Ragnar Frisch
1932

BEITRÄGE ZUR ÖKONOMISCHEN THEORIE

IN VERBINDUNG MIT

ALBERT AFTALION, PARIS / L. V. BIRCK, KOPENHAGEN / HANS MAYER, WIEN / ARTHUR C. PIGOU, CAMBRIDGE / F. W. TAUSSIG, CAMBRIDGE, MASS. / GUSTAVO DEL VECCHIO, BOLOGNA / F. DE VRIES, ROTTERDAM

HERAUSGEGEBEN VON

EMIL LEDERER UND JOSEPH SCHUMPETER

3

New Methods of Measuring Marginal Utility

by

Ragnar Frisch

Professor of Economics in the University of Oslo

With 19 Charts and 1 Plate

VERLAG VON J. C. B. MOHR (PAUL SIEBECK)

TÜBINGEN 1932

Druck von H. Laupp jr in Tübingen.

To

IRVING FISHER

the pioneer of utility measurement

ACKNOWLEDGEMENT

I want to express my thanks to Mr. Herbert Tout of the School of Business Administration, the University of Minnesota, who has been kind enough to read the manuscript. He has corrected the English expressions in several places and has also suggested some improvements in the manner of presentation.

May 1931. Ragnar Frisch.

CONTENTS

1. INTRODUCTION.

Is marginal utility a measurable thing or is it not? Much metaphysical and confused discussion took place around this question a generation or two ago and even today we witness some belated echoes of this discussion, showing how surprisingly long a time it has taken for this part of economic theory to be penetrated by the constructive and rigorous thinking that long ago came to dominate in the natural sciences and which is one of the chief factors of true progress in any science.

Dupuit and the economists of the Austrian School based much of their reasoning on the notion of utility conceived in a more or less quantitative fashion. But Jevons is probably the first who discussed to any extent the question of statistical measurements of utility. His analysis is highly interesting from several points of view. However, his discussion of the logical background of the quantitative definition of utility did not penetrate very deeply. And besides, he never got so far as to elaborate a practical method of actually carrying the utility measurement through.

The real pioneer work in the field was done by Irving Fisher. His doctoral dissertation published in 1892 is a vigorous attempt at introducing a new spirit into the discussions of value. He says: "The truth is, most persons, not excepting professed economists, are satisfied with very hazy notions. How few scholars of the literary and historical type retain from their study of mechanics an adequate notion of force! Muscular experience supplies a concrete and practical conception but gives no inkling of the complicated dependence on space, time, and mass. Only patient mathematical analysis can do that. This natural aversion to elaborate and intricate analysis exists in Economics and especially in the theory of value. The very foundations of the subject require new analysis and definition."

The most important part of Fisher's dissertation was probably his attempt to give a rigorous and quantitative definition of utility. Although Fisher does not use the expression "theory of choice," he is really the founder of this theory; the main ideas of the theory of choice at least in their static formulation are actually contained in Fisher's dissertation. He discusses, for instance, specifically the amount of indeterminateness that attaches to the total utility function.

Later Pareto developed these ideas more extensively. It is true that Pareto somewhat changed the emphasis. He took great pains to point out the fundamental difference which he claimed should exist between the marginal utility approach and his own approach. However, one who likes to see the organic connection between the ideas of various epochs, will consider Pareto's approach rather as an attempt at elaborating the marginal utility theory in a more precise fashion than as an attempt at developing a new theory in opposition to the utility theory. What Pareto called ophélimité is the same as what Fisher called utility.

Besides the abstract speculations on utility definition given in his dissertation, Fisher also developed a method of actually carrying the utility measurement through by utilizing statistical data. His aim was not only to give a logical proof of the measurability of utility but also to give a more practical and decisive proof: that of actually measuring utility. Unfortunately, this method was not published until much later. It first appeared in print in "A Statistical Method for Measuring 'Marginal Utility' and Testing the Justice of a Progressive Income Tax" (Economic Essays Contributed in Honor of John Bates Clark, 1927). Long before that time Fisher had, however, developed this method in his courses at Yale University. A mimeographed statement of the method had also been privately circulated. The reason for not releasing a printed statement of the method at an earlier stage was, as Fisher himself explains in a circular letter enclosed with the reprints of "A Statistical Method . . .," that he wanted to be able to give at the same time plausible results of actual measurements. And the results he had obtained so far were not satisfactory.

I approached the problem of utility measurement in 1923 during a stay in Paris. There were three objects which I had in view:

(1) To point out the choice axioms that are implied when we think of utility as a quantity, and to define utility in a rigorous way by starting from a certain set of such axioms;

(2) To develop a method of measuring utility statistically.

(3) To apply the method to actual data.

