefore completely, collectivist economy as identical with the ideal of a society organized in the social interest still permits a similar procedure to be followed with slight modifications, as described in Chapter 1.

It is almost impossible for me to say now exactly in what respects this work shows true originality. Most of it doubtless was absorbed from my teachers at the London School of Economics. To Professor Lionel C. Robbins, Professor Frederick A. Hayek, Professor J. R. Hicks, and Professor D. H. Robertson I am indebted for my original training in handling the tools of economic analysis. To Professor Arnold Plant and to Professor William C. Hutt of the University of South Africa I owe much for their special insistence, long resisted by me, on the possibility of approaching social problems understandingly from the free-enterprise starting point. Professor Harold J. Laski and Professor Maurice H. Dobb are responsible for helping to direct my interests to the topic on which I have written. Mr. J. M. Keynes's influence is of course seen throughout, as in practically all modern writings. I owe a special debt to Mr. R. F. Kahn and Mrs. Joan Robinson for the great pains they took in getting me to overcome my prejudices against Mr. Keynes's great advancement of economic understanding. Not least is my debt to Dr. M. Kalecki and Professor Oscar Lange, not only for their keen criticisms of my thoughts and theories, but for their incessant and valuable reminders to me of

the larger problems of social organization into which the purely economic issues have to be fitted. In the long interval since I started on this work I have had the privilege of meeting economists all over England and the United States as well as in Canada, and no doubt many have left their mark on me. I regret that it is impossible to mention them all.

Thanks are due to the publishers of the following articles of mine for permission to make use of some of the materials in them.

"The Diagrammatical Representation of Elasticity of Demand," *Review of Economic Studies*, October, 1933.

"The Diagrammatical Representation of Elasticity of Substitution," *Review of Economic Studies*, October, 1933.

"Capital Investment and Interest," *Group Meetings 1936–37*, Manchester Statistical Society.

"Some Swedish Stepping Stones in Economic Theory," *Canadian Journal of Economics and Political and Social Science*, November, 1940, University of Toronto Press.

"From Vulgar Political Economy to Vulgar Marxism," *Journal of Political Economy*, August 1939, University of Chicago Press.

I am grateful to Professors Carl T. Devine of Johns Hopkins University and George J. Stigler of the University of Minnesota for reading the manuscript and correcting a number of errors. An earlier draft of the manuscript was accepted by the University of London as a thesis in partial fulfilment of the requirements for the Ph. D. (Econ.) London.

Abba P. Lerner

NEW YORK

TABLE OF CONTENTS

THE ECONOMICS OF CONTROL

Abbreviations

ac average cost
af average quantity of factor
afc average fixed cost
ap average product
apc average private cost
apr average private revenue
ar average revenue
avc average variable cost
e_d elasticity of demand
e_s elasticity of supply
F composite factor
M marginal substitutibility
mc marginal cost
mei marginal efficiency of investment
mf marginal quantity of factor
mp marginal product
mpc marginal private cost
mpr marginal private revenue
mr marginal revenue
msb marginal social benefit
msc marginal social cost
P composite product
p price (of the product)
pf price of the factor
vap value of the average product
vmf value of the marginal quantity of factor
vmp value of the marginal product
σ elasticity of substitution

CHAPTER 1. INTRODUCTION
THE CONTROLLED ECONOMY

The fundamental aim of socialism is not the abolition of private property but the extension of democracy.

In calling this book "The Economics of *Control*" there is, of course, implied a contrast with the economics of *laissez faire*, but control does not necessarily mean collectivism. It suggests the deliberate application of whatever policy will best serve the social interest, without prejudging the issue between collective ownership and administration or some form of private enterprise. The *laissez faire* with which *control* is contrasted is not the emancipation from the dogmas of mercantilism and the interests of private monopolies. That emancipation can itself be considered as an application of *control* in the social interest. Certainly Adam Smith saw it as such. The economics of control is contrasted rather with the attitude which would have the government leave things alone just because it is the government and as such has no right to interfere with business. This dogma of the right (sometimes inspired by private interests) is based on an *asocial* attitude that fails (or refuses) to see economic activity as a means of satisfying the needs of the people but regards a business as a purely private way of making a living or a fortune to which the discoverer or the conqueror has an inalienable right—a right that is sometimes identified with democracy itself.

This is obscured by dogmas of the right and of the left.

Against the dogma of the right which says in effect that government should never interfere with profit-making business there is a dogma of the left which would establish 100 per cent collectivism

and outlaw any private profit-seeking enterprise as immoral. Our task is to steer a path between the two dogmas, counting neither private enterprise nor state ownership as the only good—to consider a State which uses its control to enable that method to prevail in each particular case which best serves the public interest.

The benefits of both the capitalist economy and the collectivist economy can be reaped in the controlled economy.

Our procedure will be to show how the various economic problems could be solved in a completely collectivist society that was organized in the social interest. This will be followed by a study of how the same problems might be solved or remain unsolved in a purely capitalist society in which economic activity is run only by private enterprise for profit. This society is called capitalist and not individualist because the collectivist society with which it is being contrasted is completely individualist in taking the satisfaction of the needs and wishes of the individual citizens as the ultimate criteria of its effectiveness. Then will come a discussion of what can be learned from both of these contrasting "pure" forms of economy, and how these lessons can be utilized in an "impure" or "mixed" economy in which both dogmas are absent and where the best devices of both "pure" forms can be utilized. We shall call this the *controlled economy*, to contrast it with the actual world and the United States in particular, which we shall call the *uncontrolled economy* because its strong anticollectivist bias causes it to be left to chance whether or not the social interest is best served by the existing economic institutions. To speak thus of the actual world and of the United States in particular as *uncontrolled* may occasion two kinds of misunderstanding and surprise which it is best to correct in anticipation.

Control must be distinguished from regulation.

First, it is not meant to deny that the actual economy is in fact regulated in many ways in the social interest. We have regulation of public utilities, pure food laws, income taxes, and social

security, and in hundreds of ways the various governments, local and national, limit and control and themselves undertake economic activities in the general interest. Yet we may refer to the actual economy as "uncontrolled" because all these activities are partial and haphazard and are not organized as they would be if it were a recognized responsibility of the government to control the resources of society to see that they are utilized in the best possible manner. Instead the responsibility is left in some unclarified way to the managers of private enterprise whose quite proper concern for the profit of the investors in the enterprise is forgotten when they are spoken of collectively and mystically referred to as "business" or "industry."

The three principal problems to be faced in a controlled economy are employment, monopoly, and the distribution of income.

There are three primary tasks that a controlled economy would carry out before anything else. It would make use of all the available resources and in particular of all men who seek work. It would, in the United States, at once abolish all dire poverty and then take other steps to diminish the tremendous inequality of income and wealth. It would put an end to monopoly throughout the economy and the accompanying exploitation and economic waste. However numerous the ways in which relatively minor interventions are made by the state, even if they all really work in the general interest, unless these primary tasks are accomplished the economy is "uncontrolled."

Second, the term does not refer to the number or complexity of the regulations that harass businessmen. The controlled economy would certainly be much simpler and have much fewer regulations than those to which businessmen and managers have to submit at present in the United States, to say nothing of Germany or Russia. Complexity of regulation is a result of a piecemeal legislation with some relatively small problem in mind each time. In the controlled economy regulation of economic activity is indeed more *thorough* than in an uncontrolled economy. But in the interests of efficiency *a minimum of complexity* in regulation is

one of the principles of control, and with the establishment of the principles of control in general it is possible to streamline the regulations. The uncontrolled economy may be likened to an automobile without a driver[1] but in which many passengers keep reaching over to the steering wheel to give it a twist while complicated regulations prescribe the order and degree to which they may turn the wheel so as to prevent them from fighting each other about it. The controlled economy has a driver, so these regulations are unnecessary.

Liberalism and socialism can be reconciled in welfare economics.

Pragmatic as contrasted with dogmatic collectivism is very close to the point of view of the liberal capitalist who is in favor of state activity wherever the liberal capitalist ideal of perfect competition cannot be made to work. So close indeed is the *rapprochement* between the two that the differences are to be sought outside of the institutional order that is advocated by both the pragmatic collectivist and the liberal capitalist. The former suggests that collective organization be applied except where competitive enterprise works better in the social interest. The latter favors the restoration of free competition wherever possible and would permit collective organization when for technical reasons this should prove impossible. Both come to the same thing.

Because the controlled economy here studied disregards the dogmas both of pure capitalism and of pure collectivism, it is likely to earn the enmity of the devotees of both dogmas. The pure capitalist will call it socialism and the pure collectivist will call it capitalism. The controlled economy is not a very good name, but we have no good single name for the rationally organized democratic society which strives to be free from both dogmas.

The term "mixed economy" is sometimes used to designate something like our controlled economy which has elements of collectivism as well as elements of private enterprise for profit. This is a very bad name because it suggests the absence of any

[1] See A. P. Lerner, "The Economic Steering Wheel," *The University Review*, June 1941.

single controlling principle but a confusion of different and perhaps contradictory principles. The fundamental point of the controlled economy is that it denies both collectivism and private enterprise as *principles* for the organization of society, but recognizes both of them as perfectly legitimate *means*. Its fundamental principle of organization is that in any particular instance the means that serves society best should be the one that prevails. Perhaps a better name would be the *service economy*, since it is the question of which method *serves best* that determines which is to be used.

Socialists can claim with some historical justification that their ideal of a socialist society originally had no dogmas about collectivism but was a controlled economy in our sense in which the public interest was the criterion for judging whether any industry should be collectivized or whether it should be left in the hands of private enterprise. But the word "socialism" has so frequently been stolen by others who used it for different purposes that it is probably best to abandon its claim to freedom from the dogmas.

The *rapprochement* between socialism and capitalism also brings the theory of the economy closer to orthodox economic theory. Free enterprise and how it works must be considered not only when contrasting the capitalist with the collectivist society, but also as one of the main instruments to be utilized in our consciously controlled economy. Most economic treatises start with a theory of the objective prices and quantities determined in the equilibrium of the markets, and sometimes add to this a consideration of the welfare aspects. Here we shall primarily be concerned with how our economy could be controlled in the interest of the welfare of the population, and only incidentally shall we see how or to what extent, if at all, these considerations of welfare are in fact satisfied in the existing society.

The two streams of thought, one deriving from liberal capitalism and the other from liberal socialism, run together and become indistinguishable as far as concerns the concrete natures of the economic institutions recommended. Yet a difference can be discerned between the two streams, just as the water from a mountain stream is said to be distinguishable a long time after it

has joined that of the river of the plain. There are not only verbal differences that serve as clues to their origin. There are likely to be differences in estimation of the political possibilities. The school deriving from liberal capitalism will stress the restoration of competition. The other school will stress the state control or operation of the inevitable monopolies. What can and should be done first by the state in any concrete situation will of course depend on the political situation and on what resistance can be offered by interests that might be adversely affected by action in the general interest. The ex-capitalist school will usually be in favor of a policy of appeasement, and the ex-collectivist school will be in favor of drastic action that would destroy the roots of the opposing private interests. In this study we shall not go into the merits of this political issue. We shall assume a government that wishes to run society in the general social interest and is strong enough to override the opposition afforded by any sectional interest. In this way we shall be able to concentrate on what would be the best thing that the government can do in the social interest—what institutions would most effectively induce the individual members of society, while seeking to accomplish their own ends, to act in the way which is most beneficial for society as a whole. Any particular politician or statesman may have to compromise. Here we shall merely attempt to show what is socially desirable. Except when they radically affect the working of the system, we must leave to the politicians the political problems of compromise.

CHAPTER 2. THE OPTIMUM DISTRIBUTION OF GOODS

The quantitative problem in distribution is how much shall go to each individual.

The many problems with which the controlled economy is confronted are closely interrelated so that they really form one single highly complex economic problem. It is therefore inevitable that when these problems are separated for study they will appear very artificial. Such a separation is nevertheless unavoidable for purposes of exposition, and the artificiality will persist until the different problems are integrated as the study is completed.

The first problem to be separated out for study is the best way of distributing the available goods among the individuals or families that constitute the consumers of our society. Of course the amount of goods available will itself depend on the way the goods are distributed, if only because of the effects on the incentives to produce. But we do not wish to consider these matters until later. *We shall therefore assume for the time being that these goods have already been produced.*

This problem we must again artificially subdivide into two parts. One part is concerned with the *quantitative* distribution of goods among the consumers—the question of *how much* in general is to go to each consumer—and we shall leave it until the next chapter. The other part abstracts from how much in general goes to each consumer but concentrates on *qualitative* distribution—*what kinds* of goods shall make up the portion of each consumer. We shall see that the problems of the quantitative and qualitative distribution of goods are so closely tied together that it is impossible to solve one completely without solving the other—but it will simplify our analysis if we separate them in this way.

The qualitative problem is how the different kinds of goods shall be allocated among the different individuals.

The second part of the distribution problem is, however, not purely qualitative. It involves not merely what *kinds* of goods shall make up the portion going to any consumer, but also in what *proportion* (or quantities) these goods shall be provided. Consequently it is better to rename the first part of the problem which deals with how much in general shall go to each consumer (the subject of the next chapter) as the problem of the *division of income*, leaving that term to be more carefully defined later, and to call the second part the problem of the *allocation of goods*. This we shall deal with in the present chapter.

The first principle that must govern the allocation of goods is to let each consumer have that which is relatively more useful to him than to other consumers.

The stress here is on the word "relatively." By this is meant the usefulness to an individual of the particular good *relative to the usefulness of other goods* to the same individual; not whether the *absolute* usefulness of a good is greater to one individual than to another individual. There is no way of telling objectively whether a particular unit of a particular good is *absolutely* more useful to one consumer than to another. Each may claim that he is the one who needs it most or who can get most satisfaction out of it, and there is no objective way of judging between them. Furthermore, the simple question of whether to give a good to Peter or to Paul is a question of whether Peter shall have more in general or whether Paul shall have more, so it belongs to the *division of income* which we are leaving to the next chapter.

Certain assumptions have to be made about human satisfactions or welfare, including the principle of *diminishing marginal substitutability* between goods.

However, if we can compare the *relative* usefulness or *relative* valuation of different goods to different consumers we shall have a criterion for the optimum allocation of goods. If two (or more)

goods have different *relative* usefulness to two (or more) consumers, it is possible to improve the situation by a reallocation of goods between consumers without hurting anyone. In this way we may keep away from the problem of division of income.

If Victor and Mark both have moderate supplies of both meat and fruit, but Victor is a vegetarian while Mark eats meat in large quantities, then the relative usefulness of fruit as compared with meat is greater for Victor the vegetarian than for Mark the meat eater. If then Victor were to get less meat but more fruit while Mark were given more meat and less fruit, both could benefit as a result of the improvement in the allocation of goods. It is important to notice that only the *relative* usefulness of meat and fruit was compared as between Victor and Mark. It may be that Mark enjoys both meat *and* fruit more than Victor does, or it may be the other way around. No comparison was made between the *absolute* usefulness of either meat or fruit for Victor and for Mark. But before we can apply this objective criterion of the best distribution of goods we must make several assumptions.

First, we must assume that consumers do enjoy satisfactions from having things like food and clothing and entertainments and all the other things that we observe people trying to obtain for themselves, and that they suffer pain when they are deprived of those things. This seems so obvious that it is possible not to realize that it is an assumption which we all make about other people by analogy with our own feelings of pain and satisfaction. It is not possible to *observe* other people actually enjoying or suffering anything. But we can see the expressions on their faces or hear the noises that they make in speech or ejaculation and can *infer* their pain or pleasure which we *assume* to be in some way similar to our own.

Second, we must assume that in general consumers try to obtain that which gives them more satisfaction rather than that which affords them less satisfaction whenever they are permitted to choose between alternatives. We do not believe this to be universally true and for that reason we do not allow children and sick people to choose things that we think would be harmful to them, but have parents or guardians to choose for them. In such cases

it is these guardians who can be regarded as the *consumers* with whom we are concerned.

Our third assumption is that there is in general a diminishing effectiveness in the substitution of one good for another. This we may call the *principle of diminishing marginal substitutability.*

The relative usefulness that is significant for our purpose is not that of the total supplies of various commodities to any consumer but the relative usefulness of small increases (or decreases) of various goods. This is because we are concerned with reaching the best allocation of goods between different consumers by *shifting* goods—just as between Victor and Mark—from points where they are relatively less useful to others where they are relatively more useful, such adjustments being made by small increments until the best or optimum allocation of goods is reached.

The relative usefulness of small increments (or decrements) of various goods to a consumer is called the *marginal substitutability* of one good for another. It is measured by the number of units of the other good for which one unit of the good increased can be substituted without making the consumer either better off or worse off. The more useful an increment of any good is to an individual the greater is the quantity of other goods for which it can be substituted without making the individual worse off (or better off), and the greater is its *marginal substitutability.* Putting the same thing the other way round, we can say that the greater the *marginal substitutability* of a good the more difficult it is to replace it—the greater is the amount of other goods that must be substituted for a sacrifice of a unit of this good if the individual is not to be made worse off by the substitution. *Marginal substitutability* is a very long as well as a very ugly name, and we shall have to use it many times. We shall therefore call it *M* for short and speak of the *principle of diminishing M.*

If we suppose that our Victor is not a very strict vegetarian so that he does have some use for his supply of meat we may imagine that he is just willing to give up 4 pounds of his supply of meat if he could get in their place another basket of fruit. (Of course he would be even more willing to get a basket of fruit at the sacrifice of *less* than 4 pounds of meat, but he is *not* willing to give up *more*

than 4 pounds for another basket of fruit.) The *M* of fruit (for meat) is 4 because 1 basket of fruit can be substituted for 4 pounds of meat. The *M* of meat (for fruit) is the inverse of this, namely ¼, since a pound of meat is substitutable for only ¼ of a basketful of fruit if it takes 4 pounds of meat to substitute for 1 whole basket of fruit.

Mark, on the other hand, being relatively more fond of meat, is willing to give up a basket of fruit for only 1 additional pound of meat. Clearly a reallocation of meat and fruit as between Victor and Mark could benefit both. Victor can give up 4 pounds of meat for an additional basket of fruit without being any worse off. This additional basket of fruit could be obtained from Mark who could provide it for only 1 of the 4 pounds of meat that Victor would give up and still be no worse off than in the beginning. This will leave 3 pounds of meat as a pure surplus that could be divided between Victor and Mark and thus make both of them better off than before.

(The surplus can just as well be expressed in terms of the other commodity, fruit. Mark is willing to give up about 4 baskets of fruit for another 4 pounds of meat. Victor would give up 4 pounds of meat for only 1 additional basket of fruit. There is available a pure surplus of about 3 baskets of fruit to be divided somehow between the two.)

In this example, the *M* of pounds of meat for baskets of fruit was ¼ for Victor and 1 for Mark. The *M* of baskets of fruit for pounds of meat would, of course, be the inverse of these figures, 4 for Victor and 1 for Mark.

The optimum allocation of goods involves the equalization of marginal substitutability (*M*).

As long as there is a divergence between the *M*'s of the two goods as between Victor and Mark, it is possible to improve on the allocation of goods, making both men better off. If the divergence between the *M*'s is less than 300 per cent (which it is in our example where the ratio is 4 : 1) the benefit from reallocation will not be as great, but as long as there is some divergence

between the *M*'s there is some gain to be had from an appropriate reallocation of the goods, moving each good from where its *M* is less to where it is greater—meat from Victor to Mark and fruit from Mark to Victor.

This beneficial reallocation may go on until Victor runs out of meat or Mark runs out of fruit. As soon as this happens the reallocation of goods must come to an end. It is no longer possible to recompense each party for what he gives up. It might still be desirable to continue transferring goods from one consumer to another, but that is now necessarily a matter of giving more to one consumer and less to another and belongs to the problem of the *division of income* between different consumers which we are leaving to the next chapter.

Another thing that can put an end to the possibility of further beneficial reallocation of goods is the equalization of the *M*'s as between the different consumers in accordance with the principle of diminishing *M*. As Victor gives up a great deal of his meat and gets more fruit, he is likely to become less willing to give up yet more of his diminished supply of meat for additional baskets of the fruit with which he is now plentifully supplied. He will no longer be willing to give up 4 pounds of meat for another basket of fruit. He may now be willing to give up only 3 pounds of meat for another basket of fruit. If the reallocation still goes on, Victor will after a while be willing to give up only 2 pounds of meat for another basket of fruit. As he gets more fruit and is left with less meat the *M* of fruit for meat (the number of pounds of meat for which a basket of fruit is acceptable as a substitute) falls from 4 to 3 to 2. The *M* of meat for fruit (the amount of fruit, measured in basketfuls, for which 1 pound of meat is acceptable as a substitute) increases from $\frac{1}{4}$ to $\frac{1}{3}$ to $\frac{1}{2}$.

At the same time, Mark is getting more meat and giving up fruit so that he becomes less willing to give up still more fruit for yet more meat. He is no longer willing to give up a basket of fruit for only 1 more pound of meat but is still willing to do so for $1\frac{1}{2}$ pounds of meat. If reallocation still goes on and his supply of meat continues to grow while his supply of fruit dwindles further, he will be still less willing to give up fruit for meat and will hold

out for 2 more pounds of meat before he gives up another basket of fruit. As Mark gets more meat and is left with less fruit the *M* of fruit for meat rises from 1 to $1\frac{1}{2}$ to 2, and the *M* of meat for fruit falls from 1 to $\frac{2}{3}$ to $\frac{1}{2}$.

At this point the *M*'s (2 for fruit or $\frac{1}{2}$ for meat) are the same for Victor as for Mark. No further benefit from reallocation of the goods is now possible. Victor is just willing to give up 2 pounds of meat for another basket of fruit, and Mark is just willing to give up a basket of fruit for another 2 pounds of meat, but neither would gain anything from this reallocation—there is no longer any surplus available. Furthermore, if the reallocation of meat from Victor to Mark and of fruit from Mark to Victor were to go on beyond this point, the *M*'s would continue to change so that they would become unequal but in the opposite direction. There would then be a net *loss* instead of a surplus from the reallocation while a reallocation in the reverse direction, meat from Mark to Victor and fruit from Victor to Mark, would now show a surplus.

The ideal allocation of fruit and meat between Victor and Mark is thus reached when the *M* of each good is at least as high to the one who has it as to the other. If both Victor and Mark have some of each good, then neither's *M* can be greater than the other's, so that they must be equal to each other. This is the exact formulation of the common sense principle that goods should go where they are most useful.[1]

[1] In our example we used the expression *M* or marginal substitutability for two different things. At the outset Victor was willing to permit 1 basket of fruit to be substituted for the 4 pounds of meat which he was willing to give up, and so the *M* of fruit for meat was 4. We also said that the *M* of meat for fruit was $\frac{1}{4}$, the inverse of the *M* of fruit for meat. However, it is possible, and even likely, that, although he is willing to give up 4 pounds of meat for another basket of fruit, he would *not* be willing to give up a whole basket of fruit for another 4 pounds of meat. He might be willing to give up only about $\frac{1}{8}$ of a basket of fruit for this increase in the amount of meat he has. (Or, which comes to the same thing, he could be induced to give up a whole basket of fruit only if he could get more than 4 pounds of meat, say 5 pounds.)

Thus there are two different *M*'s (or marginal substitutabilities) between fruit and meat. It is 4 (or $\frac{1}{4}$ if we measure it the other way) in connection with a reallocation that would give Victor more fruit but less meat. It is 5 (or $\frac{1}{5}$) in connection with the reverse reallocation.

This is because Victor's *M* of meat (in terms of fruit) *diminishes* as he gets more meat (and less fruit), in accordance with the principle of diminishing *M*. The *M* or marginal substitutability splits up into a *marginal relative attachment to the meat he has* as compared to additional fruit that could be substituted for it (which is equal to

We may leave our rather lengthy study of Victor and Mark and generalize the argument to apply to all the consumers of society and all of the thousands of different kinds of goods that they consume:

The best allocation of the different kinds of goods among the different consumers is reached only if the marginal substitutability (M) of every good A for every other good B is not less for every consumer who has some of A than it is for any consumer who has some of B.

If it were less for any consumer it would be possible to improve on the allocation by transferring some of *A* from him to the other consumer and to transfer back some of *B* to him from the other consumer. There would be a surplus which could improve the position of both consumers just as in our initial case of Victor and Mark where *A* was fruit and *B* was meat. This possibility of improving on the initial allocation shows that it could not have been an optimum allocation of goods.

From this it follows that every good which enters into the consumption of more than one consumer must have the *same M* for all those who consume it (since neither's *M* may be smaller than the other's), though it may have a smaller *M* for those consumers who do not have any of it (the *M* being measured in each case in terms of another good which the consumers have in common).

This is automatically reached by free exchange, but is upset if there is any monopolistic exploitation.

Now that we have our principle clarified and developed, how can it be put to practical use? It would obviously not be seriously proposed that government inspectors or psychologists examine

$\frac{1}{4}$) and a *marginal relative eagerness for more meat* which would be obtained by his giving up some fruit (which is only $\frac{1}{3}$). This difference between the *marginal relative attachment* to what he has and the *marginal relative eagerness* for more of it is a result of our large margin (4 pounds of meat). A large difference in the amount of meat and the amount of fruit possessed changes the *M*, and this change shows itself in the difference between the marginal relative attachment and the marginal relative eagerness. We can take smaller units, single pounds or even ounces of meat and correspondingly smaller amounts of fruit, and if we do so this difference can be made as small as we wish. The marginal relative attachment to an ounce of meat will not be significantly different from the marginal relative eagerness for an additional ounce because there is no significant difference in Victor's situation before and after the reallocation. It is this possibility of taking our units as small as we like that gives us the right to speak of one marginal substitutability or *M* between goods, irrespective of the direction in which reallocation is contemplated—whether meat is to be substituted for fruit or fruit is to be substituted for meat.

every consumer's *M* between every pair of goods that is available in the economy and transfer goods from consumers whose *M*'s are less to consumers whose *M*'s are greater.

Fortunately the optimum allocation of goods can be reached automatically without the intervention of any government inspectors. Victor is willing to give as much as 4 pounds of meat for a basket of fruit. Mark is willing to take as little as 1 pound of meat for a basket of fruit. They would therefore both be willing to *trade* at any rate of exchange between 4 and 1 pounds of meat for a basket of fruit. Victor will get more fruit and Mark more meat, and their *M*'s will thereby be moved toward each other. As long as there is a divergence between their *M*'s there will be a corresponding range of rates of exchange at which both parties will be willing to trade. Trade can therefore go on until the *M*'s have been equalized by the automatic redistribution from the trading (or until Victor runs out of meat or Mark runs out of fruit).

The rates at which two people in isolation will exchange are arbitrary (anywhere between 4 : 1 to 1 : 1 as an initial rate of exchange in our example) and will depend upon such things as the relative greed and cunning of the traders and their relative susceptibility to bluff. The surplus from the reallocation may be almost entirely appropriated by one of the parties to the exchange. This will make him better off compared to the other and will affect his *M* and the future rates of exchange at which further trade will take place when one of the parties does not wish to do any more trading at the original rate of exchange. However, this is a matter of the relative amounts of both goods left to the two individuals at the end of the series of exchanges, and this again impinges closely on the division of income between different individuals that we are leaving to the next chapter. It will still be true, if they continue to exchange as long as they can both benefit from exchange, that trade will go on until the *M*'s are equal (or one runs out of the good for which his *M* is less). The final position will always be one of optimum allocation of goods even though different bargaining advantages may have affected the relative well-being of the different consumers and in this way the division of income.[1]

[1] See Alfred Marshall, *Principles of Economics*, Macmillan & Co., London, 8th Ed., 1920, Appendix F, "Nuts and Apples."

More closely concerning us now is the consideration that free exchange need not always lead right up to the optimum allocation of goods. If one of the parties is a superior bargainer and takes into account the way in which the rate of exchange he offers affects the amount that the other party will buy, he may fix his price or rate of exchange at the level which gives him a maximum of gain from trade as compared with any other price. The other party has to accept the price as given and will trade up to the point where his M is equal to the quoted price and then will not wish to trade any further. The first party's M will not be equal to the price and therefore will not be equal to the other's M, so that an optimum allocation of goods has not been reached and both could gain from further reallocation of goods by further trading at a different rate of exchange. But the first party may not be willing to do this for fear of losing the advantage that he reaps because the other takes his price as given. If he traded some more at a new rate of exchange (that would have to be more favorable to the other if he is to be induced to agree to more trading), the other might hold out for this better price on the next day. In order to maintain his monopolistic control over price he foregoes the immediate benefit from the additional trade. He may gain from this sacrifice in the long run, but the other party will lose more than he gains. In considering the effects on the relative well being of the two consumers, we again find ourselves trespassing on the subject of the division of income between individuals. This matter is brought up here, not because it is in itself of great importance in the sense of trade between the two isolated individuals, but because it is the first time we have come in contact with the very important general phenomenon of the harm done to the economy as a whole by monopoly powers exercised in the interest of individuals. It is therefore worth while to make this principle clearer by an arithmetical example.

Suppose Mark the meat eater is an aggressive trader who is able to intimidate Victor by his greater obstinacy. He may set the price at 3 pounds of meat for 1 basket of fruit even though he would not lose in satisfaction if he received only 1 pound of meat for the basket of fruit. Victor accepts the price and trades meat for

fruit until his *M* has fallen from 4 to 3. It will not be worth his while to trade beyond that point because an additional basket of fruit would not be sufficient to compensate him for the 3 pounds of meat he would have to give up for it at the price fixed by Mark.

Mark's *M* has meanwhile been rising and the divergence between his and Victor's *M* of fruit for meat has been narrowed by both movements. However, the price was fixed nearer to Victor's *M* of 4 than to Mark's *M* of 1 so that this will not have moved all the way from 1 to 3 when Victor stops trading. Mark's *M* has risen to, say, 2 at this point; that is, he would be neither better nor worse off if he sacrificed another basket of fruit for 2 pounds of meat. At any price between 2 and 3 both Victor and Mark could gain by trading. But Mark refuses to lower his price below 3 because he knows that by so doing he will destroy his reputation for firmness or obstinacy so that next time Victor would not be willing to trade as much (or even at all) at the price set by Mark, but will hold out for better terms. As a result, trade stops at this point with Victor's and Mark's *M*'s divergent from each other, and the optimum allocation of goods is not reached. This waste is a result of Mark's monopolistic or exploitative policy.

The inconveniences of barter exchange can be avoided by the use of money.

If we leave the unrealistic case of two isolated consumers and consider a society with many consumers, no one of whom has any appreciable influence on the rate of exchange so that every consumer takes the price for granted, both of these complications disappear. There is no indeterminacy of the division of income, no monopolistic exploitation, and no divergence from the optimum allocation of goods. At each rate of exchange between any two goods each consumer is prepared to trade up to the point where the *M* of the good he acquires falls (and that of the good he gives up rises) to the market rate of exchange. The sum of these supplies from all the consumers constitutes the total supply on the market, and the sum of all the individual demands constitutes the total demand on the market. The lower the price at which any

good can be obtained, the greater would normally be the quantity of it each consumer would demand and the greater would also be the total quantity of the good demanded by all the consumers together. If the total demand for any good at any particular rate of exchange were greater than the total supply, this would make the price go up and reduce the demand until it was equal to the supply.[1] Each consumer would exchange up to the point which equated his *M* to the price, and since the price is the same for everybody the *M* will be the same for everybody (except those who get rid of any they may have had and still find their *M* less than the price). In this way the optimum allocation of all the goods is automatically reached through complete freedom of exchange among individuals where nobody exercises any conscious influence on price.

This means that it would be possible in the controlled economy to achieve an optimum allocation of a given supply of goods by simply distributing various amounts of various goods to the different consumers according to some principle of the division of income (or goods in general) among them, and then permitting consumers freely to barter or exchange the goods among themselves so that these come to rest when they are relatively most highly valued.

This method of barter, though not as objectionable as having to submit to an inspectorate for the transference of goods from points where their *M* is less to those where their *M* is greater, would still involve a great deal of trouble. Every consumer would have to find other consumers who wished to trade goods he needed for the goods he was originally provided with. The existence of organized markets for every pair of goods might help, but this would mean 499,500 different markets even if there were only 1000 different kinds of goods. (For each of the 1000 different kinds of goods there would be 999 markets, one for each of the other goods it must be exchanged for, making 999,000, but this would count each market twice, once for each of the two goods traded, so that the figure is only 499,500—still quite large.) And then

[1] The effect of price upon supply is rather more complicated and can be left till later.

there would be the problem of transportation and there would remain unsolved the problem of what to do about nontransferable goods such as haircuts or permanent waves.

Fortunately there is a much better way of solving all these problems. This is to distribute among the various consumers, not quantities of various kinds of goods according to whatever is the principle of the division of income adopted, but *sums of money* distributed according to the same principle. With these sums of money the consumers could bid for the various goods and services available. At a higher money price less of any particular good would be demanded and at a lower money price more would be demanded, and the money price could be adjusted so that the amount demanded of each good was equal to the amount available. The price of each good would be the same for every consumer, and every consumer would direct his money to the purchase of those goods which he preferred and which we assume to be those which give him the greatest satisfaction. He will buy such quantities of the various goods as will make the *M* to him of the goods that he buys equal to their relative prices.

The prices of goods can be made to reflect their *M*.

For example, if meat is 30 cents a pound and fruit is 60 cents a basket, every consumer who buys both meat and fruit will buy them in such quantities (or proportions) as will make the *M* of meat for fruit equal to $\frac{1}{2}$ and the *M* of fruit for meat equal to 2. If the *M* were not equal to the relative price ($\frac{30}{60} = \frac{1}{2}$, $\frac{60}{30} = 2$), the consumer could better his position by buying more of one of the goods and less of the other. Suppose his *M* of fruit for meat were not 2 but 3. This would mean that 1 more basket of fruit could compensate him for the loss of 3 pounds of meat. If he bought 3 pounds of meat less he would save himself 3 × 30 cents or 90 cents. He could get another basket of fruit for 60 cents, leaving him a surplus of 30 cents which he could spend in improving his position by buying more meat or more fruit or more of anything else that he liked. The consumer whose *M* was thus out of line with the relative price would gain by buying less meat and more

fruit until either he was not buying any meat at all or the *M* of fruit for meat had fallen to 2.

With everybody who purchased some of any good having his *M* (in terms of every other good he purchased) equal to the relative price, and with the relative price the same for everybody, the *M* would be the same for all except those who did not buy any of the good (and whose *M* would be less than the relative price). In this way the optimum allocation of goods can be reached with the greatest of ease. Each consumer gets his money income and spends it at the stores on what is most pleasing to him at the current prices.

The principle of diminishing *M* is not really necessary for the optimum allocation of goods to be reached in this way. In its complete absence an optimum allocation of goods could still be reached, but it is the diminishing *M* that adjusts the *M*'s of each individual and makes them equal to the relative price. If all *M*'s were constant or increasing (this would be the case if a consumer's *M* for any good *increased* as he got more of it in exchange for another so that he became *more* eager than before to continue exchanging), exchange under a barter scheme would go on until each consumer had only one good, and under the money income system of distribution he would spend all his money on one quality of one particular good. The observable fact that consumers do no such thing shows that the principle of diminishing *M* holds at least among all the goods that the consumer actually consumes. Any other goods or combinations of goods would appear to be economically irrelevant so that it is permissible and it will be found very convenient to assume that the principle of diminishing *M* holds between every pair of goods for all possible combinations of their quantities.

This solution to our problem of bringing about the optimum allocation of goods might seem to many readers to demand an apology. It seems so obviously what actually happens in the existing free market economy that all the rigamarole about marginal substitutability and barter exchange would appear quite unnecessary. There are two reasons why all this argument is not unnecessary.

The first reason is the horror that many socialists have of anything that reminds them of the existing capitalist world. This makes it necessary to show that the usefulness of money as a means of bringing about a good distribution of goods is not merely a bourgeois belief carried over uncritically from experience under capitalism but can be shown to bring about desirable ends by a consideration of fundamental principles. The first time I heard of socialism and of The Russian Revolution of 1917 they were described to me as attempts to do away with the use of money. Many have been the irrelevant aspects of society that people have tried to do away with in their attempts to build a better world—machines, factories, large cities, railways, police and armies, gold, banking, and money. This is what makes it necessary to show money as an instrument of great power which it would be at least as foolish to discard—if it were even possible—as it would be to forego the benefits of the use of electricity because it is most highly developed in the most capitalistic societies.

The second reason for the long arguments is that they will be used again and again to obtain other results that do not appear so obvious to everybody. It is much more important that the concepts and arguments be well understood than that this particular conclusion be remembered.

In some circumstances the government may interfere with the optimum allocation of goods.

This would be the end of the chapter if our second assumption—that consumers always chose what was best for them (what gave them the greatest satisfaction) were believed to be universally true or if it were believed that any deviation from "rational" behavior in this sense could be dealt with adequately by the appointment of guardians who would be put in charge of the delinquents and who would replace them as "consumers." But this is not the case. The government always wishes, in a greater or smaller measure, to intervene in people's choices and usually argues that the choices with which it wishes to interfere are not wisely made. The consumption of a particular good may be discouraged, as in the case

of whiskey, or encouraged, as in the case of education, by making the price higher for the one (by a tax) or lower for the other (by a subsidy). But this kind of intervention does not come within the scope of the present chapter, for the purpose of the intervention in these cases is the curtailment in the *production* of whiskey and the expansion in the *production* of the service of education, and in this chapter we are concerned only with the way in which a *given* supply of goods and services is distributed most effectively among consumers.

General encouragement or discouragement of the consumption of particular goods and services—even to the extent of actually prohibiting the use of goods or providing them without any charge at all and making their use compulsory—does not come into this chapter. But what is relevant here is *discriminatory* intervention when the government wishes to discourage or encourage certain consumptions among *particular parts* of the population. Then it will interfere by making the prices different for different consumers. It may prohibit the consumption of alcohol or the use of gaming machines by minors. Or it may supply certain foods at especially low prices to low income groups instead of giving them more income. In all such cases the M will not be the same for all and there will not be the best utilization of the available supplies of goods in satisfying the *desires* of the consumers. There may be very good and adequate reasons for the sacrifice of the principle of the optimum allocation of goods in any of these instances, but they must be carefully scrutinized with the full recognition of the significance of this sacrifice before they can be justified.

CHAPTER 3. THE OPTIMUM DIVISION OF INCOME

In Chapter 1 we separated the problem of the division of income among the individual consumers of a society from the problem of the allocation of existing goods of all kinds to make up each consumer's portion. We studied the latter problem, concentrating on the nature of an optimum allocation of goods, and concluded that it could most satisfactorily be brought about by giving sums of money to the different consumers in accordance with whatever principles of the division of income were adopted in our controlled economy and then permitting consumers to buy the goods they prefer in the stores at prices which are adjusted to make the demand for each good equal to its supply. We must now consider the problem of the division of income, remembering that this issue is involved in every decision or action that would make one consumer better off at the expense of another.

Money income can be used to represent real income even if prices change.

The device of paying out sums of money to the consumers and letting them bid for the various goods available goes a long way toward providing us with a mechanism for dealing with the problem of the division of income, but it is not quite self-evident that the problem of how much of goods in general shall go to any consumer is completely identical with how much *money* (money income) shall be given to him.

One consumer may be made better off at the expense of another, without any change in their money incomes, by an alteration in the prices at which they can buy goods. If the price of meat rises while the price of fruit falls, the vegetarian is made better off and the

meat eater is made worse off even though their money incomes are unchanged.

However, if there is some other independent principle that determines the prices of all the goods available, we can take the money income as a measure of the real income in actual goods and services consumed. We have such an independent principle in the principle of maintaining an optimum allocation of goods. Whatever the principle of division of income adopted, the prices of the various goods are determined at the level which makes the demand for each good equal to the supply, and so we can take the money incomes as a measure of the real incomes.

This does not mean that the prices will not change and that such changes will not make some people better off and others worse off even though their money incomes remain the same. It does mean, however, that we must disregard such changes when they are a result of changes in demand for the various goods or changes in the supply while the optimum allocation of goods is maintained. If we could say that in the old situation the actual relative well-being of the consumers was just what we wanted it to be, then, when prices changed, we would want to make some adjustment in the money incomes to offset the change in prices and leave the consumers in the same relative position as before. But we have no way of directly comparing the well-being of different consumers. Our only objective general indication is their money income, and this has not changed. We may have good reason for believing that one consumer is better off than he was before and that another is not as well off as he was before, but we have no more reason for supposing that the old situation is better than the new one (that is, is more like the situation we wish to bring about) than from supposing that the new one is better than the old one.

To obtain a criterion for the optimum division of money income we must assume that different people enjoy similar satisfactions.

This somewhat unsatisfactory state of affairs is an unavoidable result of the impossibility of measuring the satisfactions of dif-

ferent consumers on the same scale. We avoided this difficulty in Chapter 2 by having recourse to *relative* valuations or marginal substitutabilities, which are ratios between objective goods like meat and fruit. How can we overcome this obstacle in our study of the distribution of income and in our search for an optimum division of income?

We can solve this problem if we add two more assumptions to those we made in Chapter 2. We assumed there that the consumers in the economy were capable of feeling satisfaction and that whenever they could choose among two or more alternatives they chose the one that yielded the greatest satisfaction.

The first of the additional assumptions is that the satisfactions experienced by different people are *similar* in the sense that they are the same kind of thing. In other words, that it is not meaningless to say that a satisfaction one individual gets is greater or less than a satisfaction enjoyed by somebody else. When a man says, "Thy need is greater than mine," he may be right or wrong, but even if it is not possible to discover whether he is right there is no need to insist that he is delirious. That the satisfactions experienced by different people are the same kind of thing is incapable of proof. The only justification for making this assumption is that, while there are a few philosophers who argue that we cannot know this, there are no men whose behavior does not suggest the acceptance of the assumption. To reject it would in fact deny meaning even to the assertion that anyone other than myself is capable of feeling *any* kind of pain or pleasure.

This assumption gives meaning to the concept of maximizing the total of the satisfactions experienced by all the individuals in a society. The condition that has to be satisfied if this object is to be attained is that no part of the consumption goods or the income of the society shall go to any individual but the one who can obtain the greatest satisfaction from its consumption. If the income is divided among the individuals enjoying it in such a manner that this criterion is satisfied, then any change in the division would involve the substitution of a smaller (or at best an equal) satisfaction for each one that is lost, so that nothing could be gained by any change.

We must also assume that the principle of *diminishing marginal utility of income* holds generally.

Although there is now a clear enough concept of the achievement of our goal, there does yet not appear to be any conceivable way of doing anything toward bringing it about. It would seem to be necessary to discover how much satisfaction every individual would obtain from every unit of income before the appropriate distribution of income could be attained. The practical possibility of an attempt to reach the goal is brought a little nearer when we make use of the second additional assumption—*the principle of diminishing marginal utility of income.*

This principle asserts that the amount of satisfaction that every individual obtains from his income depends upon the size of his income in such a manner that he always gets more satisfaction from a larger income, *and that the extra satisfaction he gets from a given increase in his income* (the marginal utility of income) *is less if his original income is greater.* Thus an increase of income from $2000 to $2100 would increase his satisfaction by less than when his income is raised by a similar amount from, say, $1000 to $1100. The greater a man's income, the less significant to him is a given absolute rise in income. The principle is called one of diminishing utility because if we considered a man's income to be increased by consecutive equal increments, say, from $1000 to $1100, $1200, $1300 . . . etc., then the extra satisfaction attained by each *raise* (the *marginal* utility of income) would *diminish* from raise to raise because each time he would be receiving the same increase but starting from a higher initial level.

The principle of diminishing marginal utility of income can be derived from the assumption that consumers spend their income in the way that maximizes the satisfaction they can derive from the goods obtained. With a given income, all the things bought give a greater satisfaction for the money spent on them than any of the other things that could have been bought in their place but were not bought for this very reason. From this it follows that if income were greater the additional things that would be bought with the increment of income would be things that are rejected

when income is smaller because they give less satisfaction; and if income were greater still, even less satisfactory things would be bought. The greater the income the less satisfactory are the additional things that can be bought with equal increases in income. That is all that is meant by the principle of diminishing marginal utility of income.[1]

If we now consider the shifting of a small amount of income from one individual to another, we know that the loss to the one and the gain to the other would be equal to the marginal utilities to the individuals of their respective incomes. If the marginal utilities were unequal, there would always be a gain—total satisfaction would be increased—by taking a small amount of income from the individual with the lower and giving it to the individual with the higher marginal utility of income. Such a redivision of income would be in accordance with the idea of giving to him who had the greater need.

The shifting of income from an individual with a smaller to one with a greater marginal utility of income need not stop after one unit of income had been shifted. The inequality between the marginal utilities would probably not have disappeared, so the operation would be repeated and for the very same reason. The principle of diminishing marginal utility of income indicates, however, that sooner or later the procedure will come to a stop. For as the income of the receiving individual increases the marginal utility of his income diminishes in accordance with this principle. At the same time the marginal utility of the shrinking income of the other individual will increase. As this goes on, the difference becomes less and less until ultimately the marginal utilities of the incomes of the two individuals become exactly equal to each other. When such a division of income has been reached, nothing more

[1] This argument assumes that the different satisfactions obtained from the consumption of different goods are independent of the size of the income and therefore of the other goods consumed. The effects of taking into consideration these *complementarities* are considered below. The principle of diminishing marginal utility of income is not to be confused with the principle of diminishing substitutability (M) of one good for another. The former is derived from introspection and the assumption of rationality of choice. The latter rests on the firmer base of the observed phenomenon that individuals do not spend all their income on one single good. The former refers to the effects of *additions* to income, the latter to the effects of the *substitution* of one good for another.

can be done to make the total of the satisfactions of the two individuals any greater. Total satisfaction is maximized by that division of income which equalizes the marginal utilities of the incomes of all the individuals in the society. Individuals with greater capacities for satisfaction than other individuals will be given incomes sufficiently larger to bring them to the point where the diminishing marginal utility of income has made their *marginal* utility just equal to that of the individuals with smaller capacities for satisfaction (or desires or needs) and smaller incomes. Everybody's marginal capacity for enjoying income will be equalized and we will have achieved a fulfillment of the principle: To each according to his needs.

The maximization of total satisfaction by equalizing the marginal utility of income is impossible,

Here we come up against a serious difficulty. Our assumptions that individuals are capable of feeling satisfactions and that their satisfactions are the same kind of thing have given meaning to the concept of maximizing total satisfaction, while the principle of diminishing marginal utility of income has simplified the task to one of equalizing the marginal utilities of income to all the individuals in the society. But we have no means of doing this. There is no way of discovering with certainty whether any individual's marginal utility of income is greater than, equal to, or less than that of any other individual.

If any two individuals were known to have exactly the same capacity of distilling satisfaction out of income, it would also be known that an unequal division of income between them would make the marginal utility of the larger income less than the marginal utility of the smaller income and that an equalization of income would equalize their marginal utilities and maximize the total utility enjoyed by both together. If it were known that one had a greater capacity for satisfaction (at all income levels), it would also be known that an equal division of income would result in a greater marginal utility of income for the one with the greater capacity for satisfaction, so that it would require an unequal

division of income, more going to the one with a greater capacity for satisfaction and less going to the one with a smaller capacity for satisfaction, to equalize the marginal utilities of income and maximize the total satisfaction. But these things are not capable of being discovered. Every individual could declare that he has exceptionally high capacities for satisfaction and so should be given more income than anybody else if total satisfaction is to be maximized; and there is no way of testing the validity of such a claim.

but the maximization of *probable* total satisfaction is attained by an equal division of income.

In the absence of the possibility of discovering, and hence of equalizing, the marginal utilities of income to different individuals it is not possible to maximize the total of satisfactions. It is, however, still possible so to divide income as to maximize the *probable* total satisfaction, making this greater than the *probable* total satisfaction that would result from any other distribution of income. If it is impossible, on any division of income, to discover which of any two individuals has a higher marginal utility of income, the probable value of total satisfactions is maximized by dividing income *evenly*.

That this is so is seen from the consideration that a transfer of income from a richer to a poorer individual would increase total satisfaction if both individuals had the same capacities for satisfaction, for in that case the shift of income would be a movement toward the equal distribution which would make their marginal utilities equal. Such an equality of capacities for satisfaction cannot, however, be assumed to be the case. The richer individual may have either a greater or a smaller capacity than the poorer. If the poorer man has a greater capacity, the gain is increased on that account. If the richer man has the greater capacity, the gain is diminished on that account (and may even be converted into a loss). The possibility of an increase in gain offsets the possibility of the diminution of gain since they are equally likely to occur in any particular case. There remains the net gain that is seen by itself in the case of equal capacities but which becomes only a

probable gain on account of the possible increase or diminution of the gain which arises with unequal capacities.

This argument is illustrated by Figure 1. Curves AA' and BB' show the marginal utilities (measured vertically) of different amounts of income (measured horizontally from either end) enjoyed by two different individuals A and B. The curves are drawn sloping downward away from the vertical axes so as to conform to the principle of diminishing marginal utility. If an income of $200 per month is divided equally between A and B,

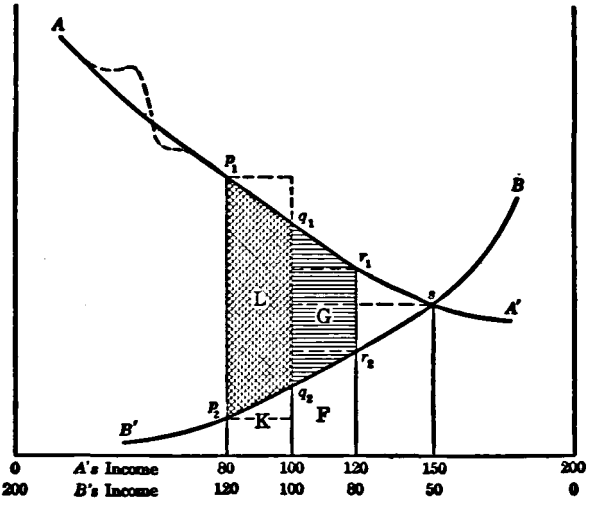


Figure 1

the marginal utilities of income to them will be represented by the height of q_1 and q_2 respectively.

A is represented as having a greater capacity for satisfaction, so his curve is drawn higher up. With the same income as B, the marginal utility of his income is greater than B's, q_1 is greater than q_2. From this it follows that a small diversion of income from B to A would increase the total of satisfactions. This is illustrated in the figure by supposing A's income to be increased by $20 while B's income is reduced by the same amount, the total income of the two being kept constant, $80 + $120 = $200. The area F + G represents the increase in A's satisfaction, while the area

F shows the decrease in *B*'s satisfaction. The net *gain* is shown by the shaded area G and is positive as long as *A*'s marginal utility is greater than *B*'s. Total satisfaction would be increased by departing further and further from the initial equalitarian distribution, shifting income from *B* to *A* until the marginal utilities are equalized at *s*, *A* having \$150 and *B* having only \$50. The total gain would be measured by the area q_1sq_2.

This cannot be done because we have no way of discovering how high the marginal utility curves are or even which of the two curves is higher. Let us again suppose ourselves to start with an equal division of income between two individuals, and consider what we *can* know about the effects on total satisfaction of a departure from the equalitarian distribution. We are now unable to use the curves in the figure and have to make guesses about them. All that we know is that they slope downward away from the vertical axes because of the principle of diminishing marginal utility of income and that one of the curves may be higher than the other.

If now a small shift is made from one individual to another, it may be a shift of the nature of the one considered above, as from *B* to *A*, that is, from an individual with a smaller marginal utility of income to one with a greater marginal utility. In that case there is a net increase in total satisfaction like that indicated by the shaded area G. But it is just as possible that the shift of income will have been from an individual with a greater to one with a smaller marginal utility of income. In that case the change would be of the nature of a shift from *A* to *B*. *B*'s gain would be indicated by the area K and *A*'s loss would be shown by area K + L, so that total satisfaction would have been *diminished* by a net *loss* indicated by the double-hatched area L.

Such a blind shift from an equal division of income is just as likely, then, to increase as to diminish total satisfaction, and if there were a very large number of cases, it might be expected that about half of the shifts would increase total satisfaction and the other half would diminish it. This would leave us indifferent as to the distribution of income in our desire to maximize total satisfaction but for one other thing that tips the scale. Although

the probability of a loss is equal to the probability of a gain, every time a movement is made away from an equalitarian division the probable *size* of the loss is greater than the probable size of the gain. This is shown in the figure where, because of the slopes of the curves, given by the principle of diminishing utility of income, the double-hatched area L (which represents the *loss*) is greater than the shaded area G (which represents the *gain*). Out of 100 million shifts away from an equalitarian distribution of income in a large population, it could be expected that about 50 million would increase total satisfaction and about 50 million would diminish it. In about 50 million cases the shift would be beneficial (from *B* to *A*), and in the other 50 million cases it would be harmful (from *A* to *B*). The total increase in satisfaction received in the beneficial shifts would come to about 50 million times the shaded area G in the figure, while the total loss of satisfaction suffered in the harmful shifts would amount to about 50 million times the double-hatched area L. There would be an almost certain social loss. From this we obtain our conclusion that *if it is desired to maximize the total satisfaction in a society, the rational procedure is to divide income on an equalitarian basis.*

Complications arising from complementarity and irrationality do not affect the general conclusion.

We can now consider a complication we have so far ignored. The argument on which we based the principle of diminishing marginal utility of income is strictly valid only on the assumption that the utilities or satisfactions that an individual derives from the consumption of different goods are independent of each other. If these utilities have complementary relationships, it is possible for the marginal utility of income to increase instead of diminishing with increased income. Items that are rejected and not bought when income is low may nevertheless have a higher utility when a higher income enables them to be bought because the fact that other things are being consumed by the individual increases the utility of these items. Thus it may be that at a low level of income an individual who would have to choose between fine clothes and

an automobile might choose the clothes rather than the automobile, not because the clothes gave a very great amount of utility but because without the fine clothes she would not dare to show herself in the automobile so that the latter by itself would be almost worthless. An increase in her income which enables her to buy the automobile as well as the clothes would then be much more significant than the increase which permits only the clothes to be bought because it enables both clothes and automobile to be fully displayed and enjoyed. The marginal utility of the clothes in the absence of an automobile might be represented by 5 units. The marginal utility of an automobile in the absence of the fine clothes would then be, say, 2. But the utility of the combination of fine clothes *and* the automobile might be 20. The utility of the first increment of income would then be 5 and the utility of the second increment of income would be 15. This would be an example of increasing marginal utility of income.

Complementarities can be positive, as in the example here given, where the possession or consumption of one good enhances the satisfaction yielded by another, or negative, when the possession or consumption of one good is a substitute for another good and diminishes the additional satisfaction that it can provide. In the one case it tends to mitigate and may occasionally even more than compensate for the principle of diminishing marginal utility of income and give rise over some range to increasing marginal utility of income. It does this by creating a kind of discontinuity (making the satisfaction obtainable from clothes plus automobile one indivisible item) which shows itself in a "bump" on the marginal utility curve, possibly even making a part of the curve slope upward with increasing income. In the other case—when it is negative—the complementarity reinforces the principle of diminishing marginal utility of income.

These two tendencies may be expected to have about equal value in the absence of more particular information, so that we would have the marginal utility of income diminishing with the same average intensity as in the absence of any complementarities but subject to irregularities which may sometimes be intense enough to make the marginal utility of income increase for some

income ranges. The principle of diminishing marginal utility therefore reduces to a probability in any particular case, but this is all that is needed as a guide to the selection of that division of income which maximizes the *probable* total of satisfactions in the society.

Second, it is not necessary for our purpose to suppose that the expenditure of income is *always* conducted in a perfectly rational manner. As long as some considerable proportion of expenditure is governed by a rational choice of items that yield a greater rather than a smaller satisfaction, the marginal utility of income will *in general* decline. There will again be irregularities, caused by irrationality in expenditure, that may either increase or decrease the rate of decline, and now and then a particularly serious irregularity may actually bring about an *increase* in the marginal utility of income over a range. This, too, does not affect the *probability* of a generally declining marginal utility of income, which is all that we need for our argument.

The acquisition through experience of capacity to enjoy income may be an argument for equalizing income gradually rather than suddenly.

Third, it might be argued, with some plausibility, that the experience of having a larger income develops a man's tastes and capacities for enjoyment, so that the marginal utility curves of rich men should be considered to be higher than those of poor men. From this it would follow that an ideal distribution of income would give more to people who have been richer. The converse of this, too, may be argued with about the same plausibility, on the ground that a man with a high income gets used to the luxuries that he can afford so that he consumes them almost automatically, hardly noticing that he does so and so getting practically no enjoyment out of what would give a great thrill to the poor man unused to these expenditures. This would offset the first argument and strengthen the case for maximizing aggregate satisfaction by equalizing incomes. However, even if the first consideration were known to outweigh the second, and even if of the two arguments *only* the first were found to have validity, it would still be true that

probable aggregate satisfaction is maximized by an equalitarian division of income. But this would now only be true in the long run. In the long run—in the absence of any knowledge of or belief in racial superiorities—different individuals' capacities for *acquiring* the power of enjoying income can be put in place of the *actual* capacity for satisfaction that we have considered above, and the same results will follow. In the long run probable aggregate satisfaction is maximized by an equalitarian distribution of income. In the short run, however, there is a difference. It would be best, if experience of higher income had the effect of *raising* the curve of marginal utility of income, temporarily to leave more than the average of income with those whose income has been higher. This consideration is of importance together with many other considerations in the *dynamic* problem which we are not here discussing, of how rapid should the *transition* from a previous unequal to an ideal equalitarian distribution of income. It cannot affect our conclusion that if it is desired to maximize satisfaction the ideal to be aimed at is an equality of income.

If acquisitiveness indicated greater capacity to enjoy income, an unequal division would be the optimum,

A qualification might appear to be necessary for differences in income due to greater effort. When an individual works longer hours or applies himself more assiduously to his work there is a presumption that he has a greater use or need for the extra income he gets from the extra effort and should get a larger income.[1] This does not follow. A man works harder only when the wage plus the attractiveness of the work itself (or minus the irksomeness of the work) in relation to his enjoyment of the alternatives of leisure or taking his work easy, his M of pay-plus-work for leisure, is greater than for others who prefer not to make the extra effort. There is no more reason for believing that he works harder because an additional dollar is worth more to him (in which case he should get a larger income) than that he works harder because the marginal disutility of work or the marginal utility of leisure

[1] I am indebted for this observation to Dr. Carl T. Devine, now of Johns' Hopkins.

is less for him (in which case he should get a smaller income). But even if this qualification should be justified it does not seriously affect our conclusions.

but concessions that have to be made for other reasons would meet this point too.

As we shall see, the principle of equalizing income does not preclude permitting those who so wish to sacrifice some of their leisure for extra income. This opportunity is especially valuable for those who value money income more and tends to correct this deviation from the pure theorem here enunciated by permitting these individuals to get a higher income by working harder. It should be remembered moreover that it is only where greater effort and not chance or inheritance causes inequality that this consideration is relevant, and in these cases it will be necessary to permit some inequality for the quite different reason that the principle of equality would have to compromise with the principle of providing such incentives as would increase the total of income available to be divided.

Our argument also assumes that satisfaction is derived only from one's own income,

Implicit in our argument for the equalitarian division of income are several other assumptions of a more negative character which amount to little more than the assumption that there is no special reason for preferring an unequal division. One of these is the assumption that each individual's satisfaction is derived only from his own income and not from the income of others. If the poor derive more satisfaction from seeing the splendors of the very rich than from an alleviation of their own poverty the argument might be upset (though even so they would be free to contribute to a Society for the Provision of Spectacular Expenditures and get the spectacles more economically). On the other hand, the pain that even those who are not poor get from the existence of poverty and the envy that the poor feel for the rich would tend to strengthen

our argument for equality. Sometimes the spectacular expenditure of the rich is identified with cultural values. This is an extremely precarious thesis and rather less plausible than its converse that spectacular expenditure by the rich tends to destroy cultural values and distort rather than develop true artistic appreciation. But in any event these considerations merely stress the "probability" nature of our conclusion. If we knew the effect in every particular case it is virtually certain that an *unequal* division of income would be the best possible, and the consideration raised in this paragraph would have to be taken into account, as well as the marginal utility curves, but in the absence of this unattainable knowledge our conclusion in terms of probabilities still holds.

and that individuals are equally sensitive to increases as to decreases in income.

Another implication of our argument for equality of income is that the marginal utility of income is the same, for any individual with a given income, whether we consider his income to be increased or decreased. But it is possible that a man feels an increase in income from $2000 to $2100 less keenly than he would feel a decrease in income from $2100 to $2000. Or he might feel the latter more keenly than he feels the former. If a diminution in income is felt more keenly than an increase, any redivision of income (which must involve a diminution as well as an increase) is harmful just because it is a change. This is clearly seen if we consider a redivision that neither increases nor decreases the inequality of the division of income. If income is shifted between two individuals in such a way that their relative position is exactly reversed, there is no difference in the degree of inequality, yet there is a net loss because the gain is felt less keenly than the loss. Even if the inequality is reduced by the shift the harm from the act of redivision may be greater than the benefit due to the superiority of the new division over the old division. This consideration might appear to upset our conclusion in favor of redivision of income from the existing unequal division to an equalitarian division. However, this is not really so.

But even if these assumptions are not warranted the general conclusions still hold.

This comes to the same as the problem already considered of the rich who were supposed to have acquired a greater sensitivity to income. Greater sensitivity to diminutions of income than to increases in income would merely lead to the conclusion that redivision would be slow. If the sensitiveness to diminutions of income is so much greater than the sensitiveness to corresponding increases of income that the benefit from an increase in income is quite negligible in comparison with the pain of the decrease in income, nobody's income should be reduced, but it would still follow from our analysis that no income above the average should be increased and that no member of the new generation should be given an income above the average, for in the determination of new incomes to new members of the population the problems from acclimatization to high income need not be allowed to arise in the first place. In less extreme cases there should be some redivision of income from rich to poor but at a slow rate, the rate of redivision being that at which the harm done by the change itself is just great enough, at the margin, to offset the gain from the improvement in the division of income. A more rapid rate of redivision would do more harm than good (at the margin) while a less rapid rate of redivision would mean the abandonment of some benefit from improved division which is greater than the harm done by the additional change that this would entail.

If consumers were more sensitive to increases in income than to decreases, the queer conclusion would follow that *any* change would be good unless it made the new division more unequal than the old; and even in that case there would remain a net gain if the change were followed by a change back to the previous division. The final division of income would be the same as in the beginning, but of the intervening increases that were followed by decreases and decreases followed by increases there would remain a net gain from the greater sensitiveness to the increases. This is indeed the exact opposite, as might be expected, of the situation where decreases were felt more keenly than increases. There the change

had to be checked and minimized even if the final results of the change were good. Here change is good in itself quite apart from the benefits from an improvement in the division of income. On this hypothesis the best thing is to have all income fluctuating as widely and as rapidly as possible though it would still be desirable for the fluctuation to be around an equalitarian base if the greatest satisfaction is to be obtained from a given total income with a given degree of change. This conclusion should be sufficiently strange to permit us to rest content, until such time as we get some knowledge on the subject, with the implication that consumers are equally sensitive to increases and decreases in income. This we can do with a clear conscience since our conclusions are not being put forward as a proposal for the immediate redivision of income on a perfectly equalitarian basis, but merely as a solution of the theoretical question of what would be the division to be aimed at if a *given* total income were to be divided in the most effective manner for the maximization of total satisfaction.

There are other arguments that are frequently adduced to support the policy of equality of income. It may be argued that an equalitarian division of income is more "fair" than any other or that it is favorable to the development of a feeling of comradeship and friendliness. It may be hailed as a corollary of the notion of the Brotherhood of Man, or it may simply be proposed as an end in itself with a direct aesthetic appeal unmarred by any rationalization or logic chopping. For most people who sympathize with the conclusions of this chapter these other more direct and more easily understood arguments usually have a greater appeal. The argument here developed is, however, not supererogatory. Its purpose is to bring to light the implications hidden in *obiter dicta* that propositions about the division of income belong to the realm of value judgments and that arguments based on the concept of maximizing total utility have no meaning, so that the economist, as such, can say nothing about an equalitarian division of income, while as a member of society all that he can say is that he likes it or dislikes it.

To reject the conclusion that the optimum division of income is an equal one is not more impartial or scientific than to accept the assumptions on which it is based.

What our analysis brings out is that the maintenance of a dignified and "scientific" impartiality on this proposition contains just as much of value judgment as the conclusion that an equalitarian division maximizes the probable total of satisfactions enjoyed by the individuals in the society. For, unless we have made a mistake in our argument, the refusal to accept the conclusion can rest only on the refusal to accept one or both of our two primary assumptions—that is, a refusal to suppose that other people have any capacity for feeling satisfactions and/or a refusal to suppose that all people enjoy satisfactions of a similar or comparable character. We may dismiss the former as so arrogant as to render it too unlikely that any sane person would really hold it (though a solipsistic philosopher might make some such statement during office hours), and the latter because it makes a distinction between individuals that could be defended only by the exponents of Nazi biology.

It is important to note that the argument given above in favor of an absolutely equalitarian division of income is not directly applicable to practical policy. It is the solution to the problem of maximizing the probable total of satisfactions that can be attained by the member of society from a *given* income. As soon as other objectives are introduced which come into conflict with this one a compromise has to be made. It will probably also be desired to produce as great an income as possible for division among the members of society, and if a greater total income would be produced if the division of income were less equalitarian, a compromise cannot be avoided.

The general argument does not rule out particular cases where some reason may be given for particularly high or particularly low needs, for instance, of invalids on the one hand or ascetics on the other. But wherever there is no good specific reason for inequality the general principle holds that if we would maximize satisfaction from a given total social income the rational procedure is to equalize individual incomes.

CHAPTER 4. DIVISION OF INCOME AND ALLOCATION OF GOODS IN THE UNCONTROLLED ECONOMY

It would be premature at this stage to criticize the actual division of income and wealth in an uncontrolled economy like the United States. In the last chapter we showed how purely welfare considerations in the distribution of a *given* income should lead us to advocate as equalitarian a division of income as possible. But we have not yet considered how we might be led to depart from this principle by other criteria and in particular by the necessity of inducing a sufficiently large income to be produced. We must therefore leave this criticism until we have completed these tasks.

Great inequalities of income create wants, and this has the same bad effect on welfare as the destruction of means for satisfying wants.

Nevertheless there are a few things to be noted here in connection with the way in which the division of income affects the allocation of goods. With the tremendous inequality of incomes and consumption and the canalizing of man's emulatory instincts in the accumulation of wealth and the spending of income, men's needs are vastly exaggerated. The greater part of the needs of all, except perhaps the very poorest, consists of things that are needed neither for physical health or comfort but in order to "keep up with the Joneses." Very often indeed important needs of physical health and comfort are sacrificed for the sake of keeping up appearances, so that there is a tremendous waste even in those resources which are directed toward producing the things that consumers demand. This is another powerful reason for more equalitarian division of income that we must bear in mind when we have prepared ourselves for the final appraisal.

It has been argued that pecuniary emulation is the least harmful form that natural human aggression can take; that the alternative is naked oppression by autocrats and bureaucracies with concentration camps and torture chambers. It would appear that this conclusion is the result more of a pessimistic outlook than of scientific reasoning. In many fields of endeavor the pecuniary interest is of relatively insignificant dimensions and there would appear no good reason where it is impossible or inadvisable to do away with it, why the game cannot be continued just as well with counters of a smaller size—why the difference between an income of $3000 and $4000 should not be at least as strong an incentive as the difference between $3,000,000 and $4,000,000.

The allocation of goods would seem at first sight to be carried out remarkable well in the uncontrolled economy. With freedom of exchange and free markets open to everybody each consumer would adjust his marginal substitutabilities to the relative prices, and with the same prices effective for everybody all the marginal substitutabilities of every good for every other good will be the same for all who consume some amount of each good, so that we would have an optimum distribution of goods.

But before we proceed to give the uncontrolled economy full marks for its allocation of goods we must note a number of ways in which it falls short of this perfection.

First there is the point already made in connection with the extremely unequal division of income. Because consumers try to live up to standards beyond their means they do not choose wisely. This is a peculiar kind of faulty choosing in that it cannot be corrected by intervention to make consumers choose differently or by appointing guardians to choose for them. Any such intervention would only make their situation still worse because they would feel that they would rather have the goods that raised their ostensible standard than the goods that the guardians considered better for them. But a diminution of the inequality of income and of the spirit of pecuniary emulation would permit consumers of their own free will to consume the good which the craving for displaying the achievement of certain ostensible standards had forced them to sacrifice.

Ignorance, certain forms of advertising, and monopoly distort the allocation of goods.

Next there is poor choice because of lack of information or because of deliberate misinformation in imperfect markets. The same or practically similar goods are sold to different people at different prices under different labels so that their marginal substitutabilities, though proportional to the different prices, are not equal to each other and the optimum allocation of goods is not reached. To some small extent this is remedied by pure food laws and by organizations like Consumers Union, but much remains to be done before consumers everywhere are given a fair chance of knowing and judging what they buy.

A great deal of this is the result of advertising. Some advertising improves the allocation of goods by providing useful information to consumers, but most of it has the effect of stressing partly or wholly imaginary differences between goods and so persuades consumers to pay a higher price for the differentiated product. (It must be remembered that here we are concerned only with the different prices paid for the same or practically the same product and not with the social usefulness of the resources devoted to advertising or with the effect on the relative quantities produced of different goods or with the effects on the division of income.)

Monopoly destroys the equality between relative prices and marginal opportunity cost so that M is not equalized among different consumers.

Finally, there is the imperfect allocation of goods because of the influence that buyers or sellers have on the price. Whenever a buyer can influence the price at which he buys anything by varying the amount he buys, he will no longer equate his marginal substitutability to the relative price; so even if prices are the same for everybody the marginal substitutabilities will not be the same and we will not have the optimum allocation of goods.

This is because each consumer, in using his income to the best advantage, equates his marginal substitutability to the relative cost to himself of various goods at the margin, and it is only if he

considers the price as unaffected by the amount he buys (whether it really is unaffected or not) that the relative marginal cost to him comes to the same thing as the relative price.

Let us consider a fairly simple arithmetical example. A consumer is considering how to distribute some part of his money income in buying meat or fruit. Meat is 30 cents a pound and fruit is 60 cents a basket and the consumer is unable to affect either price by the amount that he buys. Then the price comes to the same thing as the cost of another unit of each good or the saving in buying a unit less of each good. If he buys a basket of fruit less he saves himself exactly 60 cents, which will just enable him to buy an additional 2 pounds of meat. The basket of fruit is what he has to give up in order to get another 2 pounds of meat, so we can say that the *marginal opportunity cost* of 2 pounds of meat is 1 basket of fruit. In exactly the same way the *marginal opportunity cost* of another basket of fruit is 2 pounds of meat. The consumer acquires such quantities of fruit and meat as make his marginal substitutabilities (or *M*'s) equal to the marginal opportunity costs. The marginal opportunity cost is the same as the relative price. The relative prices are the same for everybody. All consumers of meat and fruit equate their *M*'s to their marginal opportunity costs, so all their *M*'s are equal and the optimum allocation of goods is reached.

But if a consumer is able to influence the price by varying the amount he buys, this relationship is upset. Suppose a consumer who buys 30 pounds of meat believes that if he were to buy another pound of meat the increased demand would raise the price to 31 cents. Then the cost to him of buying another pound of meat is greater than the price. To buy another pound of meat will cost him not only the 31 cents that he pays for the thirty-first pound but an additional 1 cent on each of the other 30 pounds for which he now has to pay 31 cents a pound instead of 30 cents. These 30 pennies must be added to the price to give 61 cents which is what it cost him to buy another pound of meat. Suppose he does not believe himself able to influence the price of fruit by varying the quantity of fruit that he buys. Then the marginal opportunity cost to him of another pound of meat is slightly more than 1 basket of

fruit (61 cents as compared with 60 cents) and the marginal opportunity cost of another basket of fruit would be about 1 pound of meat. So, although the relative price of fruit is 2 : 1, his marginal opportunity cost is about 1 : 1. His *M* will be 1 : 1, and different from that of any other consumer whose marginal opportunity cost did not happen to diverge from the relative prices in the same direction and to exactly the same degree as his own did. As different consumers would have different *M*'s, the optimum allocation of goods would not be reached.

The influence of *sellers* on price also works in a similar way. Its most important effect is on production, but it also has important effects on the allocation of goods that have already been produced. The most spectacular case is where sellers restrict the amount that they sell so as to get a higher price and destroy the remainder—as happened with coffee. Here no economic analysis is necessary to show that goods are not utilized to the best advantage, but it is of interest to note that it fits into our present category of waste, appearing as a limiting case. The relatively high prices of coffee to those who use it as a beverage and the very low (or zero) price to those who use it as locomotive fuel (or destroy it) indicates a great difference between these two groups in the *relative price* of coffee as compared with other goods, in its *marginal opportunity cost*, and in the *marginal substitutabilities*. Therefore there is an extreme divergence from the optimum allocation of goods.

Government crop restriction plans and the two-price stamp plan interfere with the optimum allocation of goods for the sake of helping farmers and others.

Our principle of the equalization of *M*'s between all consumers as a criterion of the optimum allocation of goods is equally applicable to all departures from the freely competitive market in the disposal of any product. It indicates poor allocation of goods not only when part of a crop is plowed under so that the rest can be sold at a better price but wherever anything (like electric current) is sold at different prices to different people or for different pur-

poses,[1] or where consumers are not free to buy the quantities they wish to buy at the ruling prices. When there are different prices to different consumers, they will (if they cannot influence price) buy such quantities as make their *M*'s proportional to the different prices and therefore unequal. Where consumers are not free to buy the quantities they wish they cannot adjust their *M*'s to the relative prices. In all such cases there is a social loss because of the departure from this optimum allocation of goods. Yet such divergences are often the result of government action intended in the social interest. It will be worth while examining more closely three different types of such intervention.

The first is the practice by the AAA of destroying part of a crop so that the farmers can get a better price from the remainder. This entails a departure from the optimum allocation of goods in that the goods destroyed are not put to that use, of all the possible uses for the good, which is most highly esteemed. The relative significance of cotton is not necessarily greater for the farmer who plows it in for manure than for the man (who may be the very same farmer) who cannot afford to buy a shirt which is thereby made more expensive. Yet it seems probable that the whole situation with part of some crops destroyed and the farmers getting enough to eat is preferable to one in which there is an optimum allocation of goods while the farmers starve.