The results of my study along these lines are contained in a paper "Sur un Problème d'Économie Pure," published in the Series Norsk Matematisk Forenings Skrifter, Serie 1, Nr. 16, 1926. In this paper, the axiomatics are worked out so far as the static utility concept is concerned. The method of measurement developed is t h e m e t h o d o f i s o q u a n t s, which is also outlined in Section 4 below. The statistical data to which the method was applied were sales and price statistics collected by the "Union des Coopérateurs Parisiens". From these data I constructed what I believe can be considered the marginal utility curve of money for the "average" member of the group of people forming the customers of the union. To my knowledge, this is the first marginal utility curve of money ever published.

An essential feature of the results obtained was that the f l e x i b i l i t y of the curve was n e g a t i v e (in conformity with the hypothesis of the decreasing marginal utility). Furthermore the flexibility proved to be d e c r e a s i n g i n a b s o l u t e m a g n i t u d e as the income increased. (Approximately, we may define the flexibility of the marginal utility curve of money as the ratio between a small percentage change in marginal utility and the corresponding percentage change in income. For a full definition see formula (2. 14) below.)

Furthermore, for the income range included in the material at my disposal, the absolute magnitude of the flexibility proved to be l a r g e r t h a n u n i t y. In a paper (in Norwegian) in Statsøkonomik Tidsskrift, Oslo, 1926, I put forward the hypothesis that the situation with respect to the money flexibility would probably be different in the United States, the flexibility here being less than unity over the income interval represented by the ordinary working man.

In this work on utility measurement I had been much stimulated by the study of Fisher's doctoral dissertation, but I did not know of his measurement method which, as mentioned above, was printed after my own paper. The fact is, also, that Fisher's method and my isoquant method are rather different in character. Fisher considers the money utility (that is, the marginal utility of money) as dependent on the entire price situation. This means that his money utility curve must be conceived of as holding good only for a given price situation. In other words, if ϱ designates the nominal income measured in dollars per unit of time (which is equal to the sum spent when no saving takes place), and $p_1, p_2 \ldots p_n$ are the prices of the various commodities, then Fisher conceives of the money utility ω as a function

$$\omega = \omega(\varrho;\ p_1, p_2 \ldots p_n) \tag{1.1}$$

of the variables $\varrho, p_1, p_2 \ldots p_n$. This function can be represented by a one-dimensional curve in (ω, ϱ) coordinates if all the individual prices $p_1, p_2 \ldots p_n$ are constant.

I started from a different assumption, namely, that the influence of price changes can, as a first approximation, be adequately taken account of by introducing only *one* variable besides the income, namely, the price of living P. That is to say, I studied the money utility ω as a function

$$\omega = \omega(\varrho, P) \tag{1.2}$$

of the two variables ϱ and P. The idea of considering the marginal utility of money as a function of these two variables was suggested to me by the study of Professor L. V. Birck's notion of a "general commodity" developed in his book "Læren om Grænseværdien"[1].

I use deliberately the expression "price of living" for P instead of "cost of living". The latter is an ambiguous term. Sometimes it is used in the true sense of a *price* of living and sometimes in the sense of an *expenditure* for living. The notion of expenditure is something that depends both on the price of living and also on whether the consumption is maintained at a high or low level. In any approach to the utility measurement, it is essential to keep these two notions quite distinct.

If we adopt the assumption (1. 2), it follows, as will be shown presently, that we must have

$$\omega(\varrho, P) = \frac{1}{P}\, w\left(\frac{\varrho}{P}\right) \tag{1.3}$$

Where $w(\varrho) = \omega(\varrho, 1)$ is the function of a single variable that expresses how the money utility varies with the income when

[1] English edition: "The Theory of Marginal Value", London, 1922.

the price of living is equal to 1. In other words, instead of studying the function of two variables $\omega(\varrho, P)$ we can study the function of one variable $w(\varrho)$. This is one of the fundamental facts on which the isoquant method of utility measurement was based.

It is obvious that the assumption (1. 2) underlying the isoquant method is not as general as Fisher's assumption (1. 1), but (1. 2) seems to be rather plausible as a first approximation, and it has the great advantage of opening up new important possibilities of actual measurement. If we adopt this assumption, the observations relating to *different price situations* may be utilized to determine points on the *same utility curve*. Fluctuations of the price of living are, so to speak, transformed into fluctuations of the real income. Therefore, violent and frequent price fluctuations do not destroy the possibility of utilizing the material for utility measurements. On the contrary, such fluctuations are to be welcomed, since they tend to produce the necessary spread in real income represented by the material at hand. In Fisher's method we must have a series of data referring to the same price situation, and this will, of course, considerably limit the possibility of actually obtaining adequate data.

There is also another important difference. The isoquant method involves only one commodity of comparison (for instance, food) while Fisher's method involves data for two such commodities (for instance, food and clothing). This is also a practical advantage of the isoquant method. In practice it is often difficult to obtain comparable data for several commodities.

On the other hand, if it were possible to obtain the necessary data, Fisher's method would be the most general. If it were possible to carry significant measurements through by his method, we would be able to verify whether or not the special assumptions underlying the isoquant method are valid.

From the beginning of 1930 I was, through Fisher's initiative, invited as a Visiting Professor at Yale University, and naturally there grew up a co-operation between us in an attempt to push the study of utility measurement further. The first thing we tried to do was to apply Fisher's method and the isoquant method to American material. We did not succeed, and for a surprising reason. The difficulty came where we least expected it, namely, from a lack of adequate price data. The United States Budget Study of 1918—19 furnished excellent material on food expenditure and also on some other expenditure groups. But we were unable te secure reliable data by which to make the g e o g r a p h i c a l price comparisons needed in our study. There exist of course plenty of time series material on prices for each of the various localities, but that kind of price comparisons between different localities at a given moment of time which we needed, were lacking. So our work slowed down, more or less, and we were contemplating confining ourselves to writing up a survey of the problem and its setting.

However, I could not get the problem off my mind, and struggling further with it, a new idea gradually took form. Was it possible to d e t e r m i n e the lacking price indexes from the budget data themselves? At first sight, the idea seems absurd, since the budget material only contains e x p e n d i t u r e d a t a and no information about prices at all. On closer examination, however, the idea proved sound, both theoretically and in practice.

The method which grew out of this is the method of translation expounded in Section 6 below. The application of this method to actual United States data is described in Section 7. The work of Section 7 was made possible through the encouragement and generosity of Professor Fisher, who has liberally put at my disposal clerical assistance for computation, drafting and other work in this connection and who has followed the progress of the work with the greatest interest. For all this I want to express my sincere thanks.

In order to give as clear a picture as possible of the nature of the procedure involved in the translation method, I have found it desirable not to start with a discussion of the most general formulae but rather to start with a brief explanation of the surface of consumption and the use which was made of this surface in my first approach (the isoquant method). From this I then proceed by steps until I reach the method of translation and finally the general flexibility equation. This latter equation, discussed in Section 8, is a tool of a still more general nature than the method of translation.

The point of view in the present study is entirely static. A first approach to the dynamic problem of utility was made

in my paper "Statikk og Dynamikk i den ökonomiske Teori" [1]. A further elaboration in this direction with particular reference to the problem of giving a rigorous quantitative definition of utility applicable not only in the static but also in a dynamic approach, will be published shortly in the transaction of the Norwegian Academy of Science.

In Sections 9—12 of the present monograph I study certain questions for which the notion of money flexibility and its numerical determination are particularly important.

---

[1] Nationalökonomisk Tidsskrift, Kbhvn., 1929, pp. 321—379.

2. MONEY UTILITY AND COMMODITY UTILITY.

Let us consider a certain well defined, physically measurable commodity such as, for instance, sugar. Furthermore, let us consider an individual that consumes a certain quantity x of this commodity per unit of time. For reasons which will presently become obvious, we may look upon the commodity considered as a c o m m o d i t y o f c o m p a r i s o n. For brevity of expression, I shall often in the following speak of "sugar" instead of using the more general expression "commodity of comparison." Frequently, I shall call x the "sugar consumption" (of the individual considered). I shall refer to a physical unit of the commodity of comparison as a "pound," and so on. Sometimes I shall use "food" instead of "sugar". However, the argument is perfectly general and applies to any commodity of comparison.

I shall not here enter upon any lengthy discussion regarding the possibility of giving a rigorous and quantitative definition of the notion of marginal utility of the commodity of comparison. In this regard the reader is referred to Professor Fisher's "Mathematical Investigations," to my paper "Sur un probleme d'économie pure," and to the forthcoming paper in the transactions of The Norwegian Academy. Here I shall only draw attention to the fact that the marginal utility of the commodity of comparison may be defined in two ways, either as a u t i l i t y m e a s u r e d p e r p o u n d or as a utility m e a s u r e d p e r d o l l a r s w o r t h of the commodity. I shall first consider the marginal utility measured per pound.

The marginal utility measured per pound u will depend on the magnitude of the consumption per year x. Conceivably it may depend also on other things, but these cases we shall not consider here. In other words, we assume that u is a function of x alone and we write it

(2. 1) $$u = u(x)$$

This assumption involves amongst others that the commodity of comparison is independent of other goods.

It should be noticed that in (2. 1) u is used in two different meanings. In the left member u stands for a m a g n i t u d e, and in the right member it stands for a f u n c t i o n s i g n. We might have used a different letter, say f, for this function sign. That is, we might have written (2. 1) in the form $u = f(x)$. But in view of the great number of symbols needed when the marginal utility theory is to be incorporated in a general equilibrium theory, it seems advisable to economize as much as possible in the use of different letters. Furthermore, there is a great memotechnical advantage in using the same letter for a certain magnitude and for the function that expresses how this magnitude depends on some other magnitude (or magnitudes). We could, of course, use small letters for the magnitudes and capital letters for the corresponding function signs, that is, we could have written (2. 1) in the form $u = U(x)$. But this is not advisable because it is very convenient to use the capital letters to designate certain quantities that arise out of an i n t e g r a t i o n of the quantities designated by the corresponding small letters. This principle of notation is particularly convenient in many parts of the dynamic theory. I have, therefore, found it advisable to use the notation expressed in (2. 1).