This is possible because the essence here is that the optimum allocation of goods is sacrificed by the AAA for the sake of an improvement in the division of income. If the sacrifice were unavoidable—if there were no other way of preventing the farmers from starving—there could be no objection to the plan. But our analysis shows that the sacrifice is not necessary. If the allocation of goods is not an optimum one it is possible to reach an optimum, thereby benefiting some (or even all, depending on how the benefit is shared) and hurting none. If it is recognized that the purpose of the policy is to raise the income of farmers, it can be seen that this can be done more satisfactorily by making the farmers a money

[1] The argument would not apply to charging different prices at different times or at different places where the current is a different good with a different marginal cost, as we shall see.

grant out of the general funds of society. The money can be raised, if taxation should be necessary,[1] by taxing all those who would have had to pay a higher price for cotton goods (if that is the item we are concerned with). More than enough could be collected in this way to benefit the farmers (and landlords and other interested beneficiaries of the AAA program) at least as much as they are benefited by the AAA program, and still have the taxpayers better off than if there had been no tax but they had had to pay a higher price for cotton goods. The net benefit will come out of avoiding the destruction of socially useful crops.

Our second case is where different prices are charged for the same thing to different consumers as by the stamp plan for distributing surplus commodities to low income groups. This is a great improvement over destroying surpluses or dumping them abroad. But the same objections hold, though to a lesser degree. Here it is even more clearly seen that the justification for the bad allocation of goods is that the action is at the same time a palliative for the bad division of income. The conclusion is the same, too. An equalization of prices together with a transfer of money income from the consumers of the good who would now get it more cheaply to the farmers who have to sell it at a lower price and to the low income groups who benefit by the stamp plan could leave every individual concerned better off than under the stamp plan, the general benefit coming from the better allocation of goods.

Direct help would be better for everybody because it would not interfere with the optimum allocation of goods.

This can be proved as follows: Suppose the stamp plan abolished and all the goods sold on the market at a single price, higher than the price to the beneficiaries from the stamp plan but lower than the price paid by the public (to whom the price had been artificially raised by taking off the market the supply that was distributed under the stamp plan). The public would gain, the stamp plan beneficiaries would lose their benefits, and the farmers would lose.

[1] That it need not be necessary will be seen in Chapter 24.

Suppose the public saves $1,000,000 by being able to buy the previous amount at the lower price, and spends $500,000 in buying the rest of the crop which had been sold to the poor for $100,000. The farmers in that case will be worse off by $600,000 ($1,000,000 loss from lower price of old sales to the public *minus* $400,000 gained from selling the rest to the public for $500,000 instead of to the poor for $100,000). Suppose now the public is taxed $1,000,000 of which $600,000 is paid to the farmer and $400,000 to the poor. The public will still be better off than under the stamp plan because, although it now loses all that it gained from buying at a lower price the amount it previously bought, it still gains from being able to buy more of the commodity at the lower price. The poor would be better off because they can still, if they wish, buy the same amount of the surplus commodity as before, but they are free if they prefer, and they almost certainly do, to spend the $400,000 on other things that do not happen to be available under the stamp plan. The farmers will be better off, not indeed as farmers, but as members either of the public or of the poor. A slightly larger tax, with a bigger benefit to the farmer and perhaps a smaller payment to the poor, could certainly leave all these classes better off with the farmer better off even as a farmer.

Any other figures can replace those in the example and the same result will be obtained. If a redistribution of income is desired it is best brought about by a direct transfer of money income. The sacrifice of the optimum allocation of goods is not economically necessary.

The same argument, leaving out the poor, is applicable in relation to destruction of crops. If the destroyed crops were marketed and the public got the larger crop for $1,000,000 less (because the price is very much lowered), the farmers would receive $1,000,000 less. If now the public is taxed $1,000,000 which is paid to the farmer, the public will still be better off by having the whole crop instead of only a part of it, while the farmer will have the same income and as a member of the public will be able to buy shirts more cheaply. A slightly larger tax and transference will improve everybody's position, even that of the farmer who uses no cotton products.

There is a difficulty in the practical application of this principle (which would lead us immediately to abolish all indirect subsidies and replace them by direct monetary subsidies). It is not practical to tax each citizen by the same fraction of what he gains from the cheapness of the good in question. A tax on the amount of the good bought would defeat its object. If the farmer could *maintain* the price he receives for the product the consumer would have to pay more by the whole amount of the tax, so he would restrict his purchases and some would be unsold. If this remainder were destroyed and the tax revenues paid to the farmer the situation would be just as in the AAA situation. If the remainder is sold at special prices to the poor we are in exactly the same position as with the stamp plan. If some of the taxes are paid to the poor, this will be some improvement over the stamp plan, but the problem of what to do with the unsold goods is not solved and we do not have an optimum allocation of goods.

On the other hand, if the farmer *lowers* his price so that all the crop is sold in spite of the tax, he must lower it enough so that the price to the consumer remains the same (or the crop will not all be sold) and the farmer pays the whole tax. When the tax revenue is paid to him he will still be no better off than before anything was done by the AAA to save him from starving.

The tax therefore has to be independent of the individual's actual expenditure on the good. Otherwise he will reckon it in the price and try to avoid it by buying less of the good. The tax might be based on some index of the importance of the good to him, such as the amount of it that he bought in some past period, but that would be a very cumbrous kind of tax. Almost the only thing left is an income tax, but nothing could be more appropriate since the fundamental purpose is to improve on the division of income.

The objection is that there are sure to be some income tax payers who buy very little or none at all of the good and who would not be sufficiently recompensed for this reduced income by the lower price of the good. Such individuals would be better off under the AAA or stamp plan and would object to the income tax or ask to be let off on the ground that the reduction in the price of the good does not benefit them enough.

The answer to this has already been given in Chapter 3, pp. 23–24. We have here a change in relative prices that makes some people better off and some worse off, but we must disregard such changes because we have no way of telling whether the change is an improvement or a worsening in the desired distribution of well-being.

Rationing also sacrifices the optimum allocation of goods in a roundabout attempt to prevent the rich from outbidding the poor.

Our third case is that of rationing—of special interest at this time when in the name of the defense emergency the government is limiting the free market more and more by priorities and rationing.

Both rationing and priorities come into effect because the authorities do not wish to allow prices to rise sufficiently to reduce demand to the available supply. The price mechanism as a means of allocating goods being thus rejected, other devices such as rationing have to be provided in its stead.

There may be very good reasons for not permitting the price mechanism to operate. If an essential consumption good is very scarce the price mechanism cannot be allowed to work in the normal way because that would permit the wealthy to buy up all of it and use it lavishly and wastefully while the poor suffered great hardships. This is the usual reason for the rationing of consumption goods that become very scarce in emergencies like war. If there is very little meat it is better for it to be rationed, everybody getting his small share, than for the rich to raise the price so that the majority of the people can hardly afford to buy any while the rich suffer no appreciable hardship. (If they are making unusually high profits they may even consume more of some scarce goods than in normal times when there is plenty for all.)

Rationing will violate the principle of the optimum allocation of goods because rations have to be more or less equal to satisfy common notions of fairness and thus consumers are not able to adjust the quantities of the different goods they consume to the different proportions which, with their different tastes and needs, would equalize the marginal substitutabilities. The tendency for

illegal sale of ration tickets for money or in exchange for other tickets is evidence of the attempt, by exchange, to improve on the unsatisfactory allocation of goods that is brought about by rationing.

Here again the fundamental trouble is the unequal division of income which becomes unbearable when some essential goods are very scarce. The losses from wrong allocation of goods could be avoided if, instead of rationing, the poor were given enough money income to permit them to compete with the rich in the purchase of scarce goods. (This would not be sufficient if the poor spent their additional income in buying not what is necessary for their health and efficiency but in fancy luxuries that attract inexperienced spenders. That would be a case where guardianship is necessary and may legitimately take the form of rationing.)

However, a distribution of money to the poor on a sufficient scale to prevent them from being deprived of necessities by the rich bidding against them on the market may be impracticable for a number of reasons. Perhaps the most important is that it would create too much spending power in the whole economy with increased demand not only for the necessities that should be shared equally but for all sorts of other goods with resulting increases in their prices. The increased prices and higher cost of living would lead to demands for higher wages. This would raise costs and lead to still higher prices and we would be in the midst of the vicious spiral of inflation. This must be avoided if at all possible and the interference by rationing with the allocation of goods may be a lesser evil.

The same situation can arise when the scarce good is not a consumption good that the poor must not be deprived of, but some material like aluminum that is essential for the defense effort. By a system of priorities and rationing or even prohibition for civilian purposes, the government can get the aluminum it needs without raising its price, and this is often considered to be a sufficient justification. Again there is a wasteful allocation of goods, which would be avoided if the prices were allowed to rise sufficiently to indicate its relative scarcity. Those who happen to have or to get some aluminum use it lavishly, now its price is low, and many who

need some for relatively urgent purposes and who would pay high prices are unable to get it because they are not sufficiently expert at unwinding the red tape that unavoidably takes the place of the price mechanism. In such cases too it may be considered inadvisable to let the price rise, partly because it would permit some people improperly to make enormous profits "out of the nation's dire need" and partly because the rise in prices may threaten inflation. And so prices are kept down by law, distribution is governed by rationing or priority certificates, and the optimum allocation of goods is sacrificed.

All the benefits can be obtained, without this loss, by *general rationing* of purchasing power.

These are insufficient reasons for abandoning the price mechanism (although rationing may be justified as a form of state guardianship to prevent foolish spending). Excessive profits can better be taken away by taxes than by keeping the price of the product down. It will be no worse for the potential "profiteer" and the benefits of the price mechanism will be retained. When there would be too much spending and danger of inflation, spending can more rationally and more equitably be kept down to the safe level by the imposition of taxes on all incomes or expenditures, so that the remaining incomes can buy the goods that are available at the old prices. The taxation will be no real hardship since it only serves to keep prices down and never decreases the actual quantities of goods and services that the taxpayers are able to buy and consume.

It may be that the supply of goods and services is so curtailed that extremely heavy taxation is necessary if spending is to be curtailed sufficiently to prevent rising prices and inflation. Such extremely heavy taxation might be considered inadvisable because of its effects on people's morale and their willingness to put forth their best efforts in the national emergency. Yet even if this is so it is not necessary to resort to rationing and priorities and the accompanying wasteful allocation of goods. In such a situation the consumption of goods will have to be low whatever the devices

employed. The only thing that can be given as an incentive is a promise of more after the emergency is over. This can be accomplished most satisfactorily by a device called *general rationing*[1] which would limit the amount of money that can currently be spent on the kind of goods that are scarce. The amount that can be spent would be the same as what would be left by the simpler but too heavy load of taxation that would prevent inflation. The difference is that some of this income is not taxed away but left with the individual though it is not available for current expenditure on the scarce goods. It may be spent on goods that are not scarce or it may be saved and released for expenditure only when the emergency is over.

The same "general rationing" is interesting as an indication of one of the two ways in which this device—which does not look very much like rationing of goods at all—was developed. A gradual improvement in systems of rationing finally leads to this which is its negation in all but name. In England and much more so in Germany the interferences with allocation from rationing many different kinds of goods became too wasteful and too troublesome. Everybody had to be given fixed amounts of hundreds of different kinds of food and clothing and so on in proportion which had to be rigid and so could not fit individual requirements. It also became too complicated. So there was developed in Germany a system of widening the items for which ration cards were issued. Instead of having different cards for coats, for trousers, for overcoats, for shirts, and for handkerchiefs, everybody was given an over-all ration for "clothes" and the different items were given different number of "points" which could be added to make up the ration in the way which pleased the consumer best. This great simplification in administration and improvement in the distribution of goods can easily be recognized as the readmission of the price mechanism in a limited and disguised form with "points" taking the place of prices and the coupon a kind of subsidiary currency.

In England similar devices were developed in a more trans-

[1] Proposed by Dr. M. Kalecki of Oxford, England. See Bulletin No. 3 of the Oxford Institute of Statistics.

parent manner still. Thus there are meat rations, not in terms of quantity of meat—which would raise very difficult problems as to qualities of meat—but in terms of money cost. The consumer is allowed to spend a certain amount of money each week in the butcher shop and can buy either a small amount of an expensive cut or a larger quantity of cheaper meat.

From this it is a small step to "general rationing" which would not fix the amount of *any* particular commodity a consumer can buy or even the amount of money spent on a particular good or on a particular branch of consumption. By the single step of limiting the amount to be spent on all kinds of scarce goods, not only are the rich prevented from outbidding the poor but the danger of inflation is met, while the optimum allocation of goods is maintained. In effect the limited amount of money that can be spent on the scarce good is more valuable than the money that cannot, and that therefore must either be saved or be spent on goods which are plentiful. The former money is really worth more than the latter, so that the device is really one of making incomes more equal by raising the buying power of a given amount of money which is about the same for a rich man as for a poor man.

The other root from which "general rationing" developed shows more clearly how different it really is from rationing proper. Actually it was prepared by Dr. Kalecki as an improvement on Mr. Keynes's plan for deferred pay.[1] To prevent inflation because of excessive spending Mr. Keynes proposed that a part of everybody's income (above certain minima) be deferred until after the war when more goods would be available and when the spending of this money could help prevent depression and unemployment. Dr. Kalecki pointed out that this would not prevent the rich from maintaining or even increasing their present consumption out of their old savings, and so proposed that everybody's expenditure be limited *directly* where it was not advisable to take away enough in taxation to force expenditure down to the required level. It is always a restriction on expenditures that is needed in the last resort to prevent inflation or to prevent consumption by the rich of goods that are needed by the poor or for the war effort.

[1] J. M. Keynes, *How to Pay for the War*, Macmillan & Co. Ltd., London, 1940.

Monopolistic interference with the best use of goods can be met by *counterspeculation.*

We have considered at some length what more rational measures could be adopted in the place of some interferences with the price mechanism that are undertaken by governments in the general interest. We may now turn to consider what might be done in the cases considered above (pp. 43–45), where the power of individuals to influence price leads to wasteful allocation of goods even though there is no intervention by the government.

All these maldistributions of goods can be remedied by taking away from such buyers and sellers the power to influence the price by their speculations. This may sometimes be done by destroying a monopoly (as the Department of Justice sometimes tries to do) or by entering into competition with a seller, but these are again problems of production. However, the power of buyers or sellers to influence price can be attacked directly by a device which may be called *counterspeculation*. The government through a special board estimates what would be the price of the good that would make demand equal to supply if there were no restriction of the kind we wish to abolish. It then guarantees this price to all the sellers in the case of a seller's restriction or to all buyers in case of a buyer's restriction. The buyers (or sellers) then know that the price will not move against them if they buy or sell more and that they will not get a better price if they restrict their dealings. The Board of Counterspeculation then buys in the free market what it has promised to sell to buyers at the guaranteed price or sells in the free market all that it has undertaken to buy from the sellers at the guaranteed price. The Board of Counterspeculation will make a profit or a loss if it makes a mistake and these may be expected roughly to cancel out. With experience it will be able to estimate more and more accurately and to guarantee for longer periods. By this means the benefits of an optimum allocation of goods may be brought about when the natural forces of competition fail to do this.

All of the devices discussed in this chapter are much more important in connection with the production of goods than with the

problem of the optimum allocation of goods already produced, but the principles are very similar, as we shall see when we come to them, so our occupation with them here will make it easier to follow their application to production. As for the actual allocation of goods that have already been produced we may after all give a very good mark to the uncontrolled economy where the allocation of goods has not been interfered with by roundabout government attempts to improve the division of income without quite realizing what it is after.

CHAPTER 5. SIMPLE PRODUCTION I
(*IN A COLLECTIVIST ECONOMY*)

In this chapter, where we first approach the problems of production, we shall consider only a very simple form of production—so simple that its artificiality will be apparent. Yet it is worth while paying considerable attention to it because it will enable us to develop important principles that are applicable to real production of any degree of complexity. If these principles are not thoroughly understood the later chapters will be very difficult to follow.

The study of production with only one factor is unrealistic, but it is a useful exercise. Factors that are not *scarce* may be ignored.

By simple production we mean the production of a single homogenous and divisible consumption good by the use of a single homogeneous and divisible factor of production. This never really happens. In all production at least one kind of labor is employed, some raw material is used, some tools or more complex instruments of production are utilized, and the operations take place in a building or at least on a piece of land so that we have at least four different factors of production. We might also count such factors as light and air and warmth that are necessary, and we could include the operation of natural forces like gravity and magnetism. We do not usually count the last two groups of influences, essential as they are, because they are freely available so that there is no need to economize in the use of the force of gravity or the air we breathe or burn up. We can use this principle to make our simple production appear more possible. We may get over the difficulty of supposing only one factor in operation by

supposing instead that all except one of the factors is freely available so that there is no sense in economizing them. Then for economic purposes we may ignore them, just as we (usually) ignore the air we breathe, and speak only of the single factor the supply of which is so limited that there *is* reason in economizing it. To such a limitation we give the technical name *scarcity*. If the amount available of anything is sufficient to satisfy all possible uses and still leave some over, it is not *scarce*. It might be rare yet have so few uses that there is more than enough and it is not scarce, but *abundant*. If there is not enough to satisfy all possible uses, it is *scarce* however great may be the quantity available. One factor may, therefore, be understood to mean one *scarce* factor.

The fundamental economic problem is the problem of *choice*.

While there are no economic problems where there are no scarce factors, the existence of a scarce factor is not sufficient to make an economic problem. If the scarce factor is able to make only one product (which must itself be scarce or else the factor could not be scarce) the whole supply of it should obviously be directed toward making this product. There is still the problem of the *manner* in which it is to work, of the *technique* to be used, but that is a technical and not an economic problem.

The economic problem arises only when a scarce factor is capable of being used to make two or more different products. Then we have the economic problem of deciding *where* to put the factor to work—what products it should be set to produce and in what proportions the factor is to be shared between the different jobs.

That there must be alternative uses for a factor before there can be an economic problem is mentioned here only in order to bring out the essential nature of the economic problem as one of *choice*. Unless there is scarcity choice is *unnecessary*. All uses can be satisfied. Unless there are alternative uses for a factor choice is *impossible*. There is only one thing the factor can do.

The necessary condition that there shall be scarce factors with alternative uses before there can be any *choice* and therefore any economic problem does not make economic problems at all rare.

This will be seen when it is realized that the possibility of using a factor for producing the same physical good at different times gives an opportunity for choice so that they must be considered as different goods distinguishable by the *time* when they become available. Similarly possible difference of *place* of the product presents an economic problem. And even if a factor can only produce a single product at a single time and at a single place, we still have an economic problem—and an important one—if there remains the choice between using labor in producing this product or refraining from production in order that the laborer may enjoy leisure. Leisure must be considered as an alternative product, permitting choice and thus giving rise to the economic problem.

The optimum division of a factor among different uses implies that the value of its marginal product is not less than the value of any alternative marginal product.

The optimum division of a factor between two products will not have been reached if it is possible, by shifting some of the factor from one product to the other, to produce something that is worth more than what is given up. If a unit of factor is taken away from the production of meat and devoted to the production of fruit, and if the result is that 1 pound less of meat and 1 basket more of fruit is produced, we cannot say whether the change is an improvement in the situation or a deterioration until we know the marginal substitutability or M of fruit for meat (or their relative valuation) which, if there is an ideal distribution of goods, must be the same for all consumers of both goods. If the additional basket of fruit is valued more highly than a pound of meat given up so that its M for meat is greater than 1, the change is an improvement. If the pound of meat is valued more highly so that the M of fruit for meat is less than 1, it is a deterioration.

The relative valuation (or M) is given by the relative price. If the basket of fruit has a price of 60 cents and a pound of meat has a price of 30 cents then the shift of a factor from producing 30 cents' worth of meat to producing 60 cents' worth of fruit is an improvement.

It should be noted that only if there is an optimum allocation of goods can one be sure that the higher priced basket of fruit indicates a greater M and is more useful to consumers than the lower priced pound of meat.

The extra pound of meat that can be produced if another unit of factor is applied to the production of meat is called the *marginal product* of the factor. If a unit less of the factor is applied to the production of meat there results a decrease in the output of meat, also of about 1 pound. This too is called the marginal product of the factor. Similarly the addition or subtraction of a unit of the factor from the production of fruit will increase or decrease the output of fruit by 1 basketful, and this basketful of fruit is called the marginal product of the factor. In our example the *marginal product* of a unit of factor is 1 pound of meat or 1 basketful of fruit. The *value of the marginal product* is 30 cents in the case of meat because the marginal product (which is 1 pound of meat) has a value of 30 cents, and the *value of the marginal product* in the case of fruit is 60 cents because the marginal product (which is 1 basketful of fruit) has a value of 60 cents. We can therefore say that it is beneficial to have a unit of the factor of production shifted from producing meat to producing fruit because the value of the marginal product of fruit is greater than that of meat.

There is then a social benefit in moving units of scarce factors with alternative uses from points where the value of their marginal product is less to others where it is greater. Such a movement should continue as long as there persists any divergence between the values of the marginal product in different uses. As the movement goes on, say from meat to fruit, the supply of meat will decrease and the supply of fruit will increase. As a result, the relative valuation (or M) of meat will rise and that of fruit will fall, in accordance with the principle of diminishing M, and their prices will move in the same way, until the values of the marginal products in the two uses are equalized. (That is, unless the quantity of the factor applied to one of the products falls to zero before this point is reached when naturally none of this good will be produced.) The price of meat will rise above 30 cents, and the price of fruit will fall below 60 cents, but the shift of factors from

producing meat to producing fruit should continue as long as the value of the fruit marginal product is greater than the value of the meat marginal product. The values will thus keep moving until they meet at some intermediate point—say at 40 cents. At this point no further benefit is to be derived from shifting factors from meat to fruit. When the value of the marginal product of each factor has thus been made the same in every use to which it is put, the optimum division of the factor between the different products will have been reached. (Of course there is no reason why the value of the marginal product of any factor should be equal to that of any other factor—it is only the value of the marginal product of the *same* factor in *different* uses that has to be equalized.)

It will be observed that the argument concerning the optimum division of a factor of production between different products is almost exactly the same in form as that given in Chapter 3 in connection with the optimum division of income among different individuals. There we found the optimum would be reached when the marginal utilities of income were the same for all the individuals among whom the income is divided. Here we find that the optimum is reached when the values of the marginal product are the same in the different uses to which a factor can be directed. In the case of the division of income we could not apply this principle directly because of the impossibility of measuring on the same scale the marginal utilities of income to different individuals, so we had to resort to an argument in terms of probabilities. Here the values of the marginal products are objectively determinable from the physical marginal products which the managers of production can estimate and the prices of the products which can be seen on the market. Consequently we have a more satisfactory direct solution and do not have to be satisfied with probabilities.

To bring this about in any real society involves an infinitely complex problem.

In the form in which we have just left the solution of the problem of the optimum division of a factor between different products, it is no more capable of practical settlement than the ideal division

of income that would equalize the marginal utilities of income to all the different individuals in the economy (though the reason for the difficulty is different). It would not be possible for the managers of production in the collectivist economy to run the economy efficiently by continually comparing values of marginal products of different factors in different branches and subbranches of production in order to shift a factor from points where the value of its marginal product is less to others where it is greater. It would not matter for this whether the managers were scattered about the various manufacturing plants in the country or whether they were collected in one vast government building and called themselves the Ministry of Economic Planning. In the former case, they would at least be able to manage their plants. In the latter, they would only get tied up in the hopeless intricacy of the problem of shifting resources to and fro. This would involve a centralized miscellaneous knowledge of all the details in all the manufacturing units in the whole economy and a conscious reshuffling of factors of production throughout the whole economy every time there was any change in needs or tastes or in technical knowledge or in the supply of any of the factors. Some approach to this seems to have been attempted with disastrous result in Russia and was adequately criticized by Trotsky who wrote:

> If there existed the universal mind that projected itself into the scientific fancy of Laplace; a mind that would register simultaneously all the processes of nature and of society, that could measure the dynamics of their motion, that could forecast the results of their inter-reactions, such a mind, of course, could *a priori* draw up a faultless and an exhaustive economic plan, beginning with the number of hectares of wheat and down to the last button for a vest. In truth, the bureaucracy often conceives that just such a mind is at its disposal; that is why it so easily frees itself from the control of the market and of Soviet democracy.[1]

It can be solved with the help of the price mechanism and a simple *Rule* that must be followed by the managers of every production unit.

The only salvation of the economy lies in utilizing the price mechanism here just as in the problem of the optimum allocation

[1] L. D. Trotsky, *Soviet Economy in Danger*, pp. 29–30, Pioneer Publishers, New York, 1931.

of goods and the optimum division of income. With the help of the price mechanism this intricate problem can be solved in a collectivist economy in the following way:

First, there must be a free market in the sale of the consumption goods so that there can be established an optimum allocation of whatever goods are produced. (The question of the division of money income among the consumers is here taken for granted. It is assumed that the division of income is either satisfactory or inevitable.)

Second, there must be a free market in the sale of the factors of production to the managers of production so that the price of any factor, payable by the manager who acquires it for use in the factory, is the same as the price paid by any other manager.

In each market, whether for factors or for products, prices are raised whenever the demand for any product or factor is greater than the supply and lowered when the supply is greater than the demand until a set of prices is reached in which each demand is equal to the corresponding supply.

Such a complete equilibrium will seldom, if ever, be reached, and certainly will not long remain unchanged. This is because as quickly as the price adjustments are made there are changes in tastes and in needs, in techniques of production and in the supply of factors, all of which will affect supply and demand for the various factors and products so that prices will have to be changed again to bring supply and demand into equality. However, at each moment the prices which momentarily bring supply into equality with demand will be playing their part in bringing about the optimum division of each factor among the different products.

The demand for each consumption good produced will be determined by the division of income, the needs and tastes of the consumers, and the prices at which alternative products are available to them. The supply of the consumption goods (and therefore also the price at which the supply of each is equal to its demand) will depend upon the quantities produced. This has yet to be determined.

The supply of the various factors of production is, in general, given by the amount of them available in the economy. The

demand for them will again depend on the quantities of the various products that are to be produced by means of these factors, and this has yet to be determined.

This brings us to the third and final step in solving the problem. The optimum division of each factor among the different products (and the amounts of the various products to be produced) can now be fixed by issuing a simple *Rule* which every manager of production must obey. The *Rule* is:

If the value of the marginal (physical) product of any factor is greater than the price of the factor, increase output. If it is less, decrease output. If it is equal to the price of the factor continue producing at the same rate. (For then the right output has been reached.)

If these three steps are taken nobody need be put in charge of attending to the details of the whole economy. The simple *Rule* carried out by each manager in his own plant, with no knowledge whatever of values of marginal products anywhere except in his own plant, will bring about the optimum division of each factor between the production of different goods.[1]

This does not mean that there is nothing for the Ministry of Economic Planning to do. If means only that the Ministry should not try to concern itself with details that can much better be attended to by the manager on the spot if the appropriate rules are provided for him in conjunction with the pricing mechanism. The business of the Ministry of Economic Planning is to establish the appropriate rules and to see that they are followed and that the price mechanism is kept in operation. It is by promulgating and maintaining the general rules consciously directed toward the optimum operation of the whole economy that the *controlled*

[1] *Cf.* Trotsky, *op. cit.*: "The innumerable living participants of the economy, State as well as private, collective as well as individual, must give notice of their needs and of their relative strength not only through the statistical determination of plan commissions but by the direct pressure of supply and demand. The plan is checked and, to a considerable measure, realized through the market. The regulation of the market itself must depend upon the tendencies that are brought out through its medium. The blueprints produced by the offices must demonstrate their economic expediency through commercial calculation." (p. 30.) "Economic accounting is unthinkable without market relations." (p. 33.)

economy is distinguishable from the *uncontrolled economy* which does not so establish the general rules. The uncontrolled economy will indeed generally show a much more luxurious and complex growth of *particular* regulations. These naturally arise from attemps to correct particular failures resulting from the absence of any general plan for the economy as a whole.

The Rule equalizes the value of the marginal product of each factor in each of its uses.

With the prices of the factors raised and lowered to make the demand for each one of them equal to its supply and with every manager of production keeping the Rule, the value of the marginal product will automatically come to be the same for each factor in all of its uses. This is because each manager expands or contracts production until the value of the factor's marginal product is equal to its price. Since the price is the same for all managers who purchase the factor, the values of the marginal products will all automatically be made equal to each other.

The equalization comes about in two ways. If a manufacturer finds that the value of his marginal product is greater than the price of the factor, the Rule enjoins him to expand production. In the first place this increases the supply of his product and tends to reduce the price at which it is sold, so that the *value* of the marginal product will fall as long as the marginal physical product remains the same. In the second place it (simultaneously) increases his demand for the factor so that *its* price tends to increase. If one additional unit of the factor which costs $1.00 permits 2 more baskets of fruit to be produced and these sell at 60 cents each, then the marginal product is 2 baskets of fruit, and the value of the marginal product is $1.20. The fruit grower is thus instructed by our Rule to purchase more of the factor and produce more fruit. As this is done by him, and by any other fruit growers who find themselves in a similar position, the supply of fruit increases while the supply of other goods previously produced by these factors decreases so that fruit tends to become cheaper. When the price of fruit has fallen to 55 cents a basket, the value of the marginal

product will have fallen from $1.20 to, say, $1.10 if the marginal physical product itself remained the same at 2 baskets of fruit.

At the same time the increased demand for the factor by the fruit growers will have raised its price. If the price rises from $1.00 to $1.10 it will now be equal to the value of the factor's marginal product in fruit, and the fruit grower will not change his scale of production again until something else happens to alter the price of the factor or the price of the product or the marginal physical product of the factor.

The private marginal opportunity cost and the social marginal opportunity cost are equalized by free consumer purchases on the market.

We have seen in Chapter 4, p. 43, that each consumer adjusts the quantities of each good that he buys in such a way that his *M* corresponds to his *marginal opportunity cost.* By the latter phrase was meant the alternative goods that he has to sacrifice in order to obtain another unit of the good in question. We saw there too that if the quantity he purchased had no influence on the price he paid (and it cannot be allowed to have such an effect if there is to be an optimum allocation of goods), the marginal opportunity cost was given by the relative price of the products. If meat is 30 cents a pound and fruit is 60 cents a basket, the marginal opportunity cost of another basket of fruit would be the 2 pounds of meat. This might be called the *private* marginal opportunity cost because it is what the individual privately has to give up in order to get the additional basket of fruit. The individual reaches the best position available to him when he makes the *private* marginal opportunity cost of each good equal to its marginal substitutability.

What we have been discussing in this chapter can be described as the *social* marginal opportunity cost. This is not what any particular *individual* has to sacrifice to get another unit of anything, but what *society* has to sacrifice when another unit of any particular good is purchased. What society has to sacrifice is the alternative product that might have been produced by the factor that was devoted to the particular good.

We may repeat that these calculations are applied to the *margin*. We do not consider what else could have been produced if the whole production of a particular good were given up, or even a very large block of it, but how much more of another product could have been produced if one unit of factor were shifted from this to the alternative product.

In this way each individual is induced, while seeking his own interest, to do that which is in the social interest.

If every manager adjusts his output so as to make the value of the marginal product of each factor equal to its price and the price of the factor is the same to every manager of production, so that the different marginal products of each factor have equal values (as we have seen, p. 65), the *social* marginal opportunity cost of any product will be measured by its price. A dollar's worth of any product will be just that amount by which output would be reduced if a dollar's worth of the factor were withdrawn from its production. This $1's worth of factor, wherever it is applied, can increase production by just $1's worth of product. Consequently, the sacrifice of $1's worth of any product will permit the production of exactly $1's worth of any alternative product. If at the same time no individual buyer has any influence on price his *private* marginal opportunity cost is also measured by the price. What our Rule does then is to equate the *social* and the *private* marginal opportunity costs so that every individual, in trying to minimize his own sacrifice of alternatives when he spends his money income to his own best advantage, is led automatically and even unconsciously to minimize the *social* sacrifice in producing what gives him most satisfaction. *This is the essential social utility of the price mechanism.* If it is appropriately used it induces each member of society, while seeking his own benefit, to do that which is in the general social interest. Fundamentally this is the great discovery of Adam Smith and the Physiocrats, although they were too optimistic in assuming that it was hardly necessary to bother about seeing that the right rules are promulgated to be sure that the mechanism works the way it should.

The Rule also works if there are many stages in production.

So far we have spoken only of factors of production that are applied to the manufacture of consumer's goods. The same analysis can be applied when production has many steps. Factors can be used to produce other factors of production which in turn can be used to produce still other factors of production, with any number of stages before we get to the final consumer's goods. If the same Rule is applied throughout the economy all that we have said will still hold. A dollar's worth of consumption good is produced, at the margin, by $1's worth of a factor *A*. This in turn is produced, at the margin by $1's worth of factor *B*, and so on as far back as one likes up to $1's worth of factor *M*. This $1's worth of factor *M* could have produced $1's worth of factor *N* and it in turn $1's worth of factor *O*, and so on until we come to factor *Z* which could have produced $1's worth of an alternative consumption good. The price of consumption goods still represents the social marginal opportunity cost, so as long as our Rule is applied throughout the economy (and the accompanying free markets are maintained) we will have the optimum division of each factor, directly and indirectly, in the production of the different final consumer's goods.

Where there is only one (scarce) factor and no indivisibilities there must be constant returns to the scale of production.

Throughout this chapter we have avoided the complications that arise from changes in the marginal product of factors and in their average product that results from changes in output. This can be excused in the present chapter where we have supposed that only one scarce factor of production is used in the production of each product (which may be either a final consumption good or another factor of production in turn). It is permissible here to assume that factor and product increase in the same proportion so that both the marginal (physical) product and the average (physical) product (the total physical product divided by the number of units of the factor) are constant. If it is required to double the product, twice the scarce factor should do it, because it is

always possible to duplicate the previous methods of production as long as all the other unmentioned factors of production are available without cost in any required quantities. This makes it impossible for average or marginal product to rise as output is increased. Nor is it possible for them to fall, because if doubling the quantity of the (scarce) factor can more than double the product it must be that a different and better mode of production is used for the larger output than for the smaller output. If there is only one factor and it is divisible (as we have been assuming), there is no reason why the better methods should not have been used on the smaller scale and so we would have constant average and marginal product or "constant returns to scale."

In the actual world increasing and diminishing returns are of course very common, but they are always due to changes in the proportion of different factors used in combination or to indivisibilities in the factors or in the product or in the techniques of production. (You cannot have a very small assembly plant for producing only one automobile as cheaply as a larger one will produce many automobiles.) We shall consider these as we come to them.

Productive speculation, as distinct from aggressive Speculation (with a capital S), is a socially most useful activity.

A word may be added here about speculation. In Chapter 4 we spoke of *counterspeculation* as a device for preventing individual buyers and sellers from influencing prices by varying the amount they buy or sell. This device counters such Speculators as are able, because they are very rich or because they can organize many people into combinations, to affect the price and thus to frustrate any attempts to bring about an optimum allocation of goods. It is these powerful bodies which are usually referred to as Speculators outside of economic theory and the stock exchanges and whose activities it is desirable to curb by devices such as *counterspeculation*. We may call this *aggressive* or *monopolistic* speculation. There is another kind of speculation, which we may call *simple* or *productive* speculation. A man who does not consider himself to have any influence on the market price but who believes

that the price is going to rise or is going to fall quite independently of his own actions, and who buys or sells in an attempt to make a profit, is a *simple* or *productive* speculator. If he guesses right he makes a profit, if wrong he makes a loss. We may call such an individual a speculator with a small "s." He is mentioned here because he fits more easily than any other producer in the actual world into the special category of producers considered in this chapter. He takes a single product available at one time and turns it into a single product available at another time. We have seen that this is to be considered as the production of one good out of another or as the use of a single factor of production to produce a single product. The same thing applies to the man who transports a good from one *place* to another. He is using the good at the first place as a factor of production for producing a good at another place.

These are perfectly legitimate production activities and must take their place in a collectivist economy in the same way as any other production and subject to the same rules. If $1's worth of aluminum today can be turned into $2's worth tomorrow, our Rule says that any manager engaged in production of aluminum tomorrow out of aluminum today should extend his activities until the price of the factor (aluminum today) is equal to the value of its marginal product (of aluminum tomorrow). In more familiar language this merely means that it is socially desirable that the metal be taken from relatively unimportant uses today to render it available for more urgent uses tomorrow. Similarly shifting goods from points where they are relatively plentiful to others where they are relatively scarce is a productive action of the greatest social use even though it was and perhaps still is a capital offense in Russia. Production includes speculation (with a small "s").

In the illustrations given in this chapter, where a concrete example of a factor of production had to be given, labor was never chosen as the example if it could be avoided. This practice will be adhered to throughout the book. Classical economists and socialists have regularly got into trouble by dealing with labor as a factor of production or as *the* sole factor of production. The diffi-

culties arise from the almost unavoidable confusion of labor as an instrument of production that is capable of producing one product or another and labor as the human beings in whose welfare the writer is interested. By keeping away from labor as a factor, except where it is unavoidable,[1] we shall attempt to avoid these confusions. This will be all the easier because we have started out with an explicit statement of our welfare interests rather than hiding them or leaving them for footnotes and appendixes or disguising them as objective "scientific" laws.

[1] We have already come across one such case (p. 59) in connection with the peculiarity of labor that it may be better not to use it at all than to use it for the production of a scarce good, if the leisure provided by not working is valued more highly than the product.

CHAPTER 6. SIMPLE PRODUCTION II
(*UNDER PERFECT COMPETITION*)
THE WELFARE EQUATIONS

Under certain conditions free enterprise leads to the optimum use of resources without any Rule expressly designed to bring it about.

Under certain circumstances it is possible for the optimum division of a factor between its different products to be reached in a capitalist economy with production carried on by business firms that are not subjected to any rule expressly devised to bring about this result. The owner (or manager) of each firm seeks to maximize his profit and this leads him to expand production whenever the extra cost involved is less than the extra revenue that results from having a larger output to sell. In the same way it would lead him to contract production if the saving from using a unit less of the (scarce) factor is greater than the loss from having a smaller amount of product to sell.

If there is *perfect competition in buying* the price of the factor is equal to the marginal cost to the buyer.

When the extra cost is equal to the extra revenue there is no profit in either expanding or contracting production.

The extra cost to the firm of employing another unit of a factor is equal to the price of the additional unit of the factor *plus* the increase in pay which the firm is forced to grant, as a result of its increased demand for the factor, to the other units of the factor in its employ. Thus if the manager of the firm is employing 100 units of a factor at $1 each, paying out $100, and the employment of one more unit would raise the price of all units to $1.01, the extra cost of employing another unit would be greater than the price. It would exceed the price ($1.01) by the 100 extra pennies that each of the other 100 units would be able to get because of

his increased demand—a total of $2.01. This can be calculated another way. The manager believes that if he were to employ another unit he would have to pay for all the 101 units at $1.01 each, a total of $102.01, which is $2.01 more than when he employs 100 units. The extra cost of employing another unit is $2.01 while its price is $1.01.

If the manager does not believe that employing another unit of the factor will increase the price (usually because he is only one of many employers of the same factor), the second item disappears in the calculation of the extra cost of employing another unit. The 101 units simply cost $101 instead of $100 for 100 units, and the extra cost of employing another unit, $1, is equal to the price of the factor.

Such an absence of any influence over price is called *perfect competition in buying.*

If there is *perfect competition in selling* the price received for the product is equal to the marginal revenue.

The extra revenue that the firm obtains from employing another unit of a factor is equal to the value of the extra or marginal product *minus* the loss in revenue that the manager believes will come from the reduction in price because of the increased supply. Thus, suppose a firm was producing and selling 200 units of product at 50 cents each, so that the total revenue was $100, and that employing another unit of factor would increase output from 200 to 202 units and this would lower the price to 49½ cents. The extra revenue would be equal to the value of the extra 2 units (99 cents) *minus* 200 times ½ cents (or $1). In this case the extra revenue is actually negative, *minus* 1 cent. This too can be calculated the other way. If 202 units were produced and sold at 49½ cents each, the total revenue would be $99.99—1 cent less than what is obtained from the sale of 200 units at 50 cents each. If the price fell only to 49¾ cents, the extra revenue would be 49½ cents. This figure is obtained by subtracting 200 times ¼ cent (or 50 cents) from 99½ cents (twice 49¾). Revenue is $100.49½ instead of $100. As long as there is some fall in price the extra

or marginal revenue is less than the value of the marginal product.

If the manager of the firm does not believe that the increase in his output will have any effect on the price, then again the second (and negative) item disappears in the calculation of the extra revenue. The 202 units will net $101 instead of $100 for 200 units. The extra revenue of $1 equals the value of the marginal product. Such an absence of influence on the price of the product is called *perfect competition in selling.*

If there is perfect competition throughout the economy individual enterprisers seeking to maximize their profits behave just as if they were following the Rule.

We have seen (p. 72) that the principle followed by the firm seeking to maximize its profit is as follows: "If the extra revenue from employing another unit of any factor is greater than the increase in cost, increase output (for that will increase the profit). If the fall in revenue from employing a unit less of a factor is less than the fall in cost from so doing, decrease output (for that will increase the profit). If the rise or fall in revenue from changing the quantity of a factor used is equal to the rise or fall in cost, continue producing at the same rate (since the change would not increase the profit)." We have just seen that if there is perfect competition in buying the change in cost from using one unit more or one unit less of the factor is exactly equal to the price of the factor, so that we may say "the price of the factor" instead of "the rise or fall in cost." We have also seen that if there is perfect competition in selling the change in revenue from varying output by employing a unit more or a unit less of the factor will change the revenues by exactly the value of the marginal product, so that we may say "the value of the marginal product" instead of "the rise or fall in revenue." If we substitute these phrases into the principles for maximizing profit it reads as follows:

"If the value of the marginal product of any factor is greater than the price of the factor, increase output. If the value of the marginal product is less than the price of the factor, decrease output. If the value of the marginal product is equal to the price of

the factor, continue producing at the same rate." This is identical with the Rule given the manager of production in the controlled economy (see p. 64) and would lead in exactly the same way to the optimum allocation of the factors among the various products.

This is illustrated by the welfare equations which must be satisfied if the optimum is to be reached.

The discussion as developed so far may be expressed by the following chain of five equations between six items, which describe the situation in each productive unit when the optimum division of each factor among its different products has been accomplished by private profit-seeking firms when perfect competition in buying and in selling exists in every productive unit throughout the whole of an uncontrolled economy. These may be called *the welfare equations*. The six items are:

1. *The Marginal Social Benefit* (*msb*). This is the benefit to society (*i. e.* the net benefit to all the members of society affected) from the particular increment of output of product considered.
2. *The Value of the Marginal Product* (*vmp*). This is the physical increment of output of product being considered, multiplied by the price paid for it by the consumer. If the increment is exactly one unit of product, the value of the marginal product will equal the price of the product (*p*).
3. *The Marginal Private Revenue* (*mpr*). This is the increase in revenue (positive or negative) received by the producer as a result of producing and selling the increment in output.
4. *The Marginal Private Cost* (*mpc*). This is the increase in cost incurred by the producer as a result of increasing the quantity of factor he purchases in order to be able to produce the increment of output.
5. *The Value of the Marginal Factor* (*vmf*). This is the physical increment of the factor of production (that is needed to make the increment of product) multiplied by the price per unit paid for it and received by the owner of the factor. If the increment is exactly one unit of factor, the value of the marginal factor will be equal to the price of the factor (*pf*).
6. *The Marginal Social Cost* (*msc*). This is the sacrifice to society from having the marginal factor used up here so that it is not available for use elsewhere. It is the "social marginal opportunity cost" referred to on p. 66. It is the alternative marginal social benefit that the marginal factor could have produced if it had been used elsewhere.

The five equations are:

(1) Marginal Social Benefit = Value of Marginal Product

This equation is satisfied if there is an optimum allocation of consumption goods and if the purchaser of the good is the only individual in society who is affected by the purchaser's use of the product. In that case the amount of money paid for it measures the usefulness of the product to the purchaser and so to the society of which he is the only individual affected.

(2) Value of Marginal Product = Marginal Private Revenue

This equation is satisfied if there is perfect competition in selling the product. The producer is then not able to influence the price of the product by varying the output, so the extra revenue he gets from an increment in output is simply the physical increment in output multiplied by its price.

(3) Marginal Private Revenue = Marginal Private Cost

This equation is satisfied if the producer maximizes his profit. This involves expanding output as long as the marginal private revenue is greater than the marginal private cost and contracting output as long as it is less; the equilibrium position where profits are maximized thus being reached only when these two values are equated.

(4) Marginal Private Cost = Value of Marginal Factor

This equation is satisfied if there is perfect competition in buying factors of production. The producer is then not able to influence the price of the factor by varying the quantity of it that he buys, so the extra cost of buying the increment of factor is simply the physical quantity of the factor multiplied by its price.

(5) Value of Marginal Factor = Marginal Social Cost

This equation is satisfied if the first four equations hold for all the other production units in the economy using the factor so that the value of the marginal factor equals the alternative marginal social benefit from using the factor in these other uses.

These five equations may conveniently be written out in abbreviated form thus:

$$msb = vmp = mpr = mpc = vmf = msc$$

The optimum allocation of factors between different uses is attained if the first of the six items is equated to the last and $msb = msc$ with only one equation instead of five. This would

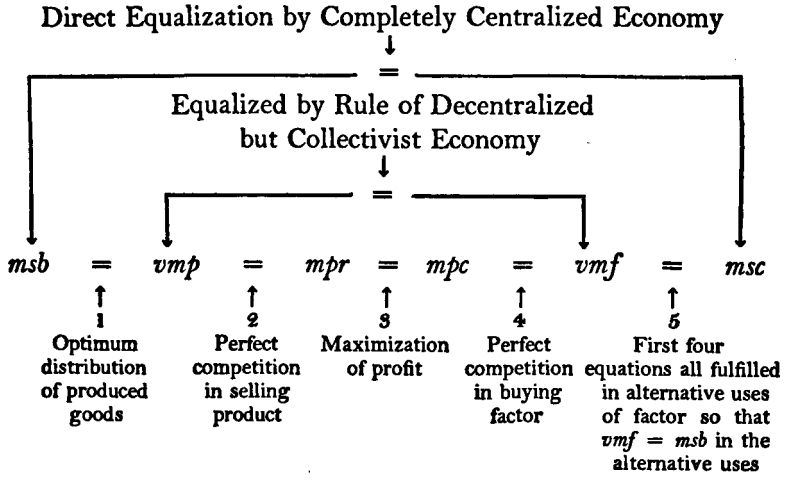
represent the task that confronts a Ministry of Economic Planning which tries to utilize the factors of production in the best possible manner by shifting resources directly from points where they are less useful to others where they are more useful until they are equally useful in all uses.

The method of the collectivist economy described in Chapter 4 can be represented by these three equations:

$$msb = vmp = vmf = msc$$

The first of these, $msb = vmp$ (equation 1, p. 76), is satisfied if there is an optimum allocation of consumption goods such as can be brought about by a free market as described in Chapter 1. The second, $vmp = vmf$, is given by the Rule that the managers of all production units must obey. The third equation, $vmf = msc$ (equation 5), is satisfied if all the other managers of production (who have to pay the same price for the factor) make this price equal to the value of alternative marginal product, since *msc* is nothing but the alternative *msb*.

The Welfare Equations



If these five conditions are satisfied, an optimum allocation of factors among the different products is reached in a capitalist economy.

CHAPTER 7. SIMPLE PRODUCTION III
(*IN THE CAPITALIST AND IN THE CONTROLLED ECONOMY*)

We may now turn to the consideration of how likely it is that an optimum division of each factor between its different products would actually be reached. In other words, how likely is it that the five equations will actually be fulfilled.

The conditions under which the welfare equations would be satisfied in a capitalist economy are very stringent and unlikely to be fulfilled.

Equation 1 ($msb = vmp$) is fairly easily satisfied. If there is a free market in consumption goods and consequently no discrimination between different consumers, the price will measure the marginal substitutability for each consumer. Accepting the given distribution of income as good or as inevitable, the value of the marginal product of a unit of factor will measure the *msb* of a unit of factor directed at the margin to each product.

Equation 3 ($mpr = mpc$) is satisfied if the firms adjust their output to maximize their profit.

The difficulties arise with equations 2 and 4 ($vmp = mpr$ and $mpc = vmf$). There is not likely to be perfect competition in buying and selling throughout the economy, and if these two equations are not satisfied everywhere, equation 5 ($vmf = msc$), which depends on all the others being fulfilled in alternative uses, will not be satisfied in any of the other cases so that the optimum division of factors will not be achieved.

The infrequency of perfect competition is shown by the importance of salesmanship.

The frequency with which competition in selling is imperfect in the actual world is reflected in our general attitude toward

selling and salesmanship as an important art. Where there is perfect competition in selling, the seller can sell as much as he wishes and could without effort sell more at the same price but prefers not to do so. If any effort is expended on salesmanship or in advertising (except perhaps to provide information as to what is available), it is because it is not possible otherwise to sell more at the same price and equation 2 is not satisfied. The prevalence of the notion that the seller has to please the buyer rather than the buyer the seller—that the customer is always right—is an indication of how natural it is for competition in selling to be imperfect. If the seller could sell as much as he wished at the current price, salesmanship would be unnecessary.

Perfect competition in buying or hiring factors of production is more common than perfect competition in selling, but it is by no means universal. Where there are large purchasers—great corporations or purchasing associations—even branches of government that unthinkingly adopt the businessman's principles of maximizing profit by minimizing the total cost of what is bought —it is natural that demand be restricted because a larger purchase would raise the price against the buyer. In other words, *mpc* is greater than *vmf*, and equation 4 is not fulfilled.

The optimum application of resources in any particular use is only *relative* to what is happening elsewhere in the economic system.

Finally, even where the first four equations happen to be satisfied, equation 5 will not be satisfied unless the first four equations hold in all the alternative uses of the factor. If this further condition is not met, *vmf* will not equal *msb* in the alternative uses and therefore will not equal *msc*. This illustrates the *relative* nature of the optimum division of a factor among its different products. If too much of a factor is employed in some uses it is inevitable that too little of it is employed elsewhere. The one defect implies the other and cannot be put right without shifting some factor from the place where too much of it is being used. This is indicated by the fifth equation.

Perfect competition must be *universal* if the optimum is to be reached.

Furthermore, an optimum division of factors in an uncontrolled economy is not only unlikely but *unstable*. If there are many firms engaged in any particular branch of production, each firm may be too small to affect the price at which it buys the factor and sells the product so that all four direct equations are satisfied. If this is true in the production of all the alternative products to which the factor can be applied, equation 5 is also satisfied and we have the optimum division of the factor. But we are still considering the use of only one (scarce) factor, and we have seen that this is accompanied by constant returns. Each firm can increase its output without changing the price of the factor or the price of the product or the *marginal physical product of a unit of factor*. This means that if the firm is making any profit it can increase this profit indefinitely simply by increasing the scale of its activities. This would immediately make at least some of the firms large enough to be able to influence the price at which they sell the product or buy the factor, so even if there should exist the condition for bringing about an optimum division of a factor among its different products this situation will tend to destroy itself.

It can be shown, with the help of the *average-marginal relationship*, that there may be no direct gain in the firms' expanding.

Things are not quite as bad as the last paragraph might suggest. This is because each firm in the original optimum position would not be making any profit (or loss), so there would be no incentive to expand (or contract) production.

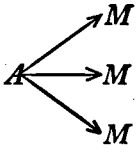
We are here assuming that there is only one scarce factor so that we have constant returns and a 1 per cent increase in the factor brings about a 1 per cent increase in the product. If 100 units of factor are used to produce 200 units of product, the addition of another unit of factor, raising the total amount of factor from 100 to 101 units, has the effect of increasing the product in the same proportion from 200 to 202. The marginal product is 2 units of

product (because an additional unit of factor brings about an increase of 2 units in the total product). The average product is also 2 ($\frac{200}{100}$) and remains at 2 after production is expanded ($\frac{202}{101}$). The marginal product is equal to the average product. This must be so whenever the average product is unaffected by the scale of output (and we have constant returns) because if the marginal product were greater than the average product an increase in output would raise the average product. If the marginal product were more than 2, the total product of 101 units would be more than 202 and the average product would be more than $\frac{202}{101}$ (that is, it would become greater than 2). Similarly, the marginal product cannot be less than the average product because then the total product of 101 units would be less than 202 and the average product would be less than $\frac{202}{101}$ (or below 2). If there are constant returns to the factor of production, and there must be where there is only one factor of production, the marginal product can be neither greater nor less than the average product and so must be equal to it.

Irrespective of the figures in any particular example we can see that if *mp* (the marginal product) is greater than *ap* (the average product), *ap* must be rising—or we can say that the greater *mp* raises *ap* toward its own level. Conversely, if *ap* is rising, *mp* must be greater than *ap*—otherwise *ap* would not be raised. On the other hand, if *mp* is less than *ap*, *ap* must be falling—the lower *mp* pulls it down toward its own level. Conversely, if *ap* is falling, *mp* must be less than *ap*—otherwise *ap* would be pulled down. If *mp* is equal to *ap*, *mp* neither raises nor pulls down *ap*, so *ap* is constant. Conversely, if *ap* is constant, as it must be in our case of constant returns, it must be equal to *mp*. This, perhaps obvious, arithmetical relationship holds in the same way between marginal and average cost and between the marginal and average measurements of anything. It was applied (on pp. 72 and 73) to marginal cost and average cost of a factor, where the average cost was the price of the factor, and to average revenue and marginal revenue, where the average revenue was the price of the product.

The average-marginal relationship may be symbolized by the following mnemonic which guards one against confusion by the

asymmetry of the relationship. The relationship is asymmetrical because it is concerned with the *position* of the marginal item (whether it is greater or less than the average item) and with the *direction of change* of the average item (whether it is increasing or decreasing or constant as the scale of operations is increased) and not the other way round.



If M (the marginal item) is higher than A (the average item), A is rising (indicated by the rising arrow pointing to the higher M). If M is lower than A, A is falling (indicated by the falling arrow pointing toward the lower M). If M is equal to A, A is constant (indicated by the horizontal arrow pointing toward the M that is on a level with A). M draws A toward itself.

Coming back to our problem, we recall that in the optimum position equation 3 is satisfied because profits are maximized so that *mpr* is equal to *mpc*. Since there is perfect competition (otherwise the position would not be an optimum one), average private revenue (*apr*) is equal to *mpr*. Since there are constant returns to scale (because only one scarce factor is being used), *mpc* is equal to *apc* (average private cost). Thus we have the three equations

$$apr = mpr = mpc = apc$$

With *apr* equal to *apc*, total private revenue is equal to total private cost too and there is no net profit or loss.

It is still true, as was noted above, that an increase in output will result in a proportionate increase in profit. A 10 per cent increase in output will be accompanied by a 10 per cent increase in profits, a 20 per cent increase in output by a 20 per cent increase in profits, and so on. But since profits are equal to zero to begin with, there is no automatic tendency for the firm to expand to increase profits and so break up the optimum situation.

But there is always an indirect gain from their becoming large enough to establish a monopoly.

However, it cannot be said that all is well. Even though there is no *direct profit* in expansion of output (or in contraction) there is nothing to keep the firm at the optimum position for there is no *loss* in expansion or in contraction. And there is an *indirect* profit in expansion. For if a firm can become large enough it *will* be able to influence price and so turn its zero profit into a positive one through being able then to get a higher price by restricting total output. It may also be able to pay a lower price for the factor as a result of restricting its demand for it. In the same way all the firms could gain by combining into a monopoly which would make a positive profit for all its members by restricting production and thus lowering factor prices and raising product prices. In this way there is a real tendency, though not so direct as is provided by the prospect of an immediate increase in profits from expansion itself, for monopoly or combination which would wreck equations 2 and 4 and destroy the optimum situation. Under constant costs there is no force to prevent this, and it is only a matter of waiting for one firm to expand and oust the others or for all the firms to come to some sort of agreement on sharing the spoils of monopolistic combination before the optimum situation is shattered even if it should arise in the first place.

The conclusion that the optimum division of a factor among its different products is not likely to be reached in an uncontrolled economy of the kind examined is unfortunate not only for the uncontrolled economy but also for the controlled economy. This is because the controlled economy is thereby deprived of a most powerful instrument.

Perfect competition has advantages over the attainment of the optimum by the Rule: the incentive to the managers is of the ideal intensity,

Where private enterprise would lead to an optimum division of a factor among its products *and is not in conflict with any of the other aims of society* is it preferable to its alternative of state enter-

prise with the managers subject to the Rule of Chapter 5. Private enterprise is preferable because there is a closer identity of the interest of the manager with the social interest. It is true that the Rule, if carefully applied, will lead more *directly* to the optimum division of factors among their products, but the incentive to apply it accurately is not so clear or so great. Every dollar that the manager of a free enterprise can save society is a dollar saved for himself, and only if he does the very best possible is he able to make a normal income for himself. Some incentives in the form of rewards (and punishment too perhaps) will have to be developed for the manager who is subjected to the Rule, and there will be a delicate problem of making them neither too weak nor too strong. In private enterprise under conditions of perfect competition all these problems are solved. Efficiency is guaranteed by the competition that eliminates all but those who use the most efficient methods possible. The incentive is of exactly the right intensity because the entrepreneur will apply his efforts up to the point where a dollar's worth of effort can be expected to bring a dollar's worth of results. He will not stop short of this ideal point, as he would if the incentive were too weak, or wastefully go beyond this point, as would happen if too strong an incentive were applied. It may seem strange to some that incentives to efficiency could be too strong, but this can be very serious. It can lead to a tyrannous disregard for the welfare of the workers and an inhuman red-tapism that would ultimately mean less and not more efficiency. Finally the principle of trying to maximize profit has the great advantage that it does not require any supervision or administration to make the enterpriser apply it.

and alternatives to government employment are a safeguard of the freedom of the individual.

More important than any of these administrative considerations is the significance of private enterprise as one of the guarantees of the freedom of the individual. There is a sound basis for this argument even if it is often distorted by fanatical capitalists who identify the freedom of the individual with the license of the capitalist

millionaire or even with the economic powers of giant corporations. Though the wealth of rich men and the power of great corporations can as well be used to limit the freedom of the individuals who directly or indirectly work for them or who depend on them for the products that they control, it still is true that the liberty of the individual obtained its first start in modern times with the freeing of private enterprise and that the possibility for the individual of finding a means of livelihood outside of employment by the state can be a check on undue subservience to the employers who represent the state. Of course this is only one of many forces that must be developed and maintained if democracy is to be preserved and by itself it can not guarantee democracy, but anything that may contribute to the safeguarding of democracy is of great value.

The controlled economy may consider that even some sacrifice of efficiency in the allocation of resources is worth while as a contribution to the safeguard of democracy, though the kind of government that would take this into account could put up adequate safeguards even if it were 100 per cent collectivist.

Where perfect competition is unable to survive—as in the case of constant return that we have examined—it might be maintained artificially by counterspeculation. Government guarantees of competitive price would nullify any attempts at monopoly by combination or by expansion of firms to a sufficiently large size to exert monopoly power over the prices paid and received. The same result could be achieved if freedom of entry were maintained in all industries so that as soon as any monopoly were formed and it raised the price of the product or reduced the price of the factor, new enterprise would rush in to take advantage of the possibility of extra profit and thus would restore prices and output to the optimum level. The difficulty here is that a large corporation or combination can often intimidate small newcomers and the only effective competition would be by the government. This brings us back to the collectivist solution, but with the government and private enterprise competing. The government agent would follow the Rule (making $vmp = vmf$) and the private firms would maximize profit.

This provides *an objective guide* in any instance whether the production unit should be operated privately or collectively.

In such a situation the better of the two forms of production could oust the other. If the private firms were more efficient than the government agencies, they would expand and raise *vmf* to the government above its *vmp*, and this would be the signal, in accordance with the Rule, for the government agencies to contract production and perhaps leave the whole field to private enterprise. If the government agencies were more efficient, the private firms would be forced to contract production or to quit the field. Paying the same price for the factor as the government and selling the product at the same price as the government (which makes *vmp* = *vmf*), any firm that is less efficient will find its *mpc* (which is equal to *vmf*) greater than *mpr* (which is equal to *vmp*), so that it would gain (or reduce its loss) by reducing output. With constant returns (which must be the case if there is only one scarce and divisible factor), *vmf* greater than *vmp* will mean that total cost (or total outlay for the factor) is greater than total income (the total value of the total product), so the firm is making a loss and would do better to close up shop.

If for any firm there is not perfect competition in buying, *mpc* will be greater than *vmf;* if there is not perfect competition in selling, *mpr* will be less than *vmp*. Consequently, there will be an even greater incentive for the firm to contract until it becomes small enough to lose its monopolistic power over price, and then, as we have just seen, unless it is more efficient than the government it will pay it to go out of business.

Complete freedom for *public enterprise* and *private enterprise* on equal and fair terms may be called *free enterprise.*

It may be that some private firms are more efficient than some government plants and that some government plants are more efficient than some private firms. If that is so, the more efficient units of either kind will remain, operating side by side permanently, and this will result in the maintenance of perfect competi-

tion and the optimum division of the factor between its different products in the controlled economy. Industry will not be the exclusive domain of either private or public enterprise, but both will operate side by side. Such freedom for both public and private enterprise deserves to be called *free enterprise*. This term is, unfortunately, often used to describe a system in which public enterprise is prohibited, but we shall call such a condition by its rightful name of *private enterprise*, keeping the term *free enterprise* to describe a state of freedom for both kinds of enterprise on fair terms which, in each particular case, permit that form to prevail which serves the public best.

It is perhaps worth repeating that many of our conclusions are closely tied to our unrealistic assumption that only one scarce factor is used in each plant to produce only one product and will no longer hold when we drop this assumption. The analysis in this chapter is valuable not as leading to any concrete conclusion that can directly be applied to the actual economy but rather as an exercise in the examination of principles to be applied to the more complex cases yet to be considered and which will ultimately lead to practical conclusions.

CHAPTER 8. COMPETITIVE SPECULATION

In Chapter 5 we touched on the question of speculation and saw that perfectly competitive speculation was in no way an anti-social action but as useful as any other form of production. By perfectly competitive speculation is meant the buying of goods for resale later or elsewhere at a higher price when the speculator does not try to influence price by varying the scale on which he carries on his speculative activities.

The social utility of competitive speculation is more certain than that of simple production.

All perfectly competitive speculation is in the social interest whether the optimum division of each factor between its different products is reached or not. It always improves on the situation, bringing it nearer to the optimum. It is strange that this should be more certainly so in the case of speculation than in the case of production in the ordinary sense which usually receives much greater social approbation. Simple production of a particular good may be perfectly competitive and yet not contribute at the margin to bringing out the best use of the factor. It may be harmful socially because there is an aberration from the optimum in the production of the alternative products and equation 5 does not hold. But perfectly competitive speculation cannot have its good works nullified by what goes on anywhere else in the economy, because it completes the whole cycle by itself in taking goods from points where they are cheaper to others where they are dearer and so from points where the value of the alternative use, the *msc*, is lower than the value of the actual use, the *msb*. Thus it always tends to bring *msc* closer to *msb*.

Suppose that the fruit growers combine to form a monopoly and restrict the output of fruit so as to maximize their profit. They make *mpr* equal to *mpc* (for that maximizes their profits), but *vmf* is less than *mpc* because when acting in combination they raise the price of the factor when they increase the quantity they buy, and *mpr* is less than *vmp* because the price of fruit is lowered against them when they increase output. Perhaps a more useful impression would be given by putting this the other way round and saying that by restricting the output they raise the price of the product and lower the price of the factor. Suppose the price of the factor is \$1 a unit, but *mpc*, the marginal private cost of a unit of the factor to the monopoly, is \$1.50. Its marginal product is 2 baskets of fruit which are sold at \$1 each (the price having been raised by the monopolistic restriction) or \$2 for the 2 baskets, but the additional revenue obtained by the fruit monopoly for selling two more units comes to \$1.50 when allowance has been made for the small reduction in the price of all the other baskets of fruit sold by the monopoly as a result of selling two more baskets.

In the production of the alternative product, meat, there is perfect competition so that the first four equations are fulfilled. The price of the factor is \$1, its marginal private cost is the same, the marginal product is 4 pounds of meat whose price has fallen to 25 cents a pound (because all the factors of production discharged by the fruit growers' monopoly turned to the production of meat and increased the supply which lowered the price) or \$1 for the 4 pounds. This \$1 is the *vmp* and is equal to the marginal private revenue (*mpr*). The total situation in the two industries is indicated by the following chart:

	msb	*vmp*	*mpr*	*mpc*	*vmf*	*msc*
Fruit	\$2.00	\$2.00	\$1.50	\$1.50	\$1.00	\$1.00
Meat	\$1.00	\$1.00	\$1.00	\$1.00	\$1.00	\$2.00

Here we see that production in the perfectly competitive meat industry is far from the socially desirable level. The *msb* is only half the *msc* because a unit of factor if taken from the production of meat and applied to the production of fruit would yield a product worth \$2 instead of \$1. The *msc* of each industry is the *msb* of the

other, and just as it is true that too little fruit is being produced it is equally true that too much meat is being produced, and this in spite of the blameless and perfectly competitive behavior of those in the meat industry. They are nevertheless playing a part in the wastage involved in producing goods worth $1 with a factor that is physically capable of producing something else worth $2.

In the case of competitive speculation this complication does not arise. Every speculator who buys cheap and sells dear improves the allocation of resources between the different products. He provides the rest of society with something that is valued more highly in place of something that is valued less highly. It is possible for the speculator to do this only when the optimum division of the factors has not been reached, because he can make a profit only when the *vmp* is greater than the *vmf* and in that case the *msb* is greater than the *msc*. But he tends to lower the *msb* and raise the *msc*, so he always brings society nearer to the optimum situation. If there is freedom for all to participate in this profitable as well as productive activity, *msb* and *msc* will be brought together, and then any further increase in this activity will be neither profitable nor socially beneficial.

It is also true of any other kind of transformation of a factor of production that when the *msb* is greater than the *msc* everybody who contributes to the expansion of the activity is performing a social service. The difference between perfectly competitive speculation and other production under perfectly competitive conditions is that the social cost of production may not coincide with private cost, as we saw in our example of the perfectly competitive meat producers.

It is beneficial to the rest of society even if the speculator is mistaken and incurs a loss,

Competitive speculation has still more to be said in its favor. Even if the speculator should be mistaken and make a loss on his deal, the other members of society will still gain unless some of his loss is imposed on others by his bankruptcy or default. Society

as a whole, including the speculator, loses because he has shifted goods from points where their price was higher to others where they are lower, that is, from points where *msb* is greater to others where it is less. The *msb* of what he has produced is less than the *msc*. But the other members of society will not have lost but rather will have gained a little at his expense—his loss being equal to the social loss *plus* a gain to the other members of society.

This is because when he buys he tends to raise the price against himself, benefiting those from whom he buys; and when he sells he tends to lower the price against himself, benefiting those to whom he sells. Indeed, it is exactly in this way that he benefits other members of society when his speculation is profitable. Let us examine more closely the nature of the benefit to the rest of society rendered by the competitive speculator when his venture turns out to be profitable.

By bidding up the price when it is low he induces other members of society to consume less, and when he sells he encourages an increase in consumption by lowering the price. If there is only one small speculator, the benefit to the other members of society will be very little since he will be able to appropriate almost the whole of the social gain to himself. (*Almost* the whole because there is some effect on price even though it is too small to influence his behavior and in this way to destroy the perfection of competition.) If there are many speculators they will turn the price against themselves so that their share of the (larger) gain is not as great (relatively; it may be much greater absolutely). If there are enough speculators to bring *msc* into equality with *msb*, the profit disappears and all the gain goes to the consumers. (The speculators however will still gain as consumers.)

This may be demonstrated on Figure 1 (p. 30) if it is suitably reinterpreted. Let the curves *A* and *B* measure the *msb* of 200 units of a good applied to two uses *A* and *B*. (These may stand for the alternatives of using a good this year and using it next year.) The optimum division of the good will be 150 to *A* and 50 to *B*, for this would equate the *msb* in the two uses, (or in other words, the *msb* in each case will be equal to the *msc*). Suppose the good is shared equally between the two uses so that the *msb's* and

prices are unequal and measured by the heights of q_1 and q_2. A speculator now shifts 20 units from *B* to *A*. In so doing he raises the price against himself from q_2 to r_2 when buying, and when selling he lowers the price against himself from q_1 to r_1 so that his profit is measured by the rectangle contained within the shaded area G. (His profit is the difference between r_2 and r_1 multiplied by 20, the number of units that he buys at the lower price r_2 and sells at the higher price r_1.) The social gain is equal to the shaded area G, and the gain to the other members of society from the speculation is shown by the triangles left when the rectangle is subtracted from the shaded area G. If the speculation only takes place on a small scale—which will be shown in the figure by drawing r_1 and r_2 very close to q_1 and q_2—the gain will be less, but almost all of it will go to the speculator. The triangles become very small. On the other hand, if speculation is free to all and is extended as long as it remains profitable, r_1 and r_2 will move over until they meet at *s*. The total social gain from speculation will be shown by the area q_1sq_2, but this will consist entirely of the two triangles which represent gain to the public, and the rectangle, which represents profit for the speculators, will have been flattened into nothing.

We can also see on this figure how even faulty (competitive) speculation that nets a loss to the speculator is beneficial to the rest of society as long as the speculator does not shift any of his obligation onto anybody else by bankruptcy or default. If the speculator speculates the wrong way and shifts 20 units of goods from *A* to *B*, the social loss is shown by the double-hatched area L. The speculator's loss is greater than this because in buying he raises the price against himself from q_1 to p_1 and in selling he lowers the price against himself from q_2 to p_2. His loss is shown by the rectangle that *contains* the double-hatched area L. The *gain* to other members of society when the speculator worsens the allocation of goods between different uses is shown by the two triangles which, together with L, make up the rectangle that measures the speculator's loss.

The triangles representing the gain from speculation to the rest of society (excluding the speculators) can be considered as

indicating the benefit from higher prices to those who sell to the speculators and the benefit, from lower prices, to those who buy from the speculators. This might appear to be an improper neglect of the loss to those who have to pay higher prices when buying in competition with the speculators and the loss to those who have to sell at lower prices when in competition with the speculators. But these losses by those who sell when the speculators sell and who buy when the speculators buy is exactly offset by the corresponding gains by those to whom these men sell or from whom these men buy. The area of the triangle shows the gain to the *additional* buyers who come into the market because the speculators lower the price by increasing the amount that is being sold and the gain to the *additional* sellers who come into the market because the speculators raise the price by increasing the amount that is being bought. These additional buyers and sellers can be considered as buying from or selling to the speculators so that there is no offsetting loss by others with whom they do business.

and even when he sells short.

Competitive speculation is socially useful even when the speculator makes a profit out of undertaking to sell things that he does not possess, however much that may savor of unholy magic or even downright trickery. For the speculator is then, in effect, indirectly shifting goods from a future use to a present (or less remote future) use, and he makes his profit out of the social benefit that comes from persuading people to consume more now rather than leave goods until next year when they will not be needed as urgently as they are needed now. The speculator, by offering to sell at a lower price for delivery next year, lowers next year's expected price. This discourages those who have the good from storing it until next year so they put it on the market, lower the price this year, and encourage consumption now. If the speculator is right he makes a profit which comes out of the social gain, leaving some gain for others. (But if there are enough speculators to equalize the price, they make no profit—all the gain goes to the public.) If the speculator is wrong he bears the whole of

the social loss (if he is rich enough) and pays an additional fine to the consumer as is shown in the reinterpretation of Figure 1.

Hostility to speculation is mistaken and arises in part from identifying productive or competitive speculators with aggressive or monopolistic speculators.

The extraordinary usefulness of speculation—its immunity from the errors into which ordinary production may be led by faulty allocation of resources elsewhere—goes ill with the hostility which people who have to work hard for their living often develop against the mysterious gains that speculators make in offices while dealing in goods which they would not even recognize. This hostility arises from two sources. One is the identification of speculation of the perfectly competitive kind, which we have been describing and which can only be beneficial to nonspeculators, with Speculation (with a capital "S") that is not perfectly competitive but monopolistic and aggressive. When very powerful individuals or combinations of individuals are able to manipulate prices so as to extract profits, all this analysis is irrelevant and their profits constitute a form of tribute that they are able to extort from the productive members of society. This tribute they extort even while their activities diminish the total product, moving society away from rather than toward the optimum position by interfering with the best use of factors. We have seen that this evil can be dealt with by the device of counterspeculation—that is, a government agency, by estimating appropriate prices and guaranteeing them to buyers or sellers, makes it impossible for any raiders to influence prices because the government can always match their monetary resources in pegging the price to keep it where it ought to be.

The profits from speculation are best eliminated by increasing the amount of speculation.

The other reason for hostility to speculation is a shifting of blame from those who are responsible for the extremes of an

uneconomic state of affairs to those who benefit while actually engaged in remedying them; much as a small child instead of correcting the cause of his illness may blame the doctor for coming into the house, giving him a nasty medicine just when he feels bad, and taking a good fee for his services into the bargain. Competitive speculators (with a small "s"), as well as any other enterprisers, are able to make a large profit only when there is a very serious maldistribution of factors among different uses. The way to prevent speculators from becoming unduly rich, and at the same time to benefit society, is not to punish them severely (as was done in Russia to those who in times of famine and for personal gain carried food from places where it was not so scarce to others where people were starving) but to bring about so good a distribution of resources that there is not much gain left for the speculators. It can do this best by encouraging speculation and by undertaking speculation itself until the profit disappears. This would be one symptom of success in the achievement of the optimum distribution of resources.

The same is true of any form of production that permits great profits to be made by the enterprisers. If perfect competition and free entry of competitors could be arranged, the maldistribution of resources which allows speculators to make great profits by partially remedying it would be eliminated. (For there is no profit when they are *wholly* remedied as at *s* in Fig. 1.) If these conditions are unattainable, an appropriate policy may be government competition which fixes prices (but by buying and selling, not by legislative interference with the price mechanism) and "artificially" prevents their manipulations. Or it may be found that private enterprise cannot survive if there is perfect competition. When this is the case we have a "public utility," and the services can be provided only at a loss if the allocation of resources is to be maintained, so special measures have to be devised.

All these different situations point to the same general solution —counterspeculation of one kind or another. What particular form this should take in the different situations that call for it we cannot discuss until we have examined some of the problems raised by more complex production.