Let p be the price of sugar (or of any other commodity of comparison), and let $\xi = xp$ be the expenditure in dollars per unit of time for this commodity. Introducing the notions of price and expenditure, we may consider also another marginal utility concept in connection with the sugar, namely, the marginal utility μ measured p e r d o l l a r's w o r t h o f s u g a r. In Pareto's terminology, μ is the "ophélimité ponderée." The distinction between u and μ is n o t a distinction between the utility of a quantity b o u g h t and the utility of a quantity a c t u a l l y c o n s u m e d. All the way through the present analysis we assume that the sugar quantity which the individual buys per unit of time is equal to the sugar consumption per unit of time so that there is no stock on hand of which to take account. The only difference between u and μ is that μ is measured per a unit which is p times smaller than the unit per which u is measured. We therefore simply have

(2. 2) $$u = \mu p$$

The formula (2. 2) does n o t mean that we determine the total utility of an expenditure of p dollars by multiplying μ by p, u is indeed not the total utility of p dollars but the marginal utility measured per that quantity which can be bought for p dollars. The point may perhaps be made clear by an illustration: If it is true that at a certain moment of time a train moves at a speed of 20 meters per second, it is equally true that at that moment of time it moves with a speed of 72,000 meters per hour. It may well be that an hour is so long an interval that we cannot assume the train to be moving with the same speed over such an interval, so that it would not be correct — even approximately — to say that the distance covered in one hour is 72,000 meters. But that is a different question; it has nothing to do with the fact that we may at a given moment of time express the speed just as well per hour as per second. And the speed measured per hour is simply equal to 3600 times the speed measured per second. This seems to be a trivial remark but it is necessary to make it because a lack of understanding of this point sometimes has given rise to erroneous objections against that kind of marginal utility reasoning that is expressed by a formula like (2. 2). The static equilibrium condition in the exchange of two (or several) commodities will always be expressed by stating that the marginal utility of the first commodity is equal to the marginal utility of the second commodity times the exchange ratio between the two commodities. Such a condition holds good regardless of whether or not one of those units per which the marginal utility is expressed may be looked upon as an infinitesimal increment in this connection. The condition that the marginal increment analysis shall be applicable is not that the unit per which marginal utility is expressed, may be considered as "small" in the particular problem at hand, but that there actually exist a possibility of making "small" transactions in the field considered.

Neither in the case of u nor in the case of μ has the utility the denomination per unit of time. In both cases is the marginal utility considered as a utility per an absolute unit; u has not the denomination utility per pound per unit time, but simply the denomination utility per pound; and μ has not the denomination utility per dollar's worth per unit of time, but simply the denomination utility per dollar's worth. As we have seen, u depends on a magnitude measured in pounds

per unit of time, and we shall presently see that μ depends on a magnitude which is measured in dollar's worth per unit of time, but that is a different question.

The marginal utility of sugar measured per dollar's worth μ, evidently depends on two variables, namely ξ and p. We might write it:

$$\mu = \mu(\xi, p) \tag{2. 3}$$

In particular, let us consider the function μ which is obtained when the price p is put equal to unity. In this case, we get a function $\mu(\xi, 1)$ of a single variable. But this function $\mu(\xi, 1)$ that expresses how the sugar utility measured per dollar's worth varies with the sugar expenditure when the sugar price is equal to unity is obviously nothing else than the function that expresses how the sugar utility measured per pound varies with the sugar consumption. That is to say, we have

$$u(\xi) = \mu(\xi, 1) \tag{2. 4}$$

This we can, of course, also write

$$u(x) = \mu(x, 1) \tag{2. 5}$$

From an algebraic point of view, there is, no difference at all between these two formulae, since the formulae only define a connection between two functions, so that the meaning of the formula does not depend at all on which letter is used to designate the variable involved, whether ξ, x or some other letter. There arises a slight difference between the two formulae only if we attach to x and ξ respectively the economic significance of a pound-quantity (per year) and a dollar-value (per year). But even in that case we see that (2. 4) and (2. 5) represent fundamentally the same thing: When the price is equal to unity, the expenditure ξ is indeed equal to the quantity consumed x.

If the price p is not equal to unity there still exists a relation between the functions μ and u, namely the following:

$$\mu(\xi, p) = \frac{1}{p} u\left(\frac{\xi}{p}\right) \tag{2. 6}$$

This formula follows from (2. 2). In fact, when we compare μ and u we assume, of course, that the expenditure ξ on which μ depends is just that expenditure which corresponds to the physical consumption x on which u depends. That is, we have $x = \frac{\xi}{p}$. Consequently, if $\mu(\xi, p)$ is the function that expresses how the marginal utility of sugar measured per dollar's worth depends on sugar expenditure and sugar price, and if $u(x)$ is the function that expresses how the marginal utility of sugar measured per pound depends on the sugar quantity, we must by (2. 2) have the relation expressed in (2. 6).

It is also possible to approach the formula (2. 6) from a different angle. Let us start by considering the marginal utility of the sugar expenditure, that is, the marginal utility of sugar measured per dollar's worth, as a function $\mu(\xi, p)$ of the two variables ξ and p. Then it is easy to see that this function must be of a particular kind. It must satisfy the following proportionality equation

$$\text{(2. 7)} \qquad \lambda\mu(\lambda\xi, \lambda p) = \mu(\xi, p)$$

where λ is an arbitrary positive factor. This equation simply expresses the fact that if the sugar price is doubled (tripled, etc.) and at the same time the sugar expenditure is doubled (tripled, etc.) so that the consumption measured in pounds remains the same, then the marginal utility of sugar measured per dollar's worth must be reduced to $\frac{1}{2}$, ($\frac{1}{3}$, etc.). But (2. 7) is only another way of expressing the relation (2. 6). It is easy to see that (2. 6) may be derived from (2. 7) and vice versa. In fact, if (2. 6) is fulfilled we have

$$\lambda\mu(\lambda\xi, \lambda p) = \lambda \cdot \frac{1}{\lambda p} u\left(\frac{\lambda\xi}{\lambda p}\right) = \frac{1}{p} u\left(\frac{\xi}{p}\right) = \mu(\xi, p)$$

Conversely, if (2. 7) is fulfilled, we simply have to put $\lambda = \frac{1}{p}$ and use (2. 5) in order to get (2. 6).

We shall now see how the marginal utility of m o n e y may be subject to a similar analysis. The nominal income ϱ is analogous to the sugar expenditure per year ξ and the deflated income $r = \varrho/P$ (P being the price of living) is analogous to the physical sugar consumption per year, x. The only difference is that ϱ and r have reference to the general commodity which we call income, while ξ and x have reference to a special commodity.

Just as we considered two notions of marginal utility for the special commodity sugar, namely, the marginal sugar utility measured per dollar's worth of sugar, and measured per pound of sugar, we now consider two notions of marginal money utility, namely, the marginal utility of money measured per dollar and measured per unit of real purchasing power. The first notion we shall designate ω and the second w, in analogy with the notation μ and u used for the special commodity. ω and w will be called respectively the n o m i n a l money utility and the r e a l (deflated) money utility. Both notions we shall refer to under the general designation money utility.

Neither ω nor w has the denomination "pr. unit of time." They are not defined as utility pr. i n c o m e unit (nominal or real) but as utility pr. unit of a b s o l u t e m o n e y s u m (nominal or real). That is why I have called these notions money utility rather than income utility. We shall presently see that ω and w d e p e n d on the income (measured pr. unit of time) but that is a different question. Conceivably we may look upon ω and w as depending also on the cash balance carried by the individual, or better on the amount of capital funds owned by him. The presence of such funds is a security element for the future and may therefore influence the present magnitude of the money utility. This question has an obvious connection with the question of saving. In the present study we shall, however, not go into a discussion of this.

Since the only difference between w and ω is that w is a utility measured per a unit, that is P times as large as the unit per which ω is measured, we must have

$$\text{(2. 8)} \qquad w = \omega P$$

This formula is analogous to (2. 2). Also in connection with (2. 8) it will therefore be well to warn against that kind of misunderstanding which was mentioned in connection with (2. 2).

In an analysis of the money utility where we do not study the effect of price variation but assume that the prices are constant, the distinction between nominal and deflated money utility is unessential. In this case we only need one single notion: that of money utility. But when we introduce changes in the price system, the distinction between nominal and deflated money utility becomes essential.

Just as we considered the marginal utility of sugar measured per dollar's worth as a function of ξ and p, we may consider the money utility ω, that is to say, the marginal utility of money measured per dollar, as a function of ϱ and P

$$\omega = \omega(\varrho, P) \tag{2. 9}$$

If we assume that the nominal money utility ω is a function of ϱ and P it is easy to prove that the real money utility, that is, the money utility measured per unit of real purchasing power, namely w, must be a certain function $w(r)$ of the real income $r = \varrho/P$. And furthermore, this function $w(r)$ is nothing else than the particular function of one variable which we obtain by putting $P = 1$ in $\omega(\varrho, P)$, that is to say, we have

$$w(\varrho) = \omega(\varrho, 1) \tag{2. 10}$$

Or which amounts to the same

$$w(r) = \omega(r, 1) \tag{2. 11}$$

In fact, we always have $w = \omega P$. Therefore w may be looked upon as that magnitude which ω attains when $P = 1$. But by (2. 9) this magnitude is a function of a single variable ϱ, which when $P = 1$, is equal to r. Therefore w is a function $w(r)$ of r, and this function may be expressed in terms of the function $\omega(\varrho, P)$ by the equation (2. 10) or the equivalent equation (2. 11). From (2. 8) and (2. 11) it follows that we have

$$\omega(\varrho, P) = \frac{1}{P} w\left(\frac{\varrho}{P}\right) \tag{2. 12}$$

As mentioned in the introduction this equation is one of the fundamental equations on which the methods of utility measurement developed in the present study are based. (2. 12) is, of course, analogous with (2. 6).