CHAPTER 9. AN ALTERNATIVE FORMULATION OF THE WELFARE EQUATIONS. EQUALITY AND PROPORTIONALITY

There is an alternative formulation of the principles we have so far examined. To give it here will serve three purposes. It will make the principles stand out more clearly by showing them from a slightly different angle; it will provide an alternative terminology that is more natural and more convenient for some of our problems; and it will connect the present formulation more closely with previous writings in the field.

Adjustment of production so as to reach the optimum may be considered in terms of the factor or in terms of the product, and there are two corresponding formulations of the Rule and of the welfare equations.

If we consider every adjustment in shifting resources from one use to another as the shifting of *one unit of factor* at a time, the marginal quantity of factor is always one unit and the value of the marginal factor is the same thing as the price of the factor. We can then substitute *pf* (the price of the factor) in place of *vmf* in the series of equations that must be satisfied if the optimum use of resources is to be achieved in a free enterprise economy. The series of equations will then appear in the form

$$msb = vmp = mpr = mpc = pf = msc$$

where *msb* and *msc* stand for the marginal social benefit and the marginal social cost where one unit of *factor* is shifted from one use to another.

It is just as legitimate to start instead with *a unit of product* and consider the effects on the economy of producing a unit more or a

unit less of the product. The series of equations will then take the form

$$msb = p = mpr = mpc = vmf = msc$$

Where p is the price of the product (per unit) and *msb* and *msc* stand for the marginal social benefit and the marginal social cost involved in producing an additional unit of product instead of leaving the resources needed for this to produce something else.

The analysis of the conditions under which these equations are satisfied follows the previous analysis very closely. The centralized bureaucratic state would try to make *msb* = *msc* in both senses, for if it is carried out in one sense it is also carried out in the other. The difference is merely one of the unit in which the adjustment is expressed. If a unit of factor could at the margin be used to produce either 2 baskets of fruit or 4 pounds of meat, the price of a basket of fruit would have to be twice the price of 1 pound of meat, say 60 cents for a basket of fruit and 30 cents for a pound of meat. The *msb* of a unit of factor applied to the production of fruit would be measured by the value of 2 baskets of fruit or $1.20. The *msc* is measured by the value of the 4 pounds of meat that have to be sacrificed because the unit of factor is producing fruit instead of meat. These 4 pounds at 30 cents a pound are worth $1.20, so *msb* = *msc*.

Exactly the same situation is described in our second formulation. The *msb* of a unit of fruit (a basketful) is measured by its price at 60 cents. Its *msc* is measured by the value of the meat that is sacrificed because the production of this basket of fruit used up some quantity of factor that might have produced meat. The quantity of factor is half a unit. Half a unit of factor could have produced 2 pounds of meat. Meat is 30 cents a pound, so 2 pounds are worth 60 cents. This is the *msc* of a basket of fruit, and *msb* still equals *msc*.

The same thing may be done the other way around starting with 1 pound of meat as the unit. Its price is 30 cents, and this measures its *msb*. One pound of meat at the margin uses up $\frac{1}{4}$ unit of factor which could have produced $\frac{1}{2}$ basket of fruit. The *msc* of 1 pound of meat is measured by the value of this $\frac{1}{2}$ basket of fruit

which is sacrificed (to permit the 1 pound of meat to be produced), and ½ a basket of fruit is worth 30 cents. Again *msb* = *msc*.

(Of course we have had to use prices to indicate *msb* and *msc*. The centralized bureaucratic state that tries to equalize them without using the price mechanism would be giving itself an unnecessarily difficult if not impossible task.)

So it does not matter which way we define *msb* and *msc*. Their equalization means the same thing on either definition. It is only necessary not to mix the two up and go from one of the meanings to the other in the middle of the analysis because that would render the argument quite invalid. However, this should not be difficult to avoid since it is always necessary to mention the item whose *msb* and *msc* is being considered.

The Rule may be expressed in terms of the marginal cost instead of the marginal quantity of factor.

A similar translation of the Rule as given to the managers of the collectivist economy is possible. The managers could be instructed to adjust their output to make $p = vmf$, which comes to the same thing as making $vmp = pf$. In either case, *vmp* is equated to *vmf*. Consider the grower of fruit who adjusts his output so that *vmp* is equal to *pf*. The price of (a unit of) the factor is $1.20. Its marginal product (*mp*) is 2 baskets of fruit, *vmp* is $1.20 (2 baskets at 60 cents each), and $vmp = pf$. Applying our second formulation to the same situation, we find that *p* (the price of the product) is 60 cents. The marginal factor (the quantity of factor needed to produce the unit of product at the margin) is half a unit of factor, and *vmf*, its value at $1.20 a unit, is 60 cents. If $vmp = pf$, $p = vmf$.

In the same way it can be shown in terms of the new formulation how perfect competition in buying and selling brings about the optimum division of each factor between its different products. In the capitalist economy $msb = p$ if there is a free market in the goods purchased and an optimum allocation of goods, $p = mpr$ if there is perfect competition in selling products; $mpr = mpc$ if the firms maximize their profits; $mpc = vmf$ if there is perfect

competition in buying factors, and *vmf* = *msc* if the first four equations are fulfilled for all the other firms so that *vmf* = the alternative *msb* which is the same thing as the *msc*.

The second formulation is more natural in some ways, and parts of it are more frequently used by economists. That is why it is sufficient to say *p* for price instead of *pp* for price of the product. Similarly *mr* for marginal revenue and *mc* for marginal cost are frequently used in the sense in which we have used *mpr* and *mpc*. The *p* for private is usually omitted because most economists are most of the time engaged only in considering private interests that may lead to equilibrium prices and outputs so that they do not find it necessary to make explicit mention of the private nature of the marginal revenue and the marginal cost they are concerned with. On the other hand, *vmf* (the value of the marginal quantity of factor) and *mf* (the marginal quantity of factor) are not commonly used terms though *vmp* and *mp* (the marginal product), its exact opposite, are very common. Yet these are perfectly symmetrical; *mp* is the quantity of product that results from applying one more unit of the factor; *mf* is the quantity of factor that must be added to produce one more unit of product.

This is misleading until it is pointed out that marginal cost really stands for the value of the marginal quantity of factor.

The neglect of *mf* in the literature has led to a slightly misleading formulation of the Rule for production in the collectivist economy. The Rule is usually given in terms of equating *p* not to *vmf* but to *mc*. Strictly speaking, this is wrong, for we have seen that it is only if output is adjusted so as to make *p* = *vmf* that the optimum allocation of factors is achieved. The error is, however, usually corrected by pointing out the implied assumption that there is perfect competition in buying factors (so that *mc* = *vmf*) or that the manager in calculating *mc* disregards any influence he may have in affecting the price of the factor. This amounts to changing the meaning of *mc* to make it mean the same as *vmf*. Consequently the formulation of the Rule in terms of equating *p* to *mc* is not actually wrong but only misleading.

Of the two formulations of the Rule to be followed for the optimum division of a factor among its different products we shall more frequently use the first. That is, we shall prefer to use the formulation in terms of *the unit of factor* that is added or subtracted from a particular use or shifted from one use to another, rather than the formulation in terms of the unit of product and the effect of producing a unit more or a unit less of product. This is the form in which the Rule was originally given on p. 64. If for some special purpose it seems desirable to take a unit of the product as the starting point, the Rule would instruct the manager of production to increase or decrease his output of the product until its price (*p*) is equal to the value of the marginal quantity of factor (*vmf*). The original formulation is to be preferred in general because it shows up more clearly the fundamental nature of the problem as one of choosing between alternative uses of a scarce factor of production.

Making the price *proportional* instead of equal to marginal cost was believed to be enough,

From the table on p. 89 and the accompanying analysis it would appear that for optimum division of a factor among its different products it is not really necessary that there be perfect competition throughout but only that there be the same degree of imperfection of competition in the alternative use of the factor. If the meat producers also combine to form an organization, they might make their *msb* twice their *pf*, and then there would be an optimum allocation of factors between fruit and meat. The table of p. 89 would now look like this:

	msb	*vmp*	*mpr*	*mpc*	*pf*	*msc*
Fruit	$2.00	$2.00	$1.50	$1.50	$1.00	$2.00
Meat	$2.00	$2.00	$1.50	$1.50	$1.00	$2.00

with *msb* = *msc* for both fruit and meat.

If this were correct the Rule for the controlled economy would have to be changed accordingly. Instead of instructing managers to adjust production until *vmp* is *equal* to *pf*, it would be sufficient

to order them to make *vmp proportional* to *pf;* the ratio to be given to all producers by the Ministry of Economic Planning. Only if this ratio happened to be unity would the Rule appear in the form we have given it. On our second formulation the Rule would be to adjust production until *p* was in some given ratio to *vmf,* the ratio again to be the same for all producers and dictated by the Ministry of Economic Planning. Whatever the ratio determined by the Ministry of Economic Planning, the different *vmp's* and *pf's* (or *p's* and *vmf's*) would be *proportional* and *msb* would be equal to *msc.*

However, this is not quite correct although it was held to be so until very recently.

It is not possible for all the industries which use a factor to do what one of them alone could do. The fruit growers can double the price of fruit by restricting output and shifting the displaced factor to the production of meat. The meat growers can double the price of meat by restricting output and shifting the displaced factor to the production of fruit. But it is not possible for both of these things to be done at the same time. If both industries attempt it, the displaced factor of production will not be absorbed but will remain unused. (Of course, there are other industries besides meat and fruit in which the factor might be absorbed, but we are using these two to represent all the different industries for which the factor of production might be used.) If the factor stays unused we shall certainly not have an optimum division of the factor among its different uses. Its *msb* will fall to zero and certainly below *msc* which is measured by the value of the meat or fruit that a unit of it could produce.

It may be possible for the price of the factor to fall until it is all employed even though in both fruit and meat *vmp* is twice *pf* (and $p = 2\ vmf$). The factor will then be ideally divided between producing fruit and producing meat. The income of the owners of the factor will be less and the income of the managers of production will be greater. If the monopoly is owned by the owners of the factor in proportion to their ownership of the factor, there will be no real difference, for each owner will get just as much more as a shareholder in the monopoly as he gets less as an owner of

the factor. If the distribution of monopoly profit is determined in any other way, some will be richer and some will be poorer, but this will be purely a matter of the division of income among different individuals and will not directly affect the division of the factor between the two uses. (It may do so indirectly if the division of income results in a change in the relative demand for meat and fruit, but this would not prevent an *optimum* division of the factor between fruit and meat for the new demand conditions.)

but proportionality cannot be universal unless it is really equality.

If the *proportionality* instead of equality between the price of each factor and the value of its marginal product were established in *all* uses, this would not distort the optimum division of the factor among its products but would only change the division of income. If *vmp* were everywhere greater than *pf* in the same proportion, less income would go to the owners of the factor and more would go to the monopolists in charge of production. These would in effect be collecting tribute from the owners of the factor. But there would still be an optimum division of the factor of production among the different products in accordance with the demand for products corresponding to the new distribution of incomes. In every firm *msb* would be equal to *msc* in the manner illustrated by the table on p. 100.

However, this is not significant because *it is not possible for the ratio* between *vmp* and *pf* (or between *p* and *vmf* or between *vmp* and *vmf*) *to be the same in all uses* of all factors unless this proportion is *unity*. In other words, they cannot *all* be proportional unless they are *equal*.

This is illustrated in the allocation of labor power between labor and leisure,

The most important of all factors is without any doubt *labor*, and labor always has the alternative of not working but providing leisure for the individual who is the owner of his own labor. For the optimum allocation of labor as between work and leisure it is

necessary that the *vmp* of a little more labor should be *equal* to *pf*, the price of the factor labor or the wage payment for the extra amount of labor. If the worker is free to work the extra hour or not, he will work up to the point where the disutility of another hour's work (which is the same thing as his valuation of another hour's leisure) is measured by the payment for the extra hour's work. He will therefore equate the *msb* of leisure to the pay he gets for his work. But this *msb* is the alternative *msb* to his working, and therefore the *msc* of employing him in industry. If then his *msb* in industry is not *equal* to his wage, *msb* will not be equal to *msc* and we will not have the optimum division of the factor labor between production and leisure.

If the *vmp* of an extra hour's labor is not equal to *pf* (the price or wage of labor), the worker will work either too much or too little. It is wasteful for him to sacrifice an hour of leisure which is worth, say, $1 to him if the net result of his work is that the product is increased by only 50 cents' worth of goods. This is wasteful because the worker is willing to give up more than 50 cents' worth of goods to take the place of what he would add to the total product and he would still be better off by not working while nobody else need be any worse off. If his *vmp* is greater than his *pf* or wage he will not work as much as is desirable. If his *vmp* is $2 and his wage is $1 an hour he will refuse to work beyond the point where his valuation of an hour's leisure rises above $1. The worker refuses to work another hour because, while the wage is $1, he values the hour at, say, $1.25. He would be glad to work another hour for $1.50, and it would be socially desirable for him to do so, for even if he were paid $1.50 for the additional hour there would still be a net benefit of 50 cents for the rest of society out of his marginal product which is worth $2. That is why it is necessary that *vmp* be not merely *proportional* but *equal* to *pf*, for labor.

and in other decisions not subject to the Rule.

The same reasoning would apply to any other factor of production which can be freely directed to producing something with-

out needing the consent of a manager who is subject to the Rule. All cases of competitive speculation would come under this category. Any hoarding of goods from one period to another or any kind of production that might be carried on by small producers on a purely competitive basis would satisfy these conditions unless the activity was subjected to the Rule. This would be unnecessarily cumbersome; would deprive the economy of the advantages of a freely competitive sector where the optimum use of resources was brought about automatically without the supervision and administrative problems necessary elsewhere. Moreover, it would be very difficult to prevent speculators from carrying on their beneficial trading activities secretly and illegally in a way that could not be prevented without a costly system of police supervision and espionage that would be of great danger to democracy. For all these reasons it is necessary to keep to the first formulation of our Rule which instructs managers to adjust production so as to make $vmp = pf$. If vmp must equal pf in these special cases (which are very important ones), it must also be made equal in all other cases if it is to be *proportional* in general.

There are two points to be added to this digression. First, the idea of a worker varying his hours to adjust his marginal valuation of leisure to his wage may seem fanciful when workers are only too glad to get any job and will work whatever hours the employer asks in order to keep their jobs. Even so, workers have many indirect ways of adjusting their hours of work or their output both individually and collectively. In an economy where full use is made of resources, either because it is a controlled economy or because full use comes about by accident in an uncontrolled economy, more satisfactory provision can be devised to give this very important freedom of choice. Technical conditions of production may dictate certain limitations in any individual factory, but it is possible, if industry should be organized with the welfare of the workers as clearly in mind as the whims of the customers, for a worker to be able to choose between different factories with different hours of work, different holiday arrangements, and so on. The choice of how much to work is important in free economies now and can be given much greater scope in a controlled

economy which considers this as important for a worker as it is for a woman to be able to choose between 200,000 different kinds of silk hose.

The other point is to guard against a misunderstanding. It might seem to be implied that if the wage is raised a worker will always work more hours and vice versa. This does not necessarily occur. If the wage is raised, the incentive for working another hour is raised by that much. But by the same act the income of the worker is raised, and that may raise even more the amount of money he is willing to sacrifice for another hour of leisure. Under these circumstances, raising his wage will not make him work harder but on the contrary may result in his working fewer hours.

This, however, does not upset our argument. Making the wage equal to the value of the worker's marginal product still improves on the division of the factor among its alternative products not only by adjusting the value of the marginal product but by varying the value of the alternative product, leisure. It is only a bad habit we all have of forgetting that leisure is just as important as any other product of labor. Perhaps we tend to lose sight of it because it is not to be seen on the transmission belt in the factory.

CHAPTER 10. COMPLEX PRODUCTION I (*FIXED PROPORTIONS BETWEEN FACTORS AND PRODUCTS*)

We now turn to complex production where more than one scarce factor of production is used or where more than one homogeneous product comes out of the factory. In this chapter we shall consider only cases where the different factors or the different products have to be combined in fixed proportions that are uniquely determined by the technique of production.

Where proportions are fixed by the technique of production, the same Rule applies as for simple production.

Considering first the case where the factors have technically fixed proportions, we find that the same Rule that was applied to simple production is still applicable. The Rule, we remember, presupposes a free market in the factors of production (so that the price of each factor is the same for all producers) and a free market in the sale of the product (so that there is an optimum allocation of the goods produced). The Rule instructs all the managers of the plants in the collectivist society to expand production when $vmp > pf$ and contract production when $vmp < pf$ and in this way to approach the optimum situation reached when $vmp = pf$ in all uses of each factor.

Strictly speaking, the marginal product of a factor is indeterminate when there are technically fixed proportions in which the factors must be combined. The addition of a unit of one of the factors by itself is useless and will not permit any increase in production at all. The marginal product would appear to be zero. On the other hand, if a unit of a factor is withdrawn from production, it will cause a large decrease in production, meanwhile rendering useless proportional quantities of the other factors that can be

used only in combination with it. This decrease in production could then be claimed as the marginal product of any one of the factors that have been thrown out of production, and it might seem that it should be shared in some way among them. We are still abstracting from indivisibilities so that any change in the scale of production makes no difference to the ratio or proportion between the quantity of the factors on the one hand and the quantity of product on the other. (The ratio or proportion between any one factor and any of the others we are assuming to be fixed by the technique of production.)

Suppose the proportions between the factors *A*, *B*, and *C* and the product *P* are 1 : 2 : 3 : 6. Then 100 *A* + 200 *B* + 300 *C* combine to produce 600 *P*. An increase in the amount of any one of the factors or in any two of them will not affect the amount of product at all. On the other hand, a reduction in the amount of any one of the factors will result in a *proportionate* reduction in the product. Thus if factor *A* is reduced by 1 per cent from 100 to 99 units, the product will also be reduced by 1 per cent though this will mean a reduction of six units from 600 to 594. This amount of product could, however, have been produced even if the other factors had also been reduced by 1 per cent. 99 *A* + 200 *B* + 300 *C* cannot produce any more than 99 *A* + 198 *B* + 297 *C*. Factor *A* might claim the credit for the six units of product which depend on the continued service of the hundredth unit of *A* if it is to be forthcoming, but factors *B* and *C* could with equal validity make the same claim, since the same decrease in product would come about if *B* were reduced by 1 per cent or if *C* were reduced by 1 per cent, even though the quantity of the other factors were maintained in each case. If such claims were allowed, *A*'s *mp* would be 6 *P*, *B*'s *mp* would be 3 *P* (because a reduction of *B* by one unit would be a reduction of 0.5 per cent and would result in the same percentage of reduction in the product, which would be three units of *P*) and *C*'s *mp* would be 2 *P* (because one unit less of *C* from 300 to 299 would reduce the product from 600 to 598).

If the price of each factor were equal to the value of the *mp* thus reckoned, the payment for *each* factor would come to the value of the total product of all the factors. One hundred units of *A*,

each paid the value of 6 *P*, would get enough to buy 600 *P*, the total product; 200 units of *B* at the value of 3 *P* per unit would get 600 *P*, and 300 units of *C* at 2 *P* each also come to 600 *P*. This shows again that on this way of counting each factor is taking credit for the contribution to production made by the others.

Although it would be fatal for any private enterpriser to have to pay each of several factors the value of the total product of all of them, this is not impossible for the collectivist economy. We shall see that the consideration of profit or loss is quite irrelevant for the economic guidance of the collectivist economy. But still there is something queer here, and the Rule cannot be applied until we have a more satisfactory measure of the marginal product.

While the marginal product of each factor is indeterminate, the Rule can be applied to the *combination* as if it were a single factor.

What we have seen indeed so far is that the marginal product is indeterminate and lies between *zero*—which is what we get when we add a unit of a factor—and the *average product* (the total product of all the factors divided by the number of units of the particular factor concerned)—which is what we lose if we take away a unit of the factor.

We can overcome these difficulties by taking as one unit, not just one of the factors but a *combination* of them, in the proportion in which they must technically be combined. Thus we take 1 *A* + 2 *B* + 3 *C* and call it one unit of *F* (the factor). The price of this combination is the price of the factor *F*. We can then apply the Rule exactly as when dealing with only one scarce factor. The manager will compare the value of the marginal product of this factor-combination *F* with its price. If *vmp* is greater than *pf* he will expand production; if it is less he will contract production; and if it is equal he will know that he has reached the proper output. If all managers follow this Rule, the division of factor *F* between its different products will be the optimum division.

The proof of this is very similar to that for simple production. The price of *F* is what other manufacturers have to pay for it, and they equalize it to *vmp* in the alternative uses. This makes the

price of *F* equal to the value of the alternative product and so to the *msc*. The Rule will again bring *msb* into equality with *msc* and so establish the optimum division of the factor.

Alternatively the Rule can be applied, with identical results, to the marginal *net* product of the individual factors.

There are two ways of showing how the new interpretation of the Rule follows logically from its simple meaning as applied to simple production.

First, it can be expressed in terms of adding one unit of the factor. Suppose the prices of *A*, *B*, and *C* are \$3, \$1.50, and \$1 respectively, and six units of *P*, their *mp*, are worth \$10. Now consider using another unit of *A*. By itself it will add nothing to production and it would be silly to consider using only *A*. The manager of the plant knows however that a unit of *A* must have two units of *B* and three units of *C* to work together with it. If they are acquired too, the increase in output will be 6 *P* which is worth \$10. But not all of this \$10 is the value of the *mp* of *A*. From it must be subtracted the cost of the additional quantity of the other factors, \$3 for the two units of *B* at \$1.50 each and \$3 for the three units of *C* at \$1 each. This leaves \$4 as the value of *A*'s marginal *net* product. This is greater than the price of *A* (\$3), so, according to the Rule, production should be expanded. The same argument can be applied to the desirability of contraction if 6 *P* are worth only \$8. With the same factor prices, the value of *A*'s marginal net product would be only \$2 and less than its price.

Second, and more directly, it can be derived by applying the same principle to combinations of the factors as to each one of them. Again supposing the value of 6 *P* is \$10. It is not desirable to add a unit of *A* nor is it prescribed by the Rule because *pf* is \$3 and *vmp* is zero for an addition of one unit of *A* by itself. Similarly it is not proper to add any of *B* or of *C* by itself or to add any two alone of the three factors. But *F*, the combination of *A*, *B*, and *C* in the proportion 1 : 2 : 3, should be added, because its *pf* (or *vmf*) is \$9 and *vmp* is \$10. Similarly if the value of 6 *P* is \$8, it is not proper to take away a unit of *A* alone because its *pf* is \$3

while *vmp* is $8 (for that is the value of the 6 *P* by which the reduction of a unit of *A* will reduce the product) and factors should be withdrawn from production only when *vmp* < *pf*. For the same reason it would not be advisable for only some of *B* or only some of *C* to be withdrawn from production or for any two alone of the three factors to be withdrawn. But a reduced amount of *F*, which stands for *A*, *B*, and *C* in the proportion 1 : 2 : 3, would be in order; because then *pf* (or *vmf*) is $9 while *vmp* is $8. In the case of technically fixed proportion it is simplest to combine the factors into a composite unit *F* and apply the Rule as in the case of simple production.

The factor prices are determined by the *differences* in the proportions in which factors are combined for different purposes.

An interesting problem arises in connection with the determination of the relative price of the factors. For the purpose of obeying the Rule to bring about the optimum allocation of resources it is sufficient to know the price of the composite unit *F* no matter how this price is made up out of the prices of its separate constituents. The prices might be $2, $2, and $1, or $8, 20 cents, and 20 cents, or any other series of prices, as long as the value of one unit of *A* plus two units of *B* plus three units of *C* comes to $9. How then are the prices of the factors determined?

No light is thrown on this by the explanation of how the Rule brings about the optimum division of each factor among its different uses because this explanation too runs entirely in terms of the alternative product of the composite unit *F* and not of the individual factors (*A*, *B*, and *C*) that make up *F*. However, some light is thrown on this theoretical puzzle in the course of considering our next problem.

Suppose that the composite factor *F* is not used anywhere else. That is, in the production of other goods the factors *A*, *B*, and *C* are used only in proportions different from 1 : 2 : 3, and perhaps they are used only in conjunction with other factors *D* and *E*. What can be meant then by the alternative product of *F* which

was the *msc?* And if it does not exist, how can we say that the optimum division of the factors has been reached?

One factor can *indirectly* be substituted for another through the substitution of products using more of it for products using less of it.

We shall find that these two problems are interdependent and are solved at the same time. The relative price of factors is determined by the very fact of *differences* in the proportions in which they are used in the production of different products. If the demand for one of the factors, say *A*, is less than the supply, its price is reduced. This lowers the value of the various composite factors (*F*'s) for the various products in which the factor *A* is a constituent but the larger the part of this *F* value represented by *A*, the more the price will be lowered. The products for which the value of *F* falls will have their production expanded relative to other products. The increase in demand for the factors accompanying *A* will raise their prices. This will more than offset the cheapening of *A* in those *F*'s where *A* is unimportant but will not do so in those *F*'s where *A* is important. The latter goods will be expanded, the former contracted, so there will result a net increase in the demand for the cheaper factor *A*.

There is thus an *indirect* way in which *A* can be substituted for the other factors when its price falls even though the proportion of the factors in the production of any single product is fixed and it is not possible to use more *A* in place of some of any other factor. The indirect substitution is through the expansion of products which use relatively more *A* because the fall in the price of *A* reduces the price of the composite factor *F* below its *vmp*.

The price of each factor determined in this way is a measure of the value of its marginal *net* product. If the price of factor *A* is $3 it must be true that, in cooperation with the other factors that are needed to make it produce at all, a unit of factor *A* can produce a product whose value is $3 greater than the value of the cooperating factors. If this were not so, the value of all the factors, including factor *A*, would come to more than the value of their

combined marginal product and according to the Rule they would not be employed, the demand for them would fall off, and their prices would fall. If the demand for *A* equals its supply at the price of $3 it must be true that the value of its marginal *net* contribution in its alternative employments equals $3. Consequently, the price of any factor measures the value of its alternative *net* contribution. This is the *msc* of using it. If every manager makes *vmp*, the value of the marginal product of his composite factor *F*, equal to *pf*, the price of the composite factor *msb* = *msc* and the optimum use of the factors is reached.

Even if the proportions are not different in different uses, the factor prices can be determined by their conditions of supply.

The determination of the price of the factor rests on the possibility of indirect substitution of one factor for another through the expansion of those products which use it in relatively greater proportion and the contraction of those products which use the factor in relatively smaller proportion. If the proportions among the factors were the same in all uses, such indirect substitution would no more be possible than direct substitution of one factor for another in the manufacture of a single product, and the price of the factors would be indeterminate. This need not cause us any great concern. First, because in such a case the relative price of the factors would play no part in the establishment of the optimum use of the factors. Indeed, from the point of view of the division of the factors among different products it is not even necessary to take cognizance of the individual factors that make up the composite factor *F*. The same composite factor being used in producing the alternatives, the value of the composite factor is directly equated to the alternative *msb* by the other managers who obey the Rule, so *msb* is seen to equal *msc* without having to go to the trouble of examining any marginal *net* products as we had to do when the proportion between the factors was different in the production of different goods.

Second, the indeterminacy is of only academic importance. If the supply of the constituent factors responded at all to the price,

the relative prices would have to be those which made the various supplies equal to the demand (and of course in the proportions in which they are technically required). The indeterminacy persists only if supply and demand are *both* absolutely insensitive to the relative prices of the factor.

If this is the case and the fixed supplies are not available in exactly the same proportion as that in which they must be combined for production, every factor whose supply is relatively in excess will be valueless and there will be only one scarce factor. This factor is fully utilized and when the various quantities of the other factors are combined with it in the technically fixed proportion some quantity of each the other factors will be left over. Such factors are available in larger quantities than can be used, so they are not scarce and their price is zero. The scarce factor whose supply limits production is now entitled to call the *mp* of the whole *F* combinations its own *mp*, and (with constant returns) its *mp* will be equal to *ap* (its average product) so that its total payment will equal the value of the whole product.

Where the supplies of factors are fixed and in the same proportion as in all production uses, the individual factor prices are indeterminate, but they are then not needed.

If two or more of the factors happen to be available in the exact proportion in which, for technical reasons, they must be combined for production, we have the indeterminacy. The indeterminacy is only in their *relative* prices. The value of the *F* of which they form a part is perfectly determined. What is indeterminate is how much of the payment for a unit of *F* shall go for factor *A* and how much for factor *B*. Since this does not affect the way in which the factors are used and since the factors all belong to the state, the manager can pay for all together or an arbitrary relative price can be fixed. It makes no difference.

The factor labor does not belong to the state even in a collectivist society. It belongs to the individual who does the work, but labor is never involved in this problem because its supply is not independent of the price paid for it.

Product and factor play symmetrical parts in production.

Complex production when there are several products which are produced in technically fixed proportions can be analyzed on exactly the same lines. Indeed, products and factors play exactly symmetrical parts in production. They differ only in sign. It is desired to *maximize* the amount of product for any given input of factors, just as it is desired to minimize the amount of factor for any given output of product. The marginal product is now not a quantity of a single product but a combination of several products in a technically fixed proportion. The *vmp* is the value of the increased quantity of these several products that results from an increase in a unit of factor. The price of the products will be determined on free market as in simple production. The Rule is applied in the ordinary way and leads to the optimum division of the factor among its different uses. If the relative demand for the products should be absolutely independent of their relative prices and if the demand should also happen to be in the same proportions as that in which they are produced, we have the same indeterminacy we have just discussed when both supply of and demand for factors of production are rigid and in the same proportion. But this is unlikely and of no significance for the collectivist economy since it will not affect the use of factors but only the relative prices of products. If necessary, an arbitrary decision could settle the matter without in any way interfering with the optimum division of the factors among these different products.

We see then that complex production, where there are technically fixed proportions between factors or between products, creates no new problem for the collectivist economy. The same Rule is applied, and the difficulty of measuring *vmp* can be overcome by interpreting it as the value of the marginal *net* product after subtracting the cost of the other factors that for technical reasons must accompany it. Even more simply, though it comes to the same thing, the Rule can be applied directly to the composite factor, each unit of which is made up of primary factors combined in the technically required proportion. Where there is more than one product, but in fixed proportion, *vmp* is the value of a com-

posite product. There is complete symmetry between factors and products.

Corresponding to *net vmp* is *net vmf*, and corresponding to *F*, the composite factor, is *P*, the composite product.

We have had no use for the counterpart, on the product side, of the value of the marginal *net* product. This is only because we have been using the first of our two formulations of the Rule—the one developed in terms of shifting a unit of factor. If we use the second formulation the Rule is expressed in terms of the relationship between p, the price of the product, and *vmf*, the value of the marginal quantity of factor needed to produce one additional unit of product. Where we have several products produced in a fixed proportion we have an apparent difficulty in segregating *mf*, the marginal quantity of factor needed for producing another unit of a product, just as we had above in segregating *mp;* and in exactly the same way we have two ways out. One way is to consolidate our product and speak of a composite unit P built up of the different individual products in the technically determined proportion. This would correspond to one unit of F as above for the composite factor of production. The other way is to keep to the individual product but interpret *vmf* as the *net vmf* which is the value of the marginal factor *minus* the value of the additional units of other products that are produced at the same time as the additional units of the product being considered. If the production of another unit of product X involves the use of an additional quantity of a factor (or of several factors) whose value is \$5, but there are produced additional quantities of Y and Z with a value of \$2, the *net vmf* for the production of a unit of X is \$3 (\$5 − \$2). The other products Y and Z are to be regarded as *negative factors* the value of which must be subtracted from \$5 just as if the increase in the production of X by one unit were accompanied by a \$2 reduction in the use of other factors, so that only \$3 more were spent in buying factors in order to produce another unit of X.

In all other respects complex production with fixed proportion is just like simple production. Since we are still assuming divisibil-

ity of factors and products, there will be constant returns to scale. In a capitalist economy, therefore, the size of the firm will be indeterminate, with the same dangers to perfect competition and with exactly the same problems as in simple production. Indeed, complex production with fixed proportions is much more like simple production than it is like complex production with variable proportions.

CHAPTER 11. COMPLEX PRODUCTION II
(*VARIABLE PROPORTIONS BETWEEN FACTORS AND PRODUCTS. NO SIGNIFICANT INDIVISIBILITY*)

In this chapter we shall consider complex production where the proportions between factors and products is variable, but we shall still retain our assumption of the divisibility of factors of production and of the processes of production.

The distinction between the *division* of resources and the *allocation* of factors is parallel to the distinction between the division of income and the allocation of goods.

With the combination of factors variable, we are confronted with a new problem. So far we have been concerned only with the optimum division of a factor among different products so that we do not produce too much of one product and too little of another. This, we saw, is analytically comparable with the optimum division of income where the purpose is not to give too much to one individual and too little to another. Now we also have to see that in producing the various products (in the proper proportions) we combine the factors constituting the resources so that we have the greatest possible total quantity of all the products that are desired. It is conceivable that there is not too much of any one good produced at the expense of any other, but the *combinations of the factors* used in the production of each good are not the best so that a rearrangement or a reshuffling of the factors among the different products would permit more of *all* goods to be produced. While the optimum division of a factor among different goods is analytically similar to the optimum division of income among different individuals, the allocation of factors so as to obtain the optimum proportion of the different factors used in the production of the several goods is analytically identical with the optimum allocation

of goods consumed by the different individuals in the society. The optimum allocation of goods is reached when it is not possible, by reshuffling the goods among the different consumers, to make any consumer better off without making some other consumer worse off. The optimum allocation of factors of production is reached when it is not possible, by reshuffling the factors of production between different firms (production units), to increase the output of any firm without diminishing the output of any other firm.[1] (The firms concerned may be producing the same product or different products, and each firm may be producing a single product or several different products—this does not affect our present argument.)

[1] The distinction here made is between *dividing* a *single* homogeneous item between a number of uses and *allocating* a number of *different* items in proper combination in each of a number of uses. The word distribution is often used in both of these senses as well as in a number of others. "The distribution of income" usually refers to the way in which the total income of a society is divided among the different individuals in the society. We have consistently called this the "division of income." The arrangement of the factors of production in appropriate combinations among the different uses to which they can be put is also often called the distribution of the factors of production. The arrangement of the various goods in appropriate combinations among the consumers is also often called the distribution of consumption goods. We have consistently used the expression "allocation" to indicate these arrangements.

The terminology we have used is not supported by etymology. A more logical use of language would be to use the words *allocation* for what we have called "division" and *collocation* for what we have called "allocation." *Allocation* would then indicate concern only for the purpose *to* which an item is directed. Whether income goes more to one individual or to another; whether resources go more to the production of one product or to the production of another product. *Collocation* on the other hand would be concerned with the way in which the different items were *combined* in the uses to which they are directed; whether the combinations of goods consumed by the different individuals in society are the optimum combinations or whether it may not be possible by rearranging the goods to make all individuals better off or at least to make some individuals better off without making any other individuals worse off; whether the factors of production are combined in the best possible way or whether it may not be possible by rearranging the factors to increase production in general, permitting more of some products to be produced without reducing the output of any products. *Collocation* indicates this simultaneous consideration of the place where the factors or other elements are placed as well as of the way in which they are combined. However, such an innovation, though more logical, is in conflict with our language habits, so it was decided to use the words "division" and "allocation" which will not offend the ears of those who are used to the established economic jargon. *Division* of income or of a factor or of resources refers to the sharing of a single item, which for the purpose in hand is regarded as homogeneous, between alternative uses. *Allocation* of goods or of factors of production refers to the place in which these are located in relation to each other in the course of serving these ends, with attention centered on their proper combination. *Distribution* is used where it is more in accordance with traditional usage if this distinction does not need to be made.

This means that the optimum allocation of factors is not reached as long as it is possible to *increase the total product* by redistributing the factors among the different products to which they are devoted. Increasing the total product here means either increasing all of the products affected, or increasing some while not decreasing any. Obviously if it is possible to do this it should be done in any society in which there is any scarcity (and one in which there is no scarcity is hardly imaginable).

Either problem is too complex to be achieved without using the price mechanism.

In a collectivist economy this might be attempted directly by the Ministry of Economic Planning, and many writers have proposed that it be done this way, even claiming that such centralization would be very efficient in planning everything to fit into everything else. This would require a centralized knowledge of what is going on in every factory, what are the changes from day to day in the demands and supplies at all possible prices of all goods and services and factors of production at all places in the economy, as well as the latest changes in technical knowledge in all branches of production. Obviously this calls for the Universal Mind of LaPlace, as Trotsky has suggested, and this is not practical. If it is tried it will have to be discarded even if it receives lip service while the managers by various subterfuges and unofficial arrangements keep the economy going by haphazard devices which, however inefficient, at least can be made to work somehow.

Again the solution is to call in the price mechanism which will permit the specialized knowledge of the managers of each production unit to be harnessed to the organization of the whole economy. For this it is necessary to express the conditions for the optimum allocation of the factors in terms of the value of their marginal products.

With the proportion between the factors variable, an increase (or decrease) in one of the factors by one unit, while the quantity of all the other factors is unchanged, will normally bring about an

increase (or decrease) in the total product. The variation in the total product is called the marginal product of this factor.

If any two (or more) factors are used in two (or more) production units the optimum allocation of the factors will not have been reached unless the ratio between the marginal products of the two factors is the same in both production units. Let us call the two factors *A* and *B* and the two production units *X* and *Y*. Suppose that in *X* the marginal product of the factor *A* is two while the marginal product of *B* is four units of product. The ratio is 1 : 2; the *mp* of *B* in *X* is twice that of *A*. Now suppose that in *Y* the ratio is not 1 : 2 but any other ratio, say 1 : 3, so that *A*'s *mp* is five and *B*'s is 15 units of product. Then a rearrangement of *A* and *B* between *X* and *Y* will permit more of both goods to be produced. Let two units of *A* be shifted from *Y* to *X*. This would increase the product in *X* by 4 and decrease the product in *Y* by 10. Now let one unit of *B* be shifted in the opposite direction from *X* to *Y*. This will reduce the output in *X* by 4 back again to the original output and will increase the output in *Y* by 15, leaving it five units above the initial output. As a result of this rearrangement of the factors we have a pure gain of five units in *Y*.

	X	*Y*
Effects of shifting two units of *A* from *Y* to *X*	+4	−10
Effects of shifting one unit of *B* from *X* to *Y*	−4	+15
Net gain		+ 5

Shifting a small amount of either factor from *Y* to *X* would leave us more in *both* *X* and *Y*. Whenever the ratio between the marginal products of two factors is not the same in two different production units, a net gain is possible by shifting some of each factor to the point where its relative marginal product is greater. In our example *A*'s relative marginal product is greater in *X* ($\frac{1}{2}$ of *B*'s as compared with only $\frac{1}{3}$ of *B*'s in *Y*) while *B*'s was greater in *Y* (three times *A*'s as compared with only twice *A*'s in *X*) and so we could gain by shifting *A* from *Y* to *X* and *B* from *X* to *Y*. Only if the relative marginal product of different factors is the same in all uses is it impossible to gain by rearranging the factors

and only in that case has the optimum allocation of the factors been attained.

The same Rule brings about both the optimum division of resources and the optimum allocation of the factors.

The optimum allocation of the factors will be reached if all the managers of production in the collectivist economy are induced to keep the same Rule that we had for simple production. By employing more of any factor when *vmp*, the value of its marginal product, is greater than its price and decreasing the employment of any factor when *vmp* is less than its price, each manager brings the price of each factor into equality with the value of its marginal product. But the physical marginal products themselves are proportional to their values. Consequently, in each production unit the marginal products of the factors will be made proportional to the prices of the factors. The factor prices are the same for all managers, so this will bring about the equality of the relative marginal products of all factors of production in all uses. In this way the optimum distribution of factors is brought about by the application of the Rule.

	X	*Y*
vmp (value of the marginal product) of *A*	$10	$10
vmp (value of the marginal product) of *B*	$25	$25
mp (marginal product) of *A*	2	4
mp (marginal product) of *B*	5	10

The Rule calls for shifting each factor of production from points where its relative marginal product is less to others where it is greater until these are equalized. In our example on p. 120 the relative marginal productivities were 1 : 2 and 1 : 3. The shifting of factor *A* from *Y* to *X* and the shifting of factor *B* from *X* to *Y* would raise the relative marginal productivity of each factor in the use from which it is taken away and lower it in the use where the supply is increased. In the table given above it is assumed that the relative marginal productivities are equalized at 1 : $2\frac{1}{2}$.

In this situation *B*'s *vmp* is $2\frac{1}{2}$ times *A*'s *vmp* in both *X* and *Y*.

If the product X is priced at \$5 per unit and the product Y is priced at \$2.50 per unit, the physical marginal productivities corresponding to the *vmp's* of \$10 and \$25 will be as shown in the table. In each use *B*'s *mp* is $2\frac{1}{2}$ times *A*'s, and the relative marginal productivity of the factors is the same in both uses. Each factor has a marginal product in Y twice as great as its marginal product in X. The optimum allocation of the factors of production is thus achieved.

The Rule does more than bring about the optimum combination of factors. It does not merely make the marginal product of different factors *proportional* in different uses; it makes the value of the marginal product *equal* to the price of the factor. In so doing it simultaneously brings about the optimum division of resources between different products. It is possible for the optimum allocation of factors among different *production units* to be achieved (by bringing about *proportionality* between the *mp's* of factors in different uses) without bringing about the optimum division of resources among different *products*. If this were the case it would be impossible to increase *total* production by rearranging the factors among the different production units. In this sense the economy would be at the peak of efficiency. Yet it might still be true that too much of some products was being produced and too little of others as judged by the prices that consumers were willing to pay for them. There is not an optimum division of resources among different *products*.

This situation would be reached if the *ratios* between the *mp's* of different factors were made the same in all their different uses without going on to make *vmp* equal to price. In an undemocratic collectivist economy which chose to disregard the demands of consumers this might be considered satisfactory. The autocratic authority could decree the division of resources between the different products (that is, it would decide in what proportion the different products should be produced). Yet even for this purpose it is possible to achieve the efficiency of production which we call the optimum allocation of factors only if a price mechanism similar to the Rule is applied. The main difference would be that, instead of permitting consumers to bid for products and so change the

prices determining what is produced, the central authority could bid for products in the proportions in which it desires them to be produced and then it would be faced with the problem of distributing them among the population. But even a completely autocratic economy, if it were at all concerned with efficiency of production, could not dispense with the price mechanism as a means of bringing about a reasonably efficient allocation of the factors among the different production units.

The optimum division of resources among products involves equating the *technical marginal substitutability* between products to their marginal substitutability in consumption.

Exactly the same analysis is applicable when there are several different products, the proportion between which can be varied. This means that it is possible by producing a unit less of one of the products to get a larger amount of another product from the same resources. Suppose that it is possible, by reducing the output of product X by one unit, to increase the output of product Y by two units (the quantity of factors being unchanged). This means that the potential unit of product X (that has to be sacrificed in order to permit the production of two units of product Y) must be considered as a factor of production whose marginal product is two units of Y. If the value of two units of Y is greater than the price of X, the Rule commands that the unit of X be sacrificed to the production of Y and that this shift be continued until either the price of X is equal to its *vmp* of Y or all the X is devoted to the production of Y (that is, no X at all is being produced). If the price of X is *greater* than its *vmp* in Y, the movement should be reversed and more of X produced and less of Y until either the price of X is equal to its *vmp* of Y or none at all of product Y is being produced.

Products sacrificed to permit the production of alternatives can be treated as factors, and factors set free can be treated as products.

It is clear that this could just as well be described the other way round with Y called a factor because a decrease in its output per-

mits a greater production of X. The amount by which X could be increased if Y were diminished by one unit could be called the marginal product of Y when devoted to the production of X. The result would be exactly the same. The quantities of X and Y produced would be adjusted so that the ratio in which one can be substituted for the other in production is equal to the ratio between the prices which is equal to the ratio in which one can be substituted for the other in consumption. Suppose the price of X is twice the price of Y and the sacrifice of one unit of X at the margin permits two more units of Y to be produced (and therefore the sacrifice of two units of Y permits one more unit of X to be produced). One unit of X and two units of Y are alternative products. One unit of X is the *msc* of two units of Y, and two units of Y are the *msc* of one unit of X. The value of one unit of X being the same as that of two units of Y, *msc* = *msb* and we have achieved the optimum division between X and Y of the resources used in the production unit.

The adjustment of the proportions between the factors can be described in exactly the same way as the adjustment of the proportion between the products. Just as a unit of product sacrificed in order to increase another product was considered as a factor of production, so the amount of one factor saved by using one more unit of another factor can be regarded as its marginal product. If adding one unit of factor A permits two units of factor B to be released while the same quantity of other factors are used and the same quantity of product is being produced, then two units of B is the marginal product of the one unit of A that is substituted for it. If the value of two units of B is greater than the value (or price) of one unit of A, then A should be substituted for B in this way (in the language of the Rule, more of it should be applied to the "production" of B) until either the value of two units of B has become equal to the price of A or B has been completely ousted from use in the production unit so that it is not possible for the substitution of A for B (the "production" of B by means of A) to proceed any further. If the price of A is greater than the value of the two units of B for which it can act as a substitute (and which

therefore can also act as a substitute for it), then the Rule prescribes a *decrease* in the amount of *A* devoted to the "production" of *B* or in other words a substitution not of *A* for *B*, but of *B* for *A*. The same thing can, of course, be expressed the other way round with half a unit of *A* called the marginal product of *B*.

The difference between factor and product is only one of sign, and there are three kinds of transformation: factor into product, product into alternative product, and factor into displaced factor.

Again we see that factors and products can be discussed in the same language, the difference being only one of sign. Factors are put into production, and it is economic to minimize them. Products are what come out, and it is economic to maximize them. In fact, there can be said to be three different kinds of transformations to which our Rule can be applied. First, the transformation of factors of production into products which we had isolated in our discussion of simple production. Second, the transformation of one product into another or technical substitution between products. Third, the transformation of one factor into another or the technical substitution between factors. In each case what is sacrificed (whether a factor or a product) is called the "factor" and what this sacrifice makes available (whether another product that is increased or another factor that is set free) is called the "product." For the optimum division of resources among different products as well as for the optimum allocation of factors among the different manufacturing units, it is necessary that the equality between the price of the "factor" and the value of its "marginal product" *in all these senses* be reached, as it would be if the Rule were applied directly throughout the collectivist economy to all three types of transformation. This would mean that the manager of every production unit would have to compare not only the price of each factor proper with the value of its various marginal products proper in terms of the actual product, but would also have to consider whether the Rule prescribes the substituting of one factor for another or a decrease in the output of one product in

order to permit an expansion of the output of an alternative product.[1]

All three are properly adjusted by the simple application of the Rule to the transformation of factor into product.

However, this degree of complication of the Rule is not really necessary although efficient managers will naturally be using these aspects of production as a check on the efficiency with which the Rule is being applied. It is sufficient if the Rule is applied only in the literal sense of adjusting the quantity of factor applied so as to make *vmp*, the value of its marginal product in each of the various products (and in any combination), equal to *pf*, its price. If this Rule is kept throughout the economy by all the managers the other relationships between factor and factor and between product and product will look after themselves or rather will be brought about automatically.

The managers who keep the Rule in the literal sense will increase or decrease the amount of each factor applied to the production of each product (as well as each combination of products) until the value of each marginal product is equal to the price of the factor. Suppose factors A, B, and C are applied to the production of products X, Y and Z. Then the price of factor A is made equal to the value of its marginal product of X as well as to the value of its marginal product of Y and of its *vmp* in Z. The same is true for the other factors B and C. This means that if the amount of product X is decreased by \$1's worth it will set free just \$1's worth of factor A (or B or C) because $vmp = pf$, and if this amount of factor is applied to the production of Y it will be able to increase the amount of Y by exactly \$1's worth because here too $vmp = pf$. In other words, the sacrifice of a small quantity of product X (say one unit) will permit an increase in the product Y (if the quantities of the factors and of all other products are kept the same) by an amount which is worth just as much as the unit of X sacrificed. The price of X (considered as a "factor") is auto-

[1] For a similar and more complete analysis of production in terms of these three different forms of transformation, see J. R. Hicks, *Value and Capital*, Oxford University Press, 1939, especially Chap. XV, "The Planning of Production."

matically equal to the value of its marginal product in the alternative product Y, and $msb = msc$.

This relationship, which in simple production is only applicable indirectly between the products of different production units, is here brought about automatically and directly within the firm by the use of the Rule in its literal sense of adjusting factors until $vmp = pf$.

The optimum combination of the factors is brought about automatically in the same way. This is as might be expected from the general symmetry between the part played by factors and by products in the productive process. The Rule makes the value of the marginal product of each factor equal to its price. Consequently, the price of each factor will be equal in value to the quantity of any other factor that it can displace while leaving the total product (and the amount of every other factor) unchanged. If the price of factor *A* is \$1 a unit and the price of factor *B* is 50 cents, *A*'s *vmp* will be \$1 and *B*'s *vmp* will be 50 cents as a result of the application of the Rule. If a unit of *A* is substituted for two units of *B*, total product will be unaffected. It will be diminished by \$1's worth by the withdrawal of two units of *B*, but this will be exactly made up by the addition of the one unit of *A* with a *vmp* of \$1. This means that *A*'s "*vmp*" in terms of factor *B* that is saved is equal to *A*'s price. So we see both the optimum division of resources among different products and the optimum allocation of factors among different production units would be achieved if only every manager of a production unit in the collectivist economy would obey the Rule which enjoins him to employ his factors of production in those quantities which equate the price of each factor to the value of its marginal product in each of the products.

The economic problem is seen more clearly in terms of input of factors than in terms of output of products.

It need hardly be said that all the analyses of this chapter can be rewritten in the alternative terminology which instead of speaking of the shifting of a unit of factor from one use to another starts out with the increase or decrease of a unit of product and

its repercussions on the production unit and on the whole economy. No attempt will be made to translate the whole analysis. This would be a good exercise for the reader. However, some study of this formulation will be made here because of the unfortunate effect it has had in making the issue appear more complicated than it really is. As we have already seen, the fundamental problem is one of the division of resources among different products and the allocation of factors among different production units. The real problem is one of *input*, and when approached in this way the issues can be seen more clearly than when concentration is directed on *output* which looks after itself if the factors of production are all properly directed. The concentration on *output* rather than on *input* is probably a result of living in an enterprise economy where every manager of production naturally concentrates on the output that he must sell to make his profit. However proper an attitude this is for the businessman-producer, it is not appropriate for the economist who should rather be concerned with the best direction of the society's productive potentialities.

The approach from the point of view of output leads to the formulation of *two rules*.

For the alternative formulation in terms of the production of an additional unit of output (where p stands for the price of the product and *vmf* stands for the value of the marginal amount of factor needed to produce one more unit of product), the correct formulation of the Rule is as follows: "*When p is greater than vmf apply more of the factor; when p is less than vmf apply less of the factor; and when p = vmf continue to produce at the same rate because the optimum position has been reached.*" With this translation we could repeat all we have said so far, the only change being the insignificant one of the size of the unit.

But this was not the way in which the Rule was developed. Instead it was expressed in terms of p and *mc* (marginal cost)—with a footnote added to point out that *mc* did not mean exactly marginal cost but what marginal cost would be if there were no influence by buyers on the price of the factors or if such influences

were disregarded in making the calculation. This is not quite a satisfactory correction. It does give *mc* a money value equal to that of *vmf*, but there remains a difference of emphasis that has confused the reader and at least one writer.[1] The expression *mc* refers to a money outlay or cost item while the expression *vmf* refers not to this cost item but to the value of a given combination of factors of production. If p is equal to *vmf* it is desirable that the producer should not reduce output even if *mc* is less than *vmf* and therefore less than p. If the producer pretends that his purchases have no effect on the price of the factors, this will involve pretending that *mc* is equal to *vmf* and so he will not reduce the output. But this instruction does not exactly meet the problem and is too general. If there is to be an adjustment of output to meet a change in conditions, the producer must plan to produce an amount which will equate p to *vmf* in the new conditions. The change in output will have an effect on the prices of the factors and therefore on *vmf*. This will have to be taken into account if the new optimum situation is to be reached without a great deal of unnecessary fumbling. It will therefore not suffice to tell the producer to pretend that he has no influence on the prices of the factors. However it was perfectly natural for this refinement to be missed because the whole emphasis was on *output* and not upon input. The Rule then was given in the form: "If p is greater than *mc* (*mc* calculated in the special way which made it equal to *vmf*) expand output. If p is less than *mc* contract output. If $p = mc$ continue producing at the same rate for that is the right output."

We have seen that this works well enough for simple production and for complex production where the proportion between the factors is technically fixed. But it is not an adequate rule where there is more than one way of producing the product (that is, when the factors may be combined in various proportions) for it does not tell in which of the various ways the product is to be expanded or contracted. In the correct translation of the Rule given on p. 128 this ambiguity does not arise, for there the particular *vmf* is compared with p and the Rule tells whether to increase

[1] A. P. Lerner, "Statics and Dynamics in Socialist Economics," *Economic Journal*, June 1937, p. 270.

or decrease or leave unchanged the amount of each factor used. This is lost when *vmf* is replaced by *mc*. Even when the special rules for calculating the special kind of *mc* are followed, so that the Rule is not *incorrect*, it does not tell the manager *how* to produce and so it is *inadequate*.

The inadequacy was remedied by calling the above rule *rule one* and adding another rule, called *rule two*.[1] This second rule instructs the manager to produce what he is producing *in the cheapest possible way*. This seems good common sense and fits in so well with the capitalist producer's idea of what is reasonable and economically sound that it continued to be used even after its meaning was modified so that it did not mean the cheapest possible way just as *mc* does not mean marginal cost. The modification was the same as in the case of *mc*. In calculating which was the cheapest way of producing the manager must assume, even if this is not true, that the current prices of the factors are fixed and will not be changed by his own purchases. If this happens to be true everything is all right, but if it is not the matter becomes quite complicated. On the one hand, he must ignore his influence on price in calculating total cost to find that method of production which makes total cost a minimum, but on the other hand he must take it into account because he has to repeat the calculation (ignoring his influence on prices) every time he changes his output or the proportion in which he combines the factors, and if he knew before he made any changes what the effects of those changes would be on the prices he could save himself a great deal of trouble. In any position with given prices his calculations might show that a different method of production was cheaper, yet when the new method was adopted the resultant change in prices might alter the situation so that the old method would appear cheaper (if the new prices were taken as given in the new calculation). The manager would then have to change back to the old method of production or perhaps to some intermediate method. A knowledge of how the prices would change as a result of the change in his demand for different factors would

[1] See A. P. Lerner, "A Note on Socialist Economics," *Review of Economic Studies*, October 1936, p. 76.

have saved all these trials and permitted the appropriate method to be chosen directly.

The purpose of this second rule (which can be quite confusing) is to bring about the optimum allocation of the different factors among different production units. If the managers do happen to be without influence on the price of the factors they will tend to bring about the optimum allocation of the factors by merely minimizing their total cost, because this policy will then have the effect of making the price of each factor (which will then be equal to the marginal cost of the factor) proportional to its marginal productivity. If factor *A*'s marginal product is twice *B*'s, but its price is less than twice *B*'s price, the manager will substitute *A* for *B*, one unit of *A* for every two units of *B* discharged, maintaining his output (since the *mp* of *A* is equal to that of two units of *B*) and reducing his costs. *A* will be substituted for *B* until either the ratio between the prices of factors is brought into equality with the ratio between their marginal products or until *B* has been entirely replaced by *A*. Rule 2 will then lead to the optimum allocation of factors among the production units and will thus answer the question *how* to produce the output determined by rule 1.

This formulation is not so satisfactory and arises from a weakness of economists for assuming perfect competition.

But where the prices are not in fact independent of the demand for them by the manager, rule 2 becomes rather difficult to handle.[1] These complications are completely avoided by keeping to our first terminology with its single Rule. The second terminology with its two rules instead of one and its confusing definitions of *mc* and of minimum total cost is an example (although a rather subtle one) of the excessive attachment of economists to the assumption of perfect competition. Instead of attacking the problem of the optimum use of resources directly, perfect competition is assumed, largely because the optimum use of resources happens

[1] Though only by the criteria applied to proposals for establishing a rational or controlled economy. It is child's play compared to the complications in ordinary business bookkeeping.

to be one of the consequences of perfect competition, and then amendments and qualifications have to be elaborated in moving from perfect competition to the real world. Such an approach makes it much more difficult to discuss the optimum use of resources than the direct approach, because the other consequences of perfect competition are irrelevant and serve only to obscure the problem.[1]

The two rules correspond to the distinction between the division of resources and the allocation of factors.

There is one advantage that the alternative second formulation has over the formulation we have preferred to use so far, and that is the way in which its two rules distinguish between the optimum division of resources in general among different goods, and the optimum allocation of the different factors so as to give the optimum combinations of factors in each production unit. Rule 1 tells the producer *how much* to produce of the particular good. When $p = mc$ and mc reflects the value of the alternative product, $msc = msb$. Rule 2 tells the producer what factors to use and makes the mp of factors proportional in different uses so that there is no waste and there is an optimum allocation of factors among the different production units. But this is a very small thing especially as the two rules are really interdependent. Neither would be able to guarantee the one optimum unless the other rule were simultaneously bringing about the other optimum. We shall therefore keep using our original formulation in terms of directing the productive forces of society in the best possible way with a single and unambiguous Rule for the managers of the collectivist economy.

In considering the same kind of complex production in a capitalist economy there is very little to add to what has already been said. Since we are still assuming divisibility of factors, products and processes we will still have constant returns in the sense that

[1] See A. P. Lerner, "Statics and Dynamics in Socialist Economics," *op. cit.*, p. 253, and "A Note on Socialist Economics," *op. cit.*, pp. 72–76. In these articles I thought I had emancipated myself from the habit of starting from perfect competition, but I now believe that my use in these articles of what I here call the second terminology and the two rules are remnants of the same weakness I there criticized in others.

changes in the scale of production will not affect the proportions between factors and products. If there is perfect competition throughout the economy in buying factors and in selling products, the proportion between the factors and the proportion between the products will be determined just as in the case of technically fixed proportions. The only difference will be that it is the relative *prices* of factors rather than the inflexibility of techniques that fix their proportions. For the firms will have to use the factors in just those proportions which minimize their costs so that they can compete. A firm using factors in any other proportion would find its costs greater than those of its rivals and with perfect competition would be promptly eliminated. In the same way the proportion between the products would be determined by their relative prices, for there will be one proportion which is most remunerative, and any firm which does not produce in this proportion will not be able to compete with those that do.

The dangers to perfect competition are of the same nature with the factors variable as in simple production.

With constant costs there is no check on the growth of firms or on the combination of firms until they are large enough to be able to influence prices so as to make monopoly profits, and powerful enough to intimidate prospective ("cutthroat") competitors who otherwise would come in and destroy their power to make any profits. The government of the controlled economy, in the interest of maintaining the optimum use of resources, would have to compete with private enterprise. In its own establishments vmp would be made equal to pf in accordance with the Rule so that $p = vmf$, and, perfect competition being maintained, p also $= mc$. With constant returns, $mc = ac$, and with $p = ac$, total revenue is equal to the cost of the factor so that there will be no profit except for firms that are more efficient than the collectively organized production units. Also the government could undertake other measures, such as outlawing monopolies and combinations in restraint of trade, or practicing *counterspeculation* to prevent any restraint of trade and to maintain perfectly competitive conditions.

Perfect competition could not be expected to come about in an uncontrolled economy even if the larger number of firms that would make perfect competition possible should happen to be established and provide the *possibility* of perfect competition. This situation could not be expected to remain for very long.

Imperfect competition can now interfere with the optimum allocation of the factors as well as with the optimum division of resources among different products.

Where there is not perfect competition, the variability of the proportion in which factors can be combined gives rise to another kind of departure from the best use of resources. In the case of simple production we saw that different degrees of competition in the sale of different goods would lead to too little being produced of the goods for which the degree of competition was least (or the degree of monopoly greatest) and too much of the goods for which the degree of competition was greatest (and the degree of monopoly therefore least). This wrong division of resources among different goods could not be avoided by having the same degree of competition in all uses because in some uses at least (such as in an individual's devoting time to leisure) competition is necessarily perfect. Imperfect competition therefore naturally led to a wrong division of resources, too many being directed to some uses (which may include unemployment as well as leisure) and too few being directed toward others.

In the absence of perfect competition in buying factors, the *mc* of factors will not be equal to their prices but greater, and unless the degree of imperfection of competition is exactly the same for each firm in buying all factors (which is sufficiently unlikely to be disregarded) *mc* will not even be *proportional* to price. The firms are still concerned to minimize the cost of what they produce because every dollar saved is a dollar added to their profits, and they do that by combining the factors in such a way that the marginal productivities of the factors are proportional not to their price, which is the same for all firms (if the monopolies have not permitted discrimination in prices), but to the *mc*'s, which will

in general differ. The marginal productivity of each factor will therefore not be the same in the different firms, and this spells an uneconomic allocation of factors as between the different firms.

This kind of waste is not possible in simple production or in complex production with technically fixed proportion between the factors, because in these cases it is not possible to combine factors uneconomically. Consequently there is a stronger argument for controlling the economy when the proportion in which factors can be combined is variable.

The same kind of inefficiency also occurs because the combinations of the *products* are variable. The firms who maximize their profits will produce their products in those proportions which make their technical substitutabilities in production proportional to their *marginal revenues* rather than to their prices. The marginal revenues will not be the same for the different firms even if the prices are and so the *vmp* of products in the production of other products will not be equal to the prices of these other products.

At first sight this looks like wrong division of resources among different products, which we already know from our examination of simple production. Too much of one product is produced and too little of another. But it is more serious than this because these divergencies are different for the different firms. One firm produces too much of product X and too little of product Y, and another firm produces too much of product Y and too little of product X. These do not cancel out any more than when one firm used too much of one factor and the other used too much of another factor. There is here a pure waste of resources rather than a disregard of the guidance provided consumers as to what should be produced. The waste is due to the *different* technical marginal substitutabilities (M) between the *products* in the different firms. More of *both* products could be produced if there were a reshuffling of products between the firms, each firm producing more of the product with the smaller technical M (relative to the other firm) and less of the product with the greater relative technical M.[1] The proof of this

[1] It should be remembered that a larger technical M means that the sacrifice of a unit of the good in question permits a greater amount of the *other* good to be produced in its place, or that an increase in the output of the good in question requires a greater

is very similar to that already given (p. 121) for the desirability of having the ratio between the marginal product of two factors the same in the production of any two goods. If in Firm I the technical M is 1 of X for 1 of Y and in Firm II the technical M is 1 of X for 2 of Y then Firm I can produce one unit less of Y and one unit more of X with the same quantities of factors as before, while Firm II can produce one unit less of X and two units more of Y with the same resources. A net gain of one unit of Y could therefore be brought about by getting each firm to concentrate on that product which has the smaller relative technical M. Firm I concentrates on X which has a relatively smaller technical M there (1 as compared to 2 in Firm II), and Firm II concentrates on Y which has a relatively smaller M there (2 as compared with 1 in Firm I). Such reshuffling of production would raise the total product until the technical M's were the same in all firms producing the same combination of products. This cannot be done in an uncontrolled economy, and thus we have a net social loss which we must add to the uneconomic allocation of factors among firms described in the previous paragraph. This strengthens still further the case for a controlled economy which would prevent such wastes.

sacrifice of the *other* good. It is therefore best for firms to concentrate on goods with a relatively smaller M because that means a smaller sacrifice of the alternative goods.

CHAPTER 12. DIMINISHING MARGINAL TRANSFORMABILITY

We have already made the acquaintance of the principle of diminishing marginal substitutability. It was introduced in Chapter 2 (p. 10 and especially the footnote on p. 13) to show how an optimum allocation of goods could be reached without every consumer spending all his income in buying just one good. It will be remembered that the marginal substitutability (M) of a good for any other good is measured by the number of units of the other good for which a unit can be substituted, while leaving the consumer neither better off nor worse off.

The principle of diminishing M (marginal substitutability) is also applicable to production.

As one good is substituted for another so that the consumer has more of it and less of the other good for which it is substituted, further additions will be less urgently needed, as compared with the other good, so that an additional unit will only be substitutable for a smaller amount of the other good than before—its M (marginal substitutability) for the other good will diminish. In this way the M comes to be adjusted to the relative price before one good has been completely substituted for all the other goods consumed by the individual.

In this chapter we shall show how the same principle is applicable to *production* where the proportions between factors or between products are variable.

The technical M between two factors is given by the ratio between their marginal products.

We have seen that the optimum allocation of factors among different production units follows the same principles as the opti-

mum allocation of goods among different consumers. The ratio between the marginal product (*mp*) of every pair of factors must be the same in each production unit in which the two are used. This ratio between the *mp*'s (marginal products) is nothing but the technical *M* of one factor for another. If *A*'s *mp* is twice *B*'s, a unit of *A* is substitutable for two units of *B* and its *M* for *B* is 2 and its *M* for any other factor is twice *B*'s *M*. If the ratio between the *mp*'s of any two factors (the first factor's *M* for the other factor) is not the same in all production units in which both are used, the optimum allocation of the factors has not been reached. It is possible to increase the total product by shifting some of each factor from points where its *M* is relatively low to others where it is relatively high. This is done in a collectivist economy by the manager of the production unit in obeying the Rule that equates the *vmp* of each factor to *pf* its price and automatically results in the equalization of the *M*'s of the factors in all the production units (see p. 121).

Each manager plays his part in this adjustment by substituting any factor whose *M* is greater in proportion to its price for the other factor (whose *M* is less in proportion to its price) until either the *M*'s have been made proportionate to price or the one factor has completely ousted the other.

The same thing is done, if there is perfect competition, by independent firms in the course of maximizing their profits.

If the *M*'s did not change as a result of the substitution of one factor for another, this substitution would have to be continued until one factor had been completely substituted for all the others in the production unit, and it alone was left to carry on production. (That excludes the unlikely case where the *M*'s happened to be proportional to price to begin with, when the ratios among the factors would be indeterminate.) We must therefore look for some reason why the *M*'s should change as a result of the substitution, so that they become proportional to price in every firm.

One explanation might be the change in the price of the factor. As factor *A* is substituted for factor *B* (because its *M* is greater in proportion to its price than *B*'s *M* in proportion to its price)

the increased demand for A will tend to raise the price of A, and the decreased demand for B will tend to lower the price of B. This change in relative prices will bring the ratio $M : pf$ for the two factors into equality so that the substitution of A for B would come to a stop. But this would only work for large productive units whose demand could affect the price paid for the factors, and this could hardly be true for every production unit where more than one factor is employed. It is true that the substitution of A for B in many small units could have the total effect of changing the prices of the factors so that they would become proportional to their M's, but in that case it would be a matter of indifference to every production unit whether it used only A and none of B, or only B and none of A, or any intermediate proportion.

This cannot therefore be the explanation of the fact that production units do not limit themselves to only one factor of production and are not indifferent as to the proportion in which they combine the factors.

It depends on the proportion in which the factors are combined.

The answer lies in the principle of diminishing M. As one factor is substituted for, and increases in proportion to, the other factor, its M diminishes. The other factor that is being displaced *decreases* in proportion, and so according to the same principle *its* M increases. The ratio between the M's depends on the ratio between the (quantities of the) factors. It is the change in the *proportion* in which the factors are combined that adjusts the M's and makes them proportional to the prices.

Possible ranges of constant or increasing M are economically irrelevant.

As in the case of the diminishing M of consumption goods for each other, there is no certainty that the principle always holds. There may be ranges where M is constant or even increasing. But such ranges—if they exist—are not economically significant because combinations within such a range of combinations could

not represent an optimum allocation of the factors. With increasing M, even if the M's of the factors are proportional to their prices, a small substitution of $1's worth of one factor for $1's worth of another in either direction would not affect the total product, but would raise the M of the *increased* factor and diminish the M of the *decreased* factor, so that a continuation of the change in the same direction would result in an *increase* in the total product. The initial position with M's proportional to prices would be a *minimum* instead of a maximum position so that a change in either direction would improve the allocation of the factors. If M were constant the substitution of $1's worth of one factor for $1's worth of another would have no effect on the total product and we would have the queer case of indifference to the proportions in which factors are combined.

There is yet another argument for the assumption of diminishing M in general. If the M between any two factors did not diminish, there would be no economic reason for distinguishing between them at all. They would, for the purpose of production, be so similar that the producer would call them one factor. At the opposite extreme is the case of fixed proportions when the factors are so different that it is not possible to have any substitution between them at all.

The optimum proportion between factors, as well as between products, is that which makes the M's proportional to the prices.

We see then that the optimum allocation of factors among production units is brought about by establishing optimum *proportions* between the factors in each production unit. This is the proportion between the factors that makes their M's proportional to their prices. The optimum proportion can be attained because of the diminishing M of each factor as its proportion to the other factors is increased.

Exactly the same relationship holds for products. Whenever two or more products can be produced in variable proportions, the optimum proportion is that which makes the technical M's of the products proportional to their prices. Here also an optimum pro-

portion can be reached because of the change in M as the *ratio* between the products changes.

The general principle of diminishing transformability shows itself as diminishing *mp*, increasing *mf*, diminishing M of factors and increasing M of products.

When an increase in quantity of a factor results in an increase in one of the products while the quantities of the other factors and products remain the same, we say that the added quantity of the factor has been transformed into the increment of product. When one factor is increased and the quantity of a second factor is thereby set free while the product and the quantities of the other factors are unchanged, we can say that the increment of the first factor has been transformed into the amount of the second factor that is set free. When a decrease in amount of one product permits a second product to be increased without affecting the quantities of the other factors or products, we may say that the first product is transformed into the second product. In all of these cases there will apply the principle of *diminishing marginal transformability.*

Where one factor is substituted for another, the principle of diminishing marginal transformability appears in the form of *diminishing M*, in which guise we have already met it. Where the transformation is of a factor into a product, the principle of diminishing marginal transformability takes the form of a diminishing marginal product. (This is the law of *diminishing mp.*) Where the transformation is of one product into another product, the principle of diminishing marginal transformability takes the form of diminishing technical marginal substitutability of the product whose output is decreased for the product whose output is thereby augmented.

Continued equal sacrifices of the first product will permit the second product to be increased by smaller and smaller increments. This can be called diminishing technical M between the products, but in doing so we will have changed the meaning of the word *substitution* to the exact opposite of what is usually meant by it. The substitution of product A for product B suggests the production of A instead of the production of B or the production of more

of *A* while producing less of *B*. It will be better to conform to the ordinary use of language and speak of the substitution of product *A* for product *B* to mean producing more of *A* in place of *B* whose output is diminished. We will now have, not diminishing technical *M*, but *increasing* technical *M*. As more of a product is produced in place of another product, the quantity of the other product that has to be sacrificed will become greater and greater because of its diminishing marginal transformability into the factor whose output is being increased. This increasing *M* of products is only another indication of the opposite *sign* here. We are interested in the *output* of products, while we are interested in the *input* of factors. It is related to the desirability of *minimizing* the quantities of the factors but of *maximizing* the quantities of the products.

This shows itself again in the transformation of factor into product when we approach it from the point of view of the product. As we saw in connection with the alternative formulation of the Rule and the welfare equations in Chapter 9, diminishing *mp* comes to the same thing as increasing *mf*. If the addition of equal increments of one unit of a factor results in diminishing increments of product (diminishing *mp*), it will take *increasing* increments of the factor to bring about constant increments of the product. This gives us the principle of increasing *mf*. Diminishing *M* of factors for each other, diminishing *mp*, increasing *M* of products for each other, and increasing *mf* are all particular aspects of the more general principle of *diminishing marginal transformability*.

The quantity of any factor applied to the production of any particular product will then be adjusted by the principle of diminishing *mp*. Adjustment will go on until the addition of the factor has reduced its marginal transformability into the product until it is equal to the ratio between the price of the factor and the price of the product. For example, if the price of a unit of factor is twice that of a unit of product the factor will be applied until an additional unit of factor can be transformed into two additional units of the product. In the more common language of production, this simply means that the factor will be applied in accordance with the Rule up to the point where *vmp* has fallen to equality with *pf*.

Constant returns to scale are not inconsistent with diminishing returns to increases in the proportion of one factor to the others.

This may appear strange because we are still assuming perfect divisibility of factors, products, and methods of production, and we have seen that as long as this is so the scale on which production is carried on should make no difference either to the average product or to the marginal product. Indeed, we called this state of affairs *constant returns* (average and marginal) and now we have the omnipresent principle of diminishing marginal transformability translated to show that diminishing marginal returns (equal to diminishing *mp*) may be considered general.

The explanation of this paradox is that the constant returns that we have met refer only to the *scale* of production. If all the factors and products but one are to be doubled, the remaining one must be doubled too and similarly for any other degree of decrease or increase in the scale of operations. This follows from the divisibility of the factors, products, and methods which permits any particular method of production, involving certain proportions between factors and products, to be repeated in exactly the same way on a larger or on a smaller scale.

Diminishing transformability does not apply where a single factor is applied to the production of a single product. In that case there is no reason why the same rate of transformation should not go on indefinitely. The principle of diminishing transformability applies when the transformation is of the nature of a *substitution* that results in a change in the proportion between the different items involved in the productive process. It may change the proportions between the factors as when one factor is substituted for another; it may change the proportions between the products, as when one product is substituted for another; or it may change the proportions between all the factors and all the products, as when one of several factors is varied in order to vary the product or when one of several products is varied as a result of a variation in the factors. There will then be increasing resistance to the continuation of such a change in the productive process and this increasing resistance is the essence of the principle of diminishing

marginal transformability in all its manifestations and metamorphoses.

When more than one factor is used an increase in one of them, *without any change in the other factors*, will mean that this factor has increased relatively to the other factors. Apart from the question of scale (which is not relevant to our present concern with proportions) the situation would be just the same if the other factors were decreased in the same proportion in which this factor is increased, or if this factor were substituted for the other factors so as to bring about the same change in the proportions between them. *The mp of a factor diminishes when the factor is increased in proportion to the other factors.* If the other factors were also increased in the same proportion there would be no substitution of one factor for the others and no diminishing *mp* because there would then be nothing but a change in scale.

Substitution always involves at least *three* items.