By virtue of (2. 12) it is easy to see that we have the p r o - p o r t i o n a l i t y equation

$$\lambda\omega(\lambda\varrho, \lambda P) = \omega(r, P) \tag{2. 13}$$

And vice versa, if the proportionality equation (2. 13) is fulfilled, then (2. 12) must also hold good. (2. 13) is analogous with (2. 7). The proportionality equation (2. 13) can also be brought into connection with the basic assumptions underlying the static theory of price equilibrium, but I shall not enter upon this here [1].

[1] This question is developed a little more fully in my paper: "Der Einfluß der Veränderungen der Preisniveaus auf den Grenznutzen des Geldes." Zeitschrift für Nationalökonomie, Wien, 1931.

The relative rate with which the function $w(r)$ changes when r is subject to a small increase, we shall call the f l e x i - b i l i t y of the marginal utility of money, or shorter, the m o n e y - f l e x i b i l i t y. We denote it $\breve{w}$, that is to say, we put

$$\breve{w} = \breve{w}(r) = \frac{dw(r)}{dr} \cdot \frac{r}{w(r)} = \frac{d \log w(r)}{d \log r} \tag{2. 14}$$

From (2. 12) we see that the money-flexibility can also be considered as the p a r t i a l rate of change of the nominal money utility, with respect to a change in the nominal income, the price of living being kept constant. That is to say, we have

$$\breve{w} = \frac{\partial\omega(\varrho, P)}{\partial\varrho} \cdot \frac{\varrho}{\omega(\varrho, P)} \tag{2. 15}$$

In certain connections it is also of interest to introduce the partial rate of change of the nominal money utility with respect to a change in the price of living, the nominal income being kept constant. This quantity we shall denote $\breve{\omega}$, that is to say, we put

$$\breve{\omega} = \frac{\partial\omega(\varrho, P)}{\partial P} \cdot \frac{P}{\omega(\varrho, P)} \tag{2. 16}$$

By differentiating the proportionality equation (2. 13) with respect to λ and then putting $= 1$, we see that for all magnitudes of ϱ and P we have

$$\breve{w} + \breve{\omega} = -1 \tag{2. 17}$$

3. THE SURFACE OF CONSUMPTION.

According to the theory of static equilibrium the various variables which we have considered in the previous section u, w, etc., are not independent variables. There exists a certain relation between them, namely, this: In the equilibrium point our individual will have distributed his expenditure over the various items in the budget in such a way that for our particular commodity, say food, the nominal money utility times the price of the commodity is equal to the marginal utility of the commodity of comparison, measured per physical unit of the commodity. That is to say we have

$$\omega(\varrho, P) \cdot p = u(x) \tag{3. 1}$$

If we introduce the expression for $\omega(\varrho, P)$ from (2. 12), (3. 1) takes on the form

$$w\left(\frac{\varrho}{P}\right) = \frac{P}{p} u(x) \tag{3. 2}$$

That is to say, if we further introduce the ratio

$$a = \frac{P}{p} \tag{3. 3}$$

which we shall call the inverted relative food price, we get the equation

$$w(r) = au(x) \tag{3. 4}$$

This we shall call the equilibrium equation. It is an equation between the three variables a, x and r, and defines consequently a surface in a three dimensional (a, x, r) space. This surface we shall call the surface of consumption. The meaning of this surface can be described in the following way. Suppose we had a number of observations of our individual in different price and income situations. Each of these observations would consist of three numbers, namely, a, x and r, and could therefore be represented by a point in the (a, x, r) space. And a number of such observations would be represented by a set of points in the (a, x, r) space. If the want constitution

Fig. 1

of our individual had not changed, these observation points would not be scattered arbitrarily in the (a, x, r) space, but would lie on a certain surface. This surface is just the surface of consumption. Equation (3. 4) is nothing else than an implicit definition of this surface.

Fig. 1 is a photograph of a model of such a surface of consumption, built on the actual numerical results obtained by the translation method (see Section 7). The model belongs to Professor Fisher who has had it made for his course on price theory, and who has kindly permitted its reproduction here.

Along one horizontal axis of the model (from the foremost corner and towards the right) is measured the deflated (real) income r and along the other horizontal axis (from the foremost corner and towards the left) the inverted relativ commodity price a. Along the vertical axis is measured the quantity consumed of the commodity of comparison x. It is of particular interest to study the curves we obtain by slicing the surface with the following three systems of planes.

(1) Vertical planes parallel to the (a, x) axes.

(2) Vertical planes parallel to the (x, r) axes.

(3) Horizontal planes, that is, planes parallel to the (a, r) axes.