Another way of putting this is to say that substitution always involves at least three items. One is increased, one is decreased, and the third remains the same. Then we can say that the first is substituted for the second in proportions that leave the third unchanged. The third item is essential, for it is only by seeing that it is kept constant that we can measure the substitutability of the first for the second. Otherwise we cannot be sure that the effect of the increase of the one item is completely offset by the decrease in the other item. In this way we can say that one unit of factor A is substitutable for two units of factor B in maintaining the product constant. We can say that one unit of product X is substitutable for two units of product Y, the quantity of factors and of the other products being kept constant. We can say that 1 basket of fruit is substitutable for 2 pounds of meat, the well-being of the consumers being kept constant, and finally we can say that one unit of factor A can be transformed into two units of product X, the quantities of the other factors and products being kept constant.

This chapter has so far been concerned only with the proportions between factors and products and their various substitutabilities

for each other *within the production unit*. We must be careful therefore not to confuse diminishing *mp* and increasing *mf* with similar phenomena that we shall meet when we have removed several more of our simplifying assumptions and are able to consider increasing and decreasing returns and costs from the wider point of view of an industry or from the point of view of society as a whole. Our analysis in its present stage is applied only to the individual production unit. Being concerned with purely technical relationships, it is equally relevant for collectivist, capitalist, controlled, or uncontrolled economies.

CHAPTER 13. THE ELASTICITY OF SUBSTITUTION AND THE LAW OF DIMINISHING RETURNS

The elasticity of substitution measures the rate at which substitutability diminishes.

As the ratio of one factor to another factor increases, its M for the other factor diminishes. The degree to which it is possible to substitute one factor for another without bringing about more than a certain decrease in its M for the other factor is called its *elasticity of substitution*. If a great deal of substitution is possible before the M changes by the given amount, the elasticity of substitution is said to be great. If only a very little substitution is possible before the same change in M comes about, the elasticity of substitution is said to be low. It can be seen that the elasticity of substitution is a measure of the variability of the proportions between the factors, of the degree to which it is possible to substitute one factor for another without running into too much resistance of the kind indicated in the principle of diminishing transformability. If the transformability diminishes very slowly the elasticity of substitution is great, and if the transformability diminishes rapidly the elasticity of substitution is small. The elasticity of substitution is the *inverse* of the rate at which the transformability diminishes. It is a measure of the *slowness* with which the transformability diminishes in response to a change in the proportion between the factors.

The elasticity of substitution should not be confused with M, the marginal substitutability, which is the rate of substitution itself. M is measured by the number of units of the other factor for which a unit of a factor can be substituted while leaving the product and the quantity of the other factors unchanged. The elasticity of substitution (which is represented by the symbol σ) is measured

by the degree to which it is possible to substitute one factor for another for any given change in M. It measures the *slowness* of the response of M to the change in the proportion between the factors of production. It is *the rate of change of the rate of substitution* as the proportion between the factors is changed.

It can be generalized to apply to all forms of transformation, measuring the rate at which substitutability increases in products.

The concept of the elasticity of substitution, like the concept of substitution itself, is applicable not only to factors of production but to products and to consumption goods. In consumption goods, the substitution of A for B consists of consuming more of A and less of B in proportions which leave the consumer equally well off. The number of units of B that must be withdrawn from consumption to make up for an added unit of A, leaving the consumer neither better off nor worse off, is the measure of the M of A for B. The slowness with which this M decreases as A is substituted for B is measured by the elasticity of substitution of A for B. In products, the substitution of X for Y consists of producing more of X and less of Y in proportions which permit the quantities of the factors and of the other products to remain the same. As more of X is produced in place of Y, the technical M of X for Y will *increase*. As we saw in Chapter 12, this is due to the signs being reversed in the case of products. As more of X is produced at the cost of decreasing the output of Y, the *diminishing* transformability of Y into X means an *increasing* M of X for Y. However, the technical elasticity of substitution of X for Y will follow the same principles. It measures the slowness with which the technical M of X for Y *increases* as the ratio of X to Y is increased. (This will be exactly the same as the elasticity of transformation of Y into X, which is the slowness with which the marginal transformability of Y into X changes as the proportion between Y and X changes.) In all cases the more slowly M changes as a result of the change in the proportion between the factors or products, the greater are the opportunities of continued substitution and the greater is the elasticity of substitution.

Zero elasticity of substitution signifies fixed proportions, and infinite elasticity of substitution indicates economically indistinguishable factors or products, the proportions between which can be varied indefinitely.

Where the proportions in which factors must be combined (or in which products have to be produced or in which consumption

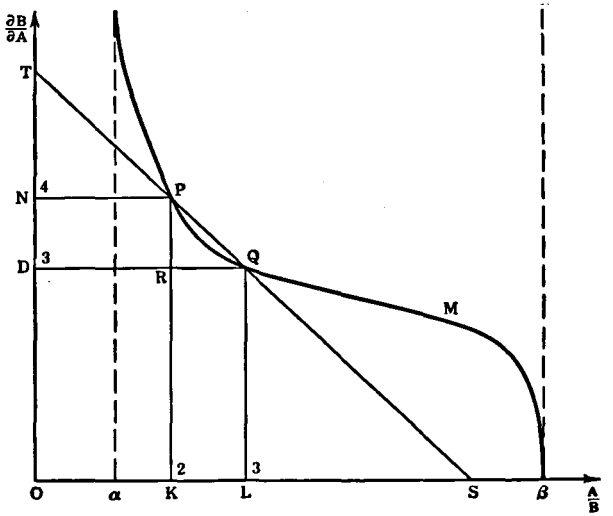


Figure 2

goods have to be consumed) are technically fixed, *M* falls precipitately from infinity to zero. As long as there is less of any factor than the fixed proportion dictates, no quantity of the other factors can make up for a decrease in this factor, so its *M* is infinitely large; but the moment its proportion to the other factors has been increased to the technically required proportion, no further increase in it could compensate for any decrease in the other factors, so its *M* has fallen to zero. Here the *elasticity of substitution* is said to be *zero*.

The opposite extreme is the situation where one factor can be substituted for another with no diminution in *M* at all. This occurs

when the two factors are economically identical and the difference, if any, is irrelevant for the purpose in hand—like dark-haired and fair-haired coal miners, or like cane and beet sugar for most purposes. The *elasticity of substitution* is then said to be *infinite.* We do not bother to distinguish between the factors and we call them one factor. As we have already indicated above, perfect substitutability of the different units for each other—more accurately expressed as infinite elasticity of substitution—is just what gives a collection of productive units the right to call themselves one factor of production.

The elasticity of substitution is illustrated in Figure 2. Horizontally is measured A/B, the number of units of A that are being used for each unit of B. Vertically is measured $\partial B/\partial A$, A's M for B, that is, the number of units of B for which one unit of A can be substituted or which can be substituted for one unit of A. The line marked M, passing through P and Q, indicates how A's M falls when the ratio of A to B increases. (This is in accordance with the principle of diminishing M.) This line may be called the marginal substitutability curve or M curve of A for B. The point P indicates that when twice as many units of A are used as of B ($A/B = 2$) one unit of A is substitutable for four units of B so that $\partial B/\partial A$ (A's M for B) is 4. The point Q indicates that when A has been substituted for B so that the ratio is three units of A for each unit of B, A's M has fallen to 3 so that one unit of A is now substitutable for only three units of B.

The diminishing M is indicated by the way in which the M curve slopes downward to the right. The further to the right we go, the greater is A/B, the ratio of A to B, and consequently the lower is A's M ($\partial B/\partial A$). Infinite elasticity of substitution would be represented by a horizontal line showing that M does not fall at all as the one factor is substituted for the other. Zero elasticity of substitution (that is, technically fixed proportions) would be represented by a vertical line drawn at the point on the horizontal axis that indicated the technically determined ratios between the factors. A normal elasticity would be indicated by an M curve sloping down to the right as in our figure.

All elasticities are measured in terms of *proportional* changes.

Now to measure the elasticity of substitution (σ). It might appear at first sight that it could be measured by the ratio between RQ (the change in the proportion between the factors) and RP (the change in M). The greater is σ (the elasticity of substitution), the more will the proportion between the factors have to change to bring about a given change in M, and the greater will be RQ/RP. If the curve is horizontal, as with perfect substitutability, RP would be zero and RQ/RP would give the correct value for σ of *infinity*. If the curve is perpendicular, as with a fixed proportion, RQ is zero and RQ/RP would equal zero, which is again the correct figure for σ.

Nevertheless, this measure for σ will not do. Applied to the example in Figure 2, these measures would make σ equal to unity, but this value will depend on the arbitrary units in which A and B are measured. If the unit in which A is measured were feet instead of yards (which should make no difference to the measure of σ) the horizontal measures (at K and L) would be, not 2 and 3, (yards) but 6 and 9 (feet), while the vertical measurement would be, not 3 and 4, but 1 and $1\frac{1}{3}$, which would be the quantities of B that can be substituted for a foot of A instead of a yard. This would increase our measure from 1 to 9 (from 1/1 to $3/\frac{1}{3}$). In the same way this measure of σ would depend on the size of the unit in which B is measured. This would be very inconvenient. We need a measure independent of the size of the units in which the factors are measured.

Such a measure is obtained by comparing not the *absolute* changes in A/B and in M but their *proportional* changes, which are independent of the unit in which the factors are measured. The *proportional change* is defined as the absolute change divided by the smaller of the two measures. The proportional change in A/B is KL/OK which comes to $\frac{1}{2}$, whether we use yards ($\frac{1}{2}$) or feet ($\frac{3}{6}$) or inches ($\frac{36}{72}$) or any other measure for A and for B. The proportional change in A's M is DN/OD, which comes to $\frac{1}{3}$ whatever the units of measurement used. The elasticity of substitution σ is therefore given by $KL/OK \div DN/OD$ which comes

to $\frac{1}{2} \div \frac{1}{3}$ or $1\frac{1}{2}$. This is the value of σ—*the proportional change in the ratio between the factors divided by the proportional change in M.*[1]

This brief treatment of σ has been entirely in terms of two factors (in the widest sense—including products and consumption goods). It is of course applicable to any number of factors and can be conceived of as referring to the substitutability between any pair of factors while the others are kept constant, or between any factor on the one hand and some or all the other factors on the other hand taken together as if they were all one factor, or between any pair of any such combinations of factors.

We have seen how the law of diminishing *mp* derives from and is a special case of the principle of diminishing marginal transformability. The principle of diminishing *mp* is sometimes called the law of diminishing returns, but that name is more often—and more conveniently—applied to the diminishing *average* product to which we now turn. We shall use the phrase "diminishing returns" only in this latter sense of diminishing *average* product.

[1] A little geometry simplifies this. Draw a straight line through P and Q to meet the axes in S and T.

$$\sigma = \frac{KL}{OK} \div \frac{DN}{OD} = \frac{RQ}{NP} \div \frac{RP}{LQ} = \text{(by similar triangles)}\ \frac{PQ}{PT} \div \frac{PQ}{QS} =$$

$$\frac{PQ}{PT} \times \frac{QS}{PQ} = \frac{QS}{PT}$$

(QS/PT, which can be taken as the measure of σ, is in Fig. 2 equal to LS/OK or to OD/NT, each of which is equal to $\frac{2}{3}$ or $1\frac{1}{2}$.)

If we suppose P and Q to be very close together, we could give a measure to σ for indefinitely small substitutions. We can suppose P and Q to be coincident (or adjacent points). We draw ST tangential to the M curve at P and σ is measured by PS/PT. This measure is applicable not only to the elasticity of substitution but to any other case where it is useful to avoid arbitrary effects of units used. It was developed first in connection with the elasticity of demand which measures the effect of the change in the price of a good on the amount demanded and is the proportional change in the amount demanded divided by the proportional change in the price. If instead of A/B we measure horizontally the amount demanded and instead of marginal substitutability we measure vertically the price of the good, M would then be the demand curve and the elasticity of demand would be measured in exactly the same way and would equal $1\frac{1}{2}$. A slight change in the figure permits the device to be applied to the elasticity of supply, which means the proportional change in the amount of a good supplied divided by the proportional change in the price that elicits this change in supply.

See Marshall *Principles of Economics*, pp. 102n, 839, 840, Lerner, "Diagrammatical Representation of Elasticity of Demand," *Review of Economic Studies*, October 1933, pp. 39–44; and Lerner, "Diagrammatical Representation of Elasticity of Substitution," *Review of Economic Studies*, October 1933, pp. 68–71.

This principle, like that of diminishing *M*, refers to changes not in the *scale* of production (which we have seen can by itself never have any effect on either marginal or average products) but to changes in the *proportion* in which factors are combined. It says that if one factor is increased and the other factor or factors kept constant the *total* product will increase in a proportion smaller than that in which the factor is increased. Its *average* product (the total product divided by the total number of units of this factor) will therefore diminish. If there are two factors *A* and *B* and *A* is increased by, say, 10 per cent while the quantity of factor *B* is unchanged, the product *P* will increase by less than 10 per cent, so that the average product, P/A, will have diminished. This seems common sense. We know that when *all* the factors are increased in a certain proportion the product will also increase in the same proportion. It seems only natural that if only one factor is increased in this proportion, the product will not increase as much as when all the factors are increased, so the product will increase in a smaller proportion than the factor, and the factor's average product will diminish in accordance with the *law of diminishing returns*.

This common-sense explanation is useful as a rough indication of the principles involved, but it goes a little further than is warranted because it contains certain implicit assumptions that are not always correct. We shall come back to these later. Here we must go more carefully into the meaning of and the justifications for the law of diminishing returns.

Both the law of diminishing returns (which is the law of diminishing average product) and the law of diminishing *mp* look like statements about technical conditions of production that we deduce from first principles while sitting in an armchair. We know that technical knowledge cannot be attained in that way and any such claim must be false. In truth these laws do not contain any information about technical relationships (which can only be derived by empirical investigation) but are only conclusions as to the nature of the combinations of factors and products that will fit in with certain assumptions we make, such as that an optimum distribution of factors is attained or that managers of firms maximize their

profits and that they do so under conditions of perfect competition. We can therefore discard our suspicions of black magic when we remember that the laws do not state any facts about the actual world supposedly achieved by intuition, but merely say that if certain conditions are fulfilled these other conditions will be found to be fulfilled too. This is a useful kind of proposition even though it can be developed in an armchair. (The question which kind of proposition is more useful is, of course, the age-old and rather senseless dispute of deduction *vs.* induction. Naturally we do not wish to touch this but will only remark that deductive armchair analysis can be of practical use only if the assumptions with which it starts have some relation to the conditions or to the objectives of the actual world.)

Constant or increasing *mp* is normally eliminated by the Rule and always eliminated by perfect competition, but it may persist under monopoly.

We can prove that there must be diminishing *mp* by showing that we must reject the hypotheses that there is increasing *mp* or that there is constant *mp*. The only possibility that remains is diminishing *mp*. There cannot be increasing *mp* in the optimum position because this would mean that the Rule has not been carried out. If it was in conformity with the Rule to use the last unit of the factor with increasing *mp*, it must also be in accordance with the Rule to use an additional unit since its *vmp* will be greater than that of the last unit actually used. Either the *vmp* was then less than *pf* so that the Rule would not have permitted its use, or else the *vmp* of an additional unit is greater than *pf* so that the Rule is being violated by more of the factor not being used. Therefore the complete application of the Rule eliminates the possibility of increasing *mp*.

There cannot be constant *mp* for the same reason. If it was in accordance with the Rule to use the last unit of the factor it must also be in accordance with the Rule for another unit to be used. Its *vmp* will be as great as for the previous unit and *pf* will be the same too, so if *vmp* exceed *pf* in the one case it will do so for the additional unit too. The complete application of the Rule

eliminates the possibility of constant *mp* too, and there must always be diminishing *mp* if the Rule is being properly applied.

The possibility that the price of the product might fall because more of it was being produced might seem able to upset this argument. The price of the product might fall more than *mp* increased so that *vmp* could fall sufficiently to make it go below *pf*. It would then be in accordance with the Rule to call a halt to the use of any more of the factor even though there might be constant or increasing *mp*.

This, however, is no real exception. It would then be proper for the manager of production to keep on substituting this factor for others without increasing the amount of product. In other words, he should use less of the other factors and more of this one, changing the proportions between the factors until a new proportion was reached where there is diminishing *mp* to all the factors. There would have to be diminishing *mp* to the other factors because increasing *mp* to all the factors would mean increasing *mp* to all of them together. This is impossible, for if all of the factors are increased together we have only a change in scale and this cannot change the *mp* of any of the factors. If there were increasing *mp* to any factor the application of the Rule must lead to substitution of this factor for the others until either the increasing *mp* had disappeared or this factor had completely supplanted all the others. There could then be no further change in proportions between the factors because there would then be only one factor left. There would be constant *mp* to increments in the amount of this factor since this would involve no change in proportions but only a change in the scale of production.

The principle of diminishing *mp* is therefore not so strong as the law of diminishing returns.

There is an exception to the principle of diminishing *mp* where the price of one factor of production rises rapidly enough compared with the prices of the others as more of it is used by the manager. This might have the effect of bringing *pf* above *vmp* even though there was increasing *mp*. The tendency to substitute this factor

for the other factors might be more than offset by its rising price. The same thing is possible on the product side when the price of one of several products falls so rapidly compared to the price of the other products that it might have diminishing *mf* (which is the counterpart on the product side of increasing *mp* on the factor side) and yet its output would not be increased at the expense of the other products because the rate at which its price fell with increasing output would more than offset the effects of the diminishing *mf*.

We know that the principle of diminishing *mp* (and of increasing *mf*) must hold in perfect competition, for a manager seeking to maximize his profits is forced to behave as if he were following the Rule, and the prices of the factors and of the products are independent of the output of the firm so that these exceptional cases cannot arise. But if there is not perfect competition the exception we have considered in the collectivist economy might be found in the private enterprise economy too.[1]

The law of diminishing *average* product (the law of diminishing returns) has a wider basis than this. It survives any degree of monopoly and can be upset only by indivisibilities (which we are still leaving out of account).

Diminishing returns follows from the necessity that every cooperating employed factor have a positive marginal product.

Let us look more carefully at the law of diminishing average product to see why it is so strongly established. We can do this best by experimentally assuming increasing (and constant) average product and seeing where this leads us. We again suppose factors *A* and *B* are combined in the production of product *P*. Increasing

[1] There is, however, an important difference. Increasing *mp* (or diminishing *mf*) in the collectivist economy does not imply any deviation from the optimum use of resources as long as the Rule is being followed. In exceptions under imperfectly competitive private enterprise, there is involved a departure from the optimum use of resources because the entrepreneur will be considering, not the rising price of the factor or the falling price of the product, but the marginal cost to him of the factor or the marginal revenue to him from increasing the output of the product. As we have seen (p. 134), this leads to a divergence from both the optimum division of resources among different products and the optimum allocation of the factors among the different firms.

average product of A would mean that an increase in A, with B unchanged, would increase P in a greater proportion than that in which A is increased so that P/A, A's average product, is increased. We know that increasing B in the same proportion as A would have increased P in this same proportion too. This means that when both factors are increased in the same proportion in which only A is increased, P increases *less* than when only A is increased. An increase of 1 per cent in both A and B increases P by exactly 1 per cent (because there are constant returns to scale), yet a 1 per cent increase in A *alone* increases P by more than 1 per cent (if there is increasing average product to A). This in turn means that the increase in B has the net effect of *decreasing* the total product. *B's marginal product is negative.* In general we can say that *if any factor has increasing average product the other factor or factors must have negative marginal products.*

Now no one who is seeking profit would keep on employing, let alone paying for, a factor that actually *reduces* the total product, least of all the monopolist to whom the marginal cost of the factor is in excess of its price. Nor will the employment of such a factor be continued in a collectivist society under the Rule, for that instructs each manager to cut down on the employment of any factor whose price exceeds the value of its marginal product even if the marginal product (and its value) is positive.

That increasing returns to any factor implies a negative marginal product of the collaborating factors can be shown in another way. The effect of the increase in A on its average product cannot be due to the change in *scale* (for that does not affect the proportion between factors and products) but is entirely due to the change in the *proportion* between the factors. This means that if the same change in the proportion between the factors were brought about in any other way the average product of A would increase just as much even though the scale of operations might be different. The change in the proportion of A to B (which we originally supposed to be brought about by increasing A while keeping B constant) could just as well be brought about by substituting some of A for some of B or simply by decreasing B while keeping A constant.

Suppose this latter is done. A is kept constant and B is dimin-

ished in quantity. There is the same increase in the proportion of A to B as when A is increased and B is kept constant so that there is the same increase in A's average product P/A. The increase in P/A while A is kept constant must mean that there is an increase in the total product P. *This increase in P is brought about by simply decreasing the amount of B employed.* B's *mp* must therefore be negative. (If a small decrease in B will increase the product, a small increase in B would decrease it.)

Since nobody will want to employ a factor whose *mp* is negative, the ratio of B to A will be diminished (and so the ratio of A to B increased) not only until the *mp* of B has ceased to be negative but until it has risen sufficiently above zero to make the value of its (positive) marginal product equal to its price (in a controlled economy whether the production unit is collectivized or in perfect competition) or to its *mc* to the firm (in a capitalist economy). Where this is the case B's *mp* will no longer be negative, and A will no longer show increasing returns.

By similar reasoning it can be shown that *constant* returns to A (constant *ap*) goes with a *zero mp* of B and is ruled out the same way. Constant returns to A would mean that a 1 per cent increase in A would bring about a 1 per cent increase in P whether B were kept constant or whether it were increased. The ratio of B to A then would make no difference to P/A so that varying B while A is kept constant would not change P. B's *mp* equals zero. As long as B's *mp* must be positive (if the Rule is kept or if profit is maximized) it cannot be *zero* or *negative*, so A's *ap* cannot be constant or increasing but must be diminishing in accordance with the law of diminishing returns.

We may summarize the relationship we have found to hold between the average product of one factor and the marginal product of its fellow factor (not to be confused with the relationship between the average and marginal measure of the same item discussed on p. 82).

ap of one factor	*mp* of the other factor
diminishing	positive
constant	zero
increasing	negative

It is easy to prove that this relationship is reversible.

mp of one factor	*ap* of the other factor
positive	diminishing
zero	constant
negative	increasing

The significance of this is clarified and a further step is made toward explaining the strength of the law of diminishing returns when it is shown to follow from the elasticity of substitution between different factors (by definition) being less than infinite (or else the two factors would be economically indistinguishable). This means that production cannot be carried on by one factor alone. A minimum proportion of some other factor is needed before any production at all is possible, and beyond this there is another higher minimum that must be reached before production can be *economically* undertaken.

There is a middle range of factor proportions where no factor is in *absolute* excess and where diminishing returns are universal.

When the proportion of factor B to factor A is so small that no production can take place, production can be made possible by either increasing the amount of B or decreasing the amount of A, as either procedure tends to rectify the proportion. When the disproportion is not great enough to prevent production altogether, it is still possible for the ratio of A to B to be so great that a decrease in A would permit the total product to *increase*. As long as this is true, A's *mp* is negative and B shows increasing returns. Just as A's negative *mp* is an indication that the ratio of A (to B) is too great, so the increasing returns to B indicate that the ratio of B (to A) is too small. At the other extreme where the ratio of B to A is too great B's *mp* will be negative and A will have increasing returns because its ratio to B is too small. Between these extremes is a middle range, bounded by the ratios at which the relatively plentiful factor stops actually diminishing the product but has not yet begun to increase it. Here $mp = 0$ and the relatively scarce factor shows constant returns. At one end of

the middle range *A*'s $mp = 0$ and *B* shows constant returns. At the other end, *B*'s $mp = 0$ and *A* shows constant returns. Within this middle range neither factor is applied in proportions that are *absolutely* too great or too small, but the ratio of one factor to another might still be *relatively* too great or too small—relatively, that is, to the prices of the factors. But even before we know the prices of the factors we know that the appropriate ratio between them must lie in the middle range. For since the prices are always positive (not counting free factors whose price is zero and which can always be disregarded), the *mp's* must be positive. This limits us to the middle range where diminishing returns reign over all factors.

The three ranges are indicated in Figure 2, p. 148. To the left of the diagram the *M* curve is perpendicular. This marks one limit to the middle range, α. To the left of α A/B, the ratio of *A* to *B*, is absolutely too small, and the ratio of *B* to *A* is absolutely too great. At α there is so much of *B* that its $mp = 0$. (To the left of this it would be negative.) No amount of *B* is large enough to make up for any decrease in *A*, and *A*'s *M* for *B* is infinitely large. The other limit to the middle range is β, where *M* touches the horizontal axis. This marks the value of A/B where there is so much of *A* compared to *B* that its $mp = 0$ and it is not substitutable for any amount of *B* at all. *A*'s *M* for *B* equals zero and *B* shows constant returns. Between these two points is the middle range with *A*'s *M* for *B* positive and less than infinite so that both *mp*'s are positive and both factors are subject to diminishing returns.

The case of fixed proportion we have seen is indicated by an *M* curve that is perpendicular throughout. This means that the middle range is narrowed to a single point—that of the fixed proportion. All other proportions would show constant or increasing returns and so are uneconomical.

The other extreme case is where the elasticity of substitution is infinite for all proportions between the factors (so that the two factors are economically identical). This is represented by a horizontal *M* curve indicating that the limits to the middle range are removed altogether. No ratio is too extreme and there are always

diminishing returns and a positive *mp*. If either factor is increased, the product will increase and it will increase in the proportion in which *the sum of both factors* is increased (since they are economically one factor and we have constant returns to scale). This means that the product increases in a smaller proportion than the one factor and we have diminishing returns.

The popular argument for diminishing returns is inadequate for several reasons.

The law of diminishing returns is sometimes supported by the argument that if it were not for diminished returns in the cultivation of land, it would be possible and economical to raise food for the whole world from a single flower pot of earth by applying enough labor and capital to its cultivation.

The argument appears much simpler than the analysis given above, but is misleading in several ways. First it would indicate constant or increasing returns not to the land in the flower pot but to the factors applied to it, for it is these that are increased and the returns refer to the result of increasing a factor.

Second, it might be possible to perform this feat with the flower pot if there were diminishing returns to the increased factors as long as the returns did not diminish more rapidly than they do on the more orthodox agricultural procedure!

Third, the illustration shows that there is a *limit* to the total product obtainable from a limited amount of land by increasing the cooperating factors, and from this it can be deduced that the average product of the increased factors must at some point begin to diminish, but it does not show that the proportions chosen for actual production will always show diminishing returns.

Fourth, it does not show that there is always an *initial* range for these factors where there are *increasing* returns because the ratio of the added factors is absolutely too small just as there is a *final* range where it is absolutely too large, its *mp* being zero or negative.

Fifth, it does not show that diminishing returns is a symmetrical affair as between the factors; that each factor would show increas-

ing or diminishing returns depending on its proportion to the other factor or factors.

Wise production does not avoid diminishing returns, it avoids *increasing* returns.

Sixth, and finally, it nearly always gives students the completely false idea that wise husbandry, in not trying to grow the world's food in a flower pot, avoids diminishing returns whenever it can. This is the exact opposite of the truth. Wise production avoids *increasing returns* wherever it can because this is a sign that the factor which yields it is absolutely too small in proportion to the other factors. Usually more of this factor will be applied as soon as this is discovered, because of the increasing average product that this would yield. Increasing only this one factor by 10 per cent will actually lead to more than 10 per cent increase in the product of all the cooperating factors. This would be continued to the point where the returns begin to diminish and beyond to the point which corresponds to the optimum combination of factors of which maximizes the profits of the firm. We have seen that the chosen point will then be well within the middle range where there are diminishing returns to all factors.

Even if it is impossible to obtain more of the factor that yields increasing returns, or if for any reason it is not desired to increase the product, the condition of increasing returns will not be allowed to persist. The proportions between the factors would be rectified by *diminishing the amount of the other factors* (*whose mp must be negative*). This would have the effect of *increasing* the total product. If that were not desired, some of both factors could be withdrawn from the use in question and made available for other purposes. The increasing returns would thus be eliminated together with the wasteful use of resources that the situation indicates.

The corresponding law for products is the law of diminishing *af*.

Once more what we have said about the relationship between factors has its counterpart in the relationship between products.

When more than one product can be turned out in the productive process and the proportion between them is variable, economical production decrees that an increase in the output of any product, the other products being kept constant, shall involve a less than proportionate increase in the factors applied. This can be expressed as a law of diminishing *af* (average amount of factor per unit of the product). Increasing *af* would mean that, in order to increase the output of this product by 1 per cent while leaving the output of the other products unchanged, the factors of production would have to be increased by more than 1 per cent. But it is possible, because of constant returns to scale, to increase *all* the products by 1 per cent while increasing the factors by exactly 1 per cent. This means that it is only in order to prevent the output of the other products from increasing that the factors must be increased by more than 1 per cent. The increase in the quantity of the factors above the 1 per cent has a negative marginal product of the other products, and so the *mf* of the other products, like the *mp* of the factors in producing the other products, is negative. Such a state of affairs indicates an uneconomic combination of factors.

It is possible to increase the output of the other products by merely decreasing the quantity of the factors. The output of the other products with a negative *mf* should be increased until *vmf* has risen to equal the price of these products. Then the *mf* of other products will no longer be negative, and there will no longer be increasing *af* for the first product.

Even if there is no use for additional output of the other products it would be advisable to expand their production (and throw away the excess) until their *mf* had risen from the negative value to zero, for this would permit an increase in the original product without any increase in the resources applied. Even if for any reason an increase in this product were not desired either, it would be possible to set free some factors of production for use elsewhere while producing the original amount of the first product and additional amounts of the others.

Increasing *af* is as clear a sign of wasteful production as increasing *ap*, and it too should be eliminated wherever possible. In Chapter 15 we shall come upon some instances where their

elimination is technically impossible, and only then can we economically have increasing *af* or increasing *ap*.

It seems a little strange that the counterpart of diminishing *ap* from the point of view of the product should be *diminishing* and not increasing *af*. It will be remembered that the counterpart of diminishing *mp* was *increasing mf*, and it seems surprising that there should not be the same change in sign in the case of diminishing (average) returns. The conditions of economic production, as seen from the point of view of the product, are then increasing *mf* but diminishing *af*; *mf* moves in the direction opposite to *mp* because diminishing transformability means that smaller and smaller increments of product result from equal increments in a particular factor and that larger and larger increments of factor are needed to bring about equal increments in a particular product; *af* moves in the same direction as *ap* because increasing one of several factors will increase the product less than if *all* the factors were increased in this proportion and an increase in one of several products necessitates a smaller increase in the quantity of factors than would be needed in order to increase *all* the products in this proportion. All the factors and all the products would be increased in the same proportion by merely changing the scale of production with all items increasing in exactly the same proportion. An increase in only one of several factors can therefore result in a less than proportionate increase in the total product, and an increase in only one of several products necessitates a less than proportionate increase in the factors of production.

CHAPTER 14. COST OF PRODUCTION

Cost and returns are not the simple inverse of each other. Cost refers to the unit of product, and returns refer to the unit of factor.

Returns and cost are sometimes spoken of as if they were the reverse of each other. That is, as if increasing returns meant the same as decreasing cost, diminishing returns meant the same as increasing cost, and constant returns meant the same as constant cost. This is not necessarily so on any clear meaning of the terms and certainly not on the meanings we have here given to returns. We have said hardly anything so far about cost.

Most of the time returns and cost are not directly comparable at all even if they are both applied to averages, which is what we shall assume here. Returns or average product, *ap*, means the total product, P, divided by the number of units of a factor, A. Whether returns increase, decrease, or stay constant is measured by what happens to the ratio P/A when the factor A is increased while the other factors, B, C, etc., are kept constant. Cost refers to the money outlay, O, when a product, X, is produced. If there are other products, Y, the revenue from the sale of these, r_y, is subtracted from the total outlay on all the factors to give the total cost of the product X. This is then divided by the number of units of X to give the expression $\frac{O-r_y}{X}$ which is the average cost (*ac*) of X. Whether we have increasing, decreasing, or constant cost is measured by what happens to $\frac{O-r_y}{X}$ (or *ac*) when the output of X is increased. Obviously there is no simple relationship in general between *ap* and *ac*.

The simple inverse relationship does hold in one situation. This is when only one factor is used to produce only one product and where the price of the factor is constant. O is then equal to $A(pf)$ (the quantity of the factor A multiplied by its price), and X is the same as P, so the two expressions for ap and ac reduce to X/A and $A(pf)/X$. Since pf is a constant these two measures must vary in inverse proportion. Any increase in A will result in an increase in X. If there are increasing returns, X/A increases as A increases and $A(pf)/X$ decreases in the same proportion, so we have decreasing cost. In the same way constant returns will be accompanied by constant cost, and diminishing returns will be accompanied by increasing cost.

Unfortunately this is not as interesting as it sounds, and this is not only because instances of a single factor producing a single product are rare. If only one factor is used to produce only one product, there must always be constant returns. There are no proportions to change, so only the scale of operations is involved. This by itself never changes the proportion between the product and the factor (X/A), so ap is unaffected. What happens to cost is then completely determined by what happens to the price of the factor as the product increases. If the price of the factor, pf, remains the same, there is constant cost. If pf rises, there is increasing cost. If pf falls, there is decreasing cost.

It is best therefore to keep costs and returns clearly apart by always adding mentally after "returns" the words "per unit of a particular factor" and after "cost" the words "per unit of a particular product." This will be a valuable safeguard against many confusions.

On our present assumptions we would have constant cost and diminishing returns throughout the economy.

On our present assumptions, with divisibility of factors, products, and productive methods and the application of the Rule in a collectivist economy, there would be diminishing (average) returns to every factor and constant (average) cost of every product in every production unit that was not large enough to influence

the price of the factors it bought or the products it sold. We have sufficiently explained the necessity for diminishing returns in the previous chapter. Constant cost results from the possibility of expanding or contracting the scale of production without changing the proportions between factors and products. If any other products are being produced besides the one we are considering, their output would be changed in the same proportion too, and since they continue to be sold at the same price everything is the same except for the scale, and the average cost is unaffected.

This makes the sum of the marginal products exhaust the whole product.

Since the average cost is constant it will be equal to the marginal cost, and since the marginal cost equals the price (whether or not this is the form in which the Rule is given to the managers) the average cost will equal the price of the product. Consequently the total cost, or the amount paid out for the factors of production, will equal the total amount received from the sale of the output. In the collectivist economy as well as in the collectivist section of a controlled economy, this is of no importance and plays no part in the Rule. The relationship between value of sales and outlay on factors might be used in tests of the efficiency of firms, but there is nothing automatically significant about it. We take note of it here only because it is relevant for the capitalist economy, as we shall see in the next chapter, and it is curious that this should come about where there does not seem to be any rule directed toward that end.

The reason is to be found in the phenomenon of constant returns to scale. If all the factors of production are increased by, say, 1 per cent, the increase in total product will also be 1 per cent. This same 1 per cent increase would be obtained if the factors were increased not simultaneously but one at a time. At each step the product will increase by an amount which is the *mp* of the 1 per cent increase in the particular factor, so that the sum of the *mp's* of 1 per cent increases in all the factors equals 1 per cent of the total product. Since the price of each factor equals the value of its marginal product (by the Rule), the amount paid to

the additional 1 per cent of factors equals the 1 per cent of the value of the total product. At the same rate of pay the amount paid to 100 per cent of the factors equals exactly 100 per cent of the value of the product.

Constant returns to scale therefore result in constant average and marginal cost, both equal to the price of the product. This makes the size of the production unit indeterminate and renders perfect competition in the capitalist economy technically possible but unstable.

From the point of view of the industry there would be increasing cost, because increased output of the industry makes some of the factors more scarce and raises their prices.

From the point of view of an industry, however, there will normally be not constant but *increasing* cost. This is because when we consider the output of a whole industry we cannot neglect the effect of an increased demand on the prices of factors. As the output of the product increases the demand for the factors used in its production will increase, and this will tend to raise their prices. If the prices of all the factors rise in the same proportion the cost will also rise in the same proportion. It is more likely, however, that some prices will rise less than others. In this case the cost will rise in a proportion somewhere between the largest and smallest factor price increase. The substitution of the factors that rise less in price (or do not rise at all) for the factors whose prices rise more will help to prevent cost from rising as much as if this substitution did not take place, but if any factor rises while none falls in price this substitution cannot completely prevent the rise in cost—unless the elasticity of substitution is infinite so that the dearer factor can be completely replaced, without running into any diminishing M, by other factors whose price does not rise even though they now have to produce the whole output themselves.

The rise in price of factors in response to an increase in the amount bought is measured by the elasticity of supply.

The rise in price of any factor when the demand for it increases can be expressed in terms of the elasticity of supply. This shows

the increase in price as dependent on the increase in the amount of the factor absorbed in the particular industry we are considering. The exact measure is given by the proportional increase in the amount absorbed divided by the proportional increase in the price needed to induce this additional supply. It is measured in terms of *proportional* instead of absolute change to avoid the arbitrary influence of the unit in which the elements (here quantity and price are measured. (This is true of all elasticities in economics; elasticity of substitution, elasticity of demand, elasticity of supply, etc.)

If a very great increase in the supply is brought about by a small increase in price, the supply is said to be elastic. If no increase in price is needed to call forth the increased supply, the elasticity of supply is said to be infinite, or the supply is said to be infinitely elastic. If the response in supply is small in relation to the increase in price, the supply is said to be relatively inelastic. If there is no increase at all in supply in response to an increase in price, the supply is said to be absolutely inelastic. The measure, of course, can also be applied in the reverse direction, measuring the degree to which the amount supplied decreases in response to a decrease in price.

The influences on the elasticity of supply are extremely complex.

The extent to which a product is subject to increasing cost will therefore depend on the elasticity of supply of the factors (which determines how much their prices have to rise to bring forth any particular increase in their supply) and upon the elasticities of substitution between the factors (which determines the degree to which the effect of rising prices of factors on cost may be mitigated by substituting factors that have become relatively cheaper for factors that have become relatively dearer). The elasticity of supply of a factor will depend upon many different things, but these can be classified into three groups of influences.

First is the elasticity of *total* supply of the factor, that is, the effect of an increase in price on the amount available for *all* purposes. Sometimes this is zero, as in the case of land when the total supply is fixed. Sometimes it is fairly high, as in the case

of labor when the wage is raised from a level at which most of the suppliers would rather go fishing to one at which it is worth while going to work to get the high pay—or as in the case of a large stock of a durable good, with a definite expected future price, when the price is raised just above the margin which makes available now a great deal of what was stored for the future. (In both these cases the alternatives of leisure or of future use of the stock of goods were left out of the *total* supply. If they were included the elasticity of supply would be zero.) Sometimes the elasticity of the total supply will be negative—as when an increase in pay for labor is directed by the laborer, in part at least, to the enjoyment of more leisure. (If we include leisure as one of the uses of labor time this would again make the total supply fixed and its elasticity equal to zero.)

Second is the elasticity of substitution of the factor for the other factors with which it is combined in other uses. If this elasticity of substitution is high, a slight increase in the price of the factor will lead to a great deal of substitution of other factors in the other uses so that a great deal of it would be set free and become available for use here. This would make the elasticity of supply great. If the elasticity of substitution in the other uses of the factor is low, very little of the factor will be replaced when its price rises, very little would be set free, and so this would contribute very little to the elasticity of supply. In any case this contribution to the elasticity of supply of the factor would disappear if the same increase in demand that raised the price of this factor also raised the prices of the cooperating factors in other uses! There would then be no substitution of other factors for this factor and none of it would be set free even if the elasticity of substitution were very great.

Third is the elasticity of demand for the factor's alternative products. If this elasticity of demand is large (that is, if the proportional decrease in the demand for them is large as compared to the proportional increase in their price) the factor will be set free in large amounts (as will the other factors used together with it in the production of the alternative products) and this will contribute to increase the elasticity of supply of the factor.

These three items, total elasticity of supply, elasticity of substitution in other uses, and elasticity of demand for the factor's alternative products, do not tell the whole story of the influences on the elasticity of supply of a factor of production. It will also depend on the elasticities of supply of the cooperating factors in the alternative uses, which in turn depend on the same three items and on the elasticity of supply of *their* cooperating factors, and so on indefinitely. In the same way the elasticity of demand for the alternative product depends on what happens to the prices of the substitutes for it, and this depends on the elasticities of supply of these substitutes and of the factors used in making them as well as on the elasticity of substitution of these substitutes for the commodity originally examined. This is the way in which an increase in the output of a particular product has its repercussions all over the economy, and this is why any attempt to run the economy from a central office must result in utter confusion although it can all be adjusted satisfactorily with the proper use of the price mechanism.

Increasing cost from the point of view of an industry must be distinguished from increasing cost from the point of view of society.

All these influences can be reduced to two elements. There is the *technical* element of the ease or difficulty of shifting resources from producing one thing to producing another, and there is the *economic* element of the ease or difficulty of inducing consumers, by raising (cost and) price against them, to give up the alternative products and set free the resources needed to expand the output under consideration. Both elements contribute to the increasing cost. With this in mind we can distinguish between two kinds of increasing cost:

First, there is the degree to which the cost of production would increase if the authorities increased the output in the absence of any change in demand for the alternative goods. Second, there is the degree to which cost would rise in response to a shift in demand when consumers decided to change the distribution of their expenditures consuming more of one good and less of some other

good. The first we may call increasing cost from the point of view of the industry; the second we may call increasing cost from the point of view of society.

Elasticity of supply will be less from the point of view of an industry than from the point of view of society because the former reflects psychological in addition to technical resistances.

From the point of view of the industry cost will be increasing for both the technical and the economic reasons. The *technical* element will reflect the increasing *amounts* of the alternative products that have to be sacrificed to permit the production of constant increments of the product in question. The *economic* element will reflect the increasing *valuation* by consumers of successive equal physical sacrifices of the scarce alternative products.

Increasing cost from the point of view of society will consist only of the technical element because the increase in demand for product X is accompanied by a corresponding decrease in demand for product Y. The smaller output of Y when its resources are shifted to the production of more of X therefore does not cause the price of Y to rise. On the contrary it is almost certain to fall. If exactly the same factors are used in producing X and Y, there is no technical resistance to the shift in production, the price of X will not rise (there will be constant and not increasing cost from the point of view of society), and the price of Y will not fall. In the absence of such a coincidence the factors cheapened by the decreased demand for product Y will not all be directed toward keeping down the price of X and so the price of X will rise in relation to the price of Y. This change in relative cost is the measure of the technical resistance to shifting resources from the production of Y to the production of X. When taken as a *proportional* change and compared with the proportional change in the ratio between their outputs, this gives us the *social* elasticity of substitution between X and Y.[1] The definition of elasticity of

[1] Since there is the same proportional change in the M of X for Y as in the M of Y for X, and the same proportional change in the ratio of X to Y as in the ratio of Y to X, the elasticity of substitution is perfectly symmetrical as between X and Y,

substitution is the same as that used in the case of *technical* elasticity of substitution. It is still the proportional change in the ratio between two quantities divided by the proportional change in their M. The M is measured by the ratio between their prices (since we are assuming there is an optimum allocation of the products among the consumers).

The concept of elasticity of substitution is applicable to the indirect technical substitution of one product for another in reallocating society's resources.

The substitution is not within the production unit, but is achieved indirectly via the shifting of factors of production from one production unit to another. It may even be that the factors released from Y were not suitable at all for the production of X but were used elsewhere to release factors for the production of X, and the reshuffling of factors may be very complex, involving many such steps. But however complex or indirect the reshuffling, it comes to the same thing. If the Rule is maintained all the time (or if the optimum division of resources and the optimum allocation of factors are maintained by perfect competition) the result will measure the diminishing M of X for Y.

Throughout all such transformations, with prices of factors and products changing as a result of a shift of demand from one product to another, each production unit (which is unable to affect prices) is operating under conditions of both constant returns to scale and constant cost. From the point of view of the industry there is increasing cost for both the technical and the economic reasons. Increasing average cost to the industry means that marginal cost to the industry is greater than average cost, but marginal cost to the industry is not of social significance. It is of interest only to an industry that is monopolized by one firm and restricts pro-

and we do not have to distinguish between the elasticity of substitution of X for Y and the elasticity of substitution of Y for X. Both elasticities have the same value under all circumstances and it is sufficient to speak of the elasticity of substitution between X and Y. See A. P. Lerner "Notes on the Elasticity of Substitution," *Review of Economic Studies*, February 1936.

duction so as to maximize its profits at the expense of the rest of the economy. From the point of view of society the marginal cost is equal to the average cost to the industry, which is also the average and marginal cost of each production unit, because that measures the value of the alternative *msb*.

CHAPTER 15. INDIVISIBILITIES I

Indivisibility may be found in the factor, in the product, or in the method of production.

We may now give up our assumption of perfect divisibility and see how our conclusions are affected when we consider that factors are often available only in large units like waterways, that products are often produced in naturally large units like ocean liners or skyscrapers, and that methods of production are also often of a minimum size even if the factors and the products are fairly divisible, like an assembly plant for automobiles or a continuous strip-steel rolling mill. The existence of any one of these indivisibilities makes it impossible for the same factors to be combined in the same way to make the same product on any scale that might be chosen. The minimum scale is that on which there is a whole unit or several whole units of each of the indivisible items involved, and even production on a larger scale with the same proportions between factors and products can take place only in multiples of this minimum. It is no more possible to produce $1\frac{3}{8}$ ocean liners by increasing all the factors by $\frac{3}{8}$ than it is possible to produce $\frac{1}{100}$ of an ocean liner by using $\frac{1}{100}$ of the quantity of all the factors and to sell the product for $\frac{1}{100}$ of the price of a whole one; or to produce 10 automobiles a year in the same way and at the same cost per unit as when 1,000,000 automobiles are made in a year.

The examples we have used, which are of very large indivisibilities, should not lead us to believe that our analysis in terms of perfect divisibility is never applicable directly to real problems. It is true that nothing is really perfectly divisible. An ear of corn is an indivisible unit, and even the electrons of which everything is composed appear to be discrete and indivisible entities. But in

the many branches of production where the indivisible units are small enough not to be important the analysis in terms of perfect divisibility can be applied directly.

It can upset the law of diminishing returns by limiting the adjustability of factor proportions.

The existence of indivisibilities upsets the applicability of the principle of constant returns to scale because it makes it impossible for the scale to be changed without changing any of the proportions between factors and products. Consequently the principles that are derived from the principle of constant returns to scale may also be inapplicable where there are indivisibilities. This is true for the law of diminishing returns.

Where one of the factors of production is indivisible, an increase or decrease in output less than sufficient to warrant an increase or decrease of the indivisible factor by a whole unit has to be brought about by varying the quantity of the other factors that are (relatively) divisible. The result is a proportion between the factors different from that which would be chosen if there were perfect divisibility. If this divergence were so great as to make the *mp* of the fixed factor negative, we should have increasing returns to the other factors.

With increasing returns it will always pay the firm (in perfect competition) to expand or to close down.

If the *mp* of factor *A*, an indivisible factor, were negative (in which case there would be increasing returns to factor *B*), the manager of the firm would do something about it. If there were perfect competition it would always pay the firm either to close down or to expand. If the firm were covering its costs, so that it did not pay it to close down, it would be able to profit by employing more of factor *B*. An increase of 1 per cent in *B* would increase the costs of the firm by less than 1 per cent, since the outlay on *B* is only a *part* of the firm's costs, while it would increase output by more than 1 per cent on account of the increasing

returns. This means a clear increase in profit. If the firm expanded in this way until the increased proportion of *B* to *A* had eliminated the increasing returns and reduced *B's vmp* to equality with its price while perfect competition still reigned in spite of the increased size of the firm, then the indivisibility of *A* would not be significant and everything would be just as if the factor were perfectly divisible.

Indivisibility is significant when it is large enough to destroy perfect competition through the expansion of the firm.

The indivisibility becomes significant when it is large enough to destroy perfect competition. To overcome the indivisibility the output of the firm might have to be so large that only one firm, or perhaps a small number of firms could satisfy the demand (at the prices which make $vmp = pf$ and so give the optimum use of resources under perfect competition). The firms, which are naturally interested in maximizing their profit, would no longer find it the most profitable thing to make $vmp = pf$ because no longer would $vmp = mpr$, and pf might cease to equal mpc; and it is only mpr and mpc that the firm consciously seeks to equate in endeavoring to maximize its profits.

We see then that indivisibility leads to an expansion in the output of the firm, and this either makes the output big enough to render the indivisibility insignificant, or it destroys the perfection of competition. Significant indivisibility destroys perfect competition.

In a collectivist economy increasing returns also leads to expansion if vmp is greater than pf. Expansion comes to a stop when these are equated. If the increasing returns are eliminated before $vmp = pf$, the situation may be just as in perfect competition. The output may be sufficiently large to render indivisibility insignificant. But if the indivisibility is significant and vmp falls to the level of pf before the increasing returns are destroyed, we have the interesting conclusion that *the productive unit must be run at a loss.*

This is because, with increasing returns (which means in-

creasing average product), the marginal product must be greater than the average product—otherwise it would not be raising the average product (see p. 82)—and *vmp* would be greater than *vap*, the value of the average product. Now the total amount paid out to factor *B* would equal *vmp* multiplied by the number of units of factor *B* employed, while the value of the total product would be *vap* multiplied by the same number. This means that factor *B* alone must be paid more than the value of the total product and any cost of the indivisible factor *A* is simply additional loss.

With significant indivisibility, perfect competition, or the application of the Rule, must result in the firm's running at a loss, so that the optimum use of resources is possible only in a collectivist or subsidized agency.

This may appear a little less surprising if we observe that it is simply another aspect of the argument used to show that if there is increasing returns and pure competition it will pay the firm either to close down or to expand. If the optimum use of resources has been reached and $vmp = pf$, then (under perfect competition) $mpr = mpc$, profits are at a maximum (or losses at a minimum), and it does not pay the firm to expand. The other alternative must therefore be true—it pays the firm to close down! Under these conditions firms under perfect competition would close down. Again we see how significant indivisibility destroys perfect competition. Where there is significant indivisibility it is only under the collectivist organization and the subsidized application of the Rule that the optimum use of resources is possible.

Increasing returns is only an extreme case of this.

It is important to note that increasing returns is only an extreme case of this kind of situation. Before $vmp = pf$, the increase in factor *B* (in proportion to the indivisible factor *A*) might bring the proportion between the factors out of the range of increasing returns into the middle range of diminishing returns, so that *vmp* was less than *vap* and the payment to factor *B* (which is equal to

B's *vmp* multiplied by the number of units of *B*) came to less than the value of the whole product (which is equal to *B*'s *vap* multiplied by the number of units of *B*); but this excess of the value of the total product over the payment to *B* might not be sufficient to cover the cost of factor *A* so that there would still be a loss at the (optimum) output determined by the Rule. This would also be the result of the indivisibility, for we have seen (Chapter 14, p. 165) that with perfect divisibility there are constant returns to scale and that this makes the payment to all the factors add up exactly to the value of the total product. What happens here is that the proportion of *A* to *B* is too great—not *absolutely* (as it is when returns to *B* are increasing) but *in relation to its price*. Its price is greater than its *vmp*, "too much" of it is used, and this excess is what causes the loss.

If the factor were divisible, this would be corrected by decreasing the quantity employed (relative to the factor *B*) until *A*'s *vmp* was equal to its price, but the indivisibility prevents that and so the optimum use of resources involves operation at a loss. Consequently the optimum use of resources cannot be reached by unsubsidised private enterprise but is no problem for the collectivist method. A productive unit does not then need to close down just because its total revenue from the sale of the product is less than the total payments to the factors employed. As long as the Rule is being observed everything is in order.

Whenever indivisibilities have this effect (making total revenue less than total outlay when each divisible factor's *vmp* equals its price) and at the same time there is perfect competition in buying factors, there will be decreasing average cost per unit of output (just as in the case of increasing returns with perfect competition). This again means that if there is also perfect competition in selling it will pay the firm either to close down or to expand.

If the average cost per unit of product is *greater* than the price of the product (which is the average revenue) the firm is operating at a loss and it pays it to close down. If the average cost is *less* than the price of the product, an expansion of output would permit the optimum proportion between the factors to be reached or at least more closely approached. This will reduce the average cost

and, with the price of the product unchanged (if there is perfect competition in selling), profit will increase. Even if average cost is greater than price, so that the firm is operating at a loss, the decreasing average cost with expansion of output may still turn the loss into a profit. In that case the firm would do better to expand than to close down. The firm will always either close down or expand, and expansion will either render the indivisibility insignificant or destroy the perfection of competition. Even if the indivisibility does not result in so great an excess of the indivisible factor as to cause increasing returns to it, there will still be *relatively*, though not *absolutely*, too much of the indivisible factor, and this will cause decreasing costs. The indivisibility will destroy perfect competition and with it the possibility of bringing about the optimum use of resources by free enterprise for profit.

The same destruction of perfect competition is brought about by indivisibility of the product and by indivisibility in production. The latter can even be analyzed in exactly the same terms as the indivisibility of a factor, by simply calling the indivisible productive item, the conveyor belt or the rolling mill, a *factor* that is made out of the factors back of it. Products, as we have already seen, can be treated in the same way as factors if one is only careful about the change in sign which indicates that they are *outputs* of the productive unit instead of *inputs*. Whatever the cause of an indivisibility, the firm in seeking to remedy it will increase in size. If the indivisibility is economically significant, expansion sufficient to eliminate it will also destroy perfect competition.

Counterspeculation is not effective by itself against monopoly established by indivisibility.

We have already discussed the danger to perfect competition from combination or expansion when there are perfect divisibility and constant returns to scale. There we saw that perfect competition could sometimes be safeguarded by government counterspeculation. In dealing with the threat to perfect competition from indivisibilities, counterspeculation is of much less avail. The government could use this device to prevent monopolistic influence

over price, but that would merely result in all the firms going bankrupt. They would be forced to make $vmp = pf$ to maximize their profit, but then they, just like the collectivist undertaking that followed the Rule, would find that when they had reached this point and it no longer paid them to expand, it would pay them to close down because they would be making a loss. Their maximized profit would be negative. The reason for this is exactly the same as the reason for the collectivist agencies' running at a loss and need not be repeated. An attempt to maintain perfect competition by counterspeculation would lead to wholesale bankruptcy. Where there are political objections to collectivist undertakings to provide the service at a loss or to paying a subsidy to private enterprises to keep the services going, it is usually better to permit the existence of monopolies which can make ends meet, and depart to some extent from the optimum use of resources, than to depart still further from the optimum by destroying the industry altogether in the vain attempt, by counterspeculation, to maintain perfect competition where indivisibilities make it technically impossible.

Small indivisibilities, which can be large in relation to the market involved, may be more important than large ones.

These conclusions are of very great importance. They are applicable not only to very large indivisibilities but to relatively small ones which are significant because the market is also small. Perfect competition can then be destroyed by moderate expansion which is insufficient to render the indivisibility insignificant. The important thing is the relationship between the size of the indivisibility and the size of the market. The indivisibility can be measured by the output of the firm needed to make it insignificant: The market can be measured by the output at which the firm becomes aware of ability to influence price. At this point the marginal private revenue of the firm, mpr, falls below p, the price of the product.

If the size of the indivisibility is greater than the size of the market, perfect competition with free enterprise is doomed. The

choice is then between sacrificing the optimum use of resources, by permitting private enterprise to make ends meet (or perhaps even make large profits) by monopolistic restriction of output to raise the price, and maintaining the optimum use of resources by *subsidizing* either collectivist agencies which are forced to run at a loss in carrying out the Rule or private enterprises which are forced to run at a loss by the artificial maintenance of perfect, "cut-throat," competition.

If the size of the market is greater than the size of the indivisibility, so that the output of the firm is sufficient to reduce the indivisibility to insignificance, perfect competition is possible. The indivisibility, by increasing the size of the firm, will have reduced the number of firms in the same market and increased the probability of their combining for the purpose of forming a monopoly that would turn the zero profit into a handsome positive profit by restricting output, raising the price of the product, and perhaps also lowering the prices paid for the factors. But this danger *could* be met by counterspeculation or by government competition, for these measures could maintain perfect competition which would still be technically possible, just as in the complete absence of indivisibilities.

When recognized, indivisibility shows itself in the problem of the public utility, where unintelligent compromise leads to unending regulation.

Industries that are subject to large indivisibilities such as make perfect competition impossible have, by a curious history, come to be called *public utilities*. It has been recognized that perfect competition in these public utilities cannot be arranged or even permitted and that it would lead to bankruptcy and the cessation of important services to the public. Monopoly is therefore permitted, though public regulation is applied in attempts to limit the degree to which the monopolies depart from the optimum use of resources in their attempts to increase their profits. This compromise between public and private enterprise leads to unending regulations and attempts to evade the regulations and more regulations to stop the evasions. In the tremendous volume of writing

on this subject there is confusion which is not entirely unconnected with a natural tendency for the great public utility corporations to try to get the public to identify their unrestricted powers (to sacrifice the optimum use of resources in restricting output and raising prices for the sake of their profits) with the democratic liberties of the citizen. An equally fertile source of confusion is the identification of the elimination of great profits with the optimum use of resources. This identification is brought about by too close a concentration on perfect competition which happens to result in both the absence of great profits and the optimum use of resources. As we have seen, perfect competition cannot be brought about in the circumstances considered, and nothing can be gained by trying to achieve one of its symptoms through legislation aiming at the establishment of another.

In the regulation of public utilities in the United States we have a classical example of how the complexity of regulations in an uncontrolled economy enormously surpasses that needed in a controlled economy. In the controlled economy, public utilities, which by definition cannot be made subject to perfect competition, would be run by public agencies instructed by the Rule to make $vmp = pf$. They would normally be run at a loss, which is justifiable in the name of the optimum use of resources.

Small indivisibilities in the actual world are perhaps more important than big ones of the public utility type, if only because they are not as easily prevented by regulation from bringing about too great a deviation from the optimum use of resources. Competition is imperfect in many small markets because the customers of particular firms are to a greater or smaller degree attached to the firm so that they would not all desert it and patronize other firms instead at the slightest increase in the price that the firm charges. They may find the firm more conveniently located, or they may believe, rightly or wrongly, that the product of this firm is better or more suitable for their purpose than the product of its competitor, or they may like the proprietor or enjoy talking to the employees of the firm, or it may be merely habit. Whatever the reason, the attachment to the firm destroys perfect competition because the firm finds that it can raise the price without losing

all its customers and that a certain amount of price raising, with the consequent restriction in its output, will be profitable. No longer does $vmp = pf$, and we do not get an optimum use of society's resources.

In the absence of indivisibility, perfect competition would be *possible* everywhere. Freedom of entry, like government regulation, can prevent excessive profits but cannot prevent the waste of resources.

If there were perfect divisibility of all factors, any such specialization on minute differences in people's needs or desires would only be to the good. There would be many more firms and smaller firms, but each would be producing in the optimum manner, the factors it employs being combined in the optimum proportions. There would be constant returns to scale and perfect competition. But there are indivisibilities, and this leads to poor combinations of factors with too much of the indivisible factors compared to the others just as in monopolistic public utilities. Freedom of new firms to enter may prevent any abnormal profits from being made by any of the firms, just as regulation by the government may prevent public utilities from making exorbitant profits, but there is the same social loss because too much of the indivisible factor is combined with too little of the divisible factors and there are too many firms. The classical example of this is the familiar set of four filling stations at a street corner, where one would be able to provide all the service at a great saving in equipment as well as in the time of idle attendants waiting for customers. (The minimum number of attendants for a filling station constitutes a fixed or indivisible factor.)

These wastes because of *imperfection* of competition are frequently called the wastefulness of competition by ingenuous planners.

Just as in the case of the public utilities this cannot usually be put right by re-establishing perfect competition. All the firms might be bankrupted. Great economies could be attained by standardization, and it is this kind of economy that is in the mind

of simple-minded planners who have not realized the intricacies involved in organizing a modern economy, and it is this waste from the absence of perfect competition that is usually called the wastefulness of competition. How best to harmonize private enterprise with the avoidance of these wastes is a tricky business that is easily confused with the more fundamental and simpler problem of how best to satisfy small differences in the tastes of consumers while making full use of the economies of standardization. This latter problem is solved simply by keeping to the same old Rule.

The economies of standardization are also based on indivisibilities and are adequately encouraged by the Rule.

The economy of standardization is only another aspect of indivisibility. If there is perfect divisibility all these economies would be available with the smallest output. If the market for a product is large enough to permit a single firm supplying it to overcome all the indivisibilities and achieve an optimum proportion of factors in spite of them, the economies of standardization are fully achieved. It is only when the market is smaller than the indivisibility that an increase in output will reduce costs, so it would seem that consumers should be especially encouraged to use such a product. This is because by so doing they will be benefiting not only themselves but their neighbors. By making these purchases they are helping standardization to be developed and permitting the standardized goods to be obtained more cheaply by other purchasers.

All this is sufficiently and accurately taken care of by the application of the Rule. In all such cases we have decreasing costs and the marginal cost is less than the average cost. Making $vmp = pf$ also makes $p = mc$ (with the qualification mentioned in Chapter 9) so that price is less than average cost. The difference can be regarded as a subsidy to the consumer for just this purpose. However, it is probably better to say that the distinctive feature of this situation is that an additional unit of product can be produced by the addition of an abnormally small increment of factor,

or that an increment of factor has an abnormally large marginal product because it moves production toward the optimum combination. This is accurately reflected in the price of the product's being low enough to make the value of the abnormally large marginal product equal to the price of the factor. The consumer is given the whole benefit of the special situation, neither more nor less, by simply following the Rule which makes $vmp = pf$.

CHAPTER 16. INDIVISIBILITIES II
("*VIRTUAL MP*" *LARGE DECISIONS*)

What is meant by the marginal product of an indivisible factor.

We may now turn to a point that has rather been slurred over so far. What is meant by the *mp* of an indivisible factor and how do we apply the Rule to an indivisible factor? Since by definition it is not possible to vary this factor by a small amount and see the effect on the product, what is meant by saying that with increasing returns the *mp* of the indivisible factor is negative and that with decreasing costs the indivisible factor's *vmp* is less than its price? And if this should be so would not this mean that the Rule, which would make *vmp* equal to the price of every factor, was not being properly applied?

It must be confessed that the statements about the *mp* (and the *vmp*) of the indivisible factors were used in Chapter 15 in a special sense which might perhaps better be described as "virtual *mp*." It is what the *mp* would be if the indivisible factor were in fact divisible, and it is obtained by considering what the actual *mp* would be, per unit of the indivisible factor, if the same change in proportions took place on a scale large enough to make the indivisibility insignificant. For example, suppose the indivisible factor *A* to be available only in units of 100 tons, and that one unit of 100 tons of *A* is combined with 50 units of factor *B*, to produce 500 units of product *X*. There is increasing returns to factor *B*. With 51 units of *B*, total product would increase to 561 units of *X*. *B*'s *ap* is increasing (from 10 to 11) and *B*'s *mp* is 61, which is greater than its *ap*. *A*'s *mp* is the change in product that would be brought about if *A* could be varied by 1 ton. This is impossible because of *A*'s indivisibility, but the same thing can be re-created

in imagination on a larger scale where the indivisibilities would be insignificant. We know that

$$\text{and} \begin{matrix} 10{,}000\ A + 5000\ B \text{ would produce } 50{,}000\ X \\ 10{,}000\ A + 5100\ B \text{ would produce } 56{,}100\ X \end{matrix}$$

These figures are obtained by increasing the scale by 100, leaving the proportion between the factors and the product unchanged. If we now divide the second situation by 102 we get

$$98.04\ A + 50\ B \qquad \text{produce} \qquad 550\ X$$

so that a *decrease* of slightly less than two units of A would *increase* the product by 50 units, and the "virtual *mp*" of A is about *minus* 25 X.

This virtual *mp* is negative when B shows increasing returns, and its value (*vmp*) is less than the price of the factor when there is decreasing cost and when the application of the Rule to the amounts of the divisible factors employed makes the value of the total product less than the outlay on *all* the factors (under perfect competition).

The Rule must be applied to the indivisible block.

The managers, in actually adjusting factors, could use this virtual *mp* and the corresponding virtual *vmp* only if they could increase or decrease A by one small unit, and this is impossible because of the indivisibility. Consequently a virtual *vmp* not equal to *pf* does not mean that the Rule has not been applied. But how *can* the Rule be applied?

The answer is that the Rule must be applied *concretely* to the issue whether the indivisible factor A should be increased by a whole block of 100 tons or decreased by such a whole block (which might mean closing down production altogether). No other issue is involved because nothing else can be done.

If the value of the extra product from adding the block of factor A is greater than the price of the block, another block should be added. If the value of the decrease in product from giving up a

block of factor *A* is less than the price of the block, it should be withdrawn from production even if this means closing down the productive unit. This would seem to be the rational way of applying the Rule, yet it has some disconcerting results.

The seeming elimination of all possible cases of decreasing cost is merely the reflection of the incompatibility of perfect competition with indivisibility.

This application of the Rule would in one sweep eliminate all cases of increasing returns and of decreasing costs. For the application of the Rule to the divisible factor resulted, as we have seen, in expansion up to a point where there was a net loss on the whole activity of the production unit. Such a loss means that the *net vmp* of the indivisible factor (after deducting the outlay on the cooperating divisible factors) is less than the price of the block of the indivisible factor so that it would seem that the Rule orders the closing down of the production unit if there is a loss. This is seen most clearly where, as in the example just given, only one block of the indivisible factor is used. When it is withdrawn because the value of its marginal net product is less than its price, the production unit must close down. The same is true when there is more than one unit of the indivisible factor because the principle of proportionality (that is, of constant returns to scale) shows that the same must hold for a decrease of each unit in turn. If there are three units of factor *A* and there is a total loss, then a reduction of *A* by one unit and of the other factors by one third their amount (so as to make their *vmp* again equal to their price) will reduce both costs and revenues by one third and losses will be cut by one third too. The price of *A* is greater than its net *vmp* by this reduction in the loss, so its withdrawal is mandatory; and similarly for the other two units.

The disconcerting conclusion, however, is only the result of an implicit assumption of the existence of perfect competition. We have already seen that as long as that is the case the expansion can go on and render the indivisibility insignificant. It is no wonder, therefore, and perfectly proper that this implicit assumption

should throw out the inconsistent one of significant indivisibility. This only confirms our analysis.

The implicit assumption of the conditions that permit perfect competition lies in supposing that there is a unique *vmp* whether the indivisible unit is applied to production or not. In so doing we were in fact assuming that the price of the product is unaffected by the change in output brought about by adding or subtracting the indivisible unit, and that the prices of the cooperating factors are unaffected by the change in the quantities of these that are bought to work with the different amounts of the indivisible factor. This is true only i fthere is perfect competition in selling the product and in buying the cooperating factors. If there were perfect competition it would pay the production unit to expand until either the competition was no longer perfect or the indivisibility had been rendered insignificant.

The values of the net *vmp* and *pf* depend on whether the indivisible unit is applied or not, so they cease to be adequate measures of *msb* and *msc*.

A significant indivisibility will affect prices, of the product or of the factors or of both, when an indivisible unit is applied. There is then no unique value of the net *vmp* of the indivisible factor. When the indivisible factor is not applied, the price of the product is greater or the prices of the cooperating factors are lower (or both). The calculated *net vmp* may be greater than the price of the indivisible factor. On our simple interpretation of the Rule this would indicate that the indivisible factor should be applied to production. But when it is applied the prices of the cooperating factors will rise or the price of the product will fall (or both). Then the net *vmp* may become *less* than the price of the factor so that the simple interpretation of the Rule would order the indivisible factor to be withdrawn. Obviously there is something wrong here.

To deal with this situation we must consider the purpose and meaning of the Rule. We developed the Rule in considering small adjustments that would not affect prices, when *vmp* represented the *msb* and *pf* represented the alternative *msb* which is the same thing as *msc*. The Rule thus equated *msb* to *msc* and brought about

the optimum situation. Now we are confronted with a larger adjustment because of the indivisibility. The price of the indivisible factor is no longer a satisfactory measure of the *msb* of alternative uses of the factor. That is to say, *pf* no longer represents *msc*. The *msb* is the value to consumers of the product here, and the net *msb* is this *minus* what the divisible factors can contribute elsewhere. This last item is still measured by the price of the divisible factors. The *msb* of the product here lies somewhere between the two values of the *vmp* at the two prices ruling before and after the indivisible factor is applied. Its *msc* lies between the two values of the factor, also calculated before and after application. Whether or not the indivisible unit should be applied depends on whether the estimated *msb* is greater or less than the estimated *msc*. For such large items, *vmp* and *pf* cease to be adequate guides to *msb* and *msc*.

Estimates of the range within which *msb* and *msc* lie can be narrowed by considering the possibilities of monopolistic discrimination.

There are ways of narrowing down the ranges within which *msb* and *msc* lie so that the Rule thus interpreted will almost always tell us whether the indivisible item of production should be undertaken or not. The *msb* is not less than what could be obtained from the sale of the product by a monopoly that can discriminate in the prices charged for the product and charges the different consumers "what the traffic can bear" (as long as the consumers are free to give up the product if they think they are being charged more than it is worth to them). The *msc* is not greater than the minimum to which the payment for the factors could be reduced by similar monopolistic exploitation of the owners of the factors. This narrows the estimates for *msc* and *msb* and helps to indicate, for any economy, whether the consumers would get more out of using the factors here or out of leaving them to be used elsewhere.

For example, suppose the calculation made at prices current when the factors are not being applied to the production of X show

$$pf = \$11{,}000 \text{ and } vmp = \$15{,}000$$

but that when the factors are applied the calculations show

$$pf = \$12{,}000 \text{ and } vmp = \$10{,}000$$

The simple application of the Rule would indicate that if the factors have not been applied to the production of X they should be applied (showing a net social gain of $4000), but if they have already been applied they should be withdrawn (saving society $2000). Obviously this is no satisfactory guide to action.

Discriminatory exploitation of factors is made possible by differences in the relative efficiency of units of factors in different uses.

If it is possible by monopolistic discrimination to get the factors for less than is received from the sale of the product, this is proof of the social desirability of applying the factors to the production of X.

It should be possible, by monopolistic discrimination, to get the factors for less than $12,000, because at the prices current before they are allocated to the production of X they earned only $11,000 in the alternative uses. As they are withdrawn from the alternative uses, their *vmp* there rises gradually, and with it the price of the factors, until they reach the rate which when applied to all the factors makes their value $12,000. But those withdrawn earlier have a smaller *vmp* than this for two different reasons.

The first reason is that technically they may be relatively less suitable for the alternative product than for X, having a smaller physical *mp* in the alternative use than the units released later. This would be a reason for their shifting to the production of X before the others and before *pf* has risen all the way to their new values. The factors that are released earlier for this reason could be obtained more cheaply by monopolistic discrimination.

The second reason is that the earlier withdrawals correspond to reductions in the alternative products before the prices of their alternative products have been raised all the way to the new level, so that even if their *mp* is not lower their *vmp* will be lower. These factors will not be obtainable more cheaply by discriminating monopoly because in the new position (when the factors have all been shifted to the production of X) the price of the alternative products will have been raised to the new high level.

If the factors withdrawn from the alternative product (which

we will call Y) are all equally efficient in the production of Y as compared with their efficiency in the production of X, only the second of these two influences will be operative. The amount of Y sacrificed, say 1000 units, will be worth $11,000 at the old price of $11 per unit, and $12,000 at the new price of $12. If the price rises at a constant rate as output decreases, the value of the sacrifice will be $11,500, just halfway between these figures. This figure measures the sum of the prices at which successive units of the alternative product are given up by the consumers with greater and greater reluctance as the price rises from $11 to $12. The *msc* is $11,500.

One can also say that the $500 difference between the *msc* of $11,500 and the $12,000 outlay on the factors when X is produced corresponds to the difference between the $500 that the consumers of Y lose (in no longer being able to buy for $11,000 what they were willing to pay as much as $11,500 to get) and the $1000 that the owners of the factors gain (when they get $12,000 in the production of X instead of $11,000 in the production of Y). The discriminating monopolist in X would not be able to get the factors for less than $12,000 because the equal relative efficiency of the factors in X and Y would not permit any discrimination.

If the factors do not all have the same relative efficiency in Y as in X, the first influence is felt too. The factors which earned $11,000 in Y again earn $12,000 in X and there is the same reduction in the quantity of Y produced (the quantity that is worth $11,000 at the old price), but now the price of Y must have risen less, say to $11.60, so that at the new price it is worth $11,600 and the loss to the consumers (still assuming linear continuity in the price rise) is $300. The factors get $12,000 in X which is $400 more than they could earn in Y even at the new higher price. This is because the factors that have moved from Y to X are those which are relatively more productive in X (which is only another way of saying that they are relatively less productive in Y). Discriminating monopoly could reduce their pay to $11,600 without inducing any of these factors to move back to Y, and the *msc* of employing them in X is $11,300, the sum of the prices at which the consumers of Y are just willing to give up the successive units

of Y at prices between \$11 and \$11.60. The payment to the factors of \$12,000 exceeds *msc* by \$700. This represents the difference between the \$1000 gain to the factors and the \$300 loss to the consumers of Y in no longer being able to get for \$11,000 the goods for which they were willing to pay as much as \$11,300.

If the demand in Y were infinitely elastic only the first influence could be felt. The price of Y would not rise at all, and the loss to consumers of Y would be zero. The increased earnings of factors transferred to X would be entirely due to differences in relative efficiency which enabled them to ask for more in X than their *vmp* in Y and could all be taken away by discriminating monopoly in X which need not pay them any more than the \$11,000 they are able to get in Y. The *msc* would be \$11,000.

The *msc* is never greater than the minimum amount for which the factors could be obtained by the most efficient exploitation of them by discriminating monopoly, and lies between this value and the lower *pf* which is the price of the factors when they are *not* used in the production of X, the adjustment of whose output we are considering.

Discriminatory exploitation of the consumers is conceivable if their elasticity of demand is not infinite.

The *msb* lies between \$15,000 and \$10,000, the two values obtained for the *vmp* by multiplying the product in X of applying the indivisible factor (together with the divisible factors cooperating with it) by the two prices of X. The *mp* is, say, 1000 units of X which is priced at \$15 when the 1000 units of X are *not* being produced, and at \$10 when they *are* being produced. If the demand curve for X is linear so that the price would fall at a constant rate from \$15 to \$10 if the output of X were continuously increased by 1000 units from the smaller to the larger output, the *msb* is exactly \$12,500. Then \$12.50 will be the *average* price at which each additional unit must be sold to persuade the consumer to buy it, and perfectly discriminating monopoly would be able to get as much as \$12,500 instead of \$10,000 from the consumers of X for the additional output.

Production may be socially desirable even if monopolistic discrimination is unable to cover costs.

In the arithmetical example we have just examined, the indivisible factor *should* be applied to X since the *msb* is about $12,500 and the *msc* is between $11,000 and $11,500. We can say, more generally, that if it is possible for a monopolist to cover his costs in applying the indivisible factor, even if he has to resort to extreme discrimination and exploitation to do so, it is in the social interest that the production be undertaken. The *msb* is greater than the monopolist's revenue from the increased output because perfectly discriminating monopoly, which makes every consumer pay the very utmost that he is willing to pay rather than go without the additional product, is never practicable. The *msc* is less than the monopolist's outlay on the factors for two reasons. First of all, perfect exploitation of the factors, which means paying no unit of any factor any more than it could obtain in other uses, is never practicable, and in the second place, even if the monopolist achieved perfect exploitation of the factors the payment to them would measure the value of their alternative product, Y, at the higher price of Y that is current when the factors have all been transferred from Y to X. The true measure of the *msc*, however, is indicated by measuring the valuation of the alternative products at the prices that were just sufficient to induce the consumers to give them up. In our example the *msc* is $11,300 even though perfect exploitation would not be able to get the factors for less than $11,600.

This analysis may be brought out more clearly with the aid of a diagram (Figure 3, p. 195).

In this figure, A represents the situation when the indivisible factor and the divisible factors that accompany it are being directed to the production of the alternative product Y. The *pf* is $11 per unit, or $11,000 for the whole of the indivisible factor and the others that have to accompany it. The *vmp* is $15,000 or $15 per unit. B represents the situation when the factors have been shifted to the production of X. The *vmp* has fallen to $10,000 or $10 per unit, while the *pf* has risen to $12,000 or $12 per unit. The dis-

tance between A and B represents 1000 units of X or of Y. (The size of the units of X and Y are so chosen that at A there are 1000 more units of Y and 1000 less units of X than at B.)

The curve marked *vmp* shows how p_x, the price of X, falls from \$15 per unit to \$10 per unit as its output is increased by 1000 units. It can therefore be considered as the demand curve for the product X as well as the *vmp* curve. If the relative efficiency of the

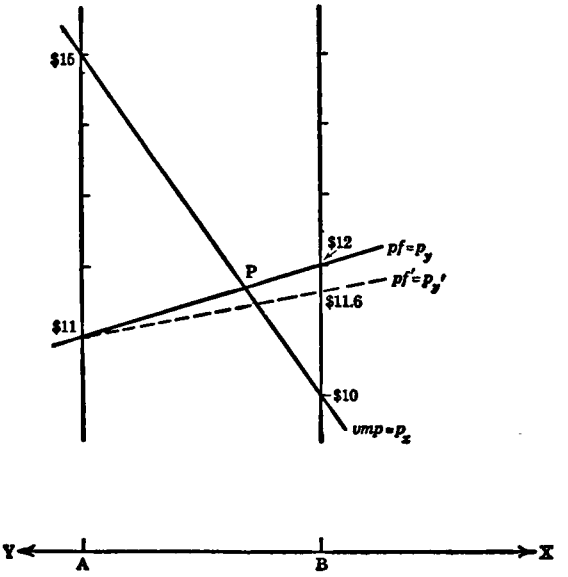


Figure 3

various units of the factors is the same in the production of X as in the production of Y, the line *pf*, which represents the price of the factors used per unit of the product X, will also represent p_y, the price of the alternative product Y, which will rise from \$11 per unit in situation A to \$12 per unit in situation B. This curve will then also be the demand curve of product Y (though drawn in the opposite direction from the demand curve for X since a movement from A to B will indicate a *decrease* of the output of Y by 1000 units while the output of X is *increased* by 1000 units).

If the relative efficiency of the different units of the factors is not

the same in the production of X as in the production of Y, the pf curve will not represent the change in the price of Y. The increase in pf will be due only partly to the increase in the price of its alternative product Y. It will be due in part to the necessity for withdrawing from the production of Y factors that are relatively less and less suitable for the production of X. We considered this case when we supposed (p. 192) that the price of Y rose from \$11 to \$11.60 while pf rose from \$11 to \$12. The price of the alternative product rises less than the price of the factor which is withdrawn from it because the remaining factors are just those whose relative efficiency is greater in the production of Y than in the production of X, and p_y' the demand curve for the alternative product, will not coincide with the pf curve but will lie below it. This is shown in Figure 3 by the dotted line which shows the price of Y rising from \$11 to \$11.60 as factors are shifted from producing 1000 units of Y to producing 1000 units of X. In this case too it will be possible for the producer of X, by the application of discriminating exploitation, to reduce the outlay for the factors by this amount, paying each factor no more than it would be able to earn in the production of Y, and then pf' in situation B would be not \$12,000 but \$11,600. The pf' curve and the demand curve for the alternative product would coincide.

Direct estimates may be made of *msc* and *msb* by considering the area under sections of the demand curves.

Considering the whole indivisible shift in output of 1000 units between Y and X, we see that in situation A, pf = \$11,000 and vmp = \$15,000. If we follow the Rule *simpliciter* we must move to situation B. At B pf = \$12,000 and vmp = \$10,000, and this orders a movement back to situation A. What the Rule really orders is a movement from A *in the direction of* B and from B *in the direction of* A, both movements being aimed at position P where $pf = vmp$ (at \$11.67 for one unit). Unfortunately the indivisibility does not permit this and so we cannot use the Rule *simpliciter*. We can only choose between situation A and situation B. We must therefore go behind vmp and pf to look for the *msb* and the *msc* which they are intended to measure. The *msb* is the

area under the *vmp* curve which represents what the consumers of *X* are just willing to pay, if necessary, for the successive units of *X*. If the *vmp* curve is a straight line this will come to $12,500. The *msc* is the area under the *pf* curve, or p_y curve, which represents what the consumers of *Y* are willing to pay for the successive units of *Y* withdrawn as the output of *Y* is reduced to permit the output of *X* to be increased. If *pf* is a straight line and its increase is due entirely to the increase in the price of the alternative product as less of it is produced, this will amount to $11,500 or 1000 times the average price which is halfway between the extremes of $11 and $12. This is shown by p_y. If the increase in *pf* is due in part to the differences in relative efficiency of the factors in the production of *X* as compared with their efficiency in the production of *Y*, then *msc* is less than this. On the assumption made above that the price of *Y* rose only to $11.60, the *msc* would be $11,300 or $11.30 times 1000. This is shown by p_y'.

Our conclusion may now be somewhat simplified. In the absence of further knowledge as to the shapes of the p_y and p_x curves it will be convenient to formulate a rule based on the assumption that they are linear. We can do this by simply keeping the original Rule, stipulating only that *vmp* is undertsood to mean the *average* of the two values of *vmp* in the two situations A and B and that similarly *pf* is understood to mean the *average* of the two values of *pf* at A and at B.

Where more information is available as to the shape of the curves p_x and p_y the area under these should be taken instead of the simple arithmetic averages of the extreme values, which is what the area comes to under the assumption of linearity. The substitution of the areas under the curves would mean that a more appropriate average was being used, and this would improve the accuracy of the formula.

The same analysis is applicable to all large decisions whether to produce or not.

This analysis of the indivisible factor is applicable not only to the other forms of indivisibility but also to the more general question of whether a particular product should be produced at all if the application of the Rule will result in a total payment to

the factors exceeding the total revenue from the sale of the product. The answer is given above. If there is *any* output of the product which will make the *average vmp* greater than the *average pf* in the sense here developed, that output may be considered as an indivisible unit and should be produced. But production should not stay there. It should be extended in accordance with the Rule until $vmp = pf$. The amended Rule in terms of averages shows that producing this "indivisible" amount is better than producing none at all, and the Rule itself indicates that further extension of production up to the point where $vmp = pf$ constitutes a further improvement in the use of resources. Only if there is *no* output that would warrant production when considered as an indivisible output should the production of the product be abandoned.

The necessity of making unreliable estimates is in the nature of the problem and not in the method of solving it.

Essentially what we have done in this chapter is to depart temporarily from considering only small marginal adjustments to the consideration of large indivisible decisions. In our solution there is an unavoidable reliance on estimates or perhaps even guesses of what prices would prevail in new situations that have not been tried, so the error may be very great. This is so, but to argue from this against the application of these estimates in a controlled economy, as has often been done, is quite illegitimate. The same estimates and guesses must be made in *any* economy where knowledge is imperfect and where large decisions have to be made. Our Rule does not *create* these difficulties; rather it reduces them to the very minimum. But in so doing it brings them into the open so that critics who are ignorant of the universal existence of the problems are given the opportunity of identifying them with the devices for dealing with them in the controlled economy.

The uneasiness of accepting a permanent loss is often due to identifying irrelevant aspects of perfect competition with the optimum use of resources.

Another aspect of the analysis of this chapter that many people find very disturbing is the tranquil acceptance of a permanent net

loss in a production unit or in an industry as of no consequence at all as long as the Rule is being obeyed. Anyone brought up in a capitalistic society feels "instinctively" that something is wrong when outlay exceeds income and there ought to be a rule against it. Enough has been said about the reasons for the Rule and how it is justified in terms of the best use of resources without reference to whether it results in a profit or a loss. The same issue will appear again in another guise in the next chapter. All that can be said here is that this feeling is nothing but an illogical (though easily understandable) transference from the capitalist economy, where it is in perfect order, to the controlled economy, where it is simply irrelevant.

There are two levels to this transference. On the lower level it is based simply on the individual's interest in keeping solvent that naturally guides every businessman in conducting his affairs. It is irrelevant where the purpose considered is not the profit or the solvency of any individual or group or enterprise, but the optimum use of society's resources.

At the higher level it is based on an *identification* of perfect competition with the optimum use of resources instead of recognizing it as merely *one way* of bringing about the optimum use of resources that is possible under certain technical conditions of production. All the symptoms of perfect competition are then mistakenly thought to be conditions for the optimum use of resources. One symptom of perfect competition is that there are no losses; consequently this is thought to be a condition of the optimum use of resources. On this level abnormal profits are also taboo for the same reason and it has led to the proposal of a rule to make $p = ac$. This syllogism falls to the ground because the optimum use of resources is *not* identical with perfect competition. It can be reached by the application of the Rule in conditions where it is technically impossible to reproduce all the symptoms of perfect competition. Indeed it is because of this impossibility that perfect competition often destroys itself and the optimum use of resources can then be attained only via the application of the Rule by collectivist agencies or by an artificial maintenance of perfect competition by *counterspeculation* supported by state subsidies.

CHAPTER 17. FIXED FACTORS (*EQUILIBRIUM OF THE FIRM, LONG, AND SHORT PERIODS*)

In Chapter 15 we discussed the effects of indivisibilities and saw how the principle of proportionality (which says that changes in output can be brought about by varying all the factors and products in the same proportion) was upset by indivisibilities. These prevented the indivisible factor from being appropriately adjusted with the result that its proportion to the other factors was made "too great." We saw that under conditions of perfect competition this brought about decreasing costs which led to the expansion of firms and the destruction of the conditions of perfect competition. The optimum use of resources could then be reached only by collectivist agencies which obeyed the Rule and did not depend on the conditions of perfect competition, or by an artificial maintenance of perfect competition through *counterspeculation* supported by state subsidies.

There may be too little of an indivisible factor.

In this chapter we shall consider an interference with the principle of proportionality that works in the opposite direction. This can happen when the divisible factors have been increased sufficiently to raise the virtual *vmp* of the indivisible factor *above* its price. At the point where the *vmp* of the indivisible factor is just equal to its price we have the optimum proportion between the factors and (if there is perfect competition) constant costs. Any increase in the output (brought about by increasing the divisible factor) would now result in the proportion of the indivisible factor becoming *relatively too small.* Its virtual *vmp* would rise above its price, indicating that for the optimum proportion between the factors to be reached it would be necessary for the indivisible factor

to be increased a little to maintain the same proportion, since the relative prices of the factors have not been changed. But this is impossible because of the indivisibility, and the use of a whole block of the indivisible factor is not warranted.

This will result in increasing cost, and, in perfect competition, either the firm will be making an abnormal profit or it will pay the firm to contract output.

We will now have the exact opposite of what happened when the proportion of the indivisible factor was too great. Increasing output by adding only the divisible factors means moving *further away* from the optimum proportion between the factors, and this results in *increasing* average cost. Since *ac* is increasing, *mc* must be greater than *ac*. From this it follows that either the firm will find it worth while to contract (if *mc* is greater than the price of the product) or it will be making a profit over and above the cost of the factors (if *p* is not less than *mc* it must be greater than *ac* and this means there is a profit). This is the counterpart of the condition of the firm in Chapter 15 which either made a loss (forcing it to close down) or had an incentive to expand. The firms now neither tend to go out of business nor are impelled to expand.

Perfect competition thus appears to be secure, but this is true only over a limited range.

There is no threat to perfect competition. There is not even the indirect incentive to expansion or combination that exists with perfect divisibility and constant costs. The firm is prevented from expanding because it is subject to *increasing cost*. Any firm that expanded relatively to its competitors would not have an advantage, as when an indivisible factor is in excess, or even be on the same footing, as when there is perfect divisibility, but would be penalized. It would find its costs higher. In this way perfect competition would be neither destroyed by forced expansion nor left to the hazard of indirectly profitable expansion but would be safeguarded by a penalty on the expansion of any production unit.

Perfect competition would be secure and we would rely on the optimum use of resources even in an uncontrolled economy.

Such optimism would be premature. The increasing costs on which it rests are valid only over a limited range. Beyond a certain moderate increase in output it will be profitable to acquire a second unit of the indivisible factor, and this will at once reverse the situation. The proportion of the indivisible factor will become too great instead of too small (as compared with the optimum proportion), and there will be decreasing instead of increasing cost.

As the firm grows larger the significance of the indivisibility diminishes.

The acquisition of a second unit of the indivisible factor which puts an end to the range of increasing cost will probably become profitable appreciably before output has been increased as much as 50 per cent beyond the point where the increasing cost began. This is because when output has been increased to 50 per cent above the point of optimum proportion with one unit of the indivisible factor (which is where cost began to increase) the output is only 25 per cent below that which could give the optimum proportion with two units of the indivisible factor. The acquisition of the second indivisible unit will considerably diminish the deviation from the optimum proportion. It is *certain* that the second unit will be acquired *before* the output has increased 100 per cent, for that would give the exact optimum proportion with two units.

There will be another range of increasing costs beyond this point until it becomes profitable to acquire a third unit, and so on. Between the outputs that correspond to optimum proportions with different numbers of the indivisible factor, there will always be a range of increasing cost followed by a range of decreasing cost. But these variations in cost quickly become very unimportant. When four units of an indivisible factor are being used, the range of increasing cost will probably come to an end before output has been increased by $12\frac{1}{2}$ per cent and the rate of decreasing cost will be correspondingly small. The same is true for the range of increasing cost before the increase in output has restored the opti-

mum proportion of factors with five indivisible units. What we are observing is the diminishing significance of the indivisibility as the firm expands. If there is still perfect competition at this point, the ups and downs of average cost can be neglected and we are back with constant returns to scale, just as with perfect divisibility, and the same hazards to perfect competition from monopolistic combination or expansion (unless perfect competition is safeguarded by counterspeculation or other government activities directed to this purpose). The indivisibilities have become insignificant and the stability of perfect competition in an uncontrolled economy is gone.

Fixed factors tend to stabilize competition, but factors and costs that are fixed in the short period are variable in the long period.

Stability of perfect competition would be safeguarded if it were not possible for a firm to obtain further units of the indivisible factor or indeed of any other factor, divisible or not—limitation is the essence of the phenomenon. This is frequently the case in the short period. A firm has a *fixed* factor, a plant that it has built or a piece of land on which it has a lease. For an appreciable period it is unable either to obtain more or to dispose of what it holds at a satisfactory price. Within this period the fixed factor is just like a unit of the indivisible factor before another unit could be added and the size of the firm will be restricted. If there were perfect competition to begin with it would be maintained.

It is convenient in analyzing this situation to divide the costs of the firm into two parts. There is the part that is paid for the fixed factor or factors and that cannot be varied whatever happens to the output. This is called the *fixed cost*. Then there is the part of cost that is paid for the other factors that are increased to increase output and decreased to decrease output. This is called the *variable cost*.

The average fixed cost is obtained by dividing the fixed cost by the output. Since the total fixed cost does not change when output changes, the average fixed cost will always fall in the same proportion as the output increases because the total fixed cost

can then be spread over a larger number of units of output. The variable cost per unit of output will be decreasing for outputs that are very small because the proportion of the variable factors applied to the given amount of the fixed factor is *absolutely* too small. There will be increasing returns to the variable factors. If there is an increase of, say, 1 per cent in the variable factors, the

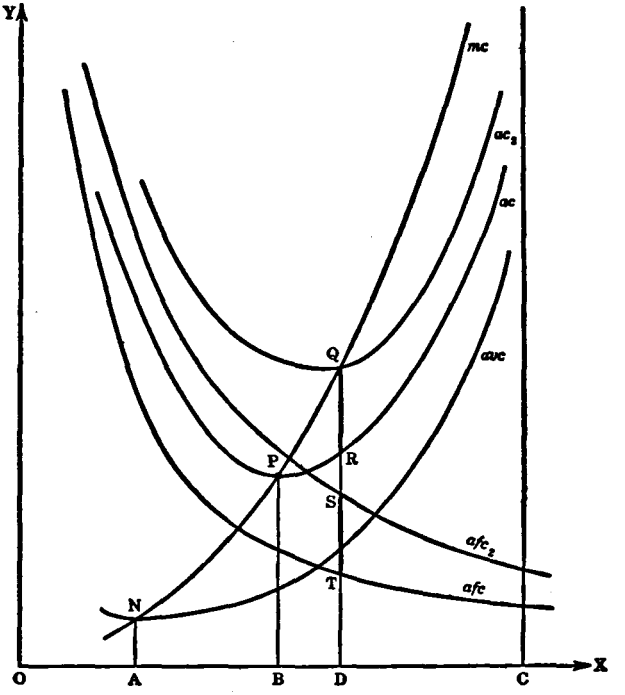


Figure 4

total variable cost (in perfect competition) will also increase by 1 per cent. The average variable cost will diminish to the degree that the product increases by more than 1 per cent. Under perfect competition production could never remain in this range because it would pay the firm either to close down or to expand production. The expansion of production is brought about by increasing the variable factors, so the proportion between these and the fixed factor is changed and the increasing returns come to an end.

This is illustrated in Figure 4. The curve *afc* represents the

average fixed cost that decreases in the same proportion as output increases. The average variable cost is represented by the curve *avc* which falls in the first stage of increasing returns and then rises from the output marked *A*. The point *C* marks the largest output that it is possible to produce by adding variable factors to the fixed factor. Any attempt to produce more than this by adding more of the variable factor would be fruitless and if carried far enough would actually diminish the product, showing a negative *mp* of the variable factors. The range of output from *A* to *C* corresponds to the middle range of factor proportion which shows a positive *mp* and diminishing returns to all the factors.

The average cost (the average of *all* the costs whether fixed or variable) is represented by the curve *ac*. It is obtained by adding *afc* and *avc* (vertically). Between *A* and *B* *afc* falls more rapidly than *avc* rises so that it more than offsets the effect of the rising *avc* on *ac* which therefore is decreasing. Beyond *B* the rate at which *avc* rises is greater than the rate at which *afc* falls so that it more than offsets this and *ac* rises. The marginal cost, which is represented by the curve *mc*, is the same as the marginal variable cost, since the increase in total cost when output is increased by a unit consists only of variable costs. (The fixed cost of course cannot vary.)

To the left of *A*, *mc* (which is equal to *mvc*) lies below *avc* because *avc* is falling; to the right of *A*, *mc* is above *avc* because *avc* is rising. To the left of *B*, *mc* lies below *ac* because *ac* is falling; to the right of *B*, it lies above *ac* because *ac* is rising. This follows from the arithmetic of the average-marginal relationship (see p. 82).

In seeking to maximize its profits (or minimize its losses) the firm adjusts its output to make $mc = mr$. In perfect competition $mr = ar = p$, so that the firm chooses the output that makes *mc* equal the price of the product. Given the price of the product, the output chosen by the firm can be read off on the *mc* curve. The price is measured on the vertical axis *OY*, and the corresponding point on the *mc* curve will give the output, measured by the horizontal distance of this point from the vertical axis.

We see again that an output less than *OA* (which could only be the response to a price less than *AN*) is ruled out because it could

maximize profit (or rather minimize loss) only when the price is less than *avc*. This means that the revenue from the sale of the product would not even cover the cost of the variable factors. It is better for the firm not to hire any of these even if it cannot avoid paying for the fixed factor. Its losses would be limited at least to the fixed cost.

Output might be anywhere between *OA* and *OC*, depending on the price, though it would require a very high price to make it worth while for the firm to produce very near to the absolute physical limit as is shown by the way the *mc* rises very high as it gets near the output *OC*. For outputs between *OA* and *OB* (which could only be the result of a price greater than *AN* but less than *BP*) *ac* is greater than *mc* and therefore greater than price (which must equal *mc*). This means that the firm is making a loss, but the revenue from sales is greater than the cost of the variable factors because *p* the price or average revenue from sales is greater than the average variable cost *avc*. Something is left over, and this is a *gross* profit over the variable costs. It is true that this is not sufficient to cover the fixed cost. That is why there is a *net* loss. But the *net* loss would be greater if any output were produced other than that which made $p\,(=mr) = mc$. If production were closed down the loss would be equal to the *whole* of the fixed cost without the benefit of the gross profit that covers part of this. At output *OB* (which is the response to the price *BP*) the gross profit is equal to the fixed cost so that there is no net loss and no net profit. At outputs between *OB* and *OC* (which are induced only by prices greater than *BP*) there is a *net* profit over and above the fixed cost. In all cases the output of the firm is determinate and there is no tendency for perfect competition to be destroyed.

The average cost plays no part in determining the *output* of a firm that tries to maximize its profits.

The average cost plays no part in the foregoing analysis. It does not determine whether the firm shall produce or not or what output it should produce. It merely helps us to calculate whether a firm in production makes enough to cover the fixed cost.

The reason for this is that fixed costs are not economically relevant. They have been incurred in the past and they involve no new sacrifice that is involved in present production. They do not form part of the *msc* of production. Only the variable factors represent withdrawals of resources from production elsewhere in order to produce here. Consequently it is natural that the fixed costs should play no part in determining output where the optimum use of resources, equating *msb* to *msc*, is brought about by private entrepreneurs under perfect competition. It is for the same reason that it plays no part in the Rule for achieving the same optimum by collectivist agencies.

In the long run the adjustment of the *number* of firms tends to make the price equal to the minimum average cost.

Things are different when we consider a period long enough for new firms to acquire or build the fixed factors and for old firms to wear out or get rid of the fixed factors. Then the *ac* curve becomes very important, especially the lowest point on it (marked *P* in Figure 4) where it is crossed by the *mc* curve. If the price is greater than *BP*, the output that maximizes profit will be greater than *OB*, *p* (which is equated to *mc*) is greater than *ac*, and a net profit is earned by the firm. This will induce others to do likewise, so that the number of firms will increase. This increases the output and lowers the price, but as long as there is any net profit above all the costs of entering this industry (including among these costs the normal reward to the businessman, normal interest on capital, and so on) new firms will keep coming in until *p* falls to *BP* and the abnormal profits have disappeared. In the same way it will be profitable for the firms already in the industry to acquire more of the same fixed factors and increase the number of their plants until the abnormal profits have disappeared.

The same phenomenon also works in the other direction. If the price is less than *BP*, the chosen output is less than *OB* and *p* (which is equated to *mc*) is less than *ac*, so there is a net loss. The gross profit over variable cost is not sufficient to cover the fixed cost. No new firm will now come into the industry and no new

fixed factors will be acquired by the firms already there. On the contrary, the existing firms will not replace fixed factors when they wear out and will not renew leases when they expire. As a result, the output of the industry will fall off and the price will rise until it is equal to *BP*.

This is because in the long run there are no fixed factors and the optimum proportion between the factors can be reached.

P represents the position that each plant tends to reach in the long period. It shows a *minimum average cost* and is nothing but the point corresponding to the optimum proportion between the fixed and the variable factors. It is reached in the long period because in the long period the fixed factors are no longer fixed but are just as variable as the other factors. The optimum proportion depends on the relative prices of the factors, and this too is shown in Figure 4 where it can be seen that if the fixed cost is greater *afc* will be higher and steeper and the *ac* curve will be correspondingly higher. The point where the *avc* curve begins to rise as steeply as the *afc* curve falls will be more to the right. This is the point where the increasing average variable cost just offsets the decreasing average fixed cost and begins to have more influence on *ac* than the latter. At this point *ac* reaches a minimum and is cut by the *mc* curve. All this means that the point of optimum proportions will be more to the right, indicating that when the fixed factor costs more it will pay, in the long run, to use more of the other factors in conjunction with a given amount of it and to produce a larger output.

The stability of competition which result from fixed factors disappears too in the long run.

In the long run, then, the *ac* curve, and in particular *P*, the minimum point on it, turns out to be the most important part of the diagram. Given the prices of all the factors and the size of the most efficient plant or "fixed" factor (which is not really fixed in the long period), the optimum output *OB* and the optimum price *BP* (which is equal to both *ac* and *mc*) are determinate. But the fixed

factor is now nothing but an indivisible factor at most and no longer a guarantee of the preservation of competition. If the expansion of the industry which takes place when p is greater than BP were all by new firms, there would be no threat to perfect competition, but we have seen that the expansion can as well consist of an increase in the number of plants owned by the existing firms. We are then in the same situation as with perfect divisibility. There is no obstacle to the indefinite expansion of firms or to their combination, because the indivisibilities are overcome in the long run by the expansion of firms. Perfect competition is possible but precarious.

Legal maxima to the size of firms could maintain perfect competition, but would interfere with efficiency. Counterspeculation is therefore preferable as when there is no indivisibility.

If every firm could have only one of the indivisible units even in the long period, perfect competition could be safeguarded. This might be done by legislation prohibiting firms of more than a certain size. If the size permitted were large enough to take advantage of all the technical economies of production this would be a satisfactory solution. Nothing but good could come to society if firms were prevented from earning such profits as come only from being large enough to influence prices. All such gains are only at the expense of other members of society who have to pay more for what they buy from the powerful firms or who get less for what they sell to them. In addition, there is the net loss to society because of the departure from the optimum use of resources. Situations may arise where this kind of limit to the size of the firm can be undertaken by the government of a controlled economy with purely or preponderantly beneficial results. But great care must be exercised in doing this because it is difficult to say exactly what the technically optimum output of a firm is. A firm managing several plants may be able to manage them more efficiently than if they belonged to separate firms, and it is difficult to separate these legitimate and socially useful economies from the other private economies that are socially harmful, such as being able to

beat down the price of factors. It is therefore better, if possible, to maintain perfect competition by other means, such as counter-speculation, and permit firms to expand to the size that they find most profitable. Again we see that the price mechanism makes its great contribution by permitting the man on the spot, who knows best, to decide where the optimum position P is, and permitting those managers who are more skillful in production to set the pace and eliminate the less skillful.

Diminishing returns to entrepreneurship, which is unaugmentable, can stabilize perfect competition, but recent developments in business organization have made this less important.

If there were a fixed factor of which each firm can get only one unit this would prevent expansion of the firm; and if the output with one unit were small enough in relation to the market, and if the firms did not combine to obtain the fruits of monopoly, perfect competition would be stabilized. We have seen that this could be brought about by government action but that where other ways of maintaining perfect competition are available the latter are preferable. However, there is in many industries a factor of production that *naturally* works in this way and nothing need or can be done about it by the government. Because of this we can have something approaching perfect competition with stability even in an uncontrolled economy.

That factor is the unit of entrepreneurship or management-cum-enterprise that is provided by the individual businessman. There is a limit to the amount of business he can handle, and if he tries to handle more than that he loses in efficiency. This follows from the same principle of diminishing returns—diminishing returns to the other factors when used together with a given amount of entrepreneurship. The peculiarity of entrepreneurship is that more of it cannot be hired. The entrepreneur can hire foremen and managers, but they can never do the same kind of work he does and make the same decisions, partly because he is operating with his own money and the manager is not (though this could be remedied by appropriate forms of remuneration) but fundamentally because the decisions made by the entrepreneur are

all interrelated and can only be made by a single individual who is aware of all the other decisions that are being made. These decisions cannot therefore be delegated, and so only one unit of entrepreneurship is available for each firm. This limits its size so that there is an optimum output of the firm, as distinct from the plant, and the same Figure 4 can be used to illustrate this. Any firm which increases its output beyond the optimum OB will lose in efficiency through cumbersomeness in operation more than it might gain in other ways, and so the expansion of the firm is checked and perfect competition is stabilized.

There are many branches of industry where this seems to fit the facts fairly well. The philosophy of *laissez faire* is based on the assumption that this is generally the case throughout all industry. It is doubtful whether this was ever true, and it certainly is not true now. Furthermore, recent improvements in the science of managerial organization, accounting, communications, indexing, and mechanical calculating tend to increase the optimum size of the firm as compared to the market and increase the jeopardy of perfect competition. These developments do not do away with the essential element of entrepreneurship as an *unaugmentable factor*, but they permit more and more subsidiary tasks to be delegated with only pure general entrepreneurship concentrated in the hands of the managing directors of corporations which can thus reach much greater size before they are overwhelmed by bureaucracy.

Nevertheless, in important parts of the economy the optimum size of the firm is still small compared to the size of the market. There is then no danger that perfect competition will be destroyed by the growth of firms or by combination of firms. There is still the danger that quasi-monopolistic institutions will be developed by trade associations, by political measures such as tariffs to prevent competition from abroad, and by licensing or other legal restrictions to hinder domestic competition. But if these are kept in check perfect competition can be maintained in such industries. If the Rule is kept in operation in the other parts of the economy where perfect competition is not stable or possible, the optimum use of resources can be obtained without any interference with private enterprise in these particular industries.

CHAPTER 18. LONG AND SHORT PERIOD. RENT AND NEGATIVE RENT

Periods are long or short relative to the time it takes to make an adjustment, and the distinction between fixed and variable factors is correspondingly relative.

In Chapter 17 we spoke about long and short periods in connection with fixed factors. These periods are not absolute or even definite periods of time like days or months or years. The distinction between the long and the short period is relative to the matter discussed. A short period is any period not long enough to permit a certain adjustment to be made. The long period is any period that is sufficient or more than sufficient to permit this adjustment to be made. The actual time needed for any particular adjustment may be anything at all from 5 minutes to a century—it all depends on the adjustment that is being considered.

Another way of expressing this is to say that there are any number of periods of different length, and the longer the period taken the greater will be the number of factors that can be adjusted (which we call the variable factors) and the fewer will be the number of factors that cannot be adjusted (which we call the fixed factors). The longer the period of time taken, the better can be the adjustment to any change in the situation.

This would appear to introduce an ambiguity into our description of the optimum use of resources, and in particular in our description of the means by which the optimum is brought about. The ambiguity applies both to free private enterprise under conditions of perfect competition and to collectivist agencies that apply the Rule (as well as to a combination of these methods in the controlled economy). The apparent ambiguity appears most clearly

if we apply the Rule in the form which instructs the adjustment of output until p equals *mc* (*mc* being appropriately qualified so that it becomes *vmf*).

Short period *mc* need not be less than long period *mc*.

The *mc* of producing another unit of any product depends on the *period* allowed. There may be a different *mc* for every different period. If very little time is allowed, the increased output can be obtained only by increasing the few factors that can be adjusted at very little notice. If more time is allowed, a different kind of adjustment may be more appropriate. There is no reason why the *mc* should always come to the same value, and if it is not the same the Rule is ambiguous. Which of the different *mc*'s should be equated to the price?

It has been argued[1] that the long-period *mc* will be greater than the short-period *mc* because in the short period only the variable factors are increased, whereas to get the long-period *mc* there must be added the cost of the additional fixed factors which become variable factors in the longer period. This is not necessarily true, if only because the long-period increase in the factors that are fixed in the short period will permit smaller increases, or perhaps even decreases, in the variable factors that enter into the short-period *mc*. It could be argued on the contrary that the short-period *mc* can never be less than the long-period *mc*, but may be greater, because if it were less the short-period adjustment would be kept up in the long period and the short-period *mc* would *ipso facto* become identical with the long-period *mc*. This is not necessarily true either because there may be a special situation in the short period—an abnormally low price of the variable factors, for instance—which results in a very low short-period *mc* which cannot be expected to continue for the long-period adjustment.

If there is complete adjustment to a certain constant rate of output and then a change in output is necessary to a new level, higher than before, which is expected to stay constant for a long

[1] See A. P. Lerner, "Statics and Dynamics in Socialist Economics" *Economic Journal*, June 1937; "Theory and Practice in Socialist Economics," *Review of Economic Studies*, October 1938; and references there to articles by Dickinson and Dobb.

time, the longer the period allowed for the adjustment, the smaller will be *mc*. In the short period the increase of output must be brought about entirely by increasing those factors which are quickly adjustable. This will lead to a deviation from the optimum proportion between the factors, and the increase in total cost will be greater than in the long period when more factors are adjustable so that a cheaper way of expanding production can be chosen. Short-period marginal cost is *greater* than long-period marginal cost.

If there is a *reduction* in output (also expected to continue for a long time), the short-period adjustment will again not be as economical as the long-period adjustment because in the long period advantage can be taken of the greater number of adjustable factors. But now this means that in the short period the reduction in output leads to a smaller *reduction* in total cost than in the long period so that the short-period marginal cost is *less* than the long-period marginal cost.

But all these are special cases. Whether the *mc* goes up or goes down or changes in any irregular way with the length of the period, the problem still remains. Which *mc* is to be compared with price in applying the Rule?

The appropriate period to use in connection with the Rule is that which refers to the date of the output considered.

Fortunately this is not a real ambiguity. The appearance of ambiguity arises from the inadequacy of this formulation of the Rule in terms of the product, which we have had occasion to criticize before. If we write it out in the more adequate form, in terms of the relationship between p, the price of the product, and *vmf*, the value of the marginal amount of particular factors necessary to produce an additional unit of product, the ambiguity disappears. The appropriate *vmf*'s for producing an additional unit of output *at a certain point of time* are definitely known. Many different factors can be applied at the margin to do this, and marginal factors may be applied at different points of time with the same result. The Rule says that all those whose value *vmf* is less than the price of the product p should be applied, and all those whose

value is greater than p should be withdrawn. The matter becomes clearer still if we use the first formulation of the Rule which we have found more convenient throughout and which is expressed in terms of input instead of output. If we do this and follow the Rule which tells us to apply factors where *vmp*, the value of their marginal product, is greater than *pf*, the price of the factor, and to withdraw them where it is less, we can see directly where to apply a factor and where not to. If the *vmp* of a factor accruing at any future time is expected to be greater than its price, the factor should be applied; if not, it should not be applied. That is all there is to it.

The marginal cost will be different for increments of output at different dates in the future. The appropriate *mc* to equate to the price of an increment of output at a particular date in the future is that which would be incurred in the course of producing an increment of output at that date. If the expected price at that date is greater than this *mc*, the cost should be incurred and the increment of output produced. If the expected price is less than the appropriate *mc*, the increment of output should not be produced. Output at that date should rather be reduced until the expected price is once more equal to the appropriate *mc*. If the price is expected to be the same at different future dates, the appropriate adjustments will equalize the different *mc*'s. If the prices are expected to be different, different *mc's* are just as appropriate for the increments of output at different dates as for different products becoming available at the same date.

A minimum adjustment period may be convenient.

Another problem arises in connection with very short periods. When the short-period marginal cost is subject to great fluctuations, should the price be permitted to fluctuate as much as is necessary to bring it into equality with the very shortest period marginal cost? Examples of this are afforded by seats at the theater or on the train. As long as there are any empty seats at a performance that is being given anyway, or on a train that is being run anyway, the marginal cost of permitting someone to use the

seat is practically zero. As soon as the seats are all taken, the marginal cost rises very considerably. It must now cover the cost of increasing the seating accommodation which may mean transferring the play to a bigger theater or running an additional coach on the train or even an additional train.

Under such circumstances, minor adjustments according to the Rule might well be sacrificed to the considerable convenience of knowing beforehand what is the price of a ticket. Once it is decided to produce the show or to run the train the price should be fixed at the level at which it is expected that all the seats will be taken. (Whether or not this price covers total cost or results in a loss is irrelevant for the optimum use of resources.) If a mistake is made (and such mistakes are unavoidable) and there are some empty seats, a reduction in price to the marginal cost would lead to a better use of resources if it led to more people seeing the performance or taking the train, but it would not be an improvement if at the short notice there was no response to the reduction in price. This gives us another principle for governing price policy in addition to the consideration of the convenience of known prices. Price changes should not be made more rapidly than the adjustments to them are made by the producers or consumers who determine how much will be bought or sold.

The only purpose of prices and of price changes is to bring about the optimum use of resources by means of these adjustments, so a price changing more rapidly than the quickest of these adjustments would be nothing but a nuisance. Rapidly changing prices may be a considerable nuisance even when they do permit some adjustment, but here no principle can be applied. The nuisance value of frequent price changes must be balanced against the improvement in the use of resources that they bring about before deciding how frequently to change prices of this kind.[1]

The average cost is adjusted to the marginal cost by the derived price of the fixed factors.

In Chapter 16 we saw that perfect competition is impossible with decreasing cost, but where there is increasing cost perfect

[1] I am indebted on this point to discussions with Mr. Wm. S. Vickrey.

competition is not only possible but stable. We also saw that the price of the product had to equal the marginal cost (in perfect competition) because the firm maximized its profit, and that it had to equal the average cost or else the firm would be making a profit or a loss and this would make more firms come in or go out of the industry. Every firm must be in a position like that indicated by P in Figure 4 (p. 204) with $p = mc = ac$.

There is something a little mysterious about the way in which *mc* is always equal to *ac*. Why could not the price be greater than BP (in Figure 4)? Suppose all the factors that are suitable for the production of a particular product are already employed in the industry. There will then be no danger of the price being reduced by competition from outside the industry even if all those engaged in it are making great profits. Why then cannot *mc* and p be greater than *ac*?

The answer is that in such a situation the entrepreneurs, all making a profit over and above the cost of the factors, would bid up the price of the fixed factor (which may be the reward of the entrepreneurs themselves). They would not bid up the prices of the variable factors for these will already be equal to the value of their marginal product and it will not pay any firm to employ more of them. But the price of the fixed factor will be bid up until the profit disappears. The *afc* curve will rise and with it the *ac* curve. As the *ac* curve rises, its lowest point will stay on the *mc* curve. This it must always do since the *ac* curve must be falling (that is, sloping down to the right) as long as *mc* is less than *ac*, and it must be rising (that is, sloping upward to the right) whenever *mc* is greater than *ac*. The *ac* curve will keep on rising until it is entirely above the horizontal price line (that indicates the higher price of the product) its lowest point just touching this line where it cuts the *mc* curve. This point (Q in Figure 4) will be just like P, showing $p = mc = ac$ and the *ac* at a minimum. The increase in the income of the owner of the fixed factor is indicated by the degree to which the *afc* and the *ac* curves are raised. (They must naturally be raised by the same amount since there is no change in the *avc* curve which is the other component of the *ac* curve.) The increase in the income of the owner of the fixed factor

can be measured by taking the increase in the height of the new *ac* curve over the old *ac* curve for any output (the increase in the height of the *afc* will be identical with this) and multiplying it by that output. In Figure 4 this can be seen in the vertical distance *RQ* (which will be equal to *TS*) multiplied by the output *OD*.

Rent (and quasi-rent in the short period) may be defined as *unnecessary payment* or *surplus*.

The extra income of the fixed factor is thus seen to be a kind of surplus which it is able to appropriate to itself when the demand for the product rises. The payment to the variable factors is kept down almost to the previous level in spite of the increased demand for the product because there is not much increase in demand for them as long as the fixed factor cannot be increased. In any case the prices paid for them cannot be greater than their earnings in alternative occupations or there would be a stream of the factor from these alternative occupations to take advantage of the greater pay, and the competition between these factors would keep down the prices of the variable factors.

The income of the fixed factor has been called *rent* by analogy with the rent of land which is by definition a fixed factor even in the long period. Many factors are fixed only in the short period, however, so payment for them is not like rent from the point of view of the long period. This payment has therefore been called *quasi-rent* to indicate the partial nature of the similarity.

The distinctive feature of rent as a payment is that it is *unnecessary* for making available the factor that receives the payment. The factors being fixed, either permanently, as is land, or temporarily as are the factors we have been considering as fixed only in the short period, they will be available however small the payment made to them. Being fixed they cannot be withdrawn and put to other uses if they are not paid what the owner considers a proper amount for their use. They have to take what is left over after paying the other, variable, factors the amounts that are necessary to get them to work in the particular place where they are needed. The variable factors must be paid what they could get in alterna-

tive occupations or they would go to these alternative occupations. The fixed factors can offer no such threat so they must be satisfied with what is left over after paying the variable factors.

This may be large or small.

This is not always so bad for them. It may mean that they are paid very little when very little is left over, but they may be paid very much if very much is left over after paying the variable factors. The distinctive feature about the payment of the fixed factors is not the amount, which may be very large or very small, but the fact that it is a *residue* or *surplus*, which is left over after making payments to other factors that are *necessary* in order to have these other factors available. The payments to the fixed factors are *unnecessary* in the sense that even if much less were paid to them, or even if practically nothing were paid to them, they would still be available in the place where they happened to be and could continue to be used in the short period within which they are "fixed factors."

This raises a number of interesting questions. The first is, "If it is unnecessary to make these payments, why are they made? Why is it that the individuals who hire these factors, the managers of the collectivist agencies or the private enterprisers who are trying to make as much profit as they can, do not reduce their payment to the minimum that *is* necessary to make them available?"

What payment is "necessary" depends on the demarcated area which constitutes the point of view taken.

We cannot say what part of the payment for them is unnecessary or consists of *surplus* until we have determined the *point of view* that we are considering. That is, we must demarcate an area within the economy and consider what has to be paid to a factor to make it available within the demarcated area. What this comes to will depend on the area we mark off in this way. A factor may be fixed from the point of view of one area and variable from that of another. We shall also find that the distinction we have made

between fixed and variable factors is not quite adequate for our purpose since it is not exactly the distinction that we have to make if we wish to find out what part of the payment for various factors is *surplus* or *unnecessary payment* from various points of view. The clarification of these matters is important as well as interesting for without it we can never get a proper understanding of the rational principles that should govern many kinds of intervention by the government with economic processes. This will be particularly important for the study of the nature and the effects of taxation. We shall see that all taxation tends to fall on *surplus* in the sense here relevant and that if this is not recognized a great deal of harm can be done to the economy by attempts to impose taxes on any other part of the income of the economy.

From the point of view of one firm there is no surplus.

If we demarcate a single firm as the part of the economy for the purpose of seeing whether any of the payment to the factors of production employed in it are of the nature of *surplus* or unnecessary, we shall find no such payments. So rarely will a businessman pay more for a factor than he has to that we may consider such payments as gifts outside of the normal course of business activity. From the point of view of the firm all payments are necessary and there is no surplus.

The wider the point of view the greater is the part of the payment that appears as surplus.

If we take a group of similar firms such as is often called an "industry," we may find that as a group they are paying more to some of the factors they employ than is really necessary for the industry to pay. Some of the factors will be getting more than they could earn if they had to seek employment outside the industry. They would be willing to stay in the industry even at a lower price. But they are able to get more than their minimum demand because the various firms in the industry compete with each other. What is paid to the factor is necessary from the point of view of the firm

because if the firm did not pay it the factor could go to another firm *within* the industry. It is this competition among the firms within an industry that enables a fixed factor (which would if necessary work for practically nothing) to appropriate the whole of the surplus that comes from an increase in demand. The competition of the entrepreneurs or the managers of the collectivist agencies raises the price of the fixed factors until they absorb all the excess value of the total product over the price of the other factors. As long as the fixed factors have not absorbed the whole of this surplus, there will be profits available. The firms will try to expand and the managers will try to increase their individual outputs and so the price of the fixed factors will continue to rise.

If the industry were to be consolidated into a monopoly, it would be able to get the fixed factors for practically nothing and keep the surplus for itself. The industry then becomes a firm (even if its monopolistic practices are disguised as the policy imposed by a trade association or the like) and so there is no *surplus*.

A redistribution of surplus does not affect the optimum use of resources.

It should be remembered that the payment of surplus by the firms or by the collectivist agencies did not interfere with the optimum use of resources. On the contrary, it was the natural result of the application of the Rule or of the conditions of perfect competition that led to the optimum use of resources. The establishment of a monopoly, if it merely led to the refusal to pay the surplus to the fixed factors but did not interfere with any other phase of production, would not interfere with the optimum use of resources. There would merely be a difference in the distribution of income. (That is, under free enterprise—in the collectivist agencies where it might merely mean that the same surplus was paid into the government treasury, in the one case from the payments to factors that belonged to the government and in the other from the profit-making managers instead.) The price of every product would still equal the marginal cost (which is always to be measured in terms of the variable factors for only these can be varied at the margin), the price of each factor would still equal its

vmp (as applied to the variable factors—we have seen that the marginal product of a fixed factor raises special problems), and so there would still be the optimum use of resources.

However, it is most unlikely that any monopoly would limit itself to such socially harmless redistributions of surplus from the fixed factors to the owners of the monopoly. It would be almost certain to undertake other activities, in seeking to maximize its profits, which would bring about the departure from the optimum use of resources that accompanies a departure from perfect competition.

The definition of the industry is very arbitrary, as appears from our first reference to it as a collection of firms. It can be defined widely, to include many firms, or narrowly, to include very few firms. The wider the definition of the industry, which means the greater the area of the economy demarcated for our purpose, the greater the part of the payment to factors that takes on the nature of *surplus*. This is because the opportunities to the factors to find employment *outside* the demarcated area are smaller, so the excess of what the factor is being paid over the minimum necessary to induce it to stay in the demarcated area is greater.

Since the width of the definition of the industry is arbitrary, there is nothing against defining it as widely as we please. The very widest definition is to embrace *everything* in the industry by making it coincident with the whole economy. Everything that is paid to any factor over and above what is necessary to make it available to society is then of the nature of *surplus*. From the point of view of society this excess is an unnecessary payment. The services of the factor would still be available to society without it.

The rent of land is a limiting case.

The item that stands out most clearly in this category is the payment for the use of land. However little is paid to its owners, the land would still be there for society to use, so that all the rent is *surplus* from the point of view of society. A corollary to this is

that a tax on the rent of land, even up to 100 per cent, would not interfere with the use of resources. It would merely take income away from the owners of the land.

It is important to remember that for this purpose land only includes the original properties of the soil and not any qualities that depend on human efforts and activities and which would cease to be available if these efforts were to come to an end. If these are taxed the efforts might become unprofitable and then there would be an interference with the optimum use of resources. It is only on land proper, in the technical economic meaning of factors whose supply is fixed from the point of view of society as a whole, that one can say that the whole payment for their use is "unnecessary" for the purpose of providing the service. The payments are made, nevertheless, because they *are* necessary from the point of view of the individual firms or managers who make the payments. If the owner of the firm or manager of the state enterprise did not make the payment he would not be able to obtain the use of the land.

From the point of view of society some surplus is to be found in nearly all payments.

From the point of view of society, then, all payments to factors whose supply is fixed are unnecessary. This means not only land, but all manufactured goods and instruments of production which have already been produced and are there to be used whether or not any payment is made to their owners, as well as the services of all individuals who would continue to give these services even if the pay for them were reduced. There are many such services and even important cases where the supply would actually be *increased* if there were a reduction in their pay. We are all familiar with the farmer who produces more food because the price is low so that he has to produce more in order to be able to get enough to pay the mortgage on the farm; and the worker who has to work longer hours in order to get enough to eat if the pay per hour is less.

Increasing cost to an industry results in rent if it is due to the movement of relatively less productive factors from other industries, but not if it is due to a higher price of the diminished alternative product.

The next thing we must notice is that we cannot draw a hard and fast line between fixed factors and variable factors for this purpose even when we have determined the area of demarcation from whose point of view we are considering whether the payment for the use of a factor is necessary. This is because some of the payment may be necessary while some of it is unnecessary or of the nature of surplus. At a certain price a certain amount of a factor will be available while at a greater price a different amount will be available. Usually a higher price will call forth a larger quantity of the factor. We cannot say how much of the payment for the factor is necessary without breaking the supply up into its different units and asking what is the minimum each unit would have to get to make it available for the area of demarcation. Some units will be getting no surplus at all. Indeed, it is to get these into the area of demarcation that it is necessary to raise the price of all the other units. There will be other units which would be available even at very much reduced pay, and perhaps some at even no pay at all. It is the necessity of having to pay higher and higher prices for all the units of supply in order to increase the quantity to the amount desired within the area of demarcation that is one of the chief reasons for the rising marginal cost to an industry of producing more of a product.

An increase in the demand for a product will lead to a greater demand for the factors used in its manufacture and to such an increase in their prices as is necessary to draw the additional factors into the field (or to choke off the extra demand by raising the cost and the price of the product). If more of all sorts of factors are available on similar terms, there will be little change in their relative price and not much change in the proportions in which they are combined in production. But if some factors are difficult to expand the greater part of the burden of increased output will fall on the other factors which have to cooperate with the relatively fixed factors. The *mc* of the product is then likely to rise significantly as additional units of the available factors are obtained. The

rising cost will be due to two different influences. First, the higher price paid for the factors as more of them have to be induced to enter the industry. This will raise the price of all the units of the factors including those which were in the industry before at the lower price. The increase in the payment for these units will be an increase in their rent or surplus from the point of view of the industry. Second, the added factors will be used in a greater proportion to the factors whose supply cannot be augmented. The marginal product of the added factors will fall and the *mc* of the product will rise even if there is no increase in the price of the factors whose quantity increased. This element in the rise in *mc* will be of the kind illustrated in Figure 4 which shows *mc* rising when more of the other, variable, factors are added to one or more fixed factors to permit production to be increased.

The excess of *mc* over *ac* is absorbed in rent.

The price of the product then rises with the rising *mc*. (It is perhaps more correct to say that the *mc* rises with the increased output to meet the greater demand at the higher price.) The *mc* (which is equal to the price of the product) is then greater than *ac*. The excess of *mc* (or price) over *ac* means that there is an element of surplus which the relatively fixed factor is able to appropriate for itself as a result of the competition of the entrepreneurs for the surplus. The excess of *mc* over *ac*, which is a result of this kind of increasing cost, is absorbed by a fixed factor as rent. We have seen that in these circumstances conditions of pure competition can be maintained, so that it can be said that the absorption of this surplus by the fixed factor plays a socially useful function. It takes away the profits that otherwise might lead to too great an entry into the industry, which would reduce the price below *mc* and so result in a faulty allocation of resources.

An excess of *ac* over *mc*, as when an indivisible factor is in excess, would call for a negative rent. This is what makes perfect competition impossible in such cases.

With decreasing cost, *mc* falls *below* *ac* so that instead of a surplus which the owner of any of the fixed factors would be

delighted to appropriate, even without the assurance that he was thereby performing a social service, we now have a *deficit*. Nobody is willing to absorb this deficit or *negative* rent or surplus and so the possibility of having an ideal use of resources breaks down unless the state is prepared to absorb this negative surplus by subsidizing the industry. In the absence of such activity by the state the decreasing cost will lead to monopoly and its deviations from the optimum use of resources. A positive rent or surplus is implied in conditions of perfect competition.

The same analysis is applicable to temporal points of view.

The extent of the surplus depends not only on the spatial economic area which is chosen as the area of demarcation but also on the *temporal* area of demarcation. That is, it depends on whether we are considering the short period or the long period. If we take a short-period point of view we can consider all manufactured goods as fixed in supply because we do not then have to consider what will happen when the existing supply wears out. That problem belongs to the long period. The whole of the payment for them is therefore *surplus*, or unnecessary from the point of view of society.

The short period corresponds more to a wide than to a narrow point of view and would better be called the "shortsighted" point of view.

This might appear a little strange since a short- as contrasted with a long-period point of view suggests a narrower point of view so that one might expect it to be more like that corresponding to a narrow definition of an industry and result in a smaller part of the payment for the factor appearing as *surplus*. However, the opposite is the case.

The true temporal analogy of the narrow point of view would consist of segregating for particular attention some limited period of time and considering as surplus the excess of what a factor is paid when used *within* this period over what would have to be paid to keep it from being used *outside* the period. Corresponding to a widening of the point of view would be an extension

in time of the segregated period. This would reduce the alternative employments available to the factor outside the period and increase the part of its income that appears a surplus. Corresponding to the point of view of the whole economy would be the view that embraced the whole of the (relevant) future within the period of particular attention so that, just as in the one instance there remains no use for the factor outside the *area* considered, so in the other there remains no use for it outside the *period* considered.

The short-period point of view, however, does not pick out a short period for particular attention—a procedure which would indeed enable us to consider the alternatives of using factors inside and outside the short period. It supposes, instead, that a short period exhausts the whole of the future that has to be considered so that there is nothing outside it. It might perhaps better be called the shortsighted point of view, which so narrows the horizon that no possible uses of a factor outside the short period can be seen. It is thus really a "wide" point of view which, by completely neglecting any alternative opportunities of earnings by the factors, maximizes the part of their earnings that appears as surplus.

The long-period point of view is longer sighted, with a horizon that permits one to consider the possibility of factors being used outside the short period. In allowing these possibilities to appear it is less exhaustive, and so "narrower," and diminishes the part of earnings that appears as surplus.

It is, of course, possible to combine a wide horizon with a very long-period point of view. This, like the short-period point of view, would also eliminate alternative uses of factors in other periods, but instead of by shortsightedness it would be by considering a longsighted plan for the whole of the future period that is seen. This, however, is not what is usually called a long-period point of view.[1]

[1] The last few paragraphs are taken almost verbatim from my article "From Vulgar Political Economy to Vulgar Marxism," *Journal of Political Economy*, August 1939, p. 563, footnote.

CHAPTER 19. SURPLUS AND TAXATION

So far we have considered only *surplus* received by the *seller* of a factor of production. This was the excess of the money payment received over and above what was necessary to get the factor of production into its actual use. This was relative to the point of view taken, which could be wide or narrow. If we took a narrow point of view, interpreting the actual use as use within the firm that used it, there would not normally be any surplus since the firm would not pay more than was necessary. If we took a wider point of view, interpreting the actual use as use within the industry, then there would appear as surplus the excess of what the factor was paid over and above what the owner would be willing to take rather than go outside the industry. The latter might be appreciably less than the actual payment which the competition among the different firms within the industry enabled the factor to get. If we defined the industry more widely, the alternative opportunities for outside employment of the factor would be further reduced and a still greater part of its actual payment would be *surplus*. The area of demarcation which we are now calling the use of the factor might be extended in many different ways. The "industry" might be extended by including the production of similar products or competing products or products that used similar materials or techniques. It might be extended on a purely geographic basis, so as to embrace all economic activity within a city or county or state or region or continent. Every time the area of demarcation was widened in any way the alternative opportunities of the factor would tend to be still further restricted, the minimum needed to get it to work inside rather than outside the area would be smaller, and the surplus, or the

excess of the actual payment over this necessary minimum, would be greater.

The area can be extended in time as well as in space. It might be necessary to give the owner of a factor a certain amount to make it available for present use in an industry because if it is not used it will be available next month. The payment for its use *now* must therefore be great enough to make it worth while not to wait for next month. Its use next month is an alternative occupation outside the area of demarcation. If the area of demarcation is extended to include the use of the factor next month (*e. g.*, if the area of demarcation is so extended as to include all uses of the factor in this industry during the next 10 years), this alternative is no longer available and the *surplus* will appear to make up a greater part of the payment to the owner of the factor.

Surplus also applied to the purchaser, and in this case it depends on the demarcation of the area *from* which he buys.

All of this analysis is applicable to the *purchaser* of anything, in perfectly symmetrical fashion. When a man buys something he may be getting a surplus in the sense of having to pay less than he would be willing to pay for the good. Again the amount of this surplus depends on the area of demarcation, though this time the area of demarcation will stand for the area from which he buys instead of that to which he sells. If the area is seen from a narrow point of view, including only one firm, there will normally be no surplus. He can buy the same thing from another seller at the same price and so there is no surplus in his being able to buy it from the firm where he actually gets it. As the area of demarcation is extended to include other possible sources of supply, the possibilities of buying the same thing at the same price, or even at higher prices, become less and less, and surplus emerges and becomes greater and greater. Further extension of the area to include the possibility of buying substitutes for the good will make it more and more important for the good to be bought and the surplus will continue to grow. The area can be extended in time

too, just as can that of the seller, each extension tending to make the amount of surplus appear greater.

All taxation falls on surplus.

All taxation falls on surplus in the sense here developed. This is because it is only if the whole of the tax can be taken out of the surplus that the transaction on which the tax is imposed will continue to be carried on. If the tax is greater than the surplus there would be a net loss on the transaction and it will be abandoned.

We have seen that the amount of surplus involved in any transaction will depend on the area of demarcation. The appropriate area of demarcation to take is that on which the tax is imposed. If a tax is imposed on sales by or purchases from a single firm, where there is no surplus, because the same things can be bought from and sold to other firms that are not subject to the tax, the only effect of the tax will be to put the firm out of business. No tax will be collected. The sellers to and purchasers from the firm will buy elsewhere at no inconvenience to themselves.

If there is some inconvenience in buying from other firms or in selling to them at the same price, there might be some loss by the individuals affected, but that would be only because the firm had not been taking full advantage of its possibilities. It could have appropriated practically the whole of the difference to itself by charging prices high enough and paying prices low enough to offset almost the whole of the advantages of dealing with it.

The firm itself would lose as a result of the tax since for it the tax is on a wide basis, the area of demarcation covering all operations which would leave it subject to the tax. If the surplus enjoyed by the firm by virtue of its buying and selling within rather than outside the area of demarcation is greater than the amount of the tax, the firm will absorb (that is, pay) the whole of the tax and continue operating just as before (unless the amount of the tax depends on the way the firm carries on business or on the amount of business done, when it would make some adjustments in trying to minimize the sacrifice).

If a tax is imposed only where the surplus is at least as great as the tax, it does not interfere with the operation of the economy. Every transaction that was worth while before is still worth while, although the benefit is not so great. The only effect is to make the taxpayer poorer by the amount of the tax, and as this presumably is what the tax is intended to bring about, everything is perfectly satisfactory from the point of view of society in general (although the taxpayer might well consider that his selection to bear this burden was unjust).

A tax greater than the surplus from the transaction on which it is based will prevent the transaction and destroy the surplus, bringing about a social loss.

However, it is practically impossible for any tax on a transaction to fall only where there is a surplus greater than the amount of the tax. There will be some people called upon to pay the tax who do not make as much surplus as this. These people will rather forego the surplus they are getting. Whenever this happens there is a net social loss. The surplus thus foregone is lost to the people affected and is not offset by any gain to the government because there is no tax collected. The tax is collected from those who make a surplus equal to or greater than the amount of the tax, and any surplus which is not enough to support the tax is destroyed. It is this destruction that constitutes the harm in taxes where they are not directed carefully so as to fall only on surplus.

The tax may be shared between the buyer and the seller. Where the sum of their surpluses comes to more than the tax, the transaction will continue and the surplus will not be destroyed. The tax may even be shared among many people. The buyer of the taxed article may be able to shift some of the tax to those who buy the product he makes out of it (if his competitors are also induced to raise their price), and the seller may be able to shift some of the tax onto the people from whom he buys the factors out of which he makes the taxed article (if his competitors are also induced to lower their price for these factors). But the transaction and the production of the surpluses will continue only if

the sum of the surpluses comes to more than the tax. Wherever the sum of the surpluses from taxed activity is less than the amount of tax, the activity is stopped and the tax is destructive of the surplus.

The land taxers were fundamentally right in stressing the "surplus" nature of land rent.

The moral of this is that only such taxes should be imposed which fall directly on surplus so that they do not result in this net loss to society. (That is, unless it is the real purpose of the tax to eliminate the activity itself.) This is the idea of the land taxers who point out that a tax on land falls entirely on surplus and so does the least harm to the economy, and that no other taxes should be imposed at least until the whole of the rent of land has been taxed away and this has been found not to be enough. A tax on pure rent of land (in the economic sense of the original and indestructible powers of the soil) will always fall entirely upon surplus and will never interfere with the use of resources or bring about the destruction of surplus that any tax on transactions of any kind will entail.

Fortunately we have developed another form of taxation, which is even better than the tax on land in minimizing the amount of surplus it destroys. It is true that a tax on pure land would *never* destroy any surplus since it would fall entirely where there was enough surplus to pay the tax. (Surplus is destroyed only when a tax is imposed where there is not enough surplus to pay the tax.) But it is difficult to disentangle pure land from the manufactured qualities provided by human activity and depending on more human activity to keep it going. Any actual law would to some extent (though not very greatly) destroy some of the surplus available from activities directed to products, such as increased fertility of the soil, which come to be classified as land by the tax collector.

Furthermore, a heavy tax on the rent of land would amount to a tax on the owners of a particular kind of property, and there does not appear to be any particular reason for desiring to make

these particular property owners poorer or to reduce their claims on the social output more than the claims of others who happen at the moment to be holding their property in forms other than land. These may have just sold their land and this does not seem to be a very good reason for permitting them to keep all their wealth while penalizing those who happen to have exchanged other wealth or savings for land.

But in assuming that the government *needs* the revenue they are reduced to the weak negative argument that the land tax would do a minimum of harm.

Underlying the argument of the land taxers is the thesis, by no means peculiar to them, that the government *needs* to raise a certain amount of money by taxes. Granted this, the argument is very strong for making use of a method of raising the money which destroys as little as possible of surplus, and a tax on land satisfies this requirement very well. Against the sentimental argument that it might fall very hard upon some widows and orphans who happen to have invested their savings in land, they can reply with another sentimental argument that the land was not made by anybody so that nobody had any right to it in the first place. The present owners have unfortunately acquired some stolen property. And anyway, the imposition of the law would be gradual so that the blow would not be so terrifying and if the worst came to the worst there might be arranged some sort of social security provisions to deal with cases of real hardship.

The real criticisms of the argument are two. First, there is never any need for the government of a well-established sovereign state to raise an amount of money by taxes just because it needs the money, if it does not wish to bring about the actual effects of the tax on the economy. If there is merely a need for money it is easier to borrow it and much easier to print it. A government has to tax because it wishes to reduce the wealth or the expenditure of the taxpayers, possibly in order to check the total rate of spending by the public so as to prevent inflation. A rational taxing policy is directed at the effects from the taxpayers' having to pay the money and not at the government's gain in getting it. Conse-

quently the argument for concentrating taxation where it will do least harm in the destruction of surplus becomes very weak. The harm can be avoided entirely by not having any taxes at all. Taxes may be necessary because in their absence (and the expansion of government spending on all sorts of necessary services) there would be too much spending and the danger of inflation. In that case taxes must be directed to reduce spending where it is most desirable for spending to be reduced, and this cannot be sacrificed entirely to the consideration of the avoidance of the destruction of surplus by taxing only the surplus to be found in the income from the ownership of land.

Land is not the most important source of surplus.

The other criticism is that land is by no means the only or even the most important source of surplus. From the point of view of society (and it is only from the point of view of society that the rent of land is all surplus), all the income over and above what is necessary to keep the population healthy enough to be able to keep on working constitutes surplus. Perhaps three-quarters of the income of a rich country like England or the United States constitutes surplus, and the rent of land is only a small part of this. From a short-period point of view all income from property is surplus in the same way as the income from the ownership of land, since the property is there and is available for society to use, however little is paid for it. Nearly all income above subsistence is surplus since almost every man would be willing to work for subsistence if he had no other opportunity of keeping alive. This permits the government to impose its taxes, whenever taxes are necessary, on surplus, while at the same time directing the tax so as to reduce income or wealth or spending only where it is socially desirable to reduce these.

Where taxation is necessary the personal income tax seems the least harmful.

The instrument for doing this is the personal income tax. All other taxes could be eliminated except where they were needed

for discouraging some particular form of activity such as drinking or smoking or speculating in land in a way which resulted in the withdrawal of useful land from social use.

The marginal income tax should not be greater than one hundred per cent.

An income tax, since it does not depend on how an individual spends his income, will not interfere with particular expenditures and so will not destroy any of the surpluses available from these transactions. Any activity that was profitable in the absence of the income tax is still profitable even if some of the profit is taken away by the tax collector. That activity which gave the greatest income before the income tax is deducted will still give the greatest income as long as the *marginal* income tax is less than 100 per cent, so it will be profitable for every income taxpayer to do all that it was profitable to do in the absence of the income tax.

Income tax does not fall entirely on surplus in connection with saving ("the double taxation of saving"), leisure, and risky investments.

There are two ways in which the income tax can interfere with the use of resources and destroy some surplus. This is where the income tax departs from its general principle of not being affected by the way in which income is directed. The exceptions are in the decisions between spending income in general or saving it and between spending income on leisure (by working less and getting less income) or on the goods that can be bought with money income.

In the first case there is a deviation because most income tax laws are so designed that interest on saved income is regarded as income again and so is taxed again. This has been called "double taxation of saving" and means that the return to the saver is less than the current rate of interest. If the rate of interest is 10 per cent and the rate of income tax is 20 per cent, an individual saving $100 gives up $100's worth of consumption. He gets $10 a year interest on this, but has to pay $2 in taxes on the interest, so the

net return on his saving is only 8 per cent. The effect is exactly as if there were a 10 per cent tax on saving in the absence of any income tax. Whenever the surplus on saving is less than this, the saving will be prevented by this operation of the income tax and this surplus will be destroyed.

The "double taxation of saving" may be partially justified as a measure for equalizing wealth.

This is not a very serious weakness of the income tax. It can be remedied by provisions to exempt savings from the income tax, making it an expenditure tax. It may even be justified by the consideration that the purpose of the income tax is not only to reduce spending but to diminish the inequality of wealth, and this objective could be evaded by high saving if saving were exempted. A further justification is that in fact it is not certain that people will save less if the net rate of interest is lower. This will depend on whether the *need* to save more, when the rate of interest is lower, in order to make a certain provision for the future, will more than offset or less than offset the *discouragement* that a lower interest rate is to those who save for the sake of the interest they would get. It seems probable that the total rate of saving will not be affected much by small differences in the rate of interest, so there will be no great destruction of surplus.

These justifications are perhaps not quite adequate. As long as the net rate of interest to the individual saver is not equal to the rate of interest on the market, there is a deviation between *msb* and *msc*, and the loss from this is what we have been speaking about in terms of the destruction of surplus. If it is desired to take measures for the equalization of income, it might be better to deal with that through an inheritance and gift tax. But these are matters of secondary importance. It does not seem very likely that there would be any very great destruction of surplus from this "double taxation of saving."

The second deviation from the optimum use of resources as a result of income tax arises from the possibility of an individual's deciding to work less and enjoy leisure instead of income. He can

then be said to be devoting his potential income to the purchase of leisure. The disturbance arises from the circumstance that before he can devote his income to the purchase of anything other than leisure he has to pay the income tax, while if he decides to sacrifice his potential income-earning efforts to the enjoyment of leisure he does not earn any discernible money income and thus avoids paying the income tax.

The result is that leisure is especially favored just as saving is especially discriminated against by a simple income tax. The effect of this is to induce less work to be done and more leisure to be enjoyed than if the tax were to fall completely on surplus and not affect the use of resources at all. This could be corrected by adding to a man's income, for taxation purposes, that part which he was able to earn but did not because he preferred to work less than some standard amount, just as in England the rent payable for a house occupied by the owner is added to his income for income tax purposes. There would still remain the deviation from the optimum use of resources inasmuch as individuals would be free to choose less arduous occupations at the cost of a decrease in their income, and in so doing they would be avoiding the income tax on the potential income that they were sacrificing for the sake of the easier work. There would therefore be an excessive tendency for these kinds of work to be sought after. However, these deviations would not be of very great magnitude.

The immunity of leisure from income tax does not mean that there will be more leisure and less work.

Just as we could not be sure that the "double taxation of saving" would decrease the amount saved, so we cannot be sure that the immunity of leisure from the income tax will increase the amount of leisure enjoyed and so reduce the amount of work done. The greater relative attractiveness of leisure because it escapes taxation might be more than offset by the need to earn more money to take the place of the income taken away by the tax collector. One can only say that there would be a tendency for the amount of leisure to be greater than if the income tax were so

amended that the part of income sacrificed for the sake of enjoying leisure were taxed in exactly the same degree as the income devoted to any other purpose.

Closely related to this exception is the case where a steeply progressive income tax, which taxes income in higher brackets much more than it taxes income in lower brackets, discourages investment in risky enterprises. The investor will not be inclined to risk his capital if the prize for which he is hoping and for the sake of which he is taking the risk will be forfeited in large measure if he should turn out to be lucky. It will be like buying tickets in a lottery where the prize is taxed away so that it becomes a case of "Heads I lose, tails you win."

The effect of progressive income taxes in discouraging risky investment is easily exaggerated.

This objection to a progressive income tax is valid if applied to an individual who is faced with incurring such a large loss or getting such a large gain that in the one case he will perhaps pay no income tax at all and in the other will pay at the millionaire rate. This kind of risky undertaking is discouraged by a progressive income tax. The objection does not hold, however, for small risky investments which leave the investor in the same income tax bracket. In that case the income tax has no effect whatever on his willingness to invest. It is true that if the venture turns out successful he will have to turn in a large part of the gain to the government, but on the other hand, if the venture turns out to be a failure, his income is less by the amount of the loss and he will save a corresponding amount in taxes. If he gets only 50 per cent of the gain, he risks only 50 per cent of the loss. The government gets half of his gain, but it also bears half his loss. In effect what the tax does is to keep capitalists on a kind of commission basis. They play the same game but for smaller stakes. They play exactly the same game because it is still true that the greater their profit the greater the amount left for them after paying the income tax.

This is true only if the taxpayer stays in the same income tax range, so that increments and decrements of income are deflated

by the income tax in exactly the same amount. If his losses are deflated less than his gains, they will appear greater to him and he will become too careful. If his gains are deflated less than his losses he will become too careless. It is therefore necessary, first that he should be able to deduct losses from profits before making his tax return, and second that in the range of income affected the *marginal* rate of tax should be constant.

The second consideration seems to be contrary to the idea of a *progressive* income tax. However, it is only in very high income levels that this consideration is at all significant. It is only the decisions of very rich men to make new investments or not to make new investments that are significant for the level of activity and the efficiency of the economy. Even fairly well-to-do middle class people are not affected very much by the differences in tax rates applicable for the different income levels in which they are likely to find themselves. For these very rich men the *marginal* rate of tax can very well be constant. However high it is, even if it is 90 per cent, what was profitable in the absence of any income tax will still be profitable even though they now get only 10 per cent of the profit and risk only 10 per cent of the losses. And even though the *marginal* rate of tax is constant, the *average* rate is still progressive. The greater an income the greater is the part of it that is subject to the top marginal rate.

This arrangement works satisfactorily only as long as these investors remain in the income range where the same marginal income tax rate applies. This means that they cannot risk losing substantial parts of their fortunes on these terms. However, it is just for the purpose of providing risky enterprises with large sums of money that single individuals were not able or willing to venture that the joint stock corporation was invented and developed. This can permit each of many men to risk a small part of his wealth in many such enterprises. On each of these small investments the high rate of income tax has no harmful effects, and so it is possible for great and risky enterprises to be undertaken even if there should be a 90 per cent tax on income above a certain level.

It will still be true that the income tax at these high levels will result in a strong divergence between the money reward from

work, which is subject to the high income tax, and the return in enjoyment from leisure, which escapes the tax. But this is not of importance because the high incomes are rarely the result of work. For the most part they are the income from property, and the work involved in deciding how to invest is done by paid experts. In those few cases where individual effort is of importance the work is usually of a kind that is sufficiently interesting to bring about the socially desirable amount of work whatever the payment for it.

CHAPTER 20. PRODUCTION AND TIME

Technical marginal transformability refers to transformability over time as well as to any other form of transformation.

We have seen above that the optimum use of resources is achieved only if the marginal substitutability between products in consumption is equal to the technical marginal substitutability or transformability between them in production. So far we have applied this principle to *different* goods available at the *same* time, but it is just as applicable between the *same* technical goods available at *different* times. Such technically indistinguishable goods are economically quite distinct just *because* they are available at different times.

The Rule calls for the equalization of present and future values of goods, making allowance for the cost of transformation over time by storage or otherwise.

In our discussion of speculation we saw how it is possible to transform goods available at one point of time into goods available at another time by *storing* them until some future time. This transforms, say, wheat-today into wheat-tomorrow. If there is no wastage in storage and if no other factors are used up in the process, 1 bushel of wheat-today is technically substitutable for 1 bushel of wheat-tomorrow, and the optimum use of resources requires that these two goods have the same price. This is brought about in a collectivist economy by the Rule which directs wheat-today to be applied to the production of wheat-tomorrow until *pf* (the price of a bushel of wheat-today) is equal to *vmp* (the value of a bushel of wheat-tomorrow). When this is done the consumer,

in adjusting the expenditure of his income between the two goods, will acquire such quantities of them as make their marginal utilities to him proportional to their prices; and with the prices equal their marginal utilities will be equal too. The *msc* of a bushel of wheat-today is a bushel of wheat-tomorrow and the *msc* of a bushel of wheat-tomorrow is a bushel of wheat-today so that *msb* = *msc* and the optimum use of these resources is achieved.

Normally there is some cost of storage, some wastage, and some risk of spoilage. This means that to produce a bushel of wheat-tomorrow takes more than a bushel of wheat-today, and that other factors have to be applied in the course of the transformation—use of storage equipment, labor of storing, transportation, grading and checking, and what not, so that it is not simply a bushel of wheat-today that is transformed into a bushel of wheat-tomorrow, but the sum of all these factors. The Rule then directs that all these factors (including the bushel of wheat-today) should be applied to producing wheat-tomorrow until the value of the wheat-tomorrow thus produced (which is less than a bushel because of the wastage) is equal to the value of a bushel of wheat-today *plus* the value of the other factors that have to cooperate in the process of storage.[1] This means that where transformation of goods-today into goods-tomorrow is brought about by storage the price of the goods-tomorrow should be greater than the price of goods-today by the marginal cost of storage.

More important than storage is the indirect transformation of present goods into future goods by shifting resources from producing present goods to producing future goods.