Each of these sets of planes define a set of curves, namely, the curves of intersection between the surface and the planes in the set. Each of these three sets of curves has a definite significance which becomes clear when we refer to the equation of equilibrium.

A given curve of the first set is the curve that represents how the quantity consumed x varies as a function of the inverted relative food price when the (real) income r is kept constant. Such a curve when projected into the plane through the a axis and x axis we shall call an inverted demand curve. It is nothing else than an ordinary demand curve except for the fact that one of the axes represents the inverted price of the commodity instead of the price itself. The curve is consequently sloping upward instead of downward. To each magnitude of the income we get one such inverted demand curve. The family of these curves as they appear when projected into the (a, x) plane is represented in Fig. 2 a. The curves in Fig. 2 a are actually the projections from the model exhibited in Fig. 1.

The second set of curves we may call the "Engel curves". They represent "Engel's law", that is, they express how the consumption of the commodity in question increases with the (real) income when the price of the commodity is kept constant. Of course, Engel's law in its original form was stated in terms of the changes that the *percentage expenditure* for a given commodity (or commodity group) undergoes when the

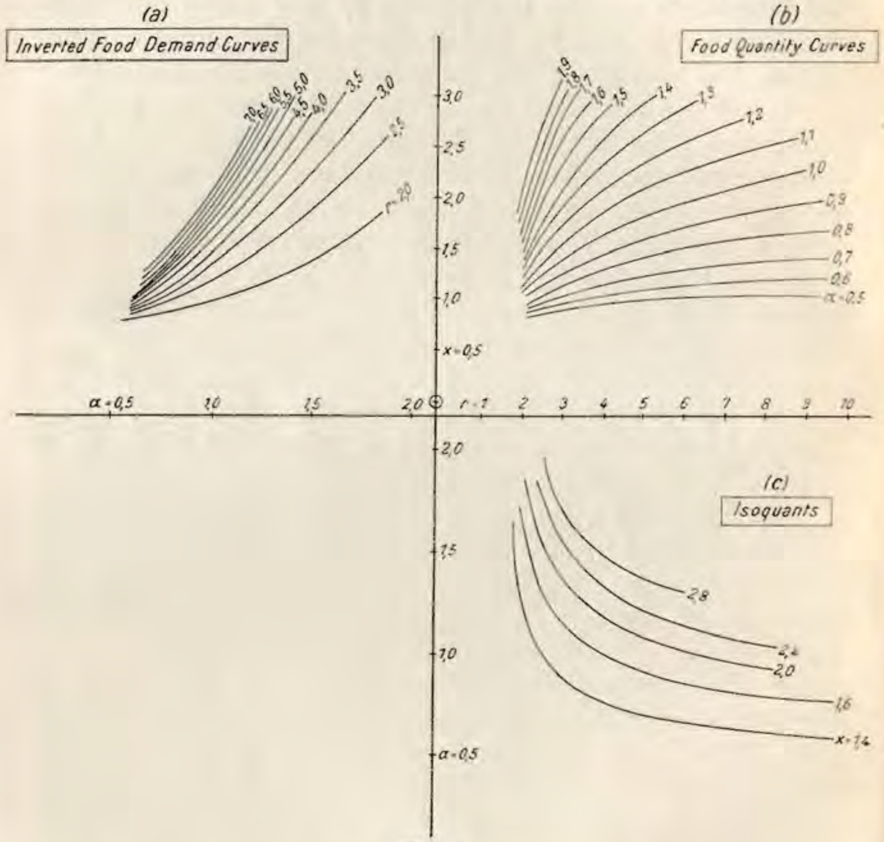

Fig. 2

income changes, while our curves in (x, r) coordinates give the *quantity* consumed as a function of the income. But the two types of curves really contain the same kind of information. If the commodity price is known, one of the curves may easily be transformed into the other. In the following we shall sometimes find it necessary to consider at the same time the curves

that express how the expenditure for the commodity of comparison varies as a function of the income, and the curves that express how the quantity consumed varies as a function of the income. The first we shall call *expenditure* curves (or more precisely: expenditure - and - income curves) and the second we shall call *quantity* curves (or more precisely: quantity - and - income curves). Thus we shall speak of food expenditure cuves, food quantity curves, sugar expenditure curves, etc. In Fig. 2b are exhibited the food quantity curves from the model.

The third set of curves obtained from the surface of consumption is perhaps the most interesting of them all. They represent contour lines, indicating the "altitude" of the surface above the horizontal plane, that is, above the (α, r) plane. These contour lines we shall call *isoquants*, because any one of them represents how α and r vary together when the quantity consumed is kept constant. They are exhibited in Fig. 2 c. *Apart from a stretch factor along the α axis, any such isoquant is simply a picture of the marginal utility curve of money $w = w(r)$.* From the equilibrium equation we have indeed

$$\alpha = \frac{1}{u(x)} w(r)$$

Therefore if x is constant, which means that $u(x)$ is constant, we have

$$(3.5) \qquad \alpha = \text{constant times } w(r)$$

But apart from the constant factor this is nothing else than the equation of the money utility curve $w = w(r)$.