Storage is, however, by no means the most important way of transforming goods in the present into goods in the future. It is only a way of making relatively minor adjustments. The important way of transforming goods in the present into goods in the future is the indirect one of shifting productive resources from producing

[1] The price of each factor will also be equal to its *vmp* if the proportion between the factors is variable, and to the *net vmp* in any case. Simply adding these prices and (net) *vmp*'s gives the same equality between the value of the marginal factors and their joint *vmp*.

the former into producing the latter. Increasing the supply of goods in the future always means consuming less in the present, but the best way is normally not simply to store the goods withdrawn from present consumption, but to produce less for the present and use the resources that are set free to improve and increase the equipment that will permit a greater output in the future.

Improvement of equipment permits future output to be increased by *more* than the present sacrifice.

If equipment of the best and most expensive kind were already available for the production of the goods in the future, and available in the greatest quantities that could be used, this possibility would not exist. No additional equipment would permit future output to be increased, and the only way to increase future output would be to store present output. But this is not the case, has never been the case in the past, and does not appear likely to be the case for some considerable time in the future. Nearly all production is carried on with inferior equipment. Improvement in the equipment would permit the same factors in the future to produce a larger or a better product, but the equipment is not perfected. There are two possible reasons why the equipment has not been perfected. One is that although better equipment would permit more to be produced in the future, the increase in potential future production is less than the current output which would have to be sacrificed to make it possible. If this were true, no present resources would be devoted to increasing or improving equipment for production in the future. The fact that some of the current supply of resources is nearly always directed toward increasing future productive capacity indicates that at least to some extent it is possible to increase future output by more than the sacrifice of present output that this involves. The failure to perfect equipment must therefore be due to the second reason, which is that to provide everybody with the very best equipment[1] would involve

[1] By the best equipment is here meant not the technically best but the economically best. Equipment is *technically* better if it permits *any* increase or improvement in output from the same resources cooperating with the equipment (or any decrease in the

too great a sacrifice of current consumption. Normally there is some sacrifice of present consumption for the sake of improving the equipment to produce for the future. We are a long way from the position where the very best equipment is already available everywhere. Consequently it is possible by sacrificing output now, to devote the resources set free to the production of goods in the future by means of better equipment and so *to increase future output by more than the present sacrifice.*

This is better than providing for the future by storage which increases future output by less than the amount sacrificed in the present. That is why the indirect way of transforming goods from the present to the future is preferred. Steel next year—or in 10 years' time—could be made available simply by storing steel out of this year's production, but it is better, instead of storing this steel, to use it to build bigger and better blast furnaces which would permit the future output of steel to be increased by *more* than the present sacrifice.

Future goods must therefore be cheaper than present goods.

The possibility of improving on our equipment means that if instead of replacing one of our existing plants as it wears out we were to build a better one we would in the future be able to produce, say, 10 units more of product per annum while using the same available resources in cooperation with the improved equipment. But to build the better equipment instead of merely replac-

resources needed to cooperate in producing the same output), but it is not *economically* better if more resources are used up in making the new equipment than are set free when the new equipment is in use. If the new equipment is expected to remain in use forever (by replacement when it wears out physically), the ultimate total saving of resources must come to more than the resources devoted to setting up the new equipment in the first place, and in that case any technical improvement would also be an economic improvement. But there is no reason for believing that the new equipment would forever continue to be the last word, so it is possible for a technical improvement not to be an economic improvement. Our argument rests on the belief that the existing equipment is not merely below the technical optimum but also well below the economic optimum. That is to say, in many parts of the economy new or additional equipment would not merely make it possible to produce in the future with less cooperating resources per unit of output, but would, over the period in which it was worth using the improved equipment, save more resources than are used up in improving the equipment.

ing the kind of equipment we had before involves using up more resources and this is possible only if we are prepared to set these additional resources free by consuming, say, 100 units less of resources this year. In effect we will be transforming 100 units of current consumption into an additional flow of 10 units of consumption per annum in the future beginning with next year.

It is also possible to make a similar transformation in the reverse direction. We can increase our consumption this year if we replace our equipment that currently is wearing out not with better equipment or even with the same kind of equipment, but with a poorer kind of equipment that uses up less resources and so sets free enough to permit us to produce and consume an additional 100 units of consumption this year. But with the poorer equipment the same cooperating resources will be able to produce 10 units less of consumption goods per annum. We would then be transforming 10 units of consumption per annum, beginning with next year, into 100 units of consumption available this year.

There is a certain awkwardness in considering the transformation of a stock of 100 units at one point of time into a flow of 10 units per annum available at another time, or vice versa, and since this is not an aspect of the phenomenon with which we are now concerned we can avoid this by a simple device. We can consider the improvement in equipment as consisting of elements or atoms of investment affecting only two years, one year and the year immediately following.

This can be done by supposing that this year we improve our equipment, sacrificing 100 units of current consumption goods and reaping an additional 10 units the next and every succeeding year, and that next year we reverse the process, increasing our consumption (next year) by 100 units and suffering a decrease of 10 units per annum in the succeeding years. The net result is then that this year our consumption is reduced by 100 units, next year it is increased by 110 units (10 units consisting of the first fruits of this year's sacrifice and 100 units consisting of next year's disinvestment), and in all following years consumption is just as if neither the investment nor the disinvestment had taken place. The scheme is shown in the following table:

Year 1	*Year 2*	*Year 3*	*Year 4*	*Year 5*
−100	+10	+10	+10	+10 etc.
	+100	−10	−10	−10 etc.
−100	+110	....	....	.... etc.

The first line shows the effects of the investment in year 1. The second line shows the effects of the disinvestment in year 2. The third line shows the combined effect of both the investment and the disinvestment, 100 units of consumption this year is transformed into 110 units available next year, all other consumptions being unaffected.

It is of course wrong to suppose that this is the kind of thing that goes on in the actual world. People do not improve equipment one year and then disinvest it the next year. But it is true that the sacrifice of 100 units this year makes *possible* an increased output of 110 units the following year. It is unlikely that the possible extra output will all be taken out in consumption the next year. It is much more likely that the equipment will be increased still more. But that can be considered as the reinvestment of the available greater consumption next year for the sake of still greater outputs in the more remote future, and we can still use as our unit of investment the transformation of 100 units of one year into 110 units of the following year.

This means that a ton of steel sacrificed now permits, say, 1.1 additional tons to be produced next year, and if the postponement is for a longer period this could in turn be devoted to improving equipment still more, so that a ton of steel sacrificed now might permit 2 additional tons to be produced in 6 years' time. This would mean that the future *mp* of a ton of steel is *more* than a ton of steel (instead of *less*, as it would be if it were provided by storage), and since by the Rule the *value* of the *mp* must be equal to the price of the factor, steel-in-the-future must be *cheaper* than steel-in-the-present, and the farther into the future we go, the greater is the *mp* and the lower must be the price.

The substitution between present and future steel is easily seen because it is relatively direct. (At least it appears direct if we do

not distinguish between different kinds of steel.) But the same holds for any other good. If fewer hats are made this year, the resources set free can directly or indirectly take the place of other resources that can be set free from producing for present consumption and turned toward the improvement and increase of hat-making machinery or to improvements in breeding sheep to provide more and better wool for hats in the future, and so the sacrifice of hats now will permit an increase in the supply of hats in the future greater than the sacrifice of hats in the present. There will therefore also have to be a falling price of hats over time, as long as the Rule is properly applied throughout the economic system, and similarly for other goods.

Even in a stationary society the *opportunity* of improving equipment must make future goods cheaper than present goods, and must be the same for all goods.

In a stationary society, in which nothing is done to improve equipment, the same output of the various goods is produced year after year and there is the same demand for them so that relative prices cannot change. There is still the *opportunity* of improving equipment. A ton of steel, if it were sacrificed, would permit more than a ton to be added to next year's product. The *mp* of a ton of steel-today is more than a ton of steel-next-year, and the price of steel must be less next year than this year. Since the relative prices do not change, the price of next year's hats must be below the price of this year's hats in the same proportion as the price of next year's steel is below the price of this year's steel, and the same relationship must hold for every other good. All prices must be falling *at the same rate.*

This may seem rather queer. The rate of fall in price of any good must correspond to the proportion in which a unit of it sacrificed this year will permit a larger output of the same product next year. How does it come about that steel and hats and everything else have the same technical marginal transformability of present output into future output? How does it happen that if 10 tons of steel sacrificed now permit 11 more tons to be produced

next year, it must also be true that 10 hats sacrificed this year will permit exactly 11 more hats to be produced next year, that 10 bushels of wheat sacrificed this year will permit exactly 11 more bushels of wheat to be produced next year, and so on for every other single good?

The opportunity of transforming present into future goods is equalized for all goods by the possibility of indirect transformation through the good that gives the greatest yield of future over present product.

The answer is that the allocation of the various factors of production among the different products in the course of attaining the optimum use of resources (whether this is brought about by the application of the Rule or by perfect competition) has the effect of equating the relative productivities of the different factors in the different uses. The factors used in the production of present hats could be directed to the production either of future hats or of future steel. Similarly the factors used in the production of present steel could be directed to the production either of future steel or of future hats. If there should be greater opportunities in the steel industry than in the hat industry, any factors set free from the production of present hats would be devoted to improving the equipment for producing future *steel.* This would permit the same amount of steel to be produced in the future with less factors of production cooperating with the improved equipment, and the factors set free would be available for the production of future hats by the old method. This would be the more efficient way of producing more future hats by the sacrifice of present hats. Whatever the particular present goods whose technical transformation into future goods is being considered, the transformation takes place via the most effective route, and that is the one where there is the greatest opportunity for improving on the equipment. Consequently the transformability of each present good into the same good in the future will be at the same rate, and all prices in the stationary economy will keep falling at the same rate.

The rate at which all the prices fall will depend on the degree to which the postponement of consumption permits an increase in

future output by setting free resources to improve equipment. In a very rich economy with very good equipment the opportunities for further improvement would normally be relatively small. Future prices would not be much lower than present prices, and the rate of price fall would be small. In a poor economy where a postponement of consumption permitted great improvements in equipment the rate of price fall would be much greater.

Factor prices would also have to be lower in the future than in the present.

The fall in price would apply just as much to the factors as to the products and for a parallel reason (as we might expect from the repeated evidence of their symmetrical relationship to the production process). The price of the *product* falls because an additional unit of factor available now is able to contribute more toward output next year (by permitting an improvement in equipment) than it can contribute toward output this year. The larger future *mp* must have the same value as the smaller *mp* this year (since the Rule makes both *vmp*'s equal to the price of the factor) and so the price of the future product is less than that of the present product. The price of the *factor* falls because a unit of factor available only next year, if applied to direct production, will be able to produce only as much next year as a unit of factor available now is able to produce this year. The value of the product will have fallen, however, so that the price of the factor, which is equated to the value of its marginal product, will have fallen in the same proportion as has the price of the product. A unit of factor available this year can also be applied to increasing the output next year if it is directed to the improvement of productive equipment. In that case it can contribute more toward next year's output than a unit of factor that is not available until next year, for the latter, if it is to affect next year's output at all, can be used only in direct production. The greater productivity, in terms of its contribution to next year's output, of a unit of factor available this year over that of a unit of the same factor available only next year will be measured by the technical possibilities of increasing next year's

output by sacrificing some of this year's output. Consequently the greater marginal product of a factor when used this year to add to next year's output will be exactly offset by the lower price of next year's output. The *vmp* will be the same whether a factor is used directly to increase this year's product or whether it is used to increase equipment to make possible a greater increase in output in the future.

The inconveniences of falling prices can be avoided by *a positive rate of interest on money.*

Such a state of affairs, with the prices of all products and of all factors falling, would not only be very inconvenient for bookkeeping purposes, but would require a perfect flexibility of all prices including rents and wages. Any inflexibility would prevent the optimum use of resources, and the increasing value of money as prices fall would provide a great incentive for individuals to hoard money. Where the supply of money is produced at practically no cost and controlled by the government in the general interest, this is not serious, for enough money could be printed to satisfy this desire without permitting the hoarding to interfere with the flow of money expenditure necessary for the health of the economy, but even then it may occasionally be troublesome.

All these inconveniences and possible inefficiencies can be overcome by a very simple device. A rate of interest on money is established equal to the marginal yield from the postponement of output (which is the rate at which prices would fall in the absence of the device). If a ton of steel sacrificed this year permits 1.1 tons of steel to be added to next year's steel output, the marginal yield from postponing the consumption of steel is $\frac{1}{10}$ or 10 per cent per annum. If the marginal yield from postponing output (or the marginal yield from anticipating input by applying a factor a year earlier) is 10 per cent per annum, all prices would fall in this proportion.[1]

[1] This would be a price fall of slightly less than 10 per cent per annum. A fall of $9\frac{1}{11}$ per cent is the same proportional change as an increase of 10 per cent because 100 is 10 per cent greater than $90\frac{9}{11}$.

An interest policy can minimize necessary price changes, thereby avoiding resistances to price changes which would interfere with the optimum use of resources.

Interest at this same rate of 10 per cent per annum charged to all producers on money borrowed to pay for factors in anticipation of the sale of their product, would prevent the prices from falling. It will still be true that 11 tons of steel next year can be produced by the same resources that would yield only 10 tons this year, but when the 10 per cent additional interest has been added to the cost of producing steel for next year with this year's resources, the cost per ton will have been raised to the same level as steel this year.

This is the way the matter is expressed in terms of production. In terms of the transformation of one product into another we can express the substance of the matter by saying that 10 tons of steel today are technically transformable into 11 tons of steel next year, but the 10 per cent interest that is payable when output is postponed for a year makes the cost of the 11 tons next year 10 per cent more than the cost of the 10 tons this year, so the cost per ton (and also the price per ton) is the same this year as next. Similarly with the factors. To produce the same product next year requires eleven units of factor available next year as against only ten available this year. The greater productivity of this year's factor over next year's (the marginal yield from anticipating input or applying factors earlier) is exactly offset by the 10 per cent interest that has to be paid for acquiring the factor a year earlier. The value of the marginal product after the interest has been deducted (the *discounted* value of the marginal product) is the same for the factor in both years and so the price of the factor will be the same in both years.

What the rate of interest does, in effect, is to make the value of money fall at the same rate as the value of goods so that the value of goods stays constant in terms of money. The fall in the value of goods as well as of the money must be measured in terms of the goods or money at some base period. Ten tons of steel are still exchangeable for 11 tons next year. Instead of steel having a lower price per ton next year ($\frac{10}{11}$ of this year's price) it has the same price as this year, but if 10 tons of steel are to be sold this

year the money proceeds would now earn 10 per cent in interest in the course of the year so that they would be able to purchase 11 tons of steel next year.

The device of charging interest establishes the same relationship between *money* at different times. Ten dollars this year is worth $11 next year because of the interest it can earn. Next year's dollar is worth less than this year's dollar by the interest that a dollar this year can earn before it becomes one of next year's dollars. The rate of interest depreciates the dollar at the same rate at which products and factors depreciate, and this prevents the money prices of products and factors from falling.

The rate of interest raises the marginal *value* yield from postponing output or anticipating input.

Another way of expressing this is to say that the introduction of the rate of interest (which means increasing the rate of interest from zero to some positive rate) has the effect of raising the marginal *value* yield from postponing output or anticipating input. If there were no rate of interest (which means a zero rate of interest) the 10 per cent greater output of steel next year would have the same value as the smaller physical amount sacrificed for it this year. The marginal *value* yield from postponing output or anticipating input would be zero. By checking the fall in prices the rate of interest raises this marginal *value* yield from zero to whatever is the rate of interest. A rate of interest of 10 per cent keeps future prices from falling at all so that the sacrifice of $100's worth of steel now permits additional production of $110's worth next year. The price of steel being the same, the greater quantity of steel will have a correspondingly greater *value*. The marginal value yield will be 10 per cent and equal to the marginal physical yield from postponing output or anticipating input.

Certain general relationships must hold between the marginal physical yield (from postponing output or anticipating input), the rate of change of prices, and the rate of interest.

It follows from this that a rate of interest less than the marginal (physical) yield from postponing output or anticipating input

would not be sufficient to prevent the prices from falling, but would reduce the rate at which they fell, so that the marginal physical yield would be roughly equal to the rate of fall of prices *plus* the rate of interest. A rate of interest greater than this marginal physical yield would cause prices to *rise* by roughly the excess of the rate of interest over the marginal physical yield.[1]

This seems contrary to our experience in the actual world where an increase in the rate of interest has the effect of *reducing* prices, but this is due to a difference in some implicit assumptions that will have to be brought into the light when we come to deal with unemployment in Chapter 22.

With relative prices variable only some arbitrarily chosen index number can be stabilized.

When we leave the stationary economy things become somewhat less neat, but the same fundamental principles still hold. Relative prices do not remain the same. There are changes in the relative supplies of factors and in the relative demand for products. But the optimum use of resources and the Rule still demand

[1] More accurately, the rate of *technical* transformation of present into future goods will be equal to the rate of *market* transformation of present money into future money (by lending it out at the current rate of interest) *multiplied* by the *market* rate of transformation of present goods into future goods (by exchange at their relative money values which reflects the rate of fall of their prices). If the rate of *technical* transformation of present goods into future goods is 10 : 15, and the rate of *market* transformation of present into future money is 10 : 12, then the rate of *market* transformation of present into future goods will have to be 12 : 15 because 10 : 15 = 10 : 12 × 12 : 15. In this example the marginal physical yield from postponing consumption is 50 per cent (since 15 is 50 per cent greater than 10), the rate of fall in prices is 20 per cent (since 12 is 20 per cent below 15 and prices have to fall in the proportion 15 : 12 if the amount of goods can exchange in the proportion 12 : 15), and the rate of interest is 20 per cent (since that is the rate of interest which will enable $10 to be transformed into $12 in a year by lending it out at interest—$12 is 20 per cent above $10). The last two (20 per cent and 20 per cent) do not add up to the first item (50 per cent). The discrepancy is due to the awkwardness of percentage calculations in reckoning the difference in each case as fractions (or percentages) of continually changing bases. The actual gain in interest (2) and the actual gain through falling price (3) do add up to the actual gain from the postponement of consumption (5), but these actual amounts are expressed as percentages of different bases (10, 12, and 15 respectively) so that the percentages do not add up in this way. For very small values of these items, however, the bases will not differ very much so that the discrepancy will be small and one can say that the marginal physical yield from postponing output or anticipating input will roughly equal the sum of the rate of interest and the rate of fall in prices.

equality between *pf* and *vmp* at all times, so that those products whose prices rise relative to the others will have a smaller marginal physical yield from the postponement of output to make the same, zero, marginal *value* yield for all goods. This follows from the application of the Rule in the absence of any interest. If the marginal *value* yield from postponing output is not equal to zero, the *vmp* of factors cannot be the same (and equal to *pf*) in production for this year as for next year.

In the same way the factors must show a zero marginal *value* yield from the anticipation of input, but there need be no equality of marginal *physical* yield from the anticipation of input of the different factors in the production of the different products. For example, if a factor is much more scarce next year, for any reason, its *mp* may actually be less this year than next year, while the opposite is true for other factors. But *a dollar's worth* of any factor will show an excess (in yield of next year's product) over $1's worth applied next year, and this excess must be the same for every factor and equal to the rate of interest.

In the absence of any rate of interest (which means a rate of interest equal to zero) there will be a general fall in prices—as in the stationary economy—but not all prices will fall at the same rate and some may actually rise. An index number of prices may be constructed, and the establishment of a rate of discount equal to the rate at which the index number was falling would stabilize the index number. If the index number shows a rate of fall of 20 per cent in the absence of any interest rate, it will require an interest rate of 25 per cent to prevent prices falling (as this fall is measured by this index number). This is the same as a discount rate of 20 per cent.

$$\left(\frac{100 + 25}{100} = \frac{100}{100 - 20}\right)$$

Relative prices (the ratio between the prices of different goods) will not be changed, but the rising prices will rise so much more rapidly and falling prices will fall so much more slowly (or will rise instead of falling) that at the end of a year every price will be

25 per cent higher than it would have been at a zero rate of interest. Prices that fell just 20 per cent when the rate of interest was zero will now, like the index number, not change at all. Prices that fell more than 20 per cent will still fall, but only by the excess of their previous fall over a 20 per cent decline. Prices that fell less than 20 per cent will now rise. Prices that did not change before will now rise 25 per cent, and prices that rose before will now rise more than 25 per cent.

Different goods depreciating at different rates offer alternative measurements of the same general marginal yield from the postponement of output.

Since index numbers are inevitably arbitrary, there is no single unambiguous rate of interest which will stabilize the price level and which is the only true measure of the marginal yield from the postponement of output or the anticipation of input. There is a different measure and a corresponding rate of interest for every index number that might be constructed. This merely means that the same marginal yield from postponement of output will have many different measures in terms of the different goods or combinations of goods that enter into the construction on the index number.

The marginal yield in terms of a particular good (or combination of goods) is measured by the technical rate of its transformation into the same good (or combination) in the future. If 10 units of X now are technically transformable into 11 units of X next year, the marginal yield in terms of X is 10 per cent. At the same time 10 units of another good (or combination) Y may be technically transformable into 13 units of Y next year so that in terms of Y the marginal yield is 30 per cent. There is no more contradiction involved in this than in saying that a piece of wood is 10 yards long at the same time that it is 30 feet long. Y compared to X is depreciating over time so that the value of 10 units of X bears the same proportion to the value of 10 units of Y this year as the value of 11 units of X bears to the value of 13 units of Y next year. Both measurements come to the same thing.

A 10 per cent yield in terms of X is equivalent to a 30 per cent yield in terms of Y.

If the rate of interest is zero, various measures of the yield are observable in the rates of fall in price of goods and of combinations of goods. In the example of the previous paragraph X will fall in the proportion 110 to 100 each year, and Y will fall in the proportion 130 to 100 each year, corresponding to the 10 per cent and 30 per cent that are the respective measures of the marginal yield from postponing output or anticipating input in the economy as a whole. A positive rate of interest will bring about a relative appreciation of price equal to the rate of interest, so that the marginal yield (which is a technical relationship and not affected by the rate of interest or by the general movement or prices) is no longer measured by the price fall. It will equal the rate of price fall multiplied by the rate at which money accumulates at interest. Thus, keeping to the same example, if the rate of interest is 25 per cent per annum, the price of X instead of falling in the proportion 110 to 100 will rise from 110 to 125, while the price of Y instead of falling in the proportion 130 to 100 will fall from 130 to 125. The marginal yield from postponement of output when measured in terms of X can be obtained by multiplying 110/125 (the rate of price fall) by 125/100 (the rate at which money accumulates at interest). This gives us 110/100 as the marginal technical transformability of present into future goods with a marginal yield of 10 per cent. This is the same as when the rate of interest is zero. For Y the marginal technical rate of transformation is 130/125 multiplied by 125/100. This gives 130/100 and again a marginal yield of 30 per cent in terms of Y.

The marginal yield from postponing output or anticipating input is thus *arbitrarily* divided between the rate of interest and the rate of price fall just as in a stationary economy. This is stressed here because in economic writings the marginal yield has usually been connected exclusively with the rate of interest. We see now that this is permissible only if the good (or combination of goods) in terms of which the marginal yield is measured does not change in price over time. This element in the formula then reduces to unity and the marginal yield is equal to the rate of interest.

Technical progress tends to make product prices fall relatively to factor prices so that it is impossible to stabilize both.

Another cause of change in relative prices that we must consider as soon as we leave the stationary economy is improvement in technical knowledge, which not only introduces new products and factors, but shows how to produce more product out of a given quantity of factors. This tends to lower the future price of products as compared to the future price of factors. Unlike the opportunity of improving equipment by postponing output, improvement in technical knowledge does not tend to reduce all prices. It tends only to reduce product prices in terms of factor prices and to raise factor prices in terms of product prices. The same general marginal yield from the postponement of output will have a larger measure in terms of products (including consumption goods) than in terms of factors just as in our last example it had a greater value in terms of Y than in terms of X. As a result of improving technical knowledge a higher rate of interest will be needed to keep (an index of) the prices of products stable, but this would cause (an index of) the prices of factors to rise. To keep the prices of factors stable a lower rate of interest would be required than in the absence of improvements in technical knowledge, but that would cause the prices of products to fall. Which of these policies should be followed or whether some intermediate rate of interest should be chosen is a matter for the economic authorities to decide. They should consider under which policy the resistances to the necessary price changes would result in the smallest deviation from the optimum use of resources.

Investment tends to change relative prices.

Finally, when we leave the stationary economy we must take account of the fact that resources will be used to increase and improve productive equipment. In the stationary economy this was ruled out by the assumption that productive possibilities did not change, which implied that the existing equipment was just maintained, with neither increase nor deterioration. It is also possible, when we leave the stationary economy, for equipment to be used

up or inadequately renewed so that it runs down. This is unusual and can be analyzed in exactly the same terms as for an increase in equipment, so we shall not concern ourselves with this here.

The application of resources to the improvement of productive equipment is called *investment.* In a stationary economy the investment is just sufficient to make up for the wear and tear and wastage of equipment so that although the *gross* investment is considerable, the *net* investment is zero. When the equipment is increased, *net* investment is positive. When the equipment is not fully replaced, *net* investment is negative. Zero *gross* investment would indicate that nothing at all is done to replace equipment that is being worn out or stocks that are being used up. This would occur only if it were believed that the end of the world was imminent or that the city was going to be captured by the enemy or that some other catastrophe or state of bliss was coming that would put an end to all needs either by satiation or by death.

When net investment takes place some equipments are increased relative to others, and this will increase the supply of some products in relation to others. This is another reason for changes in relative prices and the impossibility of stabilizing anything but some arbitrarily chosen index number around which particular prices, and other index numbers, must continue to rise and fall.

CHAPTER 21. INTEREST, INVESTMENT, AND EMPLOYMENT I

Price policy, which comes before interest policy, should be framed so as to conflict least with price rigidities.

The authority in a collectivist economy must decide on a price policy before it can have an interest policy. Given any price policy, there is a corresponding interest policy which must go with it if at different points in time the proper relationships between the prices of goods and the optimum allocation of resources among the goods produced are to be maintained. It would be possible to decide on an interest policy first and then adopt the corresponding price policy, but an awkward price policy can be so much more inconvenient than any interest policy that we may assume that the price policy is decided first.

Prices might be expected to be less rigid in a collectivist economy.

The authorities may decide to stabilize either an index of consumption goods—a cost of living index—or some index of factor prices. Where wages are difficult to adjust it might be better to stabilize these and avoid the inefficiencies and frictions that result from resistance to the price changes appropriate to the price policy chosen. In a collectivist society, where wages do not form the whole income of the workers and where lower wages mean a larger social dividend to be distributed among the workers, there is some reason for believing that wages will not be as inflexible as in a capitalist economy. To the extent that this is so it might be better to stabilize some index of the cost of living. The rise in real income as techniques improve and equipment is augmented will then be seen as increased money income. But whatever the policy,

an index number can be chosen which it is decided to stabilize. The rate of interest would then be made equal to the marginal yield in terms of this index number. If the index number tends to fall it is because the rate of interest is less than the marginal yield. Increasing the rate of interest will increase the cost and the price of all the constituents of the index for next year (as well as of every other good that is made for delivery in the future) so that it will stop the index from falling. In the same way the index can be prevented from rising by lowering the rate of interest. When the index is stable the rate of interest will be equal to the marginal yield in terms of the index.

The rate of interest is related not to the price level but to *the rate of change* in the price level.

It has already been pointed out, and it may here be repeated, that this interest mechanism is in direct contrast with the familiar one in a capitalist economy which lowers price by raising the interest rate. What we have said here seems rather to be in accord with the proposition, often howled out of court by economists, that raising the rate of interest makes the price level higher by increasing the cost. It is not really the same proposition because what is done here by raising the rate of interest is not to raise prices to a new level, but to increase the rate per annum at which prices in general are rising (or decrease the rate at which prices are falling). The explanation of this must be left to the next chapter.

The marginal yield from the postponement of output depends on the degree to which output is being postponed.

We come to the heart of the present problem when we note that, even in the collectivist economy which has decided on its price policy, the rate of interest cannot be decided as automatically as was indicated in the paragraph before last. The marginal yield from postponement of output (and from the anticipation of input) is not given once and for all by the objective conditions. It was so given in the stationary economy that we examined above be-

cause there, by definition, net investment was given and equal to zero. But now this is no longer true, and the marginal yield from the postponement of output depends on the degree to which output is being postponed. That is, it depends on the rate of investment. The greater the rate at which output is being postponed, the greater will be the quantity of resources transferred from the production of present consumption goods to the improvement of equipment which will permit a greater supply of consumption goods in the future. This must be accompanied by increasing cost of the new equipment (from the point of view of society) unless the technical elasticity of substitution between new equipment and current consumption is equal to infinity (which may be ruled out as practically impossible.)

This increasing cost from the point of view of society is not to be confused with the increasing cost from the point of view of the "industry" of manufacturing new equipment. Unlike the latter, it takes into account the decreased demand for current consumption goods which sets free resources that may be used for the production of the new equipment. The prices of the factors need not rise in the way they would if this "industry" expanded in the absence of a decreased demand for the alternative products of the factors. But there will still be increasing cost as long as there is some difference in the relative efficiency of different units of factors in the two uses. This is practically certain to occur so the technical elasticity of substitution between current consumption and new equipment is almost sure to be less than infinite (see Chapter 13).

The *marginal efficiency of investment* is a function of the rate of investment.

The increasing cost means that, as investment increases, the cost of new equipment rises in relation to the price of current consumption goods. A policy of keeping the price level of current consumption goods stable will entail an absolute rise in the cost of additional new equipment, so that the marginal yield from the postponement of output of consumption goods will fall. The authorities must therefore decide upon the rate of investment in the course

of determining the rate of interest. A greater rate of investment will mean a lower marginal yield from the postponement of consumption, so there will have to be a lower rate of interest. A smaller rate of investment will mean a greater marginal yield from the postponement of output and a higher rate of interest.

The marginal yield from postponing output is therefore a decreasing function of the rate of investment. It is, indeed, nothing but the effectiveness of a unit of investment. The effectiveness or the *efficiency* of investment is measured by comparing the excess of the value of the future marginal product with the value of the factors applied. If a present investment of $1,000,000 results in additional output next year of $1,100,000, the efficiency of the investment is said to be 10 per cent per annum. The factors worth $1,000,000 could instead be used to produce goods worth $1,000,000 for immediate consumption. We can therefore say that the yield from postponing these consumption goods until next year is 10 per cent per annum. For a small increase in the rate of investment the efficiency of the investment is equal to the *marginal* value yield from postponing output and can be called the *marginal* efficiency of investment. This serves to shift the emphasis from the negative to the positive aspect, concentrating attention not on what is sacrificed but on what is done with the resources set free. The *marginal efficiency of investment* is identical with the marginal value yield from postponing output and the marginal value yield from anticipating input. From now on we shall prefer to speak of the marginal efficiency of investment, *mei.*

The determination of the rate of investment is unavoidably political.

The collectivist authority is faced with a declining *mei.* It has to make the important decision of dividing the resources of society between satisfying current consumption and increasing productive equipment. When it has made this decision, the rate of investment is given and then there is also a definite *mei* (or marginal yield from postponement of output) which determines the appropriate rate of interest.

It must be stressed that the decision of how much investment

there should be is a *political* decision and cannot be otherwise. There is no certain way, in a collectivist economy, of permitting the consumers, *as consumers*, to make this decision via the price mechanism. The question may be decided by democratic procedures, but these will have to be political—by the public voting between high investment and high consumption parties. Many devices may be applied to discover the people's sentiments—polls of public opinion, discussion in the press, even observation of the saving habits of individuals in response to varying interest rates. But as long as the instruments of production belong to the state decisions as to the rate of investment must be political. In a later chapter we shall consider the degree to which this is an interference with the freedom of the individual and how the same decision is made in noncollectivist economies.

In applying the Rule, *pf* and *vmp* must be discounted to the same point in time.

The way in which the matter is carried out would superficially seem to reverse the order of determination, where we saw the rate of investment is prior to the rate of interest. The responsible authority would not, or rather need not, make public any figure as to the total rate of investment (per annum). It would estimate the rate of interest corresponding to the desired rate of investment at the chosen price policy. It would then establish this rate of interest. Any manager of production would then in accordance with the Rule undertake any investment whose efficiency was greater than the rate of interest. The rate of interest must be counted as a cost, either added to the price of the factors or subtracted from *vmp*. Values are now no longer independent of *time*. A dollar's worth of product has not the same value now as it will have a year hence. The *vmp* must be equated to *pf* when both *vmp* and *pf* refer *to the same time*.

It does not matter which point of time is taken for the comparison as long as it is the same for both. Perhaps the most convenient formulation would be that which concentrates on the time when the factor is put into the productive process. The Rule may then

be interpreted in terms of the *discounted vmp*. Factors are to be applied whenever the *vmp*, discounted back at the proper rate of interest to the time when the factor is being applied, is greater than *pf*. Factors are to be withdrawn from production if the discounted *vmp* is less than *pf*. The optimum position is reached when discounted *vmp* everywhere equals *pf*. If the rate of interest is 10 per cent per annum the *vmp* of $1's worth of factor must be $1.00 if the product is immediately available, $1.10 if it is available only after an interval of 1 year, $1.21 if it is available only after an interval of 2 years, and so on.

The rate of interest being given and the Rule applied in this way by all managers of production, all investments with a yield greater than the rate of interest will be undertaken and no investments will be undertaken whose yield is less than the rate of interest. The judges in each case will be the men on the spot who know best. Again the price mechanism permits this specialized knowledge to be harnessed to the general interest.

Private saving could be used as an index in making the political decision about the rate of investment.

There are several aspects of this matter that raise serious problems for an uncontrolled economy though their solution is simple enough in the collectivist or even in the mixed but controlled economy.

First, there is the establishment of the rate of interest at the required level. In the completely collectivist economy this is simply *announced* and the managers have to include it in their calculations just as they have to include other payments for factors of production. The rate of interest must also apply to consumers to the extent that they are permitted to postpone or anticipate consumption of their income. A consumer who postpones consumption permits output in the future to be increased by more than the amount that he gives up in the present. Giving up $100 of consumption this year will increase next year's output by $110 if *mei* (the marginal efficiency of investment) is 10 per cent per annum. The *msc* of $100's worth of goods this year is $110's

worth of goods next year, and the *msc* of $110's worth of goods next year is $100's worth this year, so that these two should have the same price to the consumer if there is to be an optimum allocation of resources between goods this year and goods next year.

This is what happens literally, as we have seen, when the rate of interest is equal to zero and prices are permitted to fall at a rate which makes $110's worth of goods next year (at this year's prices) sell for $100 at the lower prices which then prevail. The same effect can be obtained while keeping the price level stable by permitting the consumer who wishes to postpone consumption to deposit his savings with the state bank where they will receive a rate of interest equal to the *mei*. If a consumer wishes to anticipate consumption, consuming some of next year's income this year, he should be made to pay the same interest and for the same reason.

The degree to which consumers anticipate or postpone the consumption of their incomes could serve as one of the indices to consumers' personal preference as between present and future consumption and could be utilized as a guide to the government's investment policy. A tendency for borrowing to exceed lending would indicate that the consumers desire more present goods in preference to future goods. A tendency for consumers to save and lend more than they borrow (in order to reap the interest) would indicate a willingness to sacrifice more out of present income for the future at the current marginal yield from the postponement of consumption. The rate of investment undertaken by the government (that is, by the individual managers guided by the rate of interest fixed by the government) should be adjusted accordingly. If this criterion were accepted as adequate, the rate of investment could be regulated quite automatically. If consumers borrowed more than they deposited, the rate of interest (charged to both consumers and managers) would be raised. This would discourage investment to the level which raised *mei* to equality with the higher rate of interest. If consumers deposited more than they borrowed, the rate of interest would be lowered. The final and optimum position would be reached when consumers neither borrowed nor saved on balance and the degree of investment could be said to be automatically adjusted to their desires.

The government must adjust consumption and investment so as to prevent inflation and unemployment.

The government cannot simply lower the rate of interest or raise it in accordance with this criterion, or indeed any other criterion, and leave everything else alone. A reduction in the rate of interest would lead to an increased demand for factors of production for the new investments in which *mei* is below the old rate of interest but above the new rate. There may also be an increased demand for consumption goods by consumers who are discouraged from saving or encouraged to borrow by the reduction in the rate of interest. This increase in demand will raise the prices of the factors of production (if there was full employment of the factors in the beginning) as well as of the products that are made from these factors. This will disrupt the price policy of stabilizing the cost of living or some other price index. Furthermore the income from labor (which is the same as the prices paid for the factor labor) will increase as a result of this and give rise to a further increase in demand which will raise prices again. Rising prices raise income, and rising income raises prices. Thus a vicious spiral of inflation is put into operation by a reduction of the rate of interest.

It is even worse when the rate of interest is raised. Investment and consumption will fall off and prices will fall. Labor income will fall too and with this there will be a further fall in demand and in prices. Now we have the vicious spiral of *deflation*, with not only the disruption of the price policy but increasing unemployment. The more rapidly factor prices fall in their attempt to find employment when demand falls, the more rapidly will the price level be falling. The greater the rate per annum at which prices are falling the less will be the value of the future product made possible by the investment, the less profitable will the investment appear, and the lower will be the *mei*, the rate of investment, and the level of employment.

A conscious policy by the government for avoiding the evils of inflation and the evils of deflation we shall call *functional finance*.

The payment of a *social dividend*, which enables this to be done, must be independent of the amount of work done by the recipient.

To prevent these catastrophes the government must bring about a decrease in consumption every time it increases investment and an increase in consumption every time it decreases investment. In this way total demand can be maintained at a level sufficient to give employment to all the factors of production but not sufficient to bring about inflation. Nor is this enough. Spontaneous changes in investment or in consumption must be offset too if inflation and depression are to be prevented. In fact the government is faced with the task of continuously maintaining a proper total demand for factors, through consumption and investment, so that there is just enough demand to give full employment but not enough to start an inflation.

In the collectivist economy this can be done in two ways. The first is through adjustments in the rate of interest. This affects both the rate of investment and the rate of consumption. Second, and more important, is the direct effect of government action on income and through income on consumption.

The consumers receive part of their income from their work in payment for their labor by the managers of production, who hire labor in accordance with the Rule. The rest of the income of consumers comes to them from the government. This can be considered as the citizen's share of the earnings of the factors of production other than labor, but however it is considered, the government must distribute just enough to induce consumers to spend the right amount which, together with the investment demand for factors, will provide full employment.

The distribution of this "social dividend" may follow any principle that pleases the government. The only proviso that must be made in the interest of the optimum use of resources is that the amount paid out to any individual should not in any way be affected by the amount of work he does. This is because of the desirability of having the wage equal to the *vmp* of labor (which is what the manager will be paying the worker quite apart from any "social dividend") so as to induce neither too much or too little labor. In

the name of the optimum division of income it can be argued that the distribution of the social dividend should not be very unequal. My personal inclination is for an equal share to be given to each member of society as his right as a citizen, with no questions asked and no exceptions. There could be no better safeguard of the freedom and independence of the individual.

The social dividend might be negative, that is a *tax*.

This social dividend could be adjusted very readily, even from week to week, in accordance with the state of demand. If there is too little demand this can be remedied by increasing the social dividend. If demand is too great the social dividend has to be decreased. By this simple means both inflation and depression can be prevented. If the reduction of the social dividend to zero still leaves too much spending (which does not seem likely but is not impossible) it will mean that the money spent out of the wages and salaries earned is able at the current prices to buy more than can be produced by all the factors of production available. In this case there will have to be a negative social dividend—a tax—which reduces demand to the proper level. Of course a high rate of interest would also help by discouraging investment and perhaps encouraging saving and so discouraging expenditure on consumption.

The adjustment can be made automatic but there are many complications.

We can now see how the rate of interest, in conjunction with the variable social dividend, can be used to bring about the adjustment between consumption and investment. If the government (as a political decision) accepts the criterion of zero *net* saving or borrowing by the consumers as an indication of the right rate of investment, the entire adjustment is automatic. An increase in thrift by the public will increase consumer saving. The rate of interest is then lowered and investment is increased. If the consumers in their greater thrift reduce their demand by as much as

investment is increased, total demand remains the same and there is no need to vary the social dividend. But in that case net saving will be greater than zero. Consumers will be saving and lending to the state bank more than they are borrowing, and this indicates that there is still not as much social provision for the future as the consumers want. The rate of interest must therefore be reduced further so as to induce more investment (which is more social provision for the future). The greater rate of investment induced by the lower rate of interest, together with the spending on consumption (which may even have been increased by the lower rate of interest), will constitute a total demand greater than before. Total demand would be too great and threaten inflation. The social dividend will therefore have to be reduced before the new position is reached with a lower rate of interest, more investment, a lower income, and current consumption equal to income. In the same way a decrease in thrift will lead to a higher rate of interest, less investment, a greater social dividend, and a higher consumers' income which is all consumed.

If the government adopts any other principle for determining the level of investment, the same adjustments in social dividend must be made to keep total demand at the proper level.

The automatic principle for governing the rate of investment should not be applied without a much more careful study of the significance of saving and borrowing by consumers. In a collectivist economy there would normally be restrictions on consumer borrowing and lending because of the danger of private accumulation of property and power that are inimical to the equalitarian and democratic ideals of the society. The excess of borrowing over lending (or vice versa) might reflect these restrictions rather than the consumers' relative desires to consume goods now or in the future. In a capitalist society the excess of saving over borrowing probably reflects the great inequality of income rather than the public attitude about the social provisions that should be made for the future. In a society with a growing population an excess of saving over borrowing might merely reflect inadequate provision of social security for old age, with more people saving for their old age than there are old people living on their past

savings; or it might be the other way round with a declining population. Many problems of this kind will have to be considered and so it is perhaps best not to insist on this automatic regulation of the rate of investment and of the corresponding rate of interest but to leave it to the normal democratic political organizations to decide.

The rate of interest also affects the amount of money people want to hold.

Another thing to be noted is that, when the rate of interest is lower, the interest lost in holding actual cash instead of lending it out is less and so it is possible that the convenience of holding cash will be more fully satisfied. This will mean that the government must be prepared when the rate of interest is lowered to issue more money and when the rate of interest is raised to find itself receiving more cash than it issues. This is not likely to be very extensive in the collectivist economy, nor is it of great importance since it is easy for the government to print some more money or to take in more money than is paid out. We shall see, however, that in a capitalist as well as in a controlled but non-collectivist economy this is a matter of vital importance.

CHAPTER 22. INTEREST, INVESTMENT, AND EMPLOYMENT II

Full employment may be achieved automatically in a capitalist economy.

There is a mechanism by which an equilibrium situation with full employment tends to be reached in a purely capitalist economy. In this equilibrium the rate of interest, the rate of investment, and the level of consumption are so adjusted that there is full employment of resources. Unfortunately the mechanism, as we shall see, is not one on which much reliance can be placed. In the present chapter we shall describe the way in which the mechanism operates and in the next chapter we shall point out the several parts in it which do not always work well enough to be able to prevent inflation, deflation, unemployment, and the trade cycle.

The level of employment depends on the money demand for goods and services.

We may begin by analyzing the determinants of the actual volume of employment in any given situation. We shall assume perfect competition here, both in buying and in selling, since in the absence of perfect competition there would not be an optimum use of resources even if full employment were attained.[1]

We may assume that in a capitalist society with perfect competition any factor other than labor which has a positive *mp*, and so could earn something if employed, will be hired out by its owners for whatever it can earn rather than be left idle to earn nothing. This assumption is indeed implied in our device of measuring the

[1] The next few pages are taken almost verbatim from an article by the author, "Some Swedish Stepping Stones in Economic Theory," *Canadian Journal of Economics and Political Science*, November 1940, pp. 575–580.

level of economic activity by the number of people employed. The volume of output is determined by the number of people who are put to work with the *given* equipment. It is only because of the assumption that all productive equipment that is worth using is used instead of being left idle that we can say that the equipment with which the employed people cooperate is *given*.[1]

The number of people who find employment will depend on the profitability of employing them. The profitability depends on the money demand for the goods and services that those seeking employment are able to produce. We may say, therefore, that the level of employment depends on the total demand for goods and services of all kinds, or on total expenditure.

Net income is equal to net expenditure because each dollar of expenditure creates a dollar of income.

The total income of society is equal to the total demand for goods and services or the amount of money spent on them. This is simply because income cannot be received by anybody unless someone else is paying it out. Total payment and total receipts are merely different names for the same transactions, indicating merely whether they are viewed from the paying or the receiving end.

It is true that not all receipts can be called income. Some—

[1] This distinction is made between labor and the other factors of production because (1) we are concerned with the human problem of workers not finding work and with the suffering and frustration involved more than we are with the failure to use other factors to the utmost—that would only reduce the total social product; (2) labor has a wage that is rigid in money terms; (3) labor has a minimum real wage which is necessary for it to be able to work, while other factors have no such minimum reward below which they would not be available. But the most important reason for the distinction is (4) the other factors are able by lowering their prices to find employment only because the price of labor is relatively less flexible. If wages were as flexible as the other prices, the unemployment of the other factors would not be alleviated by the fall in their prices. All prices would fall together indefinitely until something happened to change the underlying conditions—such as a fall in the rate of interest (which might be the result of *lower* prices) if this reduction were enough to offset the aggravating effect of *falling* prices both on the rate of interest and on the rate of investment and consumption at any given rate of interest. For a discussion of different degrees of price flexibility of different factors and the effects of this on employment, see Chapter 23 and A. P. Lerner, "The Relation of Wage Policies and Price Policies," *American Economic Review, Supplement*, March 1939.

indeed most—of the receipts from the sale of goods and services constitute not *income* for the seller but merely reimbursement for *costs* incurred by him in producing the goods and services. Only the excess of receipts over costs is his income. Starting out with total payments and total receipts (which must be equal because they are different aspects of the same transactions), we can subtract from the receipts those which do not constitute income because they merely cover costs, and from the payments we can leave out those which constitute these same costs. We will then be left with a *net* total of payments for *finished* goods and services which is necessarily equal to total income. We may therefore say that employment is determined by total *net* expenditure which is equal to total *net* income.

A certain total income can be (and must be) earned because expenditures of the same amount are being made. The income is created by the expenditure. Where does the expenditure come from?

Investment, defined as expenditure other than on consumption, together with expenditure on consumption constitutes total expenditure and so total income.

The greater part of expenditure comes from the normal purchases by consumers of the goods and services they use. But this cannot account for *all* the income that is created because people normally spend *less* than their whole income, unless they are very poor, so that the sum of these expenditures is less than the total income. Some of the income, the excess of total over the part of income that is created by expenditure for consumption, must be created by some other kind of expenditure. This, naturally, must be expenditure that is *not* for current consumption, for the latter has already been counted.

Expenditure that is not for current consumption is usually for the purchase of goods that are kept and added to one's wealth (possibly to permit an increase in consumption in the future) and is called *investment*. But *any* expenditure in addition to expenditure on consumption will have the same immediate effect of creating

income (though it may affect *future* consumption differently) so that it will be convenient, when we speak of *investment*, to include any other nonconsumption expenditure—such as government expenditure on relief projects—which is regarded as income by the recipients thereof. (This last qualification is to exclude expenditure on capital goods or other assets because the seller then regards his receipts merely as substitutes for the assets he has given up and not as additions to his income that he is free to spend without impoverishing himself.)

If there were no investment, income would be stabilized at a very low level.

Consumption and investment, therefore, between them create total income. If the investment were to cease while consumption remained at the same level, income would immediately be diminished by that part of it which was created by (and was equal to) the investment. But if that happened and income fell, consumption would *not* remain the same but would be reduced as soon as the decrease in income was felt. This would mean a further fall in income and yet another reduction in consumption. Income and consumption would keep on declining until income had fallen so low that all the income earned in the society had to be consumed. If some members of society were still rich enough to live on less than their whole income, others would be so poor that they would be forced to consume more than the whole of their earnings. When this sad situation was reached there would no longer be any attempt to keep consumption below total income. Consumption (and income) then would fall no further.

This low stable level of income and consumption could be reached by a slow movement if the people reduced their consumption slowly because they were not fully aware of the extent to which their income had fallen or was falling. It could be reached very rapidly if consumption were reduced very rapidly because the decline in income was anticipated. The *path* by which the stable level of income would be reached is much more difficult to analyze because it depends so much on fluctuating expectations in a chang-

ing situation, but in speaking of the stable level of income which tends to be reached we are on relatively firm ground.

For every level of investment there is a corresponding level of income which is reached when the investment just fills the gap between income and *equilibrium consumption.*

When net investment is equal to zero, the stable or *equilibrium* level of income, as we have seen, is that at which the members of society find it necessary, as a whole, to spend 100 per cent of their income. There is a corresponding equilibrium level of income for every other level of investment. This is because for every level of income there is a corresponding level of consumption which is reached when the level of income has been maintained long enough for people to realize what their income actually is and to adjust their level of consumption accordingly. At a certain low income level the corresponding equilibrium level of consumption is equal to the whole income. This is the level we have just described and which is reached if net investment is equal to zero. At lower income levels than this, equilibrium consumption is greater than total income. At higher income levels equilibrium consumption is less than total income. The greater the level of income, the greater will be the equilibrium consumption; but as income increases equilibrium consumption increases by a smaller amount. A table for the United States might be constructed something like the following:

(in billions of dollars, hypothetical figures)

1	2	3
Total Income	*Equilibrium Consumption*	*Investment Needed* (1 *minus* 2)
120	90	30
100	75	25
90	72	18
80	68	12
70	63	7
60	57	3
50	50	0
40	42	−2

The third column shows the amount of investment that is just sufficient, together with the consumption, to create the level of income that will sustain this level of equilibrium consumption and so maintain itself. For example, if investment is undertaken at a rate of 7 billions per annum, this, together with 63 billions of consumption, will make up an income of 70 billions. With an income of 70 billions, 63 billions will be consumed, and this, together with the 7 billions of investment, will continue to create income at the rate of 70 billions per annum. Income and consumption stay at the level of 70 and 63 billions.

An investment of 7 billions per annum could not maintain an income of 80 billions because with such an income consumption would be 68 billions, and this, together with the 7 billions of investment, would create a total income of only 75 billions. With a 75 billion income consumption would be less than 68 billions and then income would fall below 75 billions. Income and consumption would keep on falling until the 70 billion income level was reached.

On the other hand, an investment of 7 billions would not be compatible with an income of *less* than 70 billions. If income were 60 billions, consumption would be 57 billions, and this, together with the 7 billions of investment, would create an income of 64 billions. With this income, consumption would be more than 57 billions and so income would rise above 64 billions. Income and consumption would keep on rising until income reached 70 billions. Consumption would then be 63 billions and this, together with the 7 billions of investment, would maintain income at 70 billion dollars per annum.

The propensity to consume is determined primarily by the distribution of income.

Given the functional relationship between income and equilibrium consumption (the propensity to consume), the level of income is determined by the level of investment. The equilibrium level of income is that for which the gap between income and equilibrium consumption is exactly filled by the investment.

In a purely capitalist society the propensity to consume (the

relationship, at every level of income, between the total income and the corresponding equilibrium consumption) is determined primarily by the distribution of income which practically forces the rich to save and prevents the poor from saving. Since an attempt to redistribute income on a sufficient scale to make an appreciable difference to the propensity to consume really belongs to the controlled economy, and would be denounced by many capitalists as "socialism," we may here take the propensity to consume as given. The level of employment then depends on the rate of investment. We may here also disregard government investment for the purpose of raising employment as a noncapitalist measure, so we are left with private investment as the significant determinant of employment.

Employment is determined by investment, which is determined by the rate of interest, which is determined by liquidity preference and the supply of money.

Private investment is determined by the rate of interest and the businessman's estimate of the profitability of investment. The estimate of the profitability of investment will, under perfect competition, correspond to the "efficiency of investment," and businessmen will undertake all investment whose efficiency is greater than the rate of interest and leave untouched all investment with an efficiency less than the rate of interest. Investment will thus be carried to the point where its *marginal* efficiency (*mei*) just equals the rate of interest.

The rate of interest is what is paid by the people who borrow money to those who have the money to lend. Looked at the other way, it is what people who have money obtain for lending it to other people instead of holding it themselves. All this wealth—from the point of view of the individual, it is wealth even if it is only paper money—does not earn anything for its owners. Everyone who keeps possession of some money thereby sacrifices the interest he could have had by lending the money to someone else at the current rate of interest. Nevertheless all the money in existence is always being held by somebody or other.

Those who hold the money do so because the convenience and security they derive from holding it are worth more to them than the interest sacrificed. This desire or need to *hold* money, which is called *liquidity preference*, depends on a number of things. It will be greater when more business is being done, because more money then has to be held in connection with the greater sum of money payments that have to be made. The more infrequent and the more irregular are these money payments and receipts, the greater will be the liquidity preference. It will be greater when prices are higher, because this will involve greater money payments; and it will be lower when prices are lower. Yet it will be higher when prices are *falling*, because it is then better to hold money and buy later; and it will be lower when prices are *rising*. It will tend to be greater when times are bad and men are afraid that some of the payments due to them will not be made on time, so that it is good to have more money in hand to meet such emergencies; and it will tend to be lower, for the contrary reason, when there is confidence that all payments will be made when due.

There are many other influences on the *demand* for money to hold or liquidity preference. Against all these must be set the *cost* of holding money which is the rate of interest. When the rate of interest is low, people will want to hold more money (all the other influences remaining the same); and when the rate of interest is high, holding money is more costly and people will try to manage with holding less of it.

There is one rate of interest at which the total amount of money people wish to hold is exactly equal to the total amount of money actually in existence. At a higher rate people will wish to hold less than this amount of money. They cannot succeed because the total amount of money in existence *must* be held by somebody, but they will try to decrease their money holdings. Similarly at a lower rate of interest they will try to *increase* their money holdings. Again they cannot succeed because that would mean holding more money than there is in existence. An individual can increase the amount of money he holds if he can find somebody else who will let him have some of *his* money, and an individual can diminish his stock of money by transferring some of it to another individual

who is willing to permit *his* stock of money to increase. But it is not possible for all the individuals together to change the total amount of money they hold as long as the total amount of money in existence does not change; and if the total amount of money in existence does change the amount held by all the members of society together will have to change in the same direction and by exactly the same amount irrespective of any preferences of the individuals.

Surplus cash is likely to be loaned out and thus to lower the rate of interest.

There are two ways in which an individual can set about decreasing his cash holdings. (Exactly the reverse of these will apply to the individual who wants to increase his cash holdings.) The individual can either *spend* the extra money or he can *lend* it. If the individuals tried to get rid of their superfluous cash by spending it on a special spree, all at the same time, it would not help because they will be receiving more money at the same rate as they are spending more money so that the attempt to decrease their money holding will be frustrated. (Incidentally they might also find that their attempts to have a specially good time were also frustrated. Unless there is an increase in the total output of consumption goods or a reduction of inventories the increased spending all round would merely result in higher prices so that the larger money outlay would all be absorbed in buying the same goods as before. The flow of consumption goods, just like the stock of money, might be fixed in total so that one individual could get more only to the extent that other individuals could be made to take less.)

It is possible that, as everybody received more money, spending would be increased still more in the attempt to reduce cash holdings in spite of the additional receipts, so that spending and receipts would increase indefinitely as a result of the futile attempts of everybody to transfer the spare money to everybody else. But with the greater level of money payments, people will find that they need to hold more money than they needed to hold before,

and when the total money payments have increased by a certain amount the quantity of money in existence will no longer be excessive. At this point the rise in spending will come to an end. To the extent that the rise in money payments was due to a rise in the prices of all factors and products (and that the prices tended to stay at their new level) this would be a final solution. Larger money incomes would be spent in buying the same goods as were previously bought by the smaller incomes.

However, it is unlikely that all or even a significant part of the surplus cash would be directed toward increasing current consumption. The public would regard the surplus cash as a part of their wealth which it would be improvident to consume just because it was in a particular form—namely money—in which they preferred not to hold so much of their property. They would convert it into other forms by exchanging it for interest-yielding pieces of property such as land or mortgages or securities or promissory notes. All of these are included in the alternative way of decreasing cash holdings which we called "lending." We shall therefore suppose that lending—in this broad sense—is what is undertaken by individuals who find that the cash they hold is more than enough to satisfy the need and desire for liquidity at the current cost in terms of the interest sacrificed by holding cash.

We can justify this assumption further by pointing out that the possible adjustment through a rise in the prices of all goods (which raised the need for cash to the level of the amount of cash actually available) presupposes a given output of goods, and a given level of employment. Otherwise prices would not rise in the same proportion as the increase in spending. In that case the determination of the volume of employment and output, which is what we are here trying to explain, would be assumed away. Indeed what such an argument usually assumes is that full employment is somehow maintained and this full employment is what prevents the output from increasing when there is an increased money demand for output. Obviously we cannot assume full employment in endeavoring to explain what it is that determines the level of employment.

It is possible that the increased spending would bring about an

increase in output. This would also increase the need for money even if there were no change in prices. But that would not explain why the given investment (or the propensity to consume) should have changed, as must happen if real income in goods and services is to change. And we shall see how the effects of increased spending on price and output all fall into place in the explanation which this assumption facilitates.

When individuals *lend* their surplus cash, this lowers the rate of interest. When they find their cash is less than the amount it is convenient for them to hold, they decrease their lending or "*borrow*" in the wide sense which includes selling securities or property of any kind for cash. This tends to raise the rate of interest. The rise or fall in the rate of interest decreases or increases the demand for money to hold, and a position of equilibrium is reached at that rate of interest which equates the demand for money to hold to the supply of money in existence.

We may say that the demand for money to hold (the liquidity preference) in conjunction with the supply of money in existence determines the rate of interest. The rate of interest in conjunction with the estimated profitability of various investments (the schedule of the marginal efficiency of investment) determines the rate of investment. The rate of investment in conjunction with the propensity to consume determines the level of employment.

This is, of course, a simplification of the picture, for these items affect each other in many other ways. In the general economic system everything affects everything else. But this is a convenient framework into which all the important influences can be fitted.

There are eight stages in the mechanism whereby unemployment is automatically eliminated and inflation automatically checked.

Having analyzed the determinants of the actual level of employment, we may now examine the nature of the forces which tend to bring about full employment in the purely capitalist economy. We may divide this process into eight stages.

First, if there is not full employment the price of labor will fall because of the competition of the unemployed for jobs.

Second, the prices of all the other factors of production will fall in the same proportion as wages, since any factor which falls in a smaller proportion will be replaced to some extent by the relatively cheaper factor labor. The demand for the other factor will thus fall below its supply and so its price will fall further. A factor whose price falls more than wages will tend to be substituted for labor so that the demand for it will increase. Since all the factors other than labor are assumed to be fully employed to begin with, this will make the demand greater than the supply and the price will rise until it is no further below the original level than the new level of wages is below its original level.

Third, as a result of the equal fall in the prices of all the factors, competition between the producers, at the same degree of intensity as before, will reduce the prices of the products until they too have fallen in the same proportion.

At this stage there is still no increase in employment. Prices of the products have fallen in the same proportion as the prices of the factors so that any employment that it was profitable to provide before the change is still profitable and any further employment that was not profitable before is still unprofitable.

Fourth, there then results a reduction in the demand for money to hold. This is because when all prices (and therefore also all incomes) are lower a smaller amount of money is sufficient to carry out all the transactions and to provide everybody with the same amount of liquidity. At the lower prices a smaller amount of money represents the same purchasing power and the rest is surplus.

Fifth, the decrease in the need for money to hold leads to *lending* on the part of all the people who now find that some of their cash holdings are not necessary and might as well be exchanged for other forms of wealth that yield an interest.

Sixth, this has the effect of bidding up the prices of the interest-yielding forms of wealth, thus reducing their yields together with the particular yield on money loans that we call the rate of interest.

Seventh, the reduction in the rate of interest will make it profitable to increase investment. All investments that yield less than the old rate of interest but more than the new, lower, rate will

now be profitable and will be undertaken. This reduces the *mei* from the level of the old rate of interest to that of the new. Here for the first time in our series of effects we have some increase in employment.

The lower rate of interest may also induce consumers to save less and consume more out of their income, and this too would lead to an increase in output of consumption goods and an increase in employment. This, however, is a specially doubtful element. It might even work the other way if the lower rate of interest made people save more so that they might be able to make the same provision for their old age in spite of the lower interest they earn on their savings.

Eighth, and finally, the increase in the income of the people employed by the increase in investment will lead to more consumption and the creation of still more income and employment in accordance with the propensity to consume. Income will increase up to the point where the net investment just fills the gap between income and equilibrium consumption as indicated on p. 275.

If there is still some unemployment, the whole process repeats itself. This goes on until it is brought to an end by the achievement of full employment. Wages will then have no tendency to fall. This is the initial step in the process and when that stops the whole process comes to an end.

The whole process also works in reverse. If there should be an increase in total demand beyond that which is necessary for full employment, it will lead to an increase in prices all round. The increase in prices increases the total of transactions in terms of money. Greater money payments have to be made for the same real transactions in goods and service, and so there will be a need for more money to hold in connection with the greater payments. The increased need for money to hold will result in an increased eagerness to borrow and a decreased willingness to lend. Securities will be offered for sale, and the rate of interest will rise. The higher rate of interest discourages investment, and this decreases income and employment, both directly and, what is more important, by the indirect influence on income through decreased consumption, until income has fallen to the lower level where the

smaller rate of investment is sufficient to fill the smaller gap between total income and the corresponding equilibrium consumption.

The eight stages in the argument illustrating the method by which full employment is reached need not, of course, take place in time in the neat order in which they are here presented. All of them are taking place during all the time of adjustment. The eighth stage may be completed in some parts of the economy before the first is begun in another part. The neat arrangement is only for the purpose of providing a framework which clearly displays the most important influences and will permit us, in the next chapter, to see where they are likely to break down.

CHAPTER 23. UNEMPLOYMENT AND THE TRADE CYCLE

At each step in the automatic elimination of unemployment (and inflation) the machinery is likely to stall.

The mechanism by which full employment tends to be reached and by which inflation tends to be checked is likely to be halted by impediments at every one of the eight steps. The interruptions in the process may be short or long or permanent, depending on the particular interruption and the particular situation. In going over these we shall limit ourselves to the much more important case of the movement from unemployment to full employment. The impediments to the checking of inflation will simply consist of the reverse of every one of these.

Wages may refuse to fall.

Right at the outset the movement may be checked by a refusal of wages to fall. Unemployed workers, especially when starvation is not the only immediate alternative, often refuse to offer their services at less than the standard wage that is being earned by their fellows who are in employment. Workers' organizations to prevent undercutting of wages are often denounced as short-sighted (if not actually wicked) on the ground that by refusing to let wages fall workers only prevent themselves from getting work. The reduction in wages, according to the analysis at the end of the last chapter, would not even reduce *real* wages because prices would fall as much as wages and permit the workers to buy the same goods with the smaller money wages.

The workers have a number of rebuttals to such attacks on their policy. They may believe that the mechanism for bringing about

full employment will stall at some other point in the process so that their sacrifice will be real and in vain. They may believe that the result will actually be not an increase but a decrease in employment, and we shall see in a moment that this is quite possible. More often the workers are not thinking of the wage of labor as a whole but of the relative wage of their own trade as compared with others. The greater the relative wages in their own trade, the greater will be the quantity of the goods produced by other workers which they will be able to buy. The only way the workers in any trade can maintain their relative wages is by maintaining or raising their *absolute* money wages. The resultant rigidity of wage rates is thus an institutional and apparently unavoidable result of the attempt to maintain relative standards. But whatever the cause, rigidity of wages in the face of unemployment is often a fact, and whenever it exists the mechanism for bringing about full employment by the automatic capitalistic method does not even get a chance to begin.

If wages do fall this will not immediately eliminate unemployment.

There is common belief, especially among employers who do not have to sell their product to wage earners, that if only wages were forced to fall when there is unemployment (by destroying trade unions, minimizing unemployment benefit, and cutting relief so that workers are starved into accepting lower money wages) all would be well. This belief is partially due to a failure to realize the possibility of re-employment being stalled at later steps in the process, but it mainly rests on an illegitimate application to the economy as a whole of principles that are valid only when applied to a small section of it. If the wages payable by a particular employer are reduced and everything else remains the same, it will usually pay him to expand output and increase the number of men he employed. This cannot be generalized to the whole economy because by "everything else remaining the same" is meant that everybody else pays the same wages as before. The particular employer will therefore have an advantage over his competitors while his customers will be receiving money incomes based on the

higher level of wages that is generally being maintained. This will not occur if wages are reduced throughout the economy. In that case the individual employer will have no advantage over his competitors, and his customers will have lower money income with which to demand his product. The argument cannot therefore be generalized to cover a *general* reduction in wages.

The prices of other factors may not follow wages.

Let us now suppose that wages *do* fall when there is unemployment and consider the second step. The prices of the other factors of production may not fall in the same proportion. Some of the other factors may be controlled by monopolistic organizations which do not permit their prices to be reduced. There will then be unemployment of the other factors, and the prices of the products using these other factors will not be able to fall as much as wages. For any level of employment the *real* wages of labor, the goods and services labor will be able to buy for its wages, will be reduced by these relatively higher prices. This is one of the reasons why labor is prepared to organize to prevent its money wages from being reduced. Analytically this is identical with the refusal of particular types of labor to reduce their wages and the effect of this refusal on relative wages.

To the extent that the other factors whose prices do not fall are competitive with labor there will be a substitution of labor for them and this will increase the employment of labor while decreasing the employment of the other factors. Where these other factors are fixed like land it is unlikely that they will remain unemployed for long while some of them hold out for higher prices, but as long as they do so the effect on the economy will be the same as if these factors did not exist and the country were so much poorer. As we shall see, this impoverishment of the economy *can* have the effect of increasing employment since unemployment is a by-product of potential wealth when society is not controlled so as to make full use of its potentialities. It is a scourge of rich rather than poor countries.

Where the other factors whose price refuses to fall are produced

instruments of production, this is likely not only to impoverish society but to be harmful to employment too. For although there will be some substitution of labor for these other instruments of production, there will be a reduction of employment in the manufacture of the instruments of production which are not being used, and this decrease in investment can, as we have seen, have a very bad effect on the level of employment.

Product prices may be held up by monopolists.

If we suppose that the prices of all factors are reduced in the same proportion, we come to the third step in the chain of effects, which was the reduction in the prices of the products. Here again there is a possibility that the prices of the products will refuse to fall. This will occur when there is a monopolistic organization of production. The monopoly can, and frequently does, follow a policy of maintaining the price in the face of a depression even though the prices of the factors have been reduced. With a small output costs per unit may very well be quite high, possibly even higher than when the prices of the factors were higher but when there was a greater output over which heavy (and often fictitious) overhead costs could be spread. Such corporations have even been known to raise their prices in a depression in an attempt to maintain profits in the face of a decrease in output. Such a procedure is especially likely when the corporation was not extracting the greatest possible profit in times of prosperity (possibly for fear of invoking state regulation if it did). It might feel justified in raising its price if it did not make any exorbitant profit but merely reduced its losses. In any case, there might be no fall in the price of the product even though the prices of the factors of production were all reduced, and so the process of automatic achievement of full employment could be stalled at the third step.

The need for cash may not fall because *falling* wages and prices can offset the effects of *lower* wages and prices.

The fourth step in the process was a reduction in the amount of money needed to hold. This might be satisfactory if all wages

and the prices of other factors and of products were reduced together and then stopped at a level which was generally expected to be maintained. This has sometimes been done by government decree but has no place in the purely capitalist mechanism. In the capitalist society wages and prices would not all be reduced by the same proportion once and for all at the same time. Even if there were none of the monopolistic restrictions and devices for maintaining prices the reduction in wages and prices would be on quite a different pattern. At first some wages fall a little—where the workers are weakest in bargaining position or least effectively organized. Then as unemployment persists more and more trade unions lose strikes and lockouts because their funds run out. When individual workers have used up their savings and their credit with the grocer and the landlord they are forced to submit to wage reductions. Wages thus fall gradually and usually at an increasing rate. Once they begin to fall prices go with them as costs are reduced and employers lose both the hope of being able to sell their stocks at a better price if they keep them a little longer and the financial ability to keep the stocks off the market.

Falling wages and prices have an effect on the demand for money to hold that is just the opposite of the effect of the *lower* wages and prices which were considered in the last chapter. When prices are *falling* it pays to hold on to money rather than to spend it quickly because waiting will enable purchases to be made later when prices are lower and one can get more for the money.

It is true that the same benefit from the appreciation of money as prices fall can be obtained when money is loaned out at interest as when it is kept idle in one's pocket while waiting for the prices to fall. If the money is loaned out at interest the interest payments will be an additional gain on top of the appreciation of the money. It might therefore be supposed that there would be an additional source of money to lend, namely the money which is withheld from spending while waiting for the price to fall further. This, however, is unlikely because it is not certain how long prices will keep on falling and the prospective purchasers prefer to have their money on hand ready to buy before they start rising again. The *increased* demand for money to hold because prices are *falling* is

likely to be more than sufficient to offset the *decreased* need for money to hold because prices are lower.

At the same time the fall in prices will be a heavy burden on firms with fixed monetary obligations. With lower prices their gross earnings will be lower too even if the variable costs have fallen in the same proportion as the prices of the products, and it will often be impossible for these firms to meet the monetary obligations fixed in terms of the previous higher level of prices. Bad debts and bankruptcies will increase, and the resultant uncertainty whether debts will be paid when due or at all will make it worth while for creditors to hold larger amounts of cash so that they will not be inconvenienced too much by such failures. This contributes to the demand for cash and helps to offset the decreased need for cash on account of the lower level of prices.

The amount of money in existence may decrease more than the need for it decreases.

We may now grant the fourth step and suppose that there finally is a reduction in the amount of money that the public wants to hold. This may come about after prices have been falling for some time and are so much lower than in the beginning that this more than offsets the effects of their continuing to fall. Also most of the bad debts may have been written off and most of the bankruptcies carried through so that there is more confidence that the remaining debts will be paid when they fall due. As a result there is not such a great demand for liquidity on that account. At any rate, whatever the reason for it, we will suppose that the fourth step is passed and now consider the fifth step, which is *lending* on the part of the public because it finds that the amount of money it has on hand is excessive.

This may not happen at all because the amount of money in existence may decrease as rapidly as the need for it, or even more rapidly. During a period of falling prices, defaults, and bankruptcies, many banks will close down and confidence and credit may be impaired so that the supply of money contracts. Even apart from any banks closing down there will be a decrease in the amount of

credit money as the need for less cash induces businessmen to reduce their obligations to the banking system at the expense of their cash balances. This leads to an automatic decrease in the amount of money. Even when there is a decrease in the need for money there may be no lending by the public. Where there is a highly developed credit system most of the money consists not of gold and silver coins, which are fairly stable in quantity, nor even of government paper money, which does not normally fluctuate so very much, but of bank money and bank credit. "Money in the bank" is the most important form of money, and it consists for the most part of figures written against people's accounts in the ledgers of the banks.

A decrease in the desire of any individual to hold money will lead to his repaying debts to the banks, and there is no guarantee that the banks will promptly try to increase their loans to other people. It is quite likely that they will be only too glad to get their money back and will want to hold on to more cash than usual in order to be able to meet their obligations if they are not able to collect other debts. The banks will then refuse to renew loans and will even try to call in loans. There is a decrease in liquidity preference on the part of the public, but it may be offset or more than offset by the decrease in the amount of money available for the public to hold.

The fall in the rate of interest may be negligible because of elasticity of liquidity preference.

The sixth step was that the increased lending by the public of its surplus cash funds lowered the rate of interest. If all the previous difficulties are overcome and the amount of money that people wish to hold decreases in relation to the supply, the fall in the rate of interest may be small or even negligible. It is impossible for the rate of interest to fall appreciably if the owners of wealth believe that the rate is "normal" and that if it should fall it will soon have to rise again to the normal level (that is, at the most within a year or two). If owners of wealth think in this way, and it seems fairly certain that they do, they will prevent the rate of

interest from falling. As soon as it falls a little below "normal" they will stop it from falling further by selling their securities which will have appreciated because of the fall in the rate of interest. They will then hold the cash they got for the securities while waiting for the rate of interest to rise again. When this happens they will be able to buy back their securities at a lower price, thus netting a profit. The owners of securities will in this way constitute a kind of reservoir that absorbs any cash that is set free from transactions and prevents it from lowering the rate of interest. Only when these people are convinced that the rate of interest is going to fall and stay down for a long time will they give up this game and permit the rate of interest to fall.

There may even be a lower limit to the rate of interest.

Even then there seems to be a lower limit to the rate of interest. At rates below this minimum the owners of wealth are about equally willing to own securities yielding this low rate of interest or to have money yielding no interest but providing a feeling of security and liquidity instead. These owners of securities will be willing to trade large amounts of securities for cash at about the current price, and this will prevent the rate of interest from falling much even if the surplus amount of money is considerable.

The effect of falling income on investment may start a spiral of deflation which would be accentuated by a cumulative increase in liquidity preference.

The seventh step was the effect of the reduction in the rate of interest in inducing a greater rate of investment. This is upset in the first place by falling prices. When prices are falling the efficiency of investment declines because the future product whose output is increased by investing present factors will have to be sold at the lower future prices. This tends to decrease investment and may more than offset the reduction in the rate of interest. Furthermore, if factor prices are falling even investments that appear profitable will be postponed not merely because of exces-

sive greed on the part of investors (who will see an even greater profit in later investments when the prices of the factors will be lower) but because the investor is afraid that a rival who takes advantage of the cheaper factor prices at a later date will be able to undercut him and destroy the profitability of his investment. This means a postponement of investments and a further decrease in current investment.

Furthermore, investment is very strongly influenced by the level of consumption. When consumption is low, business is bad and much investment is not likely to be undertaken. When consumption is high business is good and the expectations of profit from investment—the efficiency of investment—is high and a great deal of investment is likely to be undertaken. The decrease in investment and consumption because of postponement in the face of falling prices is therefore a cumulative process. The decrease in investment decreases income and consumption and prices still further, and the decrease in consumption decreases investment once more. In this way the spiral is set going. This is a common phenomenon in the downward swing of the business cycle.

This cumulative process is tied up with another cumulative process more closely related to liquidity preference. The decrease in spending by purchasers who are waiting for prices to fall results in a further decrease in prices and in employment, in less consumption and in less investment, in a further increase in bankruptcies and in bad debts, and so in a further need for cash to hold to meet the danger of debts being unpaid when due. This increases the rate of interest and decreases investment still further, adding to the downward movement in employment and income and spending and prices and to a further increase in the need for cash and in the rate of interest. This cumulative process is a phenomenon with which we are familiar in every economic crisis and which heralds intensifying economic depression rather than the restoration of full employment.

It seems probable, too, that the effect of the rate of interest on the level of investment is not really such an important item. In most fields interest is small compared with the great uncertainties

that attend investment. Consequently, even if other things did not change, the lowering of the rate of interest within the practically possible range would not bring about any great increase in investment.

We may now assume that investment nevertheless does increase and consider the eighth and last step in the process, where the increase in investment brings about a greater increase in consumption and income in accordance with the propensity to consume. Income increases up to the point where the gap between income and equilibrium consumption is equal to the greater rate of investment.

This is perhaps the most reliable step in the series. The propensity to consume is fairly stable. It is possible that the lower rate of interest will reduce the propensity to consume by inducing people to save more in order to make up for the smaller rate of interest earned on their savings, but this is likely to be about offset by increased consumption by individuals who will not think it worth saving so much just because the rate of interest is low. An increase in activity may bring about a redistribution of income which increases saving, but that could only weaken the effect a little. We may say that if the seventh step has been reached the eighth is fairly certain. It is in the first six or seven steps that the great impediments are to be found that prevent the automatic re-establishment of full employment whenever there is any departure from it.

All these considerations apply in reverse for an inflation.

All the arguments of this chapter used so far can be applied in reverse to the case of a departure from full employment in the opposite direction. There cannot indeed be much overemployment.[1]

[1] In time of war there can be a very great increase of employment over what would voluntarily be desired by workers in normal peace times. To the extent that the extra labor is forthcoming because workers are deceived by higher money payments that do not represent higher real wages it is overemployment in the sense here discussed and symmetrical with the underemployment of depressions. But it seems likely that the greater part of such extra work in wartime is the result of patriotic feeling—the desire to be doing something to help win the war—and as such it is best considered as a temporary increase in the supply of labor rather than as overemployment.

But inflation can and does take place, and the automatic mechanism fails to stop for reasons exactly parallel to those which prevent it from curing unemployment. It does not seem necessary to go over all these again in reverse.

With several rigid prices relative unemployment depends on relative prices.

So far our analysis has been almost entirely in terms of one factor, which was the only one to have a rigid price—labor with its rigid wage. All other prices were assumed to be flexible and (except in the short study of why wages are rigid) it was assumed further that there was only one kind of labor and only one wage. As this wage fell all other prices fell proportionately until an increase in investment and employment, brought about by a fall in the rate of interest, led to full employment and a stable wage and price level.

Now it is, of course, not true that there is only one kind of labor and that no price other than the price of labor is rigid. We can now bring all the factors with rigid prices into the analysis and see that our fundamental principles are unaffected. What we have said about labor is now applicable to all these factors taken together. As their prices are permitted to fall the movement toward full employment of all of them can proceed and all the same hindrances to the movement will exist.

Unemployment now ceases to be a simple thing with a single measure. It has as many dimensions as there are factors with rigid prices so that some of each of these factors is unemployed. The unemployment is shared among them in proportions determined by their *relative* prices. If one of the factors reduces its price in relation to the others it will be substituted in employment for the others. If several factor prices are reduced by the same percentage, the effect will be the same as when the price of a single factor was reduced. Unless the further steps in the process are all carried out, this will result merely in the reduction of all the flexible prices in the same proportion as the rigid prices were reduced, and employment and income in real terms will be un-

changed. If some of the rigid prices are raised and others are reduced, the first kind of factor will find less employment and the second kind will find more employment. The effect on total income and on total employment is indeterminate. An index number of the rigid prices could conceivably be constructed such that when some of these prices rose and others fell in ways that left the index number unchanged there would be no tendency for total income to change, but that would merely be the result of weighting the price of each factor by the degree to which total income tended to be affected by changes in that price. All that can be said in the general and abstract terms to which we are restricted here is that the relative unemployment of the factors with rigid prices depends on their *relative* prices, while the absolute level of unemployment depends on the absolute price level of the rigid prices only if none of the impediments investigated in this chapter is effective. If any of these impediments is operative, a tendency for the absolute level of the rigid prices to fall would not increase total employment and income and might decrease it just as in our analysis of the single rigidly priced factor which we called labor.

The fundamental cause of the business cycle is the inadequacy of demand because of the very unequal distribution of income.

The sad story we have told of impediments to the automatic achievement of full employment is not too unrealistic as a description of what goes on a great part of the time in the actual world. Most of the time there is unemployment, which means that there is not as much consumption and investment as the economy is able to provide. The inadequacy of consumption by most people can only be explained by inadequacy of incomes. They would be glad to spend more if they had it to spend. The inadequacy of investment is mainly due to inadequacy of consumption. The inadequacy of consumption follows from the extremely unequal distribution of income which prevents the poor from consuming while the rich naturally save a large part of their income.

This leads straight to the fundamental cause of the business cycle. This fundamental cause is very simple, however complex

may be the study of the details of the business cycle and however baffled may be the attempts to cure it without disturbing its fundamental cause. The fundamental cause is the inadequacy of demand.

With the given distribution of income, which is a very stable result of the multiple chance influences of a capitalist economy, the propensity to consume is such that at an income corresponding to full employment the gap between income and equilibrium consumption is very large. This gap (shown at 25 or 30 billions out of an income of 100 or 120 billion dollars in the illustration on p. 275) shows the total amount that would be saved (mostly by the rich) if there were full employment. This level of income can be maintained only if there is sufficient investment to fill the gap. But this tremendous level of investment is very much more than it is profitable to maintain for very long. If such a position of full employment should be reached, the opportunities of investment would soon begin to be used up and investment would decline. This sets in motion the cumulative processes of crisis and depression that we have just examined. Income falls and investment falls still further. Prices fall too and the cumulative movement goes on until there is a very low level of income at which the gap (between income and consumption) is very small or perhaps equal to zero or even negative. With little investment going on for a long time, opportunities for investment accumulate that are profitable even at the very low income level. When some investment starts this raises income and so more investment becomes profitable. We now have a cumulative movement upward. Many investments are undertaken that depend not only on a high level of income but on a *rising* income. The impetus of the expansion may carry it up to full employment or it may stop before that level is reached. It must stop at full employment in any case because a significantly larger level of activity is impossible. (A level of activity somewhat larger than "full employment" is possible when individuals are temporarily induced to sacrifice leisure and retirement for the sake of rewards which, because of inflation, turn out not to be as great as they appear.) When the expansion of activity comes to a stop those investments which were based on the actual fact of expansion itself come to an

end and so there is a *decrease* in activity and an additional reason for the downturn.

If and when full employment is reached, the expansion can still go on, but only in terms of prices. Prices and money income rise, but real income cannot rise. This is indicated on the table of p. 275 which shows the ratio of equilibrium consumption to income continuously diminishing as income increases from 40 to 100 billions, but in the increase from 100 to 120 billions the ratio remains the same. This is to indicate that when full employment is reached (at 100 billions) there can be nothing but a general price rise or inflation. Prices are shown increased by 20 per cent, with real income, real consumption, and real investment the same as for the 100 billion dollar income.

The study of business cycles is the study of what happens to employment when nothing is done to keep demand where it should be.

Of course there is much to be studied in connection with the business cycle that is not even hinted at in this one page account of it, but such studies are really concerned with the way in which the business cycle develops when the basis for it, the insufficiency of total demand, is permitted to continue. It is all irrelevant for our basic concern with what a controlled economy must do to remove the very possibility of trade cycles. The study of the business cycle is the study of what happens to an economy when total demand is *not* controlled.

One contributory factor to the insufficiency of total demand is itself the result of the trade cycle. The uncertainty, even when times are relatively good, whether or for how long the good times will continue, and the perfectly justified belief, in an uncontrolled economy, that the business cycle will continue, greatly increases the risk attendant on investment, lowers the efficiency of investment, and decreases the amount of investment undertaken. It also diminishes the sensitiveness of investment to the rate of interest (in swamping it by the much greater risks) so that the mechanism that tends to bring about full employment through a fall in the rate of interest is rendered even less effective.

It is natural for government activity to aggravate the business cycle, but in very severe depressions the government is usually forced to relieve the situation.

The level of economic activity is affected by the spending, taxing, and borrowing activities of the government. In a purely capitalist economy these activities will for the most part have the effect of accentuating cyclical fluctuation in economic activity. The government budget will be drawn up in the image of the budget of a business or a corporation. When times are good the revenue from taxes will be high and the government, federal and local, will believe it proper to undertake socially beneficial activities in excess of the current revenues by borrowing more for these activities on the strength of the expected continuation of the high tax yields. This will result in a net increase in total demand because the government spends the whole of the proceeds from taxes and loans, while the citizens contract their expenditure by only a part of the increased tax payments (except perhaps the very poor) and perhaps by only a very small part of the money loaned to the government.

In bad times the government will feel that it is proper to retrench, and not to go in for the luxuries of government spending on defense or parks or roads. It will be considered "unsound finance" to increase its debt in times of declining revenue from the taxes out of which the debt must be serviced. Attempts will even be made to reduce the national debt by raising higher taxes to repay some of the debt and bring it into a better "balance" with government tax revenue. The effect of such a policy of "sound finance" will be to reduce incomes still further, both by the decrease in government spending of borrowed money and by the decrease in private spending because of the higher taxes whose revenue serves to repay debt. The depression will be deepened by much more than this net *direct* decrease in government and private spending. This is because the decrease in spending acts just like a decrease in investment. Income must fall as much as will reduce the gap between income and equilibrium consumption by an amount just equal to the net *direct* reduction in spending.

For example, on the basis of the figures on p. 275, when income is 60 billion dollars, a 3 billion direct reduction in spending by government and taxpayers would reduce income by 10 billions from 60 to 50 billions. The direct loss of income in the amount of 3 billions causes further reductions in spending by those whose incomes are reduced and this again has a similar repercussion. These effects go on until the lower 50 billion income level is reached. At this point the gap between income and equilibrium consumption is less by 3 billions and so the income can be maintained in spite of the lack of the 3 billions of spending.

This is quite apart from the indirect effects of the higher taxes (other than income taxes and other taxes that fall almost entirely on "surplus") in interfering with the optimum use of those resources that are employed, by raising the prices of the taxed goods in relation to the factors employed in making them, and so raising *vmp* above the *msc*.

The process of deflation through "sound finance" is cumulative too. The greater the extent to which the depression is aggravated by the "economy" measures applied by the government (in its attempt to apply to the whole economy the proper principles for running a small business), the more urgently will the application of these principles of "economy" and "sound finance" appear necessary in order to prevent the economy from "going bankrupt," and the more will the government continue to worsen the situation. The classical example of this is provided by the "sound finance" of pre-Hitler Germany which in this way prepared the ground for the fascist revolution.

In the end the government is forced, against its will, to stop this and actually to do something to improve the economic situation. It finds that it cannot raise any more money by taxation without becoming too unpopular. It finds its citizens starving and is forced by a combination of humanitarian and political considerations to spend some money to keep them alive even though it means getting further into debt and "unsound finance." This net spending by the government has the effect of increasing income by more than the amount actually expended by the government. It creates income just as investment does. (Indeed we have defined

investment in a wide sense to include such spending by the government not on current consumption.) Income will be raised until the gap between income and equilibrium consumption has risen by the amount of this extra spending. Since the rate at which the gap changes grows smaller (algebraically) in relation to income as the income level gets lower; the effect will be greater at lower income levels. Again using the figures of p. 275 we can see that an increase in spending of 2 billions when income is at the 40 billion level will raise income by five times this amount to the 50 billion level; but it would take an increase in spending (or investment) of 6 billions to raise income from the 80 to the 90 billion level. At the bottom of a depression the government is nearly always forced, in spite of itself, to spend more than it collects in taxes and thus to raise the level of economic activity. But it is possible for a very stern and upright government, by adhering strictly to the principles of "sound finance," to carry its people *all* the way to destruction.

CHAPTER 24. INTEREST, INVESTMENT, AND EMPLOYMENT III (*FUNCTIONAL FINANCE*)

There are effective instruments in the hands of the government for maintaining full employment and preventing inflation, but their use is hindered by strong prejudices.

We are now equipped to examine the adjustment of interest, investment, and employment in the controlled economy. Here we reject both the dogma of the left that 100 per cent collectivism is necessarily in the social interest and the dogma of the right that the government of a country must keep to the fiscal principles appropriate to a grocery store. As soon as it is recognized as a duty of the government—perhaps even the primary duty of the government—to ensure the maintenance of full employment, and that any so-called principle of "sound finance" that might interfere with this task can have no possible justification, the instruments by which full employment can be maintained stand out clear and unmistakable. But the recognition of these instruments is impossible until some exceptionally powerful and firmly established prejudices have been removed. Some of these have been referred to in the previous chapters, but we must now consider them much more thoroughly.[1]

The instruments are not available until it is recognized that the size of the national debt is relatively unimportant,

The first of these is the unwillingness to see that the size of the national debt (when held by citizens of the country) is a matter

[1] The principle of disregarding all traditional conceptions of what is "sound" in finance and judging fiscal measures only by their effects or the way they *function* in society may be called *functional finance*. See A. P. Lerner "Functional Finance and the Federal Debt," *Social Research*, February 1943.

of almost no significance beside the importance of maintaining full employment. The national debt is not a burden on posterity because if posterity pays the debt it will be paying it to the same posterity that will be alive at the time when the payment is made. The national debt is not a burden on the nation because every cent in interest or repayment that is collected from the citizens as taxpayers to meet the debt service is received by the citizens as government bondholders. The national debt is not a sign of national poverty any more than the certificates of ownership of government bonds are a sign of national wealth—the two amounts exactly cancel out in any measure of the national wealth. Just as increasing the national debt does not make the nation poorer, so repaying the national debt does not make the nation richer. It is not true that the national debt "must be repaid sometime" any more than it is true that all the banks must call in all their debts and repay their depositors on some catastrophic day or all that firms and corporations will have to be dissolved someday to repay the obligations to the individuals who invested in them. Every individual buyer of government bonds must be able to get his money when it is due, but another lender can take his place when this happens (if the individual should not wish to renew his loan) and the national debt can continue—just as the forest can go on forever even though every tree in it must ultimately fall.

that the interest on the debt is not a burden on the nation,

Nor is it true that the *interest* on the internally owned national debt is a burden on the nation. The interest payments are not lost to the nation. They are merely transferred to the recipient from taxpayers or from new lenders, and if it should be difficult or undesirable to raise taxes the interest payment can be met, without imposing any burden on the nation as a whole, by borrowing the money or printing it. The prejudice against printing money and against borrowing seem to be much stronger when the purpose is to pay interest than when printing or borrowing is undertaken for any other purpose. The effects are, however, exactly the same.

and that the nation cannot be made "bankrupt" by internally held debt.

The view that the national debt is something bad that should be avoided or minimized, when in truth it is in itself neither good nor bad, is built on two great misconceptions. The first is to regard the nation, or the government that represents it, as a business concern which is likely to get into difficulties if it gets too much into debt. It may then not be able to meet its obligations and be forced into bankruptcy. This is nothing but the perfectly sound advice given by Mr. Micawber to all traders and normal business concerns. But neither the nation nor the government is a normal business concern or even an ordinary trader, like Mr. Micawber, in daily fear of the debtor's prison and the bailiffs. The government, even if it does not want to raise the money by taxes, can always meet its obligations to any citizen by borrowing from another citizen or by printing the money to pay him. The nation cannot be thrown into a debtor's prison or debarred by a bankruptcy order from continuing its business. The weird notion of a country "going bankrupt" because it has a great internal debt can only be explained as the result of private capitalists building up a conception of the state in their own image and impressing this capitalist mythology on the other members of the capitalist society.

Every debt has a corresponding credit.

The other great misconception lies in looking at only one side of the debt-credit relationship. Every debt has a corresponding credit because there must be someone to whom the debt is owed. When a business or a corporation incurs an obligation this is usually called an investment by the individual who put up the money. The sum of such investments is often held up as an indication of the wealth of the country. "The United States has invested some 200 billion dollars in business corporations; this shows how rich a country it is." When the government incurs an obligation it is looked upon not from the point of view of the purchaser of the government security, when it would appear as a credit, but from the point of view of the government where it is written on the other side of the ledger and appears as a debit or debt. It is then

held up as an indication of national poverty. It would be no whit more unreasonable to reverse the procedure and say that the United States could be a very rich country, witness the 80 billions of government securities that the citizens of the country own, were it not that the businesses and corporations are in debt to the tune of 200 billions. Of course both of these procedures are nonsensical since neither the governmental nor the private debt gives any indication of the nation's real wealth, which resides in the skill and industry of its inhabitants and in the natural resources and equipment with which these can cooperate.

Only external debt is like individual debt and impoverishes the nation.

All this is true, of course, only of internally held national debt. Increasing debt to other countries or to the citizens of other countries does indicate impoverishment of the borrowing country and enrichment of the lending country. Of this kind of debt the popular criticism is valid. When a country borrows from another country, that is something like when one man borrows from another or when one business borrows from another. The borrower is able, by this borrowing, to consume more than he produces and has to consume less than he produces later when he repays the debt. Neither of these is true of internal borrowing or the repayment of internally held debt. The country cannot by monetary manipulations consume more than it can produce, as every country is acutely aware at this time. And just as the internal borrowing does not really give the country anything that it did not have to begin with, the repayment of the debt or the payment of interest does not take away anything from the country as a whole.

For a country to borrow from another country may be foolish or wise according to circumstances, just as in the case of individual borrowing. Such debt should be limited because the repayment will constitute a real burden on the country just as the borrowing provided a real benefit quite different from any benefit that can accrue from internal borrowing. When the time comes to make the repayment there may be great inconvenience which could lead to default. But none of these considerations is at all applicable to

internally held national debt which from the point of view of the nation cancels out. The proper analogy to the incurrence of internally held national debt is not an individual borrowing from another individual but an individual borrowing money from one of his pockets to put it into another. The concern over the national debt shown by newspaper editorials and cartoonists is analogous to nervous prostration by the individual at the thought of what he owes one of his pockets and the danger of being bankrupted by this debt.

The distribution of the individual's money among his different pockets does correspond to a social phenomenon of real significance. The greater the internally held national debt the greater the amount of private property held by the members of society, either directly as private individual owners of government bonds or indirectly through the corporations and banks that own the bonds and in turn belong to the individuals who own the corporations. Government debt is the opposite of government ownership of wealth. The greater the part of the real natural wealth (land, mines, factories, roads, etc.) that is owned by the government, the smaller the scope for private property. If the government owns all the real wealth, as it does in the completely collectivist society, there is no scope for private property in the instruments of production. If the government owns none of the real wealth, all of it is private property. If the government borrows money it creates bonds—securities which constitute private property *in addition to* the real national wealth.

Because of this a large national debt might be disliked by socialists who would object to it as an extension of the field of private property. Their objection is ultimately based on the extremely unequal way in which private property tends to be distributed among the population, and they would logically hold that an increase in the scope for private property would accentuate the inequality of economic wealth and power. Nevertheless the loudest criticism of the national debt comes from the staunchest supporters of private property. This is another indication of how their own interests are clouded by the unthinking adoption and general application of capitalist mythology.

The effect of the size of the national debt and of changes in its size on the distribution of wealth and income is of considerable importance. But it does not compare in significance with the fundamental objectives of maintaining full employment or with the optimum use of resources, or with the proper division of resources between producing current consumption and adding to the equipment of society. Even its effect on the distribution of wealth and income is small compared to other influences. Increasing or decreasing the national debt, as we shall see, is one of the main instruments for achieving the fundamental objectives which must come first. But the use of this instrument does not necessarily involve any greater departure from ideals of distribution of income and wealth than may be necessary for other reasons. Steps can be taken to improve the distribution of private property in general and these will also deal with the addition to private property. The socialist's and the equalitarian's logical objections to increasing national debt apply only to an increase in private wealth in an uncontrolled economy where the distribution of wealth and income seems to follow a fairly definite law. There is no reason for supposing that this law of distribution will hold if the authorities of a controlled economy prefer another, more equalitarian, distribution of income and of private property.

The purpose of taxation is never to raise money but to leave less in the hands of the taxpayer.

The second great prejudice shows itself in the inability to see that taxation should never be imposed merely as a means of raising money for the government on the grounds that the government needs the money. The government can raise all the money it needs by printing it if the raising of the money is the only consideration.

This device is illegal for the private citizen and so it is usually regarded as somehow illegitimate for the government, by the same identification of the government with a private business that makes the government's debt look dangerous and which treats the government's revenue from taxes as equivalent to business earnings and therefore the only really proper source of money for

the government to spend. Of course, there are definite limits to the extent to which a government should pay out money (or indeed do anything whatever), but these limits must be defined in terms of the actual effects on the well-being of society and not derived from cloudy analogies with what is prudent and legal for a businessman.

The rational procedure is to judge all actions only by their effect and not by any vague notions of their propriety or impropriety. "*By their fruits shall ye know them.*" The effects of a tax are twofold. It increases the money in the hands of the government and, by decreasing the money left in the taxpayer's hands, it makes him spend less. The first effect is unimportant for the government, however important it could be to any citizen to be able to acquire money in this way, because the government can much more easily get the money it wants to have by printing it, without any fear of the police. The important effect is the second, and the question of taxing or not taxing should be governed entirely by whether this effect on spending by the individual taxpayer is desired or not. The effect, which is not easily obtained in any other way, is the basis on which a rational government uses the instrument of taxation. It will tax individuals, or a certain class of individuals, when it believes it to be socially desirable that they should not be so rich or should not spend so much. It will tax particular forms of spending (*e. g.*, on whiskey) as a means of decreasing them. It will tax more generally as a means of cutting down total spending when this is necessary to prevent excessive total demand and inflation. Taxation is important *not* as a means of raising money but as a means of cutting down private spending.

The purpose of borrowing is not to raise money but to make the public hold more bonds and less money.

The third prejudice is very similar to the second and shows itself in a difficulty in seeing that borrowing by the government should never be undertaken merely as a means of raising money, any more than taxation should be undertaken merely because the government needs to have money to spend for any purpose. As in

the case of taxation (or any other rational activity by anybody), the criterion must be found in the *effects* of the action. The primary effect of borrowing is to decrease the amount of money held by the public and increase the amount of government bonds held by the public. This will lower the value of government bonds in terms of money and thus raise the rate of interest. If the government wishes to bring about these effects it should borrow; if it does not it should provide itself with all the money it wishes to spend by printing it. Indeed if it wishes to bring about these effects on public holding of money and of bonds and on the rate of interest it should borrow even if it does not need the money raised in this way. It could store this money until such time as it would want to spend it, or, if that were more convenient, it could burn the money that is borrowed and print new money when there arises a need for spending. Borrowing money is not directly related to the current need for money to spend and must be judged by its own effects on the economy.

At the present time we are told over the radio every fifteen minutes that the government needs our money for the war effort so that any money we have left over after buying the particular product that is being advertised should be directed to buying war bonds and stamps. It is nevertheless not true that the government is short of the instruments of war because it has not received enough money from the sale of war bonds to be able to buy more tanks and planes and ships than it has already obtained. The limiting factor is not money; no government would be foolish enough (in time of war) to go without what it needs if it could obtain it by simply printing the money needed to enable it to buy the goods. The limits are in the shortage of strategic materials and skills, and no amount of money, whether printed or whether it is offered to the government by patriotic or interest-seeking bond buyers, will overcome this limitation.

The purpose of war bonds is only to make the public spend less.

This does not mean that there is no sense in the campaign to induce people to buy bonds. It only means that it is useful not

because of the government's need for money or even the soldiers' need for guns, but merely because it is necessary to reduce the amount being spent by the public on the diminishing amount of civilian goods available if the prices of these are not to be raised by the relatively excessive demand. The only important reason for the purchase of government bonds is that the money should not be used for buying other things. That is why no useful purpose is served by spending on government bonds money that otherwise would not be spent at all. This does not decrease the demand for civilian goods, it does not permit the government to buy more war materials, it merely increases the interest payments that the government will have to pay to the bondholder in the future. Fortunately, as we have seen, these interest payments are not as harmful to society as they are sometimes made out to be.

Borrowing and taxing can also be applied in reverse.

Borrowing and taxing can also be applied in reverse. If the government wishes to increase the quantity of money in the hands of the people and lower the rate of interest, it can do that by *repaying* some of the national debt. Even if there is no national debt it could bring about the same effect by *lending* money to business, establishing a *national credit.* (This would no more be an indication of national prosperity or wealth than the national debt is an indication of national insolvency or poverty.) If the government does not happen to have enough money on hand for the purpose, it can print as much as is needed, just as when it needs to spend money, and the desired effects will be brought about. If the government wishes to increase the wealth or income or expenditure either of particular individuals or classes of individuals or of individuals in general it can bring this about by reducing taxes. Where this is not sufficient to bring about the required results even when taxes have been reduced to zero, *negative taxes* can be imposed. This means that the government instead of taking money away from people gives it to them. This may take the form of relief payments, old age pensions, bonuses,

and even a social dividend when it is desired to increase consumption all round.

The effects of taxing and of borrowing overlap.

These two instruments, borrowing and lending on the one hand and taxes and bonuses on the other, are not as clearly distinct in their effects as is suggested above. Taxes and bonuses primarily affect spending, but they can also affect the rate of interest. If a tax or bonus is applied to a poor man, almost the whole effect will be on spending, but a man with a higher income is likely to decrease his spending by less than the amount of a tax and to increase his spending by less than the amount of the bonus. He will take the balance of the tax out of his savings or lendings or out of his holdings of cash, or he will add the balance of the bonus to his savings or lending or to his stock of cash. To the extent that this happens the effect of the tax is the same as that of government borrowing—it raises the rate of interest—and the effect of the bonus is the same as that of government lending or repaying debt in lowering the rate of interest. These taxes diminish lending by the taxpayer whose funds are taken away by the tax, and the rate of interest will rise. The bonus increases lending by the recipient, and this lowers the rate of interest.

Conversely the effects of borrowing and lending may have some of the effects on spending that we at first ascribed entirely to taxes and bonuses. Government borrowing, by raising the rate of interest, may induce some individuals to save more and to spend less. Lending money or repaying debt lowers the rate of interest and this may persuade some individuals to save less and spend more.

These effects are not likely to be very great. As we have seen in another connection, the effects on consumption expenditure through changes in the rate of interest can work in the opposite direction to that here indicated, so we may be content with just noting this possibility and continue to speak as if taxes and bonuses affect only consumption and as if borrowing and lending affect

only the rate of interest. An important exception to this is the current war loan drive which is an attempt to induce the public to cut down spending in order to buy war bonds and stamps. To the extent that this drive is successful in reducing spending and does not merely persuade people to buy bonds and stamps with money that they would have saved in any case, this borrowing by the government has the effect which we shall speak of generally in connection with taxation.

The indirect effect of a tax in raising the rate of interest, because of the reduction in cash that it entails, will be less than if the government had borrowed the same amount. The increased amount of government bonds would have helped to raise the rate of interest by increasing the quantity of other assets that people would try to exchange for cash, thus increasing the demand for cash to hold. In the same way a bonus does not lower the rate of interest as much as would an equal repayment of government debt because the increase in cash is not accompanied by a decrease in government bonds.

Taxing and spending, borrowing and lending, and buying and selling constitute the six fiscal instruments of the government.

We can now see how by using these different instruments to reinforce or offset each other the government can affect both consumption and the rate of interest in any way it likes. If it wishes to raise the rate of interest it will simply borrow money. Any undesired effect of this borrowing on the amount of spending can be offset by an appropriate reduction in taxes or increase in bonuses. If the government wishes only to decrease consumption, it must tax and at the same time lend out enough of the proceeds of the tax (or use it to repay national debt) to offset the effect of the tax in reducing the amount of money in the hands of the public and thereby raising the rate of interest. If it wishes to lower the rate of interest while decreasing consumption, it must impose taxes and at the same time lend out more than it collects in the taxes, and so on. The combination of these factors in government policy can become very complicated, especially when we take into

consideration the effects of changes in consumption on the need for money to hold and try to calculate the corresponding changes in borrowing or lending needed to offset the effects of the consequent changes in the rate of interest. Fortunately it is not necessary either for us or for the government to work out all these combinations. The government need merely borrow whenever it wishes to raise the rate of interest, lend (or repay debt) whenever it wishes to lower it, tax when it wishes to decrease consumption, and reduce taxes (or increase bonuses) whenever it wishes to increase consumption. This is how the instruments work. How and when they should be applied we shall see as soon as we have finished examining the instruments at the disposal of the government for the regulation of interest, investment, and employment!

The only other instruments at the disposal of the government are *buying* and *selling*. Requisitioning of goods and services can be considered as a combination of taxing the individual to whom they belong and buying the goods or services with the proceeds of the tax. Similarly the giving away of goods or services for nothing, or for a price less than the *msc*, can be considered as a combination of granting a bonus equal to the value of these goods (or the excess of their value or *msc* over the price at which they are sold) and then selling them in the ordinary way at a price equal to *msc*. There is therefore no need to count requisitions and gifts in kind as special instruments for carrying out the government's policy.

As to government buying and selling, very little can be said in general terms. The government will buy or sell anything that seems to it socially desirable to buy or to sell. It may buy battleships and airplanes because they are deemed necessary for national defense. It may buy roads or parks or hospitals to serve the nation's commerce or pleasure or health. It may buy wheat or cotton to raise their prices and increase the incomes of farmers and landlords. Even less can be said in general terms about what the government might decide to sell. There will always be particular reasons. One thing only need be mentioned—the government will never have to sell anything to its citizens merely in order to raise money. Not even bonds. It can always raise any money that it needs more easily by printing it. Of course a government may sell

things to other nations or to the citizens of other nations in order to raise foreign money with which to acquire goods from abroad. This is because the government cannot usually print and dispose of foreign money.

Printing money and destroying or hoarding money are subsidiary to these in the task of adjusting investment and consumption to give full employment.

This gives us the three pairs of primary governmental instruments for the regulation of the economy each with a positive and a negative side. The instruments are: buying and selling, borrowing and lending, taxing and giving out bonuses. Printing money is not mentioned as one of them because it is not an independent instrument. It is subordinate to these six. Without one of these it cannot have any effect at all on the economy since it will merely increase the money in the vaults of the government until it is paid out by one of these three instruments. If the money that comes in to the government treasury from selling, borrowing, and taxes is equal to or greater than the money needed for buying, lending, and bonus distribution, there is no need for any money to be printed. If the money coming in is less than the money that has to be given out and there does not happen to be enough money in stock in the government vaults, the printing press can be called upon to provide the money needed to carry out the government policies. The printing of money is not an instrument of policy. It is merely a servant of these policies, just like printing the stationery used in the various government departments.

Now that we have seen how the instruments work, we can consider how they are applicable to governmental policy in the controlled economy. First, the government decides on the buying and selling that is socially desirable for all sorts of particular reasons. Then it undertakes such taxation and pays out such bonuses as are justified by special particular circumstances; taxes on goods whose consumption it is desired to discourage without actually prohibiting them, such as whiskey; taxes that are really a form of charging a section of the population for special services, such

as automobile licenses; taxes on the very rich for the purpose of improving on the distribution of wealth and income; bonuses for the blind, relief for flood victims, subsidies to educational institutions and hospitals, and so on.

Given the total of government application of the instruments of fiscal policy, there will result definite values of the level of consumption, of the rate of interest, and of the level of investment. The government wishes to see a full use of the resources of society and a certain proportion of the resources devoted to investment. If there is insufficient total demand, so that there is unemployment, the government will lend money (or repay debt) to lower the rate of interest until the rate of investment is at the level it considers proper, and it will reduce taxes or increase bonuses until the level of consumption is enough, together with the investment, to produce full employment. The government may, in such a situation, also increase its own buying, which will contribute to the desired increase in total demand.

Spending may have to take the form of public works.

The reduction in the rate of interest may not be very effective in increasing private investment, especially in time of depression when businessmen may become so pessimistic that they would not invest even if the rate of interest were reduced to zero. The government could then undertake some investment itself as part of its own "buying" if it considered these "public works" a better use of resources than directing them to additional consumption in the present. Excessive pessimism on the part of businessmen is probably better cured by creating enough income, by reducing taxes, and by giving out bonuses and thus increasing demand for consumption goods and services than by the government's competing with business—for that is what public works will look like to the pessimistic businessman even if he is not willing to undertake the investment himself. But where it is politically difficult for the government to distribute enough in bonuses to bring about full employment, public works are beneficial even if their yield is very low or even negative. This is because they are then the only way

available to increase employment and the output of consumption goods (in response to the demand for them by the people employed in the public works). This is an eminently useful accomplishment in times of depression and worth while even if the investments turn out to be completely useless. Of course there is no reason why they should be useless unless useful investments are ruled out for the political reason that they appear to involve government competition with business. In that case useless investments would almost certainly also be ruled out by the same business interests as uneconomic in the businessman's sense, however useful from the social point of view may be their effects on employment and income in other parts of the economy.

The government should try to equalize the *msb* of public and private spending, counting also the indirect *msb* from increased employment.

The government should try, as far as possible, to equalize the *msb* of all forms of expenditure, public and private, in determining the structure of the total of all spending. With full employment this total is fixed by the consideration that an increase would bring about inflation and a decrease would bring about unemployment. The total of government spending might be fixed by political-legal restrictions on the government's deficit. Whatever the forces fixing the total of spending by the government or by the economy at large, equality of the marginal social benefits from the different directions of permissible spending is a necessary condition for the optimum use of the resources that can be employed under the restricting circumstances. When there is some unemployment it is better to undertake useless and even moderately harmful public works than to undertake none at all if this is the only way open to the government to increase total demand. The *direct msb* of the public works may be low or even negative, but the *indirect msb*, through the resulting increase in income, in demand, and in employment elsewhere to satisfy this demand, must be added to the *direct msb* to obtain the *total msb* from the public works. Reducing taxes or distributing bonuses to the same amount would be better than undertaking a harmful public works project, because the

indirect benefits would be just the same as those from the public works. But if "boondoggling" is ruled out for political reasons, even harmful public works may be better than doing nothing to increase employment.

If there is too much total demand so that there is full employment with rising prices threatening inflation, the government will reverse all these procedures. By borrowing money it can raise the rate of interest and discourage investment. By increasing taxes and reducing bonuses it can directly cut down consumption. And it may also reduce its own buying. In this way the level of interest, investment, and employment can be continuously adjusted in the optimum manner.

All items of public and private spending and taxing should be so adjusted that the *msb's* from the spending and the *msc's* of the taxes are all equal.

Just as total demand is adjusted to provide full employment, so continuous adjustment among the different elements in total demand is also necessary for the best use of the resources. The adjustment among the resources devoted to the production of different products or to different investments is brought about by perfect competition or by the application of the Rule in collectivist agencies of production. The adjustment between consumption as a whole and investment as a whole is brought about by government investment policy which is regulated through the rate of interest which in turn is adjusted by government borrowing and lending. Different expenditures by government on its different objects must be adjusted as far as possible to make the *msb* the same in different uses. Similarly the *msc* of different taxes must be equalized so as to minimize the loss by the taxpayers. Finally, the adjustment between government and private spending must be aimed at equating the *msb* of expenditure in the two fields.

This adjustment between all the different forms of public and private expenditure and the different kinds of taxes that are collected, all going on at the same time, means that the order in which we introduced the various items, first public buying and

selling, taxes and bonuses, and borrowing and lending, then private investment, and finally private consumption, was only an expository device. No item comes before any other item; all must be adjusted to each other. If there is not enough total spending, *all* forms of taxation must be reduced and *all* forms of spending increased until the *msb*'s (or *msc*'s in the case of taxes) have been reduced and are again equal to each other in full employment. If restriction of demand is necessary to prevent inflation, *all* expenditures should be contracted, and all taxes increased, until all the *msb*'s and the *msc*'s are equal to each other again at a higher level with a smaller total demand that safeguards the economy from inflation. Complete and perfect adjustment all round is, of course, impossible, and often very loose estimates of social benefit from different expenditures, of or the social cost of certain taxes, are unavoidable. But a controlled economy can at least aim at this ideal. However bad the aim, the result can hardly be as far removed from the ideal as in an uncontrolled economy where people are unaware of any uneconomic use of resources (except in blatant cases like mass unemployment) because they have not even seen the goal of optimum use.

Though there is no room for the *principle* of balancing the budget, there is a long run *tendency* for the budget to balance itself.

What many people find most disconcerting about a scheme like that sketched above is the complete disregard for, and even the absence of any reference to, the principle of balancing the budget. This is of course only another aspect of their horror of debt. It is supposed that the government should keep its books in exactly the same way as any normal business. Revenue from taxes is regarded as income. Government outlays are regarded as business expenditures. If income does not exceed expenditure, so that government debt does not increase and the government does not print any additional money to meet its obligations, the budget is said to be balanced. Otherwise the books are "in the red," and the cartoonist, drawing for the small businessman, makes hay.

We have already seen that the principle of maintaining full

employment and the chosen rate of investment completely determines the amount of borrowing or lending undertaken by the government, so that any other principle about the relation between tax revenues and expenditures, such as the budget-balancing principle, must either coincide with the policy already determined, in which case it is unnecessary, or be in conflict with it, in which case it must be rejected. No budget-balancing principle can be as important as maintaining full employment and preventing inflation.

The only reason for abiding by any principle of balancing the budget is that there is a strong prejudice in favor of such a procedure on the part of businessmen who think of the government as a business just like their own and on the part of a large section of the population in a capitalist society in whom the businessmen have been able to instil their own ideology. In a controlled economy, where a great part or even the greater part of the economy is run by businessmen, these prejudices are important and should be respected if possible. They are not more important than maintaining full employment or achieving the optimum use of resources or preventing inflation, and these objectives cannot be sacrificed to the businessman's feeling that the government should abide by "sound business principles." But if there is any way in which the budget can be balanced without giving up full employment and the optimum use of resources and without incurring inflation, it should be given the fullest consideration.

It is possible to maintain full employment while balancing the budget if demand is maintained by redistributing income.

There is such a way. A total demand inadequate to provide full employment can be increased without the government's increasing its expenditures and its bonuses as compared to its taxes; that is, without unbalancing the budget if it were balanced before. Where there is insufficient investment, the rate of interest can be lowered without printing or issuing new money. But these alternatives are not likely to be any more acceptable to the businessman.

Total demand can be increased by a redistribution of income from the rich to the poor. Increased taxes on the rich, offset by

decreased taxes on the poor or by greater bonuses to the poor, will increase total demand without unbalancing the budget. The rich will decrease their spending very little while the poor will increase *their* spending by almost the whole of the reduction in their taxes or the increase in their bonuses. The rate of interest can be reduced in a similar way by further taxation of the rich who will take some of the money to pay the taxes out of their money stocks or hoards, and then *lending* the revenues on the market or using them to repay government debt. This will even improve the budget situation while lowering the rate of interest to increase investment. In fact this policy attacks the root of the general inadequacy of demand by correcting the maldistribution of wealth that is responsible for it in the first place. However, this is not likely to be very popular with businessmen because while it spares their prejudices it hits their pockets, and these are perhaps even more sensitive than their prejudices.

Businessmen's prejudices against functional finance are best met by a determined maintenance of adequate demand.

The confidence of businessmen is important because so much of the enterprise and the investment of society is in their hands. This confidence is upset as much by a large unbalancing of the budget as it is by a much larger one that may be necessary to achieve sufficient total demand to provide full employment in the absence of a drastic redistribution of income. It is therefore the very worst possible policy to unbalance the budget sufficiently to upset the businessmen without going all the way to bring about full employment. The businessmen will reduce their investments because of their loss of confidence in the probity of the government and may reduce total demand by more than it is increased by the government. The resultant fall in income and employment will confirm their pessimism and make matters still worse, and it is likely to result in a cumulative decrease in economic activity.

The wise course for the government is the brave one of going all out and bringing about full employment no matter how much it has to resort to deficit spending. Businessmen will be hardly

more pessimistic to begin with than on the other policy, since a 6 billion dollar deficit cannot possibly make them feel much more unhappy than a 3 billion dollar deficit. On the other hand, they will see that consumers have the money to buy the goods that they are able to produce and sell at a profit. Their pockets will ultimately overcome their prejudices. There will, of course, be great grumblings, and the prosperity will be called "artificial" and "unsound" and even "illusory." But production and investment will be profitable and these epithets will disappear more quickly than the gibes at horseless carriages if the government sticks to its guns and maintains full employment for several years. Once this is done no government will be able to go back to the tender mercies of the trade cycle.

There are also some devices for making functional finance look more like traditional finance.

Attempts may be made to win the businessman for the control of prosperity by complex devices for making him believe that the budget is being balanced all the time. This has been done with some success in Sweden. By the use of many different kinds of budgets, annual budgets, capital budgets, extraordinary budgets, and budget equalization funds, a rational policy has been carried out in disguise. When necessary the details of operation of the budget-balancing principles are revised to help in this. But it does seem rather silly to have all this pretence.

The objection that functional finance interferes with free choice between saving and spending is extraordinarily empty.

In the course of applying these rational principles to the regulation of investment and employment, the government inevitably influences the rate of interest and the proportion of the nation's income that is devoted to investment. This has been critized as an interference with the free choice of the citizens between consuming their income and saving it—a freedom of choice that citizens have in the uncontrolled capitalist economy.

This is an extraordinarily empty objection. The choice between present consumption of one's income and saving it depends much more on the distribution of income than on the "time preference" on which the controlled economy is said to trespass. But the great weakness of the objection is that the individual in the uncontrolled capitalist economy who decides to save does not shift resources from producing present goods to producing goods in the future. He merely sets them free so that they become unemployed. The decrease in demand by the men who are put out of work by this and who now have no wages to spend leads to still further decreases in employment.[1] In fact the amount that people save is determined in the uncontrolled economy, not by their thrift, but by the investors who raise or lower the total income of society to the level where consumers are freely willing to save exactly the amount that the investors have invested. It is only this illusion of freedom of consumers' choice that exists in the uncontrolled economy. And even this is not lost, since the individual is still free to save or spend his own income in any proportion that pleases him. What is gained is the possibility of a careful consideration of people's attitudes as between present and future by the government when it makes a democratically controlled decision as to how much of the resources of society to devote to current consumption and how much to devote to increasing and improving society's equipment for producing goods in the future. In doing this, subsidiary devices such as those discussed in Chapter 21 can be put into service to give the consumer a real influence over the allocation of resources between present and future consumption.

[1] For example, suppose everybody decided to save an additional 5 per cent of income when income was at the 70 billion dollar level, with consumption at 63 billions and investment at 7 billions. An increase of saving from 10 to 15 per cent of income would require an additional investment of 3½ billions if income is not to fall, since the gap between income and equilibrium consumption would be increased by this amount of additional saving. In the absence of any such increase in investment income would fall to about 61 billions and consumption to about 54 billions, a fall of more than two and a half times the desired increase in saving which does not take place at all. If investment fell off, and this is the most likely effect of the decrease in demand for consumption goods, income would fall still further.

CHAPTER 25. CAPITAL, INVESTMENT, AND INTEREST

In the present chapter we shall concern ourselves with some theoretical issues that arise out of the problems connected with interest, investment, and employment dealt with in the last four chapters.

Investments are not usually consumed after one year.

We have said that the marginal yield from postponing consumption is positive because it is possible, by postponing present consumption, to set free resources which can be used to improve the equipment of society enough to permit future consumption to be increased by more than present consumption has to be sacrificed. In this discussion it is difficult to avoid the suggestion that the sacrifice of 100 units of consumption this year, which permits an extra 110 units of consumption next year as its marginal product, is actually accompanied by such an increase of consumption next year.

If this were so, the improvement or increase in equipment would only be *temporary*, nothing of it being left after the 110 units were consumed next year. This is implied in the concept of the marginal product. If there were anything left of the increased equipment after next year and this permitted any further extra output at any time in the farther future, this further extra output would have to be added to the extra 110 units available next year as part of the marginal product of the original sacrifice of 100 units this year. When the marginal product is declared to be 110 units next year, it is implied that all other present and future inputs and outputs are unchanged. Such a peculiar temporary improvement in equipment is worth more careful study.

If the sacrifice of 100 units of consumption goods this year is to permit an increase of more than 100 units next year, it is necessary that advance notice of the abstinence from consumption be given so that the 100 units of consumption goods available this year are not produced, and the resources are directed instead to increasing the equipment of society in a way that will permit a future increase in output greater than the present sacrifice. It will obviously not do merely to refrain from consuming what has already been produced. This could only lead to the product being wasted or, at the very best, stored (if the goods are not too perishable), and this would permit an increase of future consumption which is *less* than the present sacrifice by the amount of wastage and spoilage and the costs of (that is, resources used up in) the process of storage.

All replacements are really devoted to future output.

If it were known in advance that consumption this year was to be reduced by 100 units, the resources that would have been devoted to producing them could have been directed to increasing next year's output. These resources consist only in small measure of factors of production actively applied this year. The greater part of the factor services of any year is applied to the replacement of equipment worn out and the replenishing of stocks used up. In making these replacements they are really directed toward *future* consumption even though it is a common capitalistic bookkeeping practice (and quite properly so) for the firm to count the outlay on such factors as current costs to be subtracted from the revenue from this year's output to show the current profit. Current resources are for the most part devoted to replacement of equipment that wears out, including the replacement of stocks. (Stocks are an essential part of equipment since production could not be carried out smoothly without them.)

Equipment can be considered as "imprisoned" factor services.

All existing equipment (except that provided by nature) has been made by factors of production applied in the past and (since

the equipment is not yet worn out) devoted to consumption in the future. The existing equipment may be considered as incorporating these factor services applied in the past and holding them until the equipment is worn out in the course of producing the final consumption goods. At that point the past factor services will be released from the equipment where they have been imprisoned since their first application. The value of the equipment will tend to equal the value of the factor services "imprisoned" within it plus the interest for the time they have been locked up there. Otherwise, there will be a profit in increasing the equipment (if the value of the equipment is greater) or it will not pay to replace it as it wears out (if the value of the equipment is less). These increases or decreases in the quantity of equipment will tend to bring about equality between the value of the equipment and that of the factor services incorporated in it, including the interest for the time they have been locked up in it.

The relation between the quantity of capital and the flow of services corresponds to the average *time* the services are "imprisoned."

When equipment (including stocks) is looked at in this way there is seen to be a correspondence between the *ratio* of the quantity of equipment to the flow of factor services available in the economy and the *average time* elapsing from the application of the factors to their emergence in consumption goods. This correspondence is of the same nature as that between the *ratio* of the quantity of water in a lake to the rate at which water flows into it (and out of it) and the *average time* that each drop of water stays in the lake before flowing out again. If the volume of water in the lake is equal to *twice* the annual flow of water into it (and out of it), the average time a drop of water stays in the lake is *2 years*. Another analogy would be the correspondence between the ratio of the number of men in an army to the annual number of recruits (or retirements including deaths and expulsions, etc.) and the average length of time that a soldier spent in the army before he left it. If there is an equal recruitment (and retirement), every year of 1 million men, an army of 3 million men will mean that the average

army life is 3 years, and an army having an average life of 4 years would mean that the army is 4 million strong. In the same way, if the services of factors incorporated in the existing equipment (and therefore not yet escaped in the form of final consumption goods) is 4 times as great as the annual supply of these services, the average time that a unit of factor service is imprisoned is 4 years. This is called the *average period of production*. If the average period of production is 5 years, the existing equipment incorporates factor services equal to the total amount made available in the course of 5 years, and its value will tend to equal five times the value of the annual flow of the factor services plus the interest on each unit of factor service, applied at compound interest from the time it was applied in the production of the equipment.[1]

The postponement of consumption permits a lengthening of the period of production. If the average period of production were 4 years and if consumption were reduced to zero for a whole year, and if there were sufficient notice to make possible the best use of the knowledge of the interruption of consumption, and if it were believed that after this year there would again be a permanent consumption of the whole of the net product of consumption goods, then the average period of production could be increased from 4 years to 5. The equivalent of a whole year's supply of factor services would be incorporated in increasing the equipment of the economy. The greater equipment would, after the first year of fasting, make possible a *permanently* larger flow of consumption goods. How much larger the new output would be than the old would depend upon the technical possibilities of increasing production by the use of more and better equipment (that is, a longer average period of production) and upon how rich in equipment the economy was to begin with. The longer the original average period

[1] In the technical literature on this subject great difficulties were encountered in attempting to distinguish between *original* factor services which are applied to the production of equipment (but which could have been used to provide consumption goods and services directly) and *derived* factor services which are given off by the equipment later. The ratio we are here discussing was expressed as the ratio between the flow of *original* factor services and the stock of equipment considered as a reservoir of original factor services incorporated in them and not yet given off in the production of consumption goods and services. It does not seem necessary to make this distinction between *original* and *derived* factor services, so the difficult and probably insoluble problems in making the distinction need not bother us.

of production (that is, the greater the degree to which the economy is already supplied with equipment), the less effective will be additions to the existing equipment involved in a lengthening of the average period of production.

The postponement of consumption for a year may be regarded as a temporary lengthening of the average period of production.

If is, of course, impossible for an economy to go without any consumption for a year, but the same principle holds for any release of resources by a reduction of consumption if the resources thus set free are devoted to increasing or improving the equipment (that is, not if the resources free by the decrease in consumption are merely lost in unemployment). A reduction of consumption by 100 units will release resources, and these, if applied to lengthening the period of production, might permit future output to be greater by 10 units per annum for as long as the new, longer, period of production is maintained.

If the output of current consumption goods is at such a level that the factor services remaining after providing the current consumption are just sufficient to replace the equipment and stocks that are being used up, we have a stationary state in which the average period of production (or the quantity of equipment) and the flow of output do not change. If consumption is at a *lower* level, some of the factor services can be devoted to lengthening the average period of production. Consumption can also, for a time, be at a level *higher* than that which leaves just enough factor services over to replace existing equipment and stocks as they are used up. If this happens there is a deterioration of equipment and a shortening of the average period of production. Thus if consumption is 100 units above that level which uses up equipment and stocks at the rate at which they can just be replaced, there will result a shortening of the average period of production. If a further shortening of the average period of production is to be avoided, future output will have to be reduced by, say, 10 units per annum below the level that would have just permitted the old average period of production to be maintained.

We can now express the postponement of consumption for a year in terms of its effect on the average period of production. The sacrifice of 100 units of consumption this year permits a lengthening of the period of production and this permits the output to be increased by 10 units per annum for as long as the longer average period of production with its better equipment is maintained. But the longer average period of production is not maintained. Next year consumption is increased not only by the 10 extra units yielded by the longer average period of production, but by an additional 100 units that *shortens* the average period of production and reduces future outputs by 10 units per annum back to the old rate with the shorter period of production that was in effect before consumption was postponed in the first place.

We can see from this example that the marginal yield from the postponement of consumption equals the marginal yield from resources applied to the lengthening of the period of production. Both are equal to 10 per cent, and this figure is seen more directly in the permanent yield from a once-for-all lengthening of the period of production. The same is true when this is done in reverse and a *disinvestment* (of 100 units of consumption) reduces future yields by 10 per cent (or 10 units of consumption goods per annum).

The average period of production is a useful concept because it emphasizes the connection between capital equipment and the *time* aspect of production. It shows how capital is needed only because production takes time, and that more capital is needed when production takes place by methods that take a longer time. Although this is obvious to the manufacturer who has to borrow capital because he has to wait until he is paid for his product, it is in danger of being forgotten when the economy as a whole is considered. This danger is especially strong in the case of socialists to whom the notion of interest and capital always suggests exploitation, or the improper receipt of interest payments and the wielding of undue economic power by capitalists. These unpleasant thoughts hinder the study of the economic implications of time-consuming methods of production that go with the use of capital equipment, whether owned by a few capitalists, or in equal measure

by all the citizens, or by the state itself. A kind of Freudian repression has permitted many socialists to hold on to an obsolete labor theory of value that leaves out all these things.

Alternatively all investments can be treated as if they were permanent.

But useful as is the concept of an average period of production in emphasizing the role of time in production, there are serious difficulties connected with its use, which we cannot go into fully here, and many economists do not like to use it. Instead it has been proposed to speak as if all equipment were "permanent," by virtue of its replacement whenever it wears out, and when it is in existence it permits current factors of production to produce a *permanently* larger current output. If this procedure is followed we would say that the sacrifice of 100 units of consumption this year permits some additional equipment to be constructed, and this additional equipment enables the cooperating resources to produce 10 units more of consumption goods per annum. Next year the new equipment is turned into 100 units of consumption goods which are consumed together with the 10 extra units it permitted to be produced before it was turned into consumption goods. This seems to leave in some doubt the way in which a "permanent" piece of machinery is converted into consumption goods, but that belongs to the same issue as the objections to the average period of production that we cannot go into here. We may, however, suppose that all that is meant by "permanent" is that when the investment is made the intention was to maintain it permanently by replacing it when it was worn out. This does not make it impossible for society to change its mind in the future and consume the "permanent" addition to equipment by failing to replace it when it does wear out.

The way in which we have described postponement of consumption in terms of the average period of production and in terms of "permanent" investment does not mean that these latter forms of describing investment are more fundamental than a description in terms of postponing consumption for a year. The procedure can be reversed and a permanent investment even more

simply described in terms of annual postponements. A permanent investment yielding 10 per cent per annum is nothing but a continually repeated annual postponement of consumption. The postponement of 100 units this year permits an extra 110 units to be produced next year. Next year only 10 of these are consumed and the other 100 postponed again permitting 110 extra units to be produced the following year. In this way the original sacrifice of 100 units permits the enjoyment of 10 more units every year forever.

A better "atom" is the postponement of one dollar for one year.

All three forms of expression are equivalent and any might be used according to the particular purpose in hand. Probably annual postponement is the most convenient as an elementary unit out of which more durable investments may be considered to be built up. The availability of 110 units next year does not mean that they have to be consumed any more than the 100 units that were available this year had to be consumed. Rather it must be interpreted to mean that next year there will arise the option of consuming the 110 units and returning to the status quo, or consuming more than this and reducing the equipment of the economy below its original level, or consuming less than the 110 and leaving the equipment greater than in the beginning. Whichever way we do this, we get the same figure for the marginal yield from postponing consumption or the marginal efficiency of investment.

In a stationary economy the *marginal productivity of capital* is equal to the marginal efficiency of investment.

The permanent increase in output, made possible by the increase in society's equipment which we can call an increase in the average period of production, should properly be called the marginal productivity of capital by analogy with the marginal productivity of any other factor, since it is the factor "capital" that is increased. The marginal productivity of any factor is the amount by which the output of the product can be increased by using a unit more of

the factor while the quantity of the other factors is left unchanged. The marginal productivity of capital should by analogy be used to indicate the increase in the output of goods when an additional unit of capital is used together with the same quantity of other factors. It is measured as so many per cent per annum, instead of just so many units of product per unit of capital, because the increment of capital is measured by the number of units of consumption goods that are devoted to increasing it. The marginal product compared is in the same units as the factor, so the relationship between the increment of capital and its marginal product can be expressed as a ratio or percentage.

In the examples we have just considered the marginal productivity of capital was equal to the marginal efficiency of investment or the marginal yield from postponement of consumption. This does not normally occur. It was so in the examples we used because we were making the special assumption of a stationary economy to begin with, and it is only in a stationary economy that the marginal productivity of capital and the marginal efficiency of investment, *mei*, equal each other. To make this clear we shall have to distinguish between the individual and the social point of view.[1]

Only individuals (or small parts of the economy) are free by borrowing to adjust the quantity of their real capital to make its marginal productivity equal to the rate of interest.

Problems concerning the marginal productivity of capital can arise only from the point of view of a *part* of the economy that is not too large. The most important of these partial points of view is that of the individual. It will be convenient to speak of the individual point of view even though the argument is equally true if several individuals are taken together or if the point of view is that of some other small section of society. Marginal productivity of capital is relevant only to individuals (in this extended sense) because only an individual is free to increase the quantity of any

[1] The argument of the next few pages follows a paper "Capital, Investment and Interest" read by the author before the Manchester Statistical Society in April 1937. A summary is published in the Society's *Report of Group Meetings*, 1936–1937.

capital asset while leaving constant the quantity of other assets he uses so that he can see the marginal product of such an increment of a capital good. He can do this by *buying* the capital good, *borrowing* the money if necessary. He will buy assets up to the point where the marginal productivity of capital (that is, of each of his assets) is equal to the money rate of interest he must pay on his debts.

Society can adjust its capital only by investing or disinvesting and this takes time.

Society is not able to obtain assets instantaneously in this way by buying them because there is nobody from whom society can buy or borrow. In order to be able to speak about the marginal productivity of capital from the point of view of the whole economy, economists must compare two stationary economies with a small difference in the quantity of capital equipment but with the same quantity of other factors, labor, land, etc., each stationary economy being perfectly adjusted to a state of affairs that is expected to remain unchanged forever. In fact this is what we did above when we compared societies with shorter and longer average periods of production or with smaller and greater quantities of "permanent" equipment.

If this is done it is possible to speak of the difference in the outputs as the marginal product of the increment of capital. If the difficulties arising from the heterogeneity of both capital goods and consumption goods were overcome so that we could express the difference in the capital goods of the two economies in terms of a quantity of homogeneous consumption goods, it would be possible to express the marginal productivity of capital as a percentage of so much per annum. When we have done this we must admit that there is no reason for expecting the social marginal productivity of capital to be equal to the rate of interest. Only if society, like an individual, could instantaneously increase or decrease its capital by *buying* an increment of capital whenever the yield was greater than the rate of interest, or *selling* it when the yield was less would this equality tend to be brought about.

It is possible for the whole economy to increase it capital, but only by *investing*. (It can decrease its capital by *disinvesting*, that is, by consuming more than is being replaced. The whole argument is reversed for this case and need not be repeated.) Investment is no exercise in comparative statics, in which different stationary societies are imagined, but is a dynamic process that *takes time*. If there is full employment, investment involves a *present* sacrifice which has to be balanced against *future* benefits. On the other hand the individual, who can buy earning assets, has only to compare *future* earnings with *future* interest payments which he will have to pay on his borrowings to buy the capital goods (or which he could have got by lending out his own money instead of using it to buy these capital goods). The equilibrium reached for society does not decide the *quantity* of capital that it is worth holding at the current rate of interest (which is what is decided by the individual) but *the rate per unit of time* at which the capital is to be acquired, or in other words *the rate of investment*. The analogy is rather with the quite different problem confronting the individual when he attempts to adjust the *net sum* of his wealth (his capital after subtracting his debts) by saving out of his income. He does not then decide how rich he would like to be and immediately go to that position. That of course is not possible. What he can do is to decide how much he will add to his wealth per unit of time—his rate of *saving*—and immediately start saving at that rate.

The individual adjusts his rate of saving per unit of time to the rate of interest, and society will do the same through the investors, whether they are private investors or managers of collectivist agencies following the Rule. But whereas for the individual a high rate of interest means a greater reward for saving (although this may not increase the rate of saving), for society a high rate of interest is a *discouragement* to investment and so to social saving. The social adjustment of investment (and therefore also of social or total saving) to the rate of interest comes about as the result of two separate activities. The users of assets bid up their prices until the marginal productivities of capital funds devoted to buying them (from the point of view of the individual) equal the rate of interest. The producers of the assets adjust their rate of output of

the assets to the points where $vmp = pf$ and $p = vmf$. The rate of output of assets (over above replacement) is the measure of the (net) investment or the total of saving by society. The combination of the two influences enables us to relate investment to the rate of interest. For every rate of interest there is a corresponding investment, and the lower the rate of interest the greater the demand for and the output of assets and the greater the rate of investment. This relationship is *the schedule of the marginal efficiency of investment.*[1]

The marginal productivity of capital is the marginal efficiency of investment when the rate of net investment is zero.

Corresponding to the individual's adjustment of his saving per unit of time to the rate of interest is society's adjustment of its investment per unit of time to the rate of interest (although society will always invest less if the rate of interest is higher, while the individual might be induced to save more by a higher rate of interest). The individual also adjusts the quantity of capital assets in his possession to the rate of interest by buying or selling them (borrowing money for this if necessary) until the (private) marginal productivity of each capital asset is equal to the rate of interest. Can we fit a social marginal productivity of capital in here to complete the picture?

To do this we must, like the classical economists who attempted the same thing, consider a stationary economy. If we then imagine a very small increase in the amount of capital without specifying the period during which the investment involved in producing it takes place, the investment and the increase in capital come to be the same thing. This is what we did when we considered a sacrifice of one hundred units from consumption but did not mention the period in which the abstention took place. By spreading a small

[1] Mr. Keynes, in his *Theory of Employment Interest and Money*, calls this the schedule of the marginal efficiency of capital, and what we have called the marginal efficiency of investment (*mei*) he calls the marginal efficiency of capital. This is probably because any point on the schedule represents a situation in which both the *social* marginal efficiency of *investment* and the *private* marginal productivity of *capital* are equal to the rate of interest. His terminology fails to distinguish clearly between these two fundamentally different concepts.

investment over an indefinitely long period we can reduce the *rate of investment* as near as we please to zero. This approximates the stationary society where the rate of investment *is* zero. The marginal productivity of capital is therefore measured by, and can be defined as, *the marginal efficiency of investment when the rate of net investment is zero.*

With this definition everything falls into place. If the marginal productivity of capital equals the rate of interest, net investment will be zero and there will be no tendency for the capital equipment either to increase or to decrease. If the marginal productivity of capital is greater than the rate of interest, net investment will be positive and the capital equipment will be increasing. If marginal productivity of capital is less than the rate of interest, net investment will be negative and the capital equipment will be decreasing. The adjustments, however, are not instantaneous as in the case of the individual, who can immediately adjust the quantity of capital by buying or selling. They can only take place gradually as additional equipment is built or as old equipment is worn out and not completely replaced. The *rate per unit of time* at which the capital increases when its marginal productivity is greater than the rate of interest, or decreases when its marginal productivity is less than the rate of interest, is determined by the schedule of the marginal efficiency of *investment.* Investment will be carried on at that rate per unit of time which makes *mei* equal the rate of interest. A greater rate of investment would not pay, and a smaller one would indicate the foregoing of possible profits. Any rate of investment other than that which equated *mei* to the rate of interest would also be contrary to the Rule. When there is a positive rate of net investment and capital is increasing, however slowly, the marginal productivity of capital is above the rate of interest. When there is negative net investment and capital is decreasing, its marginal productivity is below the rate of interest. The difference between the marginal productivity of capital and the rate of interest is the *force* which makes the stock of equipment grow or decline. Only in a stationary society does this difference disappear, and then the marginal productivity of capital and *mei* are both equal to the rate of interest.

This can be illustrated on a three-dimensional diagram.

These relationships are illustrated in Figure 5, which shows a three-dimensional figure on which there is a surface made up of the schedules of the marginal efficiency of investment, one such curve for each quantity of capital. These are curves like *AB* or

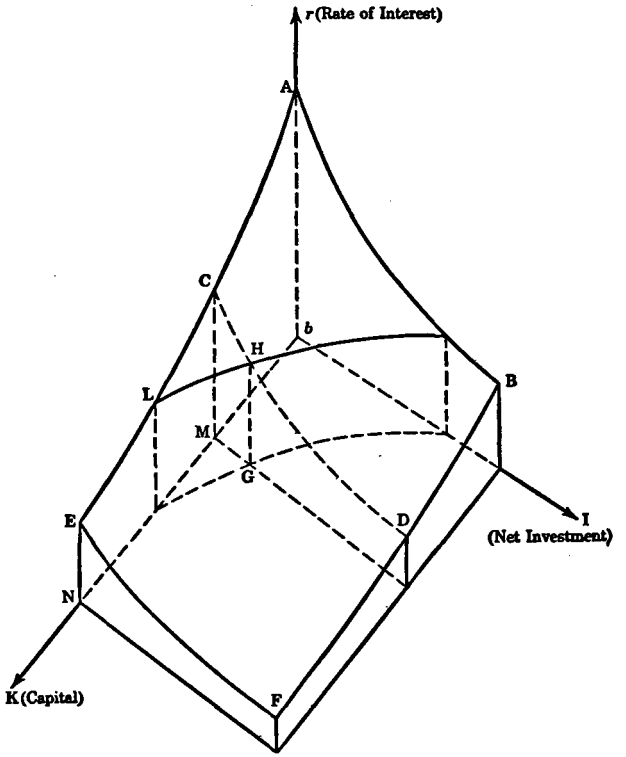


Figure 5

CD or *EF* in the figure relating the rate of investment, *I*, to the rate of interest, *r*. Each of these curves shows how great a rate of investment must take place at various rates of interest (measured vertically) given the quantity of capital in existence (measured along *bK*) if the marginal efficiency of investment is to equal the rate of interest. The whole set of these curves of the schedule

of the marginal efficiency of investment, for all the intermediate quantities of capital between *b* and *N*, constitute the surface *ABFE*. The greater the quantity of capital the lower the corresponding schedule of the marginal efficiency of investment, because there are fewer opportunities for building new equipment. The schedule of the marginal productivity of capital is shown by the curve *AE*. If the quantity of capital is *bM* (*b* is the *base* from which the quantity of capital or the rate of net investment is measured), the marginal productivity of capital is shown by the vertical measure *MC*. If the rate of interest is shown by the height *GH*, the rate of net investment will be *MG* where the *mei GH* equals the rate of interest. If the rate of interest is kept at this level and no other changes take place, the development of the economy can be indicated by a movement from *H* toward *L* along the line *HL*, which is drawn on the surface at a constant height to indicate a constant rate of interest and *mei*. At *L* there is no longer any net investment; a stationary long period equilibrium is reached in which the rate of interest equals both *mei* and the marginal productivity of capital. (*L* would in fact *never* quite be reached because the rate of movement toward it, which is given by the rate of net investment, diminishes to zero as *L* is approached.)

The same analysis and diagram are applicable to an individual (or other section of the economy). There are four possible cases. The first is that of an individual in a perfectly competitive economy with no fixed factors or indivisibilities. Here the rate of interest, the marginal productivity of capital, and the marginal efficiency of investment are unaffected by the scale of his activities. The surface *ABFE* becomes a horizontal plane on the same level as the current rate of interest, and the situation is completely indeterminate. We have seen that this is likely to lead to the establishment of a monopoly which would change the situation.

The second is where the individual has some personal ability or some other fixed asset of which he cannot get any more. The surface will then slope downward in the direction of *K* (as *AE* and *BF* do in Figure 5) indicating a decreasing marginal product as more of the variable factors are applied to the fixed factor,

changing the proportions. But the surface will be horizontal in the *I* direction. This is because there is no limit to the rate per unit of time at which he can acquire assets by buying them, borrowing money for the purpose if he has not enough of his own. The indefinitely great rate of "investment" means that he can move at once to the position corresponding to *L* by *buying* the quantity of capital goods which makes the (private) marginal productivity of capital equal the rate of interest. Once he gets there, there is no tendency for further expansion, so it would be possible for perfect competition to be maintained if it existed in the first place.

The third case is where the individual with the fixed factor is also a monopolist so that the factor services increase in price when he buys them at a greater rate per unit of time (that is, there is not perfect competition in buying). He will then limit the rate of investment to the point where the marginal yield, in spite of the rising price of the factor, is just equal to the rate of interest and will invest at a greater rate only if the rate of interest is lower. The surface of the figure slopes down both in the *I* direction and in the *K* direction just as for society and as shown in Figure 5. Both the ultimate size of the firm and the rate at which this size is approached are limited by the rate of interest.

The fourth and last case is where the individual can obtain *all* sorts of assets, none of them being fixed or significantly indivisible, but the prices of the factors are increasing functions of the rate at which they are being acquired (there is not perfect competition in buying). The surface will then slope downward in the direction of *I* but will be horizontal in the direction of *K*. There will be a limit, set by the rate of interest, at which the individual firm can expand, but the ultimate size is indeterminate and there is the same ultimate threat to the maintenance of perfect competition in selling as in the first of the four cases discussed. These four cases, in which the surface is considered to be downward sloping in both the *K* and the *I* direction, or horizontal in both directions, or horizontal in one and downward sloping in the other, or downward sloping in the one and horizontal in the other, comprise all the possible combinations.

The effects of unemployment can be brought in.

To the degree that there are unemployed resources, society is in the position of an individual with some fixed factors but with perfect competition in buying all those that can be acquired. This is the second of the cases just considered. In a time of acute depression there will be a range over which the surface of Figure 5 is horizontal in the *I* direction. This corresponds to the availability of unemployed resources of all kinds that permit investment to be carried on on a larger scale with no increasing sacrifice of competing consumption to cause the price of factors to rise and *mei* to fall. Output and investment can be varied at constant marginal cost, so that if investment were reduced to zero the marginal efficiency of investment would not change. Consequently *mei* is equal to the marginal productivity of capital even if investment is not equal to zero. This result is due to the peculiar condition of *general* unemployment when an increase in investment and output does not involve a change in the *proportion* between activities in the different parts of the economy, but only a change in the level of activity all round. More equipment is used with more of the other factors. There is no change in the *ratio* of capital to other factors and that is why there is no change in the height of the surface (or the curve) for any movement in the *K* direction. When there is no change in proportion there is no diminishing substitutability to complicate matters or diminishing returns or diminishing *mei* or diminishing marginal productivity of capital. An increase of investment is accompanied by an increase of everything else, and nothing is changed but the scale of operations. Naturally this is only a limiting case, for unemployment is rarely quite general with some of *every* factor and instrument of production unemployed and available immediately at a constant price. But the lower the level of employment the nearer is this situation approximated, and in severe depressions like that of the early thirties it is not too far-fetched as a description of the actual state of affairs.

We may now recapitulate our conclusions about the relationship between the marginal efficiency of investment and the mar-

ginal productivity of capital. In short-period equilibrium the rate of interest equals *mei*. In long-period (stationary) equilibrium the rate of interest equals both *mei* and the marginal productivity of capital. The time it takes to reach an equilibrium, long or short, depends on the point of view taken. From the point of view of the individual the short-period equilibrium is reached very quickly—as soon as he has decided the rate per unit of time at which he will acquire assets—and the long-period equilibrium does not take very long either since it is reached as soon as he has acquired the quantity of earning assets, or productive equipment, that he considers right and does not want to increase or decrease it any further.

The social short-period equilibrium is reached fairly soon—as soon as *all* the firms have adjusted their rate of investment and output to maximize profit at the current rate of interest. For the collectivist section of any economy it is reached when all the managers have adjusted the rate of investment and output to the rate of interest according to the Rule. In either case it is further implied that consumers have adjusted their consumption to their income, reaching the equilibrium consumption we discussed in Chapter 22 (p. 275). It is perhaps a matter of a few months before *mei* and the rate of interest are brought into approximate equality.

The social long-period equilibrium is reached when the equipment of society has been raised to the level at which no net investment is undertaken at the current rate of interest. Apart from the special case of general unemployment, this would not be reached for decades or maybe centuries even if there were no other changes (in taste and in technical knowledge) to upset it. It is only in the social long-period equilibrium, when we are all dead (or else economically prostrated by a depression that reduces net investment to something like zero), that the social marginal productivity of capital is equal to *mei* and the rate of interest.

The capital concept is essentially static. Practical problems are never concerned with capital but only with investment.

The foregoing analysis of the marginal productivity of capital brings out the *static* nature of the capital concept. Whether we

speak of the period of production or of the quantity of equipment as a measure of capital, clarity cannot be obtained (even theoretically) without assuming a stationary society in which the increment in product attributable to the existence of another unit of capital is not complicated by the effects of investment activity. In our formulation of the issue, and in fact in the very definition of the marginal productivity of capital, this is seen in the necessity of defining *net* investment. To do this it is necessary to decide just how much gross investment would be needed to maintain the existing capital equipment and so we have a stationary economy. There are many difficult and interesting problems connected with this, but they are not of any practical importance. Practical social problems are never directly concerned with capital or with the marginal productivity of capital because capital can only be affected *through* investment. If *mei* is equated to the rate of interest by the Rule or by the quest for profit under perfect competition, investment is determined at the proper rate and we have the optimum utilization of the available resources of society. There will, of course, be effects on the quantity of capital, but these all look after themselves if the rate of investment is properly regulated in the course of the optimum utilization of resources.

It may be mentioned once more that, for the purpose of establishing the optimum use of resources and in particular making *mei* equal to the rate of interest, there need be no concern with the measure of *net* investment (which raises all the problems connected with the stationary economy). In Figure 5 investment is measured from the base (b) which corresponds to zero net investment. This was necessary to show the curve AE which represents the marginal productivity of capital. But for practical purposes investment might be measured from zero *gross* investment or from any other arbitrarily chosen base. This matters no more than whether we express the temperature by the number of degrees above the freezing point of water or above zero degrees Fahrenheit or above absolute zero. For the optimum use of resources investments of all kinds must be extended to the point where their marginal efficiencies are equal to the rate of interest and so to each other and to the relative excess marginal valuation of present as com-

pared with future consumption. This will be the same whatever the arbitrary base from which we measure the investment.

The rate of interest differs with risk, liquidity, and the period considered.

Up to this point we have spoken as if there were only one rate of interest. Actually many different rates of interest are paid by individuals who borrow money, and the rates vary for different reasons. First, there is the greater or smaller risk run by the lender that he will not be paid the interest or even repaid some or all of the money loaned. To compensate for this the lender usually charges an especially high rate of interest as a kind of insurance. This extra charge should be separated from the interest proper and be regarded as a risk premium. Pure interest is what is paid for a loan when there is no fear of default.

Second, there is the risk to the lender, even if there is no danger of default, that he may need the money before the repayment of the loan is due. This is the true sacrifice that is made whenever money is loaned. If the borrower is well known and trusted, it will be relatively easy for the lender, if he should need the money before the repayment is due, to sell the promissory note, or other title to the repayment of the loan. The rate of interest charged, therefore, will be lower than if the borrower were not generally known and trusted and the lender were unable to get cash for it before the repayment of the loan is due, except at a great loss. The first loan is said to be more *liquid* than the second because the note can more easily be turned into cash which is the most liquid of all assets. If the borrower is so well known and trusted that his note is generally acceptable, he will be able to borrow money at extremely low rates of interest because his notes are very much like money in their general acceptability. Indeed, if the borrower is a bank or the government the notes may actually *be* money and may yield no interest at all to the holder. In all these cases the interest charged is less than the true or pure rate of interest. Part of the reward to the lender lies in the liquidity yielded by the note paid in recognition of the loan. The true rate of interest is what is

paid for a loan that has no liquidity, the lender not being able to sell the note to anyone or to use it to increase his own borrowing power. He can only lock it up and wait until the interest and repayment are due.

Third and finally, the rate of interest may differ for loans of different duration even when the risk and liquidity are absent or are present in the same degree. This is because the rate of interest is different at different times (except in a stationary economy), and the rate charged for a loan is an average of the rates for the constituent days that make up the period. The interest paid is the sum of the different interest payments for the different days.

Risk often seems less serious and illiquidity may really be less serious when incurred for a short period only, and this tends to make the rate of interest less as a rule for short-period than for long-period loans, but we shall neglect this complication.

Competition equalizes the *sum* of the money and liquidity yields from holding different assets. This permits the theory of the rate of interest to be generalized to explain all kinds of differential interest rates.

The second complication, namely, that arising from the different degrees of liquidity yielded by different assets (such as IOU's or bills of exchange), raises difficulties with the simplified account of the determination of the rate of interest given in Chapter 22. There is not merely one kind of earning asset and one kind of money as this simple account assumes. There is a whole series of assets, including money, which partake of the nature of earning assets and of the nature of money in varying degrees. They partake of the nature of earning assets to the extent that they earn interest and they partake of the nature of money to the extent that their possession yields a *liquidity* return in the convenience and security of knowing that one is able to cash them in for money without serious loss if it should be necessary. Competition tends to equalize the *sum* of these cash and liquidity yields from the possession of all assets including money. The yield on money proper is entirely in the form of liquidity convenience. There is no liquidity convenience in a pure earning asset, and all the yield is in the form

of money earnings. These are the two kinds of assets compared in Chapter 22. To generalize the theory of the determination of the rate of interest we merely have to change "earning asset" and "money" into "asset *A*" and "asset *B*," respectively, where asset *B* has a smaller money yield but a larger liquidity yield than asset *A*. There is a certain difference in their money yield or a *differential rate of interest* which will induce the public to hold asset *A* and asset *B* in the proportions in which they exist in the economy. The relationship between the interest rates yielded by any pair of assets is determined in this way, and the particular rate of interest yielded by any particular kind of security that is created for any particular kind of loan is merely the differential rate of interest between this asset and the asset "money." The rates of interest that are paid for partial sacrifice of liquidity are determined in the same way as the rate of interest we discussed in Chapter 22 which is paid for the total sacrifice of liquidity.

The creation of liquidity should be reserved to the monetary authority, for it permits other creators of liquidity to subsidize less productive investments and thus to depart from the optimum use of resources.

In a capitalist economy large corporations and the government are often able to borrow money at very low interest rates because the securities they offer are easily negotiable and partake largely of the nature of money. With money borrowed in this way they can acquire productive resources, withdrawing them from the process of current consumption and investing them. This ties up the resources just as effectively as if they were invested by any small corporation or unincorporated business that has to pay the full rate of interest corresponding to the relative excess valuation of present over future consumption. This leads to a maldistribution of resources, not only because the same factor has different prices for different producers, but because the price does not correspond to *msc* which is measured by the value of the alternative *msb* of the current consumption sacrificed.

Suppose the rate of interest on perfectly safe but non-liquid securities is 5 per cent, but the large corporation, because of the

liquidity yield of its securities, can borrow at 3 per cent. Consumers value $100's worth of present consumption as highly as $105's worth next year. But the corporation, which can borrow at 3 per cent, will purchase factors of production and devote them to producing for the future up to the point where the sacrifice of $100 now will yield no more than $103 next year. Consumers do not value this as highly as the present consumption that must be sacrificed (if there is full employment), so *msb* is less than *msc* and resources are being wrongfully used.

In a collectivist economy large corporations are not permitted to subsidize their investments by the creation of liquidity. That function is reserved to the monetary authorities in their determination of the general rate of interest. The same full rate is charged to all managers, in accordance with the Rule, and this permits the optimum use of resources.

In a controlled economy steps must be taken, therefore, to equate the rate of interest to all, keeping the creation of liquidity entirely for the treasury and banking system that is regulated in the social interest. This seems to be one of the objects of the Securities and Exchange Commission.

The differences in the rate of interest because of differences in the time to which the loans refer are dealt with much more simply. Every investment must make its *mei* equal the rate of interest applicable to the period between the application of the resources used in the investment and the emergence of the additional future product. This is implied in the Rule and is naturally followed by the private investor if all the other conditions for the optimum use of resources are present.

CHAPTER 26. FOREIGN TRADE I

The same Rule applies to foreign as to domestic trade.

In a collectivist economy the same Rule can be applied to foreign as to domestic trade. The managers of importing agencies will import goods from abroad up to the point where *pf* equals *vmp*. For this purpose the foreign currency that pays for the goods acquired abroad will count as a factor used in the production of these goods so that importation will be extended to the point where a domestic dollar's worth of foreign currency can purchase an amount of foreign goods that can be sold at home for a dollar (plus the marginal cost of transport and other costs of importing). Similarly the managers of exporting agencies will export goods of all kinds until *pf* equals *vmp*. This means up to the point where $1's worth of exports can be sold for as much foreign currency as can be exchanged for $1 (plus the marginal costs of transport and other costs of exporting). Producers who use imported materials or whose products are exported will either deal with importers or exporters or conduct the importing or exporting themselves. In the former case the optimum position is reached if the importers and exporters follow the Rule. In the latter case the same results are brought about if the producers translate their costs or revenues from the foreign currencies to domestic dollars at the current rate of exchange and proceed in the ordinary way to apply every factor, domestic or imported, until its price is no greater than the value of its marginal product as prescribed by the Rule.

As a part of this mechanism the price of foreign currency, like any other price, would have to be adjusted to make the supply equal to the demand. The supply of foreign currency is provided by the exporters who receive it for the goods they sell abroad and

have to exchange it for domestic money. The demand for foreign currency is provided by the importers who need it to pay the foreigners for the imports.

If the price of foreign currency is raised (which means that the foreign value of the dollar is decreased), imports are more expensive in dollars and are discouraged while exports can get a better price in dollars and are encouraged. This diminishes the amount of foreign currency demanded by the importers (who import less) and, if the exports increase in a greater proportion than their price to the foreigner is reduced by the fall in the value of the dollar, the supply of foreign currency will increase (to pay for the greater *value* of the increased exports even when measured in terms of the appreciated foreign currency). The supply of foreign currency thus becoming greater than the demand its price would tend to fall again. If the price of foreign currency were to be reduced, there would in the same way come about an excess of demand over supply which would tend to make the price rise again. Because of these effects the price of foreign currency tends to stay around an equilibrium position where its supply is equal to the demand.

Foreign currency can be used to represent *msc* and *msb*.

Because of this adjustment each $1 spent in acquiring foreign currency represents $1's worth of domestic goods that is sent abroad in creating the corresponding supply of foreign currency. This is why it is permissible to count $1's worth of foreign currency as $1's worth of domestic factor in the application of the Rule by the importers. The $1's worth of goods exported is the *msc* that this *pf* represents. Similarly the foreign currency obtained from exporting goods can count as *vmp* because it represents that value of imports which provides the demand for the foreign currency and which constitutes the *msb*.

The Rule unites the whole world in one system for the best use of resources.

If all countries follow the same procedure, we will get the optimum utilization of resources over all the countries taken

together. They form one system that follows the Rule and brings about the optimum use of resources. This implies that there is perfect competition throughout the different countries and completely free trade between them. Indeed if this were so there would be no point in distinguishing between domestic and foreign trade.

There would still be problems of interregional trade which arise because there are resistances to the movement of factors or products from place to place even in the absence of all artificial restrictions. Land, in the ordinary sense of the word, cannot be moved from one place to another. Labor often does not want to move or cannot afford to pay the costs of moving from one region to another even if there are no legal restrictions on migration. For many factors of production, materials, and final products, moving from one region to another is very expensive. All these cause the price of the same thing to be different in different places. The problem can be formally dismissed by calling the same physical good a different good when it is in a different place (which is after all not only an economic but a physical attribute) but fundamentally the problem remains, even if in changed form, of explaining the relationship between these prices. For the present purpose it is more convenient to keep closer to common usage and consider why and to what extent the same goods have different prices in different places.

One factor or product may move in substitution for the movement of others.

If all goods were freely transportable without cost, there could be no regional price differences. Any such difference would quickly be eliminated by the costless movement of goods from points where they are cheaper to others where they are dearer in accordance with the Rule or as a result of the activity of speculators under perfect competition. But this condition is more than enough. It is not necessary for *all* goods to be able to move costlessly. Many goods would not need to move to achieve this equality. Consequently it would not matter if these could not be moved

costlessly or even if they could not be moved at all. All that is necessary is that the various factors of production should be brought together so that they are able to produce, and that consumption goods and consumers should be brought together so that consumption can take place. If some of the factors cannot move, this is of no consequence provided the cooperating factors can be moved to these factors. Similarly if either the consumer goods or the consumer can move all is well. It does not matter that the mountain will not go to Mahomet as long as Mahomet is able to go to the mountain.

From this it follows that there would be no interregional price differences if only some factors and some products (or perhaps some consumers) were freely and costlessly transportable providing that these were sufficient to permit all the factors to be brought together that have to be in one place to carry on production, and to bring together the consumption goods and the consumers. If two factors have to be used together and they are not in the same place to begin with, the movement of one of them renders unnecessary the movement of the other and is a complete substitute for it. Similarly the movement of a product is a substitute for the movement of the factors to the place that the product is needed, and conversely the movement of all the factors is a substitute for the movement of the product. We have already seen that the movement of consumption goods and the movement of the consumers are substitutes for one another so that there is no need for both to be costlessly transportable to eliminate all interregional price differences.

This reduces interregional price differences to the cost of the cheapest substitute movement.

In the actual world most factors and products are transportable, but very few, if any, are *costlessly* transportable. Consequently interregional price differences are unavoidable. They cannot be greater than the direct transportation costs, but are often much less than this because other and cheaper transportation can be substituted for their direct transportation from points where they

are cheaper to others where they are dearer. If a product can be transported cheaply, it will be produced where the factors are cheaper and sent where they are dearer, increasing the demand for the factors where they are cheaper and decreasing it where they are dearer, until the difference in the *cost* of the product is no greater than the cost of transporting it. This means that the difference in the price of the *factors* in the two places will be no greater than the cost of transporting their marginal product, even though the direct transportation of the factor may be much more expensive or even impossible. The transportation of the product tends to equalize the prices of the factors in the same way as a direct transportation of the factors would. This is what is meant by saying that the transportation of the product is a substitute for the transportation of the factors.

In the same way transportation of factors, if that is cheaper, can be substituted for transportation of products and reduce the price differences among products to no more than the cost of transporting the factors. The transportation of one factor will reduce the interregional price difference of any other factors for which it is a substitute and even the consumer will be transported and reduce interregional price differences of products that are not transportable themselves or are transportable only at very great cost. This, indeed, is the essential economic aspect of travel and migration. In general those items which are most easily and cheaply transported move while those which are more difficult to move stay where they are and the interregional price differences are limited to the cost of the cheapest substitute movements.

A higher *mei* in one region than another must be accompanied by correspondingly higher rate of interest, or a relatively falling price level (or a combination of these).

Exactly the same principles are applicable to the movement of *capital* between different regions which is usually called *foreign investment.* If the marginal efficiency of investment is greater in region *B* than in region *A*, the rate of interest would have to be

correspondingly higher in *B* if the relationship between the price levels in the two regions is to stay constant. If the rate of interest is the same in both regions, say equal to the *mei* in *A*, there could be no equilibrium unless prices in *B* were falling at a rate per unit of time equal to the excess of the *mei* in *B* over the rate of interest, as we saw in Chapter 20.

These alternative possibilities reflect the same ultimate price ratios between the goods at different times and in the different regions, but we shall assume this last, most convenient, possibility to be maintained by the monetary authorities who keep the rate of interest higher in *B* than in *A*, equal in each case to the local *mei*, and so remove the need for prices in *B* to fall as compared to prices in *A*. All the real consequences we are about to derive from this assumption would follow just as well as from the other case or from a compromise when the higher *mei* in *B* shows itself to some degree in both of these symptoms instead of showing itself *only* in a higher rate of interest or *only* in a price level's falling relative to the price level of *A*.

In a single collectivist economy, operating according to the Rule, the higher *mei* in region *B* would result in all investment being directed to region *B*, no investment being undertaken in *A* until the *mei* had been equalized for the two regions. This concentration of investment will be recognized as formally identical with the concentration of investment in a particular industry which showed greater investment opportunities as discussed in Chapter 20 in terms of the marginal yield from the postponement of output in the steel industry as compared with the hat industry (p. 247). Because of this concentration of investment where the *mei* is greater and the resultant equalization of *mei* in all regions as in all industries, the differences in the rate of interest, or the changes in relative local price levels, are not necessary.

The same concentration of all investment in *B* would take place in a capitalist economy. The higher *mei* in *B* would have the effect of attracting investment from *A* whether it was the rate of interest or the price level which was kept level in both countries or whether some compromise between these was adopted. If the

rate of interest were kept the same in both regions, the falling price level in *B*, which is necessary for equilibrium where the rate of interest is less than *mei*, would permit an equilibrium in trade between the two regions only if it were accompanied by continual appreciation of *B*'s currency (in relation to *A*'s) equal to the rate of fall of prices in *B* (compared with prices in *A*). This is necessary to prevent an increasing flow of goods from *B* to *A* with a diminishing flow of goods in the opposite direction. (Indeed such an unequal flow of goods would mean an increasing demand for *B*'s currency to pay for the increasing flow of goods from *B* to *A* while there is a diminishing demand for *A*'s currency to pay for exports from *A* to *B*. This would automatically tend to raise the value of *B*'s currency in terms of *A*'s.) The continual appreciation of *B*'s currency in terms of *A*'s will attract investment from *A* to *B* to just the same degree as would the higher rate of interest. Investors would earn the same money yield in *B* as in *A*, but the appreciation of their investment by the time they reconvert it from *B*'s currency into *A*'s would give them the same greater net yield as when the price levels of the two countries are kept even by a higher rate of interest in *B* to correspond with the higher *mei*. Whichever policy is followed all investment would be concentrated in *B* where the *mei* is greater.

The concentration of investment in *B* could be stopped only by preventing the appreciation of *B*'s currency in relation to *A*'s, or, if the price level is stabilized in *B*, by forcing *B*'s currency to *depreciate* in terms of *A*'s at a rate equal to the excess of the *mei* in *B* over that in *A*. Such a depreciation would mean that capitalists in *A* who invest in *B* would lose as much from the depreciation of their capital as they would gain from the higher *mei*. Thus the shift of investment from *A* to *B* would be checked. As we have just seen it would be difficult to keep *B*'s currency from appreciating (or to force it to depreciate on the other plan) because of the excess this would create of goods flowing from *B* to *A* over those flowing in the opposite direction. The concentration of investment in *B* is as unavoidable (without direct prohibition) as it is desirable in order to take advantage of the greater yield from investing where the *mei* is greater.

The concentration of investment so as to equalize *mei* leads to international indebtedness.

When *A* and *B* are different countries with different currencies and separate accounts, the concentration of investment in *B* will mean that country *B* becomes indebted to country *A*. Taking the economic system as a whole (including both countries), total investment must equal total saving because total investment is the income other than that created by consumption expenditure and total saving is the excess of total income over total consumption expenditure. But now in country *B* investment will be greater than saving while in country *A* saving will be greater than investment by the same amount.

The excess of investment over saving in country *B* is made possible by an excess of imports over exports (in value). This excess consists partly of goods that are directly used in the investment and partly of goods that are used to maintain consumption in *B*, taking the place of local resources that are devoted to the investment and permitting the investment to take place without calling upon any sacrifice in further saving by the population of *B*. The excess of imports into *B* over its exports would mean an excess of the supply of *B*'s currency (to pay for the imports) over the demand for *B*'s currency (by the people of *A* to pay for *B*'s exports). This is corrected by the additional demand for *B*'s currency by the government or the people of *A* who need this money to acquire the claims over the investments in country *B*. The claims thus acquired are equal to the excess savings in *A* over the investment in *A* and so are just sufficient to equalize the supply and demand for *B*'s currency (and therefore also the demand and supply of *A*'s currency) since that in turn is equal to the excess of *B*'s investment over *B*'s saving.

Another way of expressing this is to say that *B* exports *claims* on its new investments equal to the excess of its imports over the other exports and so total exports *including claims* are equal to imports. Country *A imports* these claims. *A's* total imports, including the claims, are equal in value to its exports. In real terms (that is, not counting the claims) exports are greater than imports by

the amount of the net importation of claims. If we count the purchase of claims abroad as investment by *A*, and the selling of claims abroad as disinvestment by *B*, then saving is again equal to investment in each country.

The Rule provides an objective principle for collaboration of different nations as of the individuals within the state.

As long as *mei* is completely equalized in both countries by means of a capital movement as just described, no specific foreign trade problems arise. If both countries apply the same Rule, everything works out as if they were parts of the same country, both in the matter of the shift of real goods from one to another and in the matter of capital movements or the importing or exporting of *claims*. The Rule works *outward* beyond the area of the state as well as inward from the state to the individuals who comprise it. Just as it integrates the economic activities of the members of society to maximize their total welfare as consumers, so it integrates the economic activities of different states for their common good. It provides an objective principle for the collaboration of independent states, all gaining from the collaboration and none exploiting or being exploited by any of the others. It is this result of the universal application of the Rule that lies behind the internationalistic ideal of the classical free traders. Their tragic failure is traceable in no small measure to their insufficient separation of the optimum use of resources that is provided by the Rule from the other concomitants of *laissez faire* and their insufficient recognition of the tendency of perfect competition (which would bring about all the beneficial effects of the Rule) to destroy itself. It is not too much to say that any permanent peace among nations, which must avoid the exploitation of any nation by other nations, will have to be based on a general agreement to abide by a Rule of the same nature as that on which our whole analysis is built. Without such an objective Rule nations will not be able to resist the temptation to use their influence over the prices of imports and exports in attempts to gain from "beggar my neighbor" policies at each other's expense.

Specific international trade problems arise only from artificial barriers to the movement of goods, of money capital, and of people.

This working outward of the Rule to integrate all the nations of the world in harmonious cooperation presupposes the complete absence of all artificial restrictions on movement of all kinds. There must be no tariffs or quotas, whether open or disguised as health measures and the like, and no local differentiation in the application of the Rule which would have the same effects. Interregional price differences will then be the result of nothing but the natural and technical resistances to the movement of factors, products, and consumers. There must be no restriction on foreign investment so that *mei* and the rate of interest are equalized everywhere since the cost of movement of the claims (which are nothing but the receipts for the foreign investments) is negligible. Varying risks of investments in different parts of the earth's surface must be allowed for before arriving at a pure *mei*, which must be equated to the rate of interest, just as risks are eliminated in obtaining the pure domestic rate of interest.[1] And finally, and most important of all from the point of view of international justice and stability, there must be complete freedom of migration with exceptions only where there is a genuine danger of overpopulation of the whole world from areas where emigration, by removing the pressure of population, only permits more of the native population, to survive so that they remain almost as crowded as if emigration were not permitted.

These freedoms of movement of goods, of money capital, and of people do not exist, and it is only in connection with the artificially imposed barriers to the movement of goods, people, and capital that the specific problems of international trade arise.

[1] There may be considerable costs of movement in transporting the goods that constitute the foreign investment that has come to be capital movement, but this does not bring about any difference between the *mei* and the rate of interest. The yield of investment is based on the cost of the factors at the place of investment, when any costs of transporting goods there have already been included in the calculation of the *mei*.

CHAPTER 27. FOREIGN TRADE II

There are rational as well as irrational temptations for a nation to discriminate between domestic and foreign goods, capital, and people.

There are three ways in which a collectivized national economy might rationally be tempted to depart from the universal application of the Rule and instead to practice discrimination between domestic and foreign goods, investments, and people, thus transgressing the collectivist counterparts to these three freedoms of movement.

An appropriate restriction of imports *and exports* (below what would be indicated by the Rule) can benefit a country (while imposing a loss on the foreigner greater than this gain).

First, there is the possibility of increasing the wealth of the country at the expense of other countries by aggressively carrying on trade on monopolistic principles, directed toward maximizing the gains of the *country* instead of following the Rule which would maximize the total benefit of all the countries together. Such a deviation in the national interest will appear possible whenever the foreign demand for any of the country's exports or the foreign supply of any of the country's imports is less than infinitely elastic. If a country is of any size, this condition will be satisfied for almost every one of the goods exported or imported.

The application of the Rule to foreign trade without discriminating between domestic and foreign goods is based on the same principle of taking the prices of imports and exports as given, and disregarding any possibility of changing their prices by varying the quantity bought or sold. This is an integral part of the Rule.

We have seen how in a capitalist economy the absence of perfect competition, which means the presence of power in the hands of the monopolist to affect the price at which he buys or sells, leads to a departure from the optimum utilization of resources. The economy as a whole is made worse off by this while the monopolist gains a part of what the rest of society loses. Now the country is in the situation of the monopolist, and it can gain part of what society as a whole loses from the disruption of the international application of the Rule.

If the country chooses to disregard entirely the effects on the rest of the world economy, it imports and exports less than the Rule would direct. Instead of importing goods up to the point where the price of the import (including the cost of transport) equals the domestic price, it will import only up to the point where the marginal cost to the country of importing another unit—the increase in the total amount spent on the importation of the goods—equals the domestic price. In this it behaves like any monopolist who does not face perfect competition in buying. If the elasticity of supply of a good is infinite, so that the price does not rise at all when more of it is bought, the marginal cost to the nation is equal to the price per unit and it does not pay the country to depart from the Rule. Or one could say that the transformation of the Rule, putting the marginal cost to the country in the place of the price of the goods, does not make any difference because these happen to have the same value. The country is in the same situation as a firm under perfect competition which is induced, while seeking to maximize its own profit, to cooperate in bringing about the optimum utilization of resources. But if the country's purchases have some influence on the price, so that when it buys more the price, which is the *average cost*, rises, then the *marginal cost* is greater than the price, or average cost, and there is a divergence from the optimum utilization of resources.

How much the marginal cost to the country is greater than the price will depend on how much is being bought and on how much the price rises in response to an increase in the amount bought. If the price does not change at all (that is, if the elasticity of supply is infinite), the marginal cost to the country is no greater than the

price. If the elasticity of supply is equal to unity, so that the price rises in the same proportion as the quantity increases, the marginal cost will be twice as great as the price. In general the relationship between the price and the marginal cost can be seen in the simple formula $mc = p + p/e_s$ where mc stands for marginal cost to the country, p stands for the price, and e_s stands for the elasticity of supply of the good. The matter is illustrated in the following example:

If an increase from 100 to 101 units in the amount bought will raise the price in the same proportion (say from $1.00 to $1.01), the marginal cost of the 101st unit would equal its price of $1.01 *plus* an additional cent for every one of the 100 units previously bought and whose price will now be $1.01 instead of $1.00. The marginal cost thus comes to $2.01. This is slightly more than twice the old price of $1.00 and slightly less than twice the new price of $1.01. If the amount bought increased in a smaller proportion from 1000 to 1001 and the price rose in the same (smaller) proportion from $1.00 to $1.001, the marginal cost would be $2.001 (two dollars and one mill) which is still slightly more than twice the old price and slightly less than twice the new price, but closer in each case to twice the price. By making the change as small as we please we can make the difference as small as we like, so we may neglect it and simply say that if the price rises in the same proportion as the quantity bought, the marginal cost will be twice the price or 100 per cent above the price.

If the elasticity of supply is greater than unity, the price rise will be proportionately smaller and the excess of the marginal cost over the price, which is nothing but the original quantity bought multiplied by the price rise, will also be proportionately smaller. If the elasticity of supply is twice unity, the price rise and the excess of the marginal cost over price will be only half as great as when the elasticity of supply equals unity. If the elasticity of supply is one-half, the price rise and the excess of the marginal cost over the price rise will be twice as great. This gives us the general formula which tells us that the excess of marginal cost over price is equal to the price itself divided by the elasticity of supply. $mc - p = p/e_s$.

This formula measures the degree to which power over the buying price brings about a deviation from the Rule which would maximize the general good. Instead of buying goods up to the point where $vmp = pf$, purchase is stopped at the point where $vmp = pf - pf/e_s$; pf/e_s or p/e_s is the measure of the deviation from the Rule.

This is just as applicable to purchases by a firm as to imports by a country. In the case of the firm, $vmp = msb$ and $pf = msc$ so that pf/e_s, which is the difference between vmp and pf, measures the excess of msb over msc, but this is of no interest to the firm since, although it means a divergence from the optimum use of resources in society as a whole, the firm does maximize its own profits. In the case of the country which limits its purchases of imports to the point where the domestic price (the vmp) exceeds p, the price paid for the imported good (including transport costs), by p/e_s, this excess of domestic price over the foreign price measures the excess of msb over msc from the point of view of international society, but it is not of direct interest to the country since in spite of this its own benefit is maximized at the point where the domestic price of the imported good is equal to the marginal cost to the country of importing. The marginal cost to the country is equal to the marginal benefit to the country, and the benefit *to the country* from its international trade is maximized.

A restriction of the same nature also applies to exports, which constitute the *sales* of a country. The Rule instructs that any output should be carried to the point where p, the price of the product, is equal to vmf, the value of the marginal quantity of factors needed to produce a unit of product. But where a seller can influence the selling price, so that the price rises when he sells less and falls when he sells more, the marginal revenue is less than the price of the product, and it is the marginal revenue that is of interest to the firm or the country that is thinking only of itself and is not concerned with following the Rule and does not care if the maximization of its own gain involves a diminution in the welfare of all the firms or of all the countries taken together.

In this problem, which involves contraction of the product, it is more convenient to use the second set of terms, given in Chapter

9, which we have usually rejected. In these terms we note that the Rule calls for equating p, the price of the product, to vmf, the value of the marginal quantity of factors. The ability to influence p, which is the same as the average revenue, so that p falls as sales are increased, means that the *marginal revenue*, which is what the self-seeking firm or country is really interested in, is less than price or average revenue. This follows from the arithmetic of the average-marginal relationship (see p. 82). The rest of the analysis follows exactly the same lines as that just worked out for the ability to influence the buying price.

By how much marginal revenue is less than the price will depend upon how much the price falls in response to an increase in sales. If the price does not change at all (that is, if the elasticity of demand is infinite), the marginal revenue is not less than price at all. If the elasticity of demand for the product is equal to unity, so that price falls in the same proportion as sales are increased, the marginal revenue will be just 100 per cent less than price (that is, it will be equal to zero). This is illustrated in the following example.

If an increase in the amount sold from 100 to 101 lowered the price in the same proportion (say from $1.01 to $1.00) the marginal revenue for the 101st unit would equal the price ($1.00) *minus* the loss of 1 cent on each of the previous 100 units that must now be sold at $1.00 each instead of at $1.01. The marginal revenue thus comes to zero, or 100 per cent less than the price.

If the elasticity of demand is greater than unity, the price will fall proportionately less, and the deficiency of marginal revenue below price, which is measured by the original quantity sold multiplied by the fall in the price, will also be proportionately less. If the elasticity of demand is two (twice unity), the fall in price, and the deficiency of marginal revenue below price, will be only half as great as when the elasticity of demand is unity. If the elasticity of demand is one-half, the price fall and the deficiency of marginal revenue below price will be twice as great as when the elasticity of demand is unity, and the marginal revenue will be a *negative* quantity as great as the price. In our numerical

example it would mean the price fell in a proportion twice as great as the 1 per cent increase in sales, say from $1.02 to $1.00. The increase in sales would *reduce* the total revenue by $1.00 from $102 (from the sale of 100 units at $1.02) to $101 (from the sale of 101 units at $1.00 each) giving a marginal revenue of *minus* $1.00. Alternatively this can be expressed by saying that the marginal revenue is equal to the price of the additional unit ($1.00) *minus* the loss involved in having to sell the previous 100 units at $1.00 each instead of at $1.02, which comes to a loss of $2.00. Subtracting this loss from $1.00 again gives us *minus* $1.00 as the marginal revenue.

In general we can say that the deficiency of marginal revenue below price is equal to the price itself *divided* by the elasticity of demand. If *mr* stands for marginal revenue, *p* stands for price at which the goods are sold to foreigners, and e_d stands for the elasticity of demand, we have the formula $mr = p - p/e_d$ or $p - mr = p/e_d$.

This formula measures the degree to which the power over selling price results in a deviation from the Rule which would maximize the general good. Instead of selling up to the point where $p = vmf$ (when $vmp = pf$), sales are stopped at the point where $mr = vmf$, or $vmf = p - p/e_d$. The deviation from the Rule is measured by p/e_d.

This is just as applicable to a firm selling to other firms or agencies as to a country selling exports to other countries. The firm's $p = msb$, and $vmf = msc$, so the p/e_d which is equal to $p - vmf$, means the excess of *msb* over *msc*, but this is of no interest to the firm, since in spite of the social loss its own profits are maximized by this procedure. In the case of the country which limits its exports to the point where the foreign price exceeds the domestic price by p/e_d this quantity measures the excess of *msb* over *msc* from the point of view of international society, but this is not of direct interest to the country since, in spite of this, its own benefit from international trade is maximized. A dollar's worth of exports (at the domestic price which is less than the price at which the goods are being sold abroad by p/e_d yields the country exactly one extra dollar (since that is the marginal revenue from ex-

ports from the point of view of the exporting country) with which it can get an extra dollar's worth of imports. The marginal benefit to the country is equal to the marginal cost, so its gain is maximized.

We see then that a collectivist country that wished to maximize its own gain and was not concerned with the welfare of other countries would depart from the Rule by restricting its imports and its exports by varying degrees, depending on the elasticity of supply of its various imports and the elasticity of demand for its various exports. Instead of carrying on trade until domestic and foreign prices were equalized (apart from transport costs) it would leave the domestic price of imports *above* the foreign price by p/e_s and the domestic price of exports *below* the foreign price by p/e_d, p in each case standing for the foreign price and e_s and e_d for the foreign elasticities of supply and demand respectively.

This invites retaliation which would make everybody lose.

By such a policy the country maximizes its own benefit but must cause the other countries to lose more than it gains, since the *total* product of all the countries together is diminished by the deviation from the optimum utilization of resources that is obtainable from the general application of the unamended Rule. A self-interested country is not likely to be influenced by this consideration, but it may have to take into account the possibility of *retaliation*. Our analysis of the possible gain to a country from these restrictive practices, in speaking of the elasticity of demand for the country's exports and of the elasticity of supply of the country's imports, implicitly assumed that no other country's policy was influenced by what this country did. However, it is possible for more than one country to play the same game, and if this is done by all the countries, the particular benefits cancel out while the general loss through the restriction of international trade remains. If all countries persist in trying to *exploit* each other in this way, international trade would be completely destroyed with all the benefits that could have come from it. All parties will stand to gain by a restoration of trade and then there will be endless haggling as to the terms on which it should be carried on. The only

objective solution to the quarrel is that provided by a general adherence to the Rule which would maximize the total benefit. Any other solution is inferior in the objective sense that any country which made an exceptionally good bargain could be adequately compensated for what it would lose by the general adoption of the Rule. This could be brought about by payments in goods or money by the other countries, which would still be better off under the Rule. The Rule maximizes the cake that is to be shared among the countries, and if there is a bully prepared to destroy some of it in the course of grabbing a bigger share for himself, the others would still be better off if they gave him all that he could get by grabbing and avoided the further loss to themselves from the waste involved in the departure from an orderly distribution of the benefit according to the Rule.

Fear of retaliation might prevent a country from applying the monopolistic restrictions on its trade, but it would not be persuaded to refrain from them if the other countries were attempting this exploitation in any case. That would lead to competitive restriction, the destruction of international trade, and perhaps an attempt to come to terms again. A country might feel sufficiently friendly to others to refrain from such aggression or to apply it only to a limited degree. It might even wish to reverse the procedure for the benefit of a good neighbor, buying from it *more* than the Rule would indicate and selling to it *more* than the Rule prescribes. (We must here keep away from the capitalistic feeling that to sell goods cheap to a country is to do it an injury.) This would never be the most economical way of helping a neighbor. It would be better for both countries if trade were limited to the amount prescribed by the Rule and if any help was given by a direct money gift that could be spent by the country that receives it either on increasing its imports or on decreasing its exports according to its own preferences.

The second temptation for a collectivist society to depart from the three freedoms that are associated with the universal application of the Rule is in connection with limitation on the immigration of foreigners. There are two different ways in which this temptation arises.

First there is the situation where the immigration of foreigners is a way of getting round the discrimination exercised by the country against foreigners in restricting purchases from them and sales to them, as described in the last few pages. This is not likely to form a very strong consideration. It could be satisfied while still permitting free immigration if the government, acting in the interest of the inhabitants who had immigrated earlier, would continue the same restrictions on the latecomers after their immigration, buying goods from them only at the lower price at which they are bought from abroad and selling goods to them only at the higher price at which they are being sold to foreigners abroad. This would still be better for the immigrants than prohibiting or limiting their immigration. However it is unlikely that the material benefits to the older inhabitants from such an exploitation of newcomers could be sufficient to make the process worth while.

Earlier inhabitants, as a whole, cannot lose by the immigration of workers who are paid no more than their marginal product.

More important is the limit on immigration because of a disinclination to share the natural advantages of the country with all comers. This is primarily a selfish argument, but a restriction of this kind can also be in the general interest if the immigrants come from a country where the population tends to push its scale of living down to subsistence level by unchecked multiplication whenever the subsistence level is exceeded. To allow free immigration from such centers will not permit standards to be raised in these countries and would only permit other countries to be pulled down to their own level.

This argument would not hold if the immigrants got no share of the social dividend and their income were limited to the value of the marginal product of their work. They would then be taking nothing from the older inhabitants and by working in their adopted country, where the value of their marginal product is presumably greater than where they come from, they would be adding more to the total product. To the extent that by their large numbers they reduce the marginal product and the wage of com-

peting labor they would also be reducing their own marginal product and wage below what it was when they first began to immigrate, and in that case they would be getting less than they add altogether to the total product. The difference goes to the other factors of production, whose *vmp* is thereby raised, permitting a higher social dividend to be distributed and a higher scale of living to be enjoyed by the older inhabitants or citizens who qualify for the social dividend. Those citizens with whom the immigrants compete more directly may find their *vmp* reduced more than their share of the social dividend is raised, but they could be more than fully compensated and still leave all the other citizens better off.

The citizens of a country do not lose by permitting immigrants who are paid only the value of their marginal product, because such a movement of labor performs the same function as the movement of their products in international trade and is a substitute for them. Like the free movement of goods in the absence of discrimination in applying the Rule, it cannot impoverish the country even though it eliminates the monopolistic gains, which can be substantial if precarious, that were discussed in connection with possible deviations from the first of the three freedoms of movement, the free movement of goods. Restrictions on immigration are, however, very likely even in a rational and democratic collectivist economy because it is not pleasant, even if it is possible, to have a section of the population *within* the country discriminated against so severely. The immigrants, if their income is to be limited to the value of their marginal product, will not only have to go without the social dividend which is the right of all full citizens, but will have to pay in taxes for their enjoyment of social services that are provided to full citizens without tax as part of their social dividend. These taxes are likely to be too hard for the immigrants to bear, while not to give them the social services would entail a callousness that is not so obvious if they are in a faraway country, and is likely to entail a health hazard for the citizens themselves. On the other hand, to give in and permit them freely to enjoy a part of the social dividend, as by letting them have the social services without paying for them, might lead to

huge and endless immigration from the breeding centers which would reduce all to their own miserable scale of living.

This problem brings us to the third way in which a collectivist economy could rationally be tempted to depart from the three freedoms of movement. The first was concerned with the freedom of movement (or the absence of discrimination) for goods, the prices of which could be influenced by varying the amount bought or sold. The second was concerned with the movement of people. The third is concerned with foreign investment. (This is often called the movement of capital, but we have seen that if we look after the investment the capital looks after itself.)

Foreign lending might well be limited for the sake of preventing ill feeling unless some scheme is adopted for equalizing the wealth of nations.

The Rule calls for the concentration of investment where *mei* is greater so as to equalize it. This will often call for the country where the *mei* is greater falling into debt to the other countries. It will have to pay interest on the money borrowed to make the investments, and this has nearly always led to bad feelings on both sides. The borrowing country feels that it is being exploited when in the course of time it has to pay back more than it borrowed (because of the interest payments) and is inclined to default on its richer neighbor when it has the opportunity. The lender resents this attitude and so there is bad feeling. It may be better to sacrifice some of the potential product for the sake of having less borrowing and lending and less bad feeling. The precept "neither a lender nor a borrower be" may have its place in guiding the affairs of nations if it is not followed too strictly.

Underlying the feeling on the part of the borrower that he is being exploited when asked to pay his debt and the interest on it (apart from sharp practices by the lenders that are not infrequent) is the notion that the inhabitants of the lending country (who are usually much richer and have been able to invest sufficiently in their own countries to reduce their *mei* below that in the borrowing country) have more than their fair share of the good things of the

earth and that it would only be right if some investment were provided free to the poorer country by the richer countries and so help to even out the world's wealth. Of course such a gift could not be expected if it would only increase population without raising standards, but some international equalization of wealth would seem to be a prerequisite for a peaceful and satisfied world just as some equalization of income is necessary for a healthy economy in a capitalist society.[1]

Where there are several socialist (that is, democratic collectivist) societies, their common aims might permit a joint policy, not on strictly business lines, for equalizing the income per capita in the several countries by concentrating unpaid investment where the *mei* is greater, not only for the purpose of maximizing total yield, but as a substitute for population movements. The ideal use of resources still calls for the equalization of *mei*, and this would be worth while even on a strictly business basis, but friendly relations seem to call for more than this when otherwise there would be considerable inequality of the per capita income in the two countries. In the absence of some such arrangement it is probably better for foreign investment to be less than the Rule would indicate.

[1] See A. P. Lerner "Economic Liberalism in the Postwar World," Chapter VII in *"Postwar Economic Problems,"* Ed. Seymour E. Harris, McGraw-Hill, 1943.

CHAPTER 28. FOREIGN TRADE III
(*IN A CAPITALIST ECONOMY*)

In a capitalist economy there are many more reasons for expecting deviations from the optimum situation which would result from free trade (and which would be identical with that achieved by a universal adherence to the Rule). quite apart from the more general deviations that result from the absence of perfect competition throughout the economy.

Particular interests are able to harm the economy as a whole by insistence on protective devices in lieu of harmless but unpopular compensation.

First, there are a series of interferences with free trade because of sectional interests which stand to gain by the tariffs and quotas and other impediments to free trade and which consequently are deaf to the irrelevant argument that the economy as a whole must lose by the restrictions. To the same category belongs the resistance of organized labor to the immigration of cheap labor or the importation of the products of cheap labor. Such immigrant labor, getting only its marginal product, can only enrich the community as a whole, but labor in direct competition with it may lose. It is true that society could afford to compensate all who suffer from such competition from foreign dumping or cheap labor, but society shows no signs of doing this and so the sectional interests go ahead and usually succeed in getting some protection. Each protection is a burden on all the other parts of the economy, so that the net result is likely to be that *everybody* is worse off than if nobody were protected, but even if they were all aware of this each interest would want the other protections to be removed before its own was touched, and so we have the kind of impasse that is seen most dramatically in disarmament conferences.

Among the arguments used as weapons by the various interests seeking protection and endeavoring to persuade everybody that the protective measure is not only their just due but really good for everybody else, there are innumerable sophisms that economists have industriously, and in general fruitlessly, exposed for generations. However, a few of the arguments are sound.

Even "taxing the foreigner" is not carried out scientifically.

One of the sound arguments is that it is possible to tax the foreigner by a tariff that would reduce the price the foreigner received for his product and so the price to the domestic buyer would rise by less than the amount of the tax—the rest of the tax is paid by the foreigner. This will be recognized as indicating an elasticity of supply of the imported good that is less than infinite so that the price falls if less is bought. This situation permitted the collectivist economy to increase its income by restricting purchases. It is, however, unlikely, even in this case, that the tariff urged at the instigation of a sectional interest will be the one which would maximize the benefit to the country from the restriction of imports. In the capitalist society, moreover, there is seldom any call for the corresponding restrictions on exports, which could benefit the country in the same way at the expense of the foreigner, because there is usually no sectional interest that would gain by this restriction. How the optimum restriction of imports and exports could be attained in a controlled economy with free enterprise, if the gain from exploiting the foreigner should be considered worth striving for, will be examined in the next chapter.

The most serious foreign trade problems of the capitalist economy are connected with employment. We have examined the process by which full employment may be reached in a capitalist economy that is complete in itself—that is, with no foreign trade—if the amount of money is given and the rate of interest is permitted to adjust itself to it, equalizing the demand for money to hold with the amount of money available to be held. In examining this process we noted a number of points at which it is likely to be stalled. When we bring in the complications of foreign trade we

find there are still other difficulties in the way of the automatic movement to and maintenance of full employment in an uncontrolled capitalist economy.

The typical capitalistic monetary system is the gold standard. Instead of a fixed amount of money, which we assumed in Chapter 22, there is a fixed gold value of the monetary unit. This is maintained by the government either by actually issuing gold coins of a fixed weight or by buying and selling gold for the currency at a fixed price. The machinery for this is fundamentally the same as that we have described for counterspeculation. The government, by offering to buy or sell unlimited quantities of any good, can keep its price constant. Such buying and selling is the only effective way of pegging a price. The passage of a law that declares any other price illegal would merely be evaded by subterfuges of unending ingenuity. A country on the gold standard, by offering to buy or sell unlimited amounts of gold at a given price, pegs the price of gold in terms of the currency and therefore also the value of the currency in terms of gold.

The conveniences of the gold standard are insufficient to make up for the hindrances it imposes on a policy for full employment.

Once the gold standard is established, there are considerable conveniences in keeping it unchanged at the same price of gold as long as the other countries do so too. This results in a practically constant price of foreign exchange which is very useful to importers and exporters, and by providing a simple rule for the management of the currency it serves as a check on the hasty impetuousness that many people believe (contrary to all experience) to be the natural characteristic of governmental action. These conveniences tend to be exaggerated until the maintenance of the price of gold is identified with the notion of upholding the national honor and even in some mystical way with national solvency.

Fear of losing gold is met least objectionably by raising interest rates.

The danger of a country's being forced off the gold standard arises from the possibility that the demand for gold (which the

government guarantees to buy and sell in unlimited quantities at a fixed price) will exhaust the government's stocks and the government will no longer be able to sell gold at the guaranteed price. The price of gold will then rise in terms of the currency, so that the value of the currency falls in terms of gold, and the national honor is besmirched. Monetary policy is therefore primarily directed toward preventing such a situation. This is done by raising the rate of interest, through a restriction of the supply of money and of credit, whenever there are any signs of the possibility of a net outflow of gold from the coffers of the monetary authority. This would happen whenever the amount of gold *demanded* from the government monetary authority in exchange for money was greater than the amount of gold being *offered* to the government in exchange for money. Gold withdrawn from the reserves of the monetary authority is for the most part sent abroad to purchase foreign money from the governments of other countries on the gold standard to pay for a balance of imports whose value exceeds the value of the exports (the rest of the foreign money being obtained from the foreign purchasers of the exports). The import balance (the excess of the value of imports over the value of the exports), which is usually called an unfavorable balance of trade, is the cause of the outflow of gold. Because the supply of foreign currency available from the proceeds of the exports is not sufficient to satisfy the demand for foreign currency to pay for the imports, the price of foreign currency, which is demanded by the importers who have to get the foreign currency to pay their foreign suppliers, rises on the foreign exchange market. This goes on until it is cheaper, instead of buying foreign exchange on the market, to obtain it by the indirect method of buying gold from the government (at the guaranteed rate), exporting it, and purchasing the foreign currency from the foreign government with this gold at the fixed rate established by the foreign country's gold standard.

The restriction of credit and the resulting high rate of interest tend to correct the outward flow of gold in three ways which we shall consider in the order of the speediness of their action. First, the high rate of interest makes it profitable for idle funds to be

kept in this country rather than in the other countries where the interest they earn is lower. Foreigners and citizens who had foreign money loaned out at interest abroad call in these loans, exchange the currency for the currency of this country, and lend it out here at the higher rate of interest. This increases the supply of foreign currency offered for our own currency on the foreign exchange market, so that the need for foreign exchange by the importers can be satisfied, the price of the foreign currency falls back to normal, and there is no need for gold to be exported.

This is only a temporary alleviation, for there is a limit to the amount of idle money that can be moved freely from one country to another in this way. When it has all moved where the rate of interest is higher, the gold movement will be resumed if there is still an excess of imports over exports.

It should be noted that this will work only if there is complete confidence in the continued maintenance of the gold standard. If the rise in the rate of interest is interpreted by the capitalists who own the idle funds as an attempt to keep on the gold standard that might fail in its purpose, they will move their capital *out* of the country rather than into it. They will endeavor not to have their wealth in terms of our currency, which is in danger of depreciation in terms of the other currencies if we are eventually forced off the gold standard at the old gold parity, but to have their wealth in terms of other currencies, or in terms of gold, which would then appreciate in terms of our currency. The high interest that may be earned in the meantime will be relatively insignificant.

This may lead to severe unemployment, which gives rise to pressure for import duties.

The second way in which the high rate of interest checks the flow of gold out of the country is by its effect on the level of economic activity in the country. The high rate of interest reduces investment. This cuts down income and so consumption and income fall in accordance with the propensity to consume until the lower level is reached where the gap between income and equilibrium consumption is equal to the lower level of investment. With this

lower level of income, consumption of all kinds is reduced, including the consumption of imported goods. This tends to correct the import balance and to check the outflow of gold. If this should not be sufficient, and gold continues to flow outward, the monetary authorities will restrict the supply of loans still further until enough unemployment has been created to reduce imports to the level necessary to check the outward flow of gold. Clearly this is not a desirable solution.

The third way in which the high rate of interest checks the gold flow works *through* the second. The lower level of activity and the unemployment tend to lower the prices of factors of production and the cost of products of all kinds, including the cost of exports. This causes exports to increase, together with the level of economic activity in general including imports, until full employment is reached again. There will be a lower level of prices in this new position which induces foreigners to buy more of our exports, and so there is enough supply of foreign currency to permit even that quantity of imports to be bought and paid for which accompanies a fully employed economy, and there is no tendency for gold to flow out any more.

The difficulties in the way of this third and final solution are the same as those we have already noted in Chapter 22, so that the second phase, with considerable unemployment, is likely to hold the stage for a considerable and even indefinite period.

This strange situation in which the maintenance of full employment is sacrificed to the relatively insignificant consideration of stabilizing the value of gold or of foreign currencies (which is classically portrayed in the economic history of England from 1925 to 1931) gives rise to a powerful argument for a general tariff. This would discourage imports and eliminate the import balance. The danger to the gold standard would be removed without plunging the economy into a depression.

Export subsidies violate the spirit of the gold standard.

The only real objection to a general tariff for this purpose is that the depression is only temporary (it is hoped) while the tariff

is likely to become a permanent barrier to trade that would have to be raised again the next time there was an import balance. This objection can be met by adding a subsidy to exports at the same rate as, and to be paid out of, the tax on imports. The subsidy would encourage international trade just as much as the tariff would discourage it. By this means the gold flow can be stopped without harming international trade and so the gold standard can be maintained without sacrificing either employment or the benefits from the international specialization of production that international trade makes possible.

But this is only a pretended maintenance of the gold standard. The tax on imports and the subsidy to exports come to the same thing as a tax on the purchase of foreign currency (to pay for the imports) and a subsidy on the sale of foreign currency (obtained from the sale of the exports), and everything is just as if foreign currency had been permitted to appreciate in terms of the domestic currency (or the domestic currency permitted to depreciate in terms of the foreign currency and in terms of gold) by the amount of the tax or the subsidy. It would be better and far simpler to have an open depreciation of the currency rather than this disguised form of it which more than anything else would appear to foreigners as unfair competition that calls for retaliation.

Capital movements can cause the same difficulties.

The same problem arises if *mei* and the rate of interest should happen to rise abroad or fall at home. We have seen that this would result in a flow of lending abroad, which would make for an excess demand for foreign currency and an outward flow of gold. In the equilibrium situation this would be offset by an excess outward flow of export goods, on balance, which, by itself, would have led to an inflow of gold. Unfortunately this equilibrium does not come about any more automatically than the maintenance of full employment. Under the gold standard, the export balance would come about only after the outward flow of gold had called forth a restriction of credit and a rise in the rate of interest which had discouraged domestic investment and economic activity in

general (in addition to temporarily checking the amount of foreign lending), and the depression had sufficiently reduced wages and cost and prices to make exports cheap enough to induce foreigners to buy enough more of them to balance the effects of the foreign lending. (In this they would be helped by the decrease in imports because domestic goods would now be relatively cheaper to buy compared with imported goods.) As before, there is a considerable likelihood of stalling in the second phase with the equilibrium in the exchange market brought about by the decrease in imports that results from the poverty of the depression artificially induced by the raised rate of interest.

The same difficulties may arise *within* a country. One region may find a decrease in the demand for its product from the other regions, or *mei* may be higher in other regions. This leads to an import balance of goods or an export balance of lending, or both, and money is drained out of this region just as it would be drained out of one gold standard country to another under like circumstances. The high rate of interest discourages investment, so economic activity declines and there is unemployment until wages and costs have been sufficiently reduced to restore the demand by the other regions for the products of this region. Any depreciation of the local currency which would avoid the necessity for unemployment and local cost reduction is, of course, impossible. If this works out satisfactorily within a country, the gold standard advocate can argue, why can it not operate internationally? Does not our argument for flexible exchanges also logically commit us to a demand for currency autonomy for every region, district, and village?

The prerequisite for a single currency (or gold standard) area is effective freedom of movement of goods, people, and capital.

The answer is that the region within the country differs from an independent country in the important respect that it enjoys the free movement of goods, of people, and of investment into the other regions of the same country. An import balance that causes money to flow out of one region means an export balance in

another region where money is flowing in, and economic activity is at an unusually high level. The inhabitants of the first region can move without restriction into the prosperous region. This does not mean that there cannot be depressed areas within a country, but it does mean that these are not likely to be very serious if it is easy to get a job by moving into the other area, and in any case the depressed area will be recognized as a responsibility of the country as a whole. This is not true for separate countries. Tariffs and quotas interfere with trade, and the movement of people from country to country is seriously limited. It is only because of these restrictions that a rigid enforcement of fixed exchange rates between the values of the currencies of different countries can be so very harmful. If there were complete freedom of movement of goods, investment, and people, an international currency system would be as sound as a single monetary system for a country within which the three freedoms of movement are realities, and a properly managed gold standard system might be one way of arranging this.

Stable exchanges are a result of this—a symptom that cannot safely be established by decree.

The conveniences of having a single monetary system might then be sufficient to make it worth while to undertake the trouble of having to make minor adjustments in the levels of local wages and costs instead of merely adjusting the value of the local unit of money. But the *first* step in bringing about this desirable state of affairs is to remove the hindrances to movement. As a result of this freedom of movement of products, investment, and people, the prices of the factors and the *mei* are made equal in the different countries, prosperity is more equally enjoyed by all countries, and there will be less need for variations in the relative values of the different currencies to maintain full employment. When these changes are sufficiently small, even though the exchanges are free to move up or down, *then* will be the time to consider the conveniences and inconveniences of fixing a single monetary system for the whole world. It will not be a very impor-

tant matter, for substantial stability of the exchanges will have been reached before the issue arises. But to try to establish it before the underlying conditions of stability (that is, the freedoms of movement) have been established is certainly to attempt to establish world unity by decreeing one of its symptoms—like attempting to establish good health by commanding everybody to display a glowing complexion.

Where demands are inelastic the automatic mechanism works the wrong way.

There are other circumstances that render the automatic maintenance of full employment still more precarious. If unemployment succeeds in reducing wages, costs and prices, the economy passes from the second to the third phase. Exports are increased because they are cheaper and imports are cut down because with the lower incomes not so many of them can be afforded and cheaper domestic goods are substituted for them. But this does not necessarily mean a correction or even an improvement in the import balance, for that is concerned not with the amounts of physical goods but only with their *values*. At the lower price of exports foreigners will buy more of them, but unless the quantity of exports increases in a *greater proportion* than their prices fall, that is, unless the foreign elasticity of demand is greater than unity, the *value* of the exports will not increase. If the elasticity of demand is less than unity, the value of exports will *decrease* and that would make the import balance greater than ever. The situation might still be saved by the decrease in imports, which *must* involve a decrease in the value of imports. But if the demand for imports is very inelastic the decrease may not be sufficient to make up for the decrease in the value of the exports. If the elasticity of demand for imports is greater than zero by the degree to which the elasticity of demand for exports is below unity, the two influences will just offset each other and the fall in domestic prices will have no effect at all on the import balance; and if the elasticity of demand for imports is less than this the import balance will actually *increase* as a result of the fall in price so that gold will flow out even faster

than before and the crisis will get worse and worse. It should be noted that at the critical point; where the elasticity of demand for imports is just as much above zero as the elasticity of demand for exports is below unity, the sum of two elasticities is unity. If this sum is greater than unity, a fall in the price level tends to correct the import balance and to check the gold flow; if it is just equal to unity, the fall in prices has no effect on the import balance; and if it is less than unity, the fall in prices *increases* the import balance and renders the situation even more critical.

The critical point is where the *sum* of the elasticity of demand for imports *plus* the elasticity of demand for exports is equal to unity.

For example, if the elasticity of demand for exports is unity, the quantity of exports increases in the same proportion as the price falls (together with the domestic price level) so that the value of the exports will remain the same. If the elasticity of demand for exports is less than unity, say one third, the quantity bought will increase only one third as much as the price falls and then the total value of the exports will fall. Suppose the price of exports falls 3 per cent. This will result in an increase in exports of 1 per cent so that total value of exports will fall about 2 per cent. Now suppose the elasticity of demand for imports to be two thirds (so that the sum of the two elasticities is equal to one). Then the decrease in income and in domestic prices of 3 per cent is equivalent to a 3 per cent increase in the price of imports (for that is their *relative* increase) and will lead to a decrease in the amount bought, and in their value, of 2 per cent (two thirds of the change in their relative prices because the elasticity of demand for imports is two thirds). The values of imports and exports move together and the import balance is unchanged. If the sum of the two elasticities is less than unity, there will be a "perverse" movement of the import balance.[1]

[1] To complete this argument it must be added that, as prices fall in this country, there is a parallel tendency for increased activity and rising prices abroad where there is an export balance and an inflow of gold. This would mean that the 2 per cent decrease in the value of the imports and exports would not be absolute, but only relative to the new prices and incomes abroad, but they would still equal each other, and so

This state of affairs could not be set right by reducing the value of the currency. That would avoid all the harmful effects of the preliminary phase with a depression at work to reduce the prices, but if the sum of the elasticities is less than unity it would not correct the first situation any more than the fall in prices does. The import balance would increase in the same way and bring with it a further depreciation in the value of the domestic currency which would only make matters still worse.

To correct such a situation it is necessary to *raise* the value of the currency instead of lowering it. The automatic adjustment by lowering prices is not only unnecessarily painful, but works in the wrong direction. It is not only nasty, but it is no medicine, for it only aggravates the disease.

The tendency to assume that elasticities are high shows undue optimism.

These possibilities are usually ignored or else pointed out as queer but practically unimportant cases. This is due in part to the tendency in economic writings to assume elasticities to be high because that would fit in better with the ideas of perfect competition on which, until recently, economic analysis was based, and partly to the related habit of assuming a world with many small countries, none of them large enough to influence prices appreciably. If this were true, such low elasticities as those here discussed, where the sum of two of them is less than unity, could be relegated to the realm of *curiosa*. But there are now large economic empires with tremendous control over the price at which they will buy, especially in the short period; and more important still is the proliferation of tariffs and quotas and exchange clearing schemes that are used as "defenses" against a country which reduces its costs or the value of its currency, or which tries to reduce its import balance by subsidizing exports. These have the effect of reducing the elasticity of demand for its export. Indeed where the exports are fixed by quotas the demand is exactly zero. When such

there would still be no change in the import balance. It is likely that import and export prices would not move as much as domestic prices because they are influenced more by conditions abroad, but this consideration affects both imports and exports symmetrically and does not affect the essence of the argument.

a country has cut its imports to the very minimum in its efforts to correct its balance of payments, the demand for the remaining imports will be very inelastic too. We then have the phenomenon just analyzed, which is not at all uncommon in times of international depression. One country after another is forced off the gold standard, finds its currency depreciating without this affording much relief, and is able to find a stable position only after a long fall in the value of its currency and the introduction of special restrictions on trade have so altered the situation that the elasticities are no longer so low.

CHAPTER 29. FOREIGN TRADE IV
(*IN THE CONTROLLED ECONOMY*)

There is no need to give up the benefits from foreign trade for the sake of insulation from disturbances.

In the last chapter we came across some of the serious problems that are added to the cares of a controlled economy by foreign trade. Some writers have been so concerned with the possible ill effects on an economy of events outside it that would interfere with its rational management that they have even proposed a minimizing of foreign trade, sacrificing the benefits of international specialization in order to be free from these complications. There seems to be no need for any such sacrifice. The controlled economy can reap the benefits from trade at least as fully as any other, and this need not conflict at all with its other aims.

It is better to be overgenerous in compensating particular sacrifices in the general interest than to forego the general benefit.

The first principle is not to have any restriction of trade in the interest of any *section* of the economy. It is much better, if necessary, to make an open payment to such interests out of the general funds of the economy rather than by tariffs or quotas to permit this segment to reap a small benefit out of the much greater harm that the restriction does to the economy as a whole. Here, as in general, there is much greater scope than is usually recognized for *compensation* as a proper means of overcoming resistance by affected interests to measures that are needed in the general good. There is a strong prejudice against compensation because of the frequent doubt whether the benefits taken away by the action in the social interest were really deserved in the first place, but such

an issue should never be allowed to hinder the socially desirable action from being taken. After the full compensation has been paid, questions of desert are assimilated to the general problem of the distribution of income and wealth and can be dealt with accordingly.

There is a simple formula for the optimum tax on imports and on exports, in perfect and imperfect competition, if it is desired to exploit the foreigner.

If the controlled economy should wish to take advantage of its monopoly powers to exploit the countries with which it trades, it can do so by imposing appropriate import and export taxes. We saw in Chapter 27 that the marginal cost to the country of a unit of an imported good is $p + p/e_s$ where p is the price paid for the import and e_s is its elasticity of supply to the country. Firms in perfect competition and collectivist agencies following the Rule would import goods up to the point where *vmp* falls to p, which means to the point where the domestic price falls to the price paid by the importer for it (including the marginal cost of importing). The importers can be induced to restrict their purchases to the point where *vmp* equals the marginal cost to the country $(p + p/e_s)$ by a tax of $1/e_s$ (the inverse of the elasticity of supply of the good). This would raise the price paid by the firm or the collectivist agency to the marginal cost to the country.

This means that when the elasticity of supply of a good to the country is equal to unity, a 100 per cent import tax is appropriate. If the elasticity of supply is one-half, a 200 per cent tax is appropriate; and if the elasticity is as high as 20, so that price is raised only 1 per cent when the total amount imported into the country is increased by 20 per cent, the proper import tax would be one-twentieth or 5 per cent. In the same way the tax on exports should be the inverse of the elasticity of the foreign demand for the exported good. Firms under perfect competition and collectivist agencies following the Rule would export goods up to the point where the domestic price of the good, pf, is equal to the price paid by the foreigner, *vmp*. But the marginal revenue to the country is

equal to $p - p/e_d$, where p is the price paid by the foreigner and e_d is the foreign elasticity of demand. Exporters can be induced to limit exports to the point where the domestic price is equal to the marginal revenue to the country from the exports by imposing a tax on exports equal to $1/e_d$ (the inverse of the elasticity of the demand for the exports). If the elasticity of demand is unity, the appropriate tax is one of 100 per cent. This amounts to a virtual prohibition of the exports, and is proper because the marginal revenue to the country is zero—an increase in exports causes a proportionate fall in the price so that there is no return at all to the country. If the elasticity is less than unity, the tax should be more than 100 per cent. This means that the exporter should be fined *more* than the whole revenue from the sales of the product because in that case an increase in exports reduces the price more than proportionately so that the total value of the country's exports falls and the marginal revenue to the country is *negative.* If the demand has an elasticity as high as 5, so that a 5 per cent increase in the amount sold would lower the price by only 1 per cent, a 20 per cent tax would be proper for the purpose of maximizing the benefit to the country from its exports.

The ideal tax on imports equal to the inverse of the foreign elasticity of supply and on exports equal to the inverse of the foreign elasticity of demand is applicable when the Rule is being followed or when the same results are being reached by free enterprise under perfect competition. When competition is imperfect there is some restriction in importing and exporting as compared with perfect competition because the elasticities of supply and demand for the firms are less than infinite. The marginal cost to the firm is greater than the price of the imports that it buys, and the marginal revenue to the firm is less than the price at which it sells the exported goods to the foreigner. But the restriction is less than that required to maximize the gain from trade by the whole country because a 1 per cent change in the amount bought or sold by the firm, since it is less than the amount bought or sold by all the firms in the country, would have a smaller effect on the price than a 1 per cent change in the total amount bought by the whole country. The deviation between the price and the marginal cost

or the marginal revenue will be less and the elasticities will be greater for the firms than for the country. The appropriate tax would then correspond to the difference between the elasticities of supply and demand confronting the firms and the elasticities of supply and demand confronting the country. The taxes would have to be just enough to raise the marginal cost of imports to the firm to the marginal cost of imports from the point of view of the country, and to lower the marginal private revenue from the sale of exports by the firm to the marginal revenue from the sale of exports from the point of view of the country.[1]

[1] To be exact, the tax would be $\frac{f/c-1}{f+1}$ on imports and $\frac{f/c-1}{f-1}$ on exports where c is the elasticity of supply or demand from the point of view of country and f the elasticity of supply or demand from the point of view of the firm. (Naturally it is the elasticities of supply of the imports and the elasticities of demand for the exports that are relevant.) For example, suppose an imported good is being bought at $1.00 a unit, the elasticity of supply to the firm is 10, and the elasticity of supply to the country is 2. Then the marginal cost to the firm is $1.10 while to the country it is $1.50. Our first formula gives us $\frac{10/2-1}{10+1}$ or $\frac{4}{11}$ as the tax. This will raise the firm's marginal cost by 4/11 of $1.10, bringing it to $1.50 which is the marginal cost to the country. In the case of exports, with the same price of $1.00 and the same elasticities, the marginal revenue to the firm is 90 cents and the marginal revenue to the country is 50 cents. Our second formula gives us $\frac{10/2-1}{10-1}$ or 4/9 as the tax. This will reduce the marginal revenue of the firm to 5/9 of what it was and makes it 50 cents as required. In each case the firm equates the new marginal revenue, after payment of the tax, to its marginal cost, and the new marginal cost to its marginal revenue, and if we overlook any other monopolistic restriction the firm may be tempted to indulge in, we will have just the right amount of restriction in international trade to maximize the gain to the country from international trade on the assumption that the country is not concerned with any damage it might do to other countries' benefits from trade by these restrictions and that it is not afraid of any retaliation.

The formula is slightly different for exports simply because the marginal revenue diverges from price in the opposite direction from marginal cost. In the special case of perfect competition this divergence disappears for the firm because the marginal cost and the marginal revenue are then both equal to the price, f is equal to infinity, and the two formulas reduce to $1/c$, that is, to $1/e_s$ and $1/e_d$ as we saw in our examination of the ideal tax to be applied under perfect competition or where the Rule is being followed.

Such a tax would reduce imports and exports to the ideal level from the point of view of the country, making the monopolistic or imperfectly competitive firms import and export just as much as would maximize the gain to the country from the international trade, exploiting the foreigner as much as possible. But the monopolistic firm will not buy and sell the proper amount of *domestic* goods, because in these cases the marginal cost and the marginal revenue to the country are nothing but the marginal cost and the marginal revenue to society and these are the *msc* and *msb* and are equal

All the discussion of the optimum tax on imports and exports is, of course, subject to the political consideration of the desirability and the ethics of such exploitation of the foreigner, to say nothing of the danger of retaliation. But whatever the final political decision in the matter, it is good to know the possibilities and the limitations of such maximization of the country's gain from foreign trade. The same principle applies to foreign borrowing and lending, the price of which is the rate of interest. These too can be taxed on exactly the same principles so as to maximize the gain to the country through getting a higher rate of interest on a smaller amount of foreign lending or through paying a lower rate of interest on a smaller amount of borrowing from abroad.

Foreign exchange values should be subservient to the maintenance of full employment.

The rate of foreign exchange should be subordinated in the controlled economy to the maintenance of full employment and the rate of interest that induces the desired level of investment. Normally this would be done by the procedure applied to all other prices, namely, to raise the price of foreign exchange when the demand is greater than the supply and lower the price when the supply is greater than the demand. If the elasticity of demand for

to the price. To make the marginal cost to the firm equal to the price would require a *subsidy* equal to the inverse of the elasticity of *supply* of the factor to the firm, and to make the marginal revenue to the firm equal to the price would require a subsidy equal to the inverse of the elasticity of demand for the product of the firm. Monopolies tend to buy and sell too little from the social point of view, and if it is not possible or convenient to *destroy* the monopoly, it may be possible to *bribe* it into buying and selling the socially most desirable amounts of factors and products. The lower the elasticities, the greater the degree of monopoly and the greater the necessary bribe. As perfect competition is approached the elasticities approach infinity and the necessary subsidies or bribes tend to approach zero. It is, of course, simplest not to have any monopolies but to eliminate them by legislation or by government competition or by counterspeculation whenever possible. If it should not be possible to eliminate monopolies, however, it is better to bribe them into utilizing society's resources in the best way than to permit these wastes to continue out of pique. The question of the distribution of wealth and income among the individuals of society can be treated separately from this.

This discussion of the treatment of monopoly is not especially germane to international trade problems. It is treated here only because the technique, like many general techniques for economic problems, is most clearly developed in connection with international trade.

the country's imports *plus* the elasticity of demand for its exports comes to less than 1, the procedure would have to be reversed in order to bring about equality between demand and supply (as we saw in the last chapter). This would be the responsibility of an authority like the Exchange Stabilization Fund Board whose main function it would be, by the device of counterspeculation, to prevent small fluctuations in demand and supply from making the exchange rate fluctuate and to prevent manipulators from disrupting the even course of foreign trade by their operations. When the Exchange Stabilization Boards of every country find that, while they are primarily concerned with maintaining full employment, the rate of exchange need not be varied much in the course of equating demand to supply, it would be possible to establish an international currency by permanently fixing the rates of exchange among the currencies of the different countries. In such a case, however, it will not matter much whether this is done or not.

When a country that suffers from unemployment depreciates its currency in terms of other currencies (whether by reducing its gold content or otherwise), it makes its exports cheaper to foreigners (who have to pay for them in cheaper money) and it makes its imports dearer for domestic purchasers (who have to pay for them in the relatively appreciated foreign money). If the elasticities are not too low, this reduces the import balance of the country so that the rate of interest can be lowered without risking the export of all the gold. The lower rate of interest encourages investment and employment. The encouragement of exports also increases employment in the export industries and the discouragement of imports increases employment in the industries that compete with imports. All these direct increases in employment increase income, and this increases spending, which increases income and spending several times. The improvement in economic activity may improve the incentives for investment so that there is another round of improvement in investment, employment, income, spending, and more income. In this way a controlled economy can use a reduction of the foreign value of the currency to improve the employment position.

However, it is essential to note that for every one of these effects

in the country whose currency depreciates, there is a corresponding *opposite* effect in the other countries whose currencies show relative appreciation. This would mean that the prosperity of the country that depreciated the currency is acquired only at the expense of the other countries whose currencies must appreciate relatively, whose exports will fall off, whose imports will increase, and whose employment, income, and investment must fall for every rise we have noted in the first country. The depreciation of the currency is a species of the "beggar my neighbor" game!

Wage and cost reductions are just as competitive with other countries for employment as the reduction of exchange values.

There is a great deal of truth in this charge. Most, if not all, of the direct benefit a country gets from it is at the expense of other countries, but this is not a sufficient cause for condemning it as an antisocial act (in the international sense) which, like robbery, only enables one to benefit at the expense of the other and whose general practice can only do harm to society at large.

First we must note that this charge, which is usually stressed by defenders of the orthodox gold standard game, is just as applicable to the mechanism for achieving equilibrium under a system of fixed exchanges such as the gold standard. Unemployment then has to be cured by falling wages and costs and this, if it happens, will bring about an increase in exports and a decrease in imports, while the opposite effect with all its evil consequences will be seen in the other countries. The same thing is true of attempts to increase employment by the imposition of import tariffs. To the extent that they succeed in cutting down imports and increasing employment in the one country, they also succeed in cutting down the exports of the other countries and reducing employment there.

But unlike the case of tariffs there remains a net benefit all around instead of a net loss if all countries expand domestic demand without worrying whether this might make their exchange rates fall.

However, it is only the third of these measures, the imposition of tariffs for the purpose of improving the employment position,

that can truly be described as a "beggar my neighbor" policy. This is seen in the *indirect* effects. The indirect effects constitute what is left if all the countries apply one of these policies so that the direct effects cancel out. In the case of the tariffs all that remains is an interference with international trade. This makes all the countries poorer except perhaps for a possible temporary improvement while new investment is being undertaken in providing equipment in each country for the manufacture of the previously imported goods, if this should come to more than the decrease of investment in the exporting industries. (This may frequently be the case because the latter cannot fall below zero.) The permanent effects for the economy as a whole can only be bad.

This is not true of price and cost reduction under fixed exchanges. The direct effects cancel out as before, but the indirect effects may be very beneficial. At the lower price level the total amount of money, if it is unchanged, can buy more goods and is to that extent more useful as money. Less money need be held in connection with the same total of real transactions at the lower prices because these have a lower total monetary value. There is, therefore, some excess money which tends to be loaned out and this lowers the rate of interest. Consequently we can get more investment and more employment all round. While it is true that any country that reduced its costs and prices relative to the others gains in employment at the expense of the others, it is also true that all the countries reducing their costs and prices together are able to enjoy some total net benefit.

Depreciating exchanges also work in this way. Of course it is not possible for all countries to depreciate their currencies in terms of each other's, any more than it is possible for all countries to reduce their costs and prices relative to each other. But if the countries suffering from unemployment are not afraid to let their currencies depreciate in terms of gold or other currencies, they will ease their monetary situation by increasing the amount of money and lowering the rate of interest, so as to increase investment and employment. This would normally result in more imports being bought out of the increased incomes and would create an import surplus which in turn would cause the currency to depreciate (or

gold to flow out). In a fixed exchange system this had to be avoided at all costs and so the supply of money was not eased, the rate of interest was not lowered and investment and employment were not improved. Now, if there is no care about permitting the currency to depreciate, the country suffering from unemployment will go ahead, not merely unafraid of a depreciation of the currency, but probably even welcoming it because of the *direct* effect of increasing domestic employment at the expense of employment in other countries. If the other countries carry out a similar policy at the same time, no relative change in the values of the currencies will take place, so that there will be none of the direct effects of any country filching some employment from its neighbors. But there will remain the indirect effects of the general increase in the supply of money, that will lower the rate of interest all round and permit more investment and more employment for all the countries just as does a general reduction in costs and prices.

The advantages of simply increasing the amount of money all round as compared with reducing costs and prices all round are exactly the same as we saw in our study of employment in a single country when we ignored international complications. There is the tendency for the prices to stay for a long time before they fall, if they fall at all; meanwhile unemployment has to be suffered. And then there are the effects of the need for money and on the supply of money that result from *falling* prices as distinct from lower prices. All such difficulties are avoided by simply increasing the amount of money directly, by printing it, rather than trying to increase the effective amount of money by engineering a price fall through a depression.

There are two ways in which the amount of money might be increased in all countries. If the rates of exchange must be maintained throughout at constant ratios there has to be a well-worked-out, internationally controlled plan by which all the countries keep in step so that no country expands its money more than another. A relatively greater expansion by one country would result in that country's currency tending to fall in relation to that of other countries. This could be checked by an outflow of gold, but if it continues until gold stocks are exhausted the country would be

forced off the gold standard and its currency would then depreciate as compared to the currencies of the countries that did not expand their money so much. A concerted action that would prevent such occurrences is very difficult to arrange, and so far all such attempts have failed. Furthermore, unless the quantities of money in the various countries happen in the beginning to exist in the proportions corresponding to the amounts needed for full employment in every country, this final position cannot be reached even if there is agreement among the countries for a parallel expansion which would keep their exchanges stable. Some countries will reach full employment before others, and then they will not wish to expand any further since that could not increase employment for them and could only lead to a price inflation. If these countries stop expanding their currencies, the other countries which still are suffering from unemployment cannot expand either for that would create an import balance and lead to their being forced to relinquish the fixed exchange rate for their currencies. They must choose between giving up the fixed rate of exchange or permitting the unemployment to remain until such time as prices and costs have fallen enough to bring about full employment at the old exchange rates.

Stable exchanges will then be the result, and the establishment of fixed exchange rates will be a reasonable if not very important issue.

The other way in which full employment can be attained in all the countries by a general expansion of money demand is that described before where the countries do not worry about their rates of exchange but all expand their monetary supplies (as part of the general scheme of increasing domestic purchasing power) until there is sufficient investment and employment in every country. When there is full employment in any country the expansion of monetary demand is checked so as to prevent inflation. The rates of exchange are left to find their own equilibrium just like any other prices. After this equilibrium is reached, the exchanges will tend to move only with the much slower real changes in technical knowledge of production, in the supplies of

productive equipment in the different countries, in other influences on the productivity of things entering into international trade or competing with internationally traded goods, and in the changes in tastes and need for various imports by the inhabitants of the several countries. If it should be found that these changes cancel each other out very closely, so that the rates of exchange do not tend to change very much in the course of keeping the supply of each currency equal to the demand for it, then the controlled economy could consider whether it was worth while entering into arrangements with other countries for the fixing of the rates of exchange on some permanent basis such as the gold standard. But until such a situation of natural stability of the exchanges is reached, the controlled economy cannot think of giving up the right to adjust its domestic policy of regulating effective demand to eliminate depression and inflation.

INDEX