This shows that all the isoquants must be similar in the sense that any one of them can be obtained from any of the others simply by multiplying the ordinate by a constant. If the isoquants are plotted with a logarithmic scale along the α axis they should consequently *all have the same shape* and only differ with regard to a displacement in the direction of the α axis. This, of course, is of considerable help in the actual statistical work.

It is obvious that if we have given the two functions $w(r)$ and $u(x)$ then the surface of consumption is uniquely determined. Its equation will then be given by (3. 4). Conversely if the shape of the surface is given, will that determine the two functions

$w(r)$ and $u(x)$? It is easy to see that the shape of the surface cannot determine the two functions $w(r)$ and $u(x)$ absolutely uniquely. Without changing the shape of the surface we may indeed multiply both sides of the equation (3. 4) with a common factor. It is therefore obvious that the two functions $w(r)$ and $u(x)$ cannot be determined otherwise than apart from an arbitrary constant factor, the same for $w(r)$ and $u(x)$. I shall show that this constant factor is the only arbitrariness left in the two functions $w(r)$ and $u(x)$. More precisely: if the shape of the surface is such that the surface may be defined by an equation of the form (3. 4), then the shape of the surface is sufficient to determine the two functions $w(r)$ and $u(x)$ uniquely, apart from an arbitrary constant factor, the same for $w(r)$ and $u(x)$.

In fact let $w(r)$ and $u(x)$ be any two functions such that (3. 4) becomes an implicit definition of the surface. If $\bar{w}(r)$ and $\bar{u}(x)$ is any other set of functions that determine the same surface, we must have

$$\frac{w(r)}{u(x)} = \frac{\bar{w}(r)}{\bar{u}(x)}$$

identically in x and r, and therefore

$$\frac{\bar{w}(r)}{w(r)} = \frac{\bar{u}(x)}{u(x)} \qquad (3.\ 6)$$

In the last equation the left member is independent of x and the right member is independent of r. The common ratio expressed by (3. 6) must therefore be a constant c independent of both x and r. That is to say, we must have

$$\begin{aligned} \bar{u}(x) &= c\,u(x) \\ \bar{w}(r) &= c\,w(x) \end{aligned} \qquad (3.\ 7)$$

which shows that the only arbitrariness in the functions $w(r)$ and $u(x)$ that define the surface of consumption is the arbitrary factor of proportionality.

The arbitrariness of this factor may be looked upon as an expression for the fact that there is nothing in the objective behavioristic criteria involved in the present analysis, which makes it possible to compare the utilities referring to one individual (or family) with the utilities for another individual (or family). This does not exclude that a method of actually making such inter-individual (or inter-familia) utility comparisons may some day be developed. But so far no such method has been developed

and I must admit that I doubt very much whether it will ever be done.

The fact that the functions $w(r)$ and $u(x)$ are affected by an arbitrary proportionality factor, different for different individuals, does not mean that a similar indeterminateness exists with regard to the shape of the surface of consumption. This shape is independent of the arbitrary factor in question. *An attempt to construct statistically the surface of consumption does therefore not involve any comparison between utilities for different individuals.* The shape of the surface is a perfectly objective behavioristic datum. And it is on this objective datum that the methods of the present investigation are built.

Since the functions $u(x)$ and $w(r)$ are affected with an arbitrary factor of proportionality, different for different individuals, while the shape of the surface of consumption is free from this arbitrariness, we see that in point of principle it is admissible to make any sort of comparisons between the shape of the surface of consumption for one individual and the shape of this surface for another individual. But there is only one particular sort of comparisons that can be made between the money utility function ${}_1w(r)$ for one individual and the money utility function ${}_2w(r)$ for another individual, namely, that sort of comparison that *is unaffected* if we multiply the function ${}_1w(r)$ by one arbitrary constant (independent of r) and multiply ${}_2w(r)$ by another arbitrary constant (independent of r). One sort of comparison that could be made is, for instance, a comparison between the relative rate of change of the two functions ${}_1w(r)$ and ${}_2w(r)$, that is, a comparison between the money flexibility ${}_1\check{w}$ for the first individual and the money flexibility ${}_2\check{w}$ for the second individual. A similar remark applies, of course, to a comparison between the function $u(x)$ for one individual and this function for another individual.

In the attempt at a statistical determination of the shape of the surface of consumption it will seldom be possible to follow the behavior of one single individual (or a single family). Most frequently it will be necessary to have recourse to data for a *whole group* of individuals (or families). Instead of studying the functional relationship between the magnitudes x and r with reference to a single individual, we will often have to study the functional relationship between the *average* consumption in the group $\bar{x}$ and the average income in the group $\bar{r}$. The question therefore arises: Will this functional relationship between the averages $\bar{x}$ and $\bar{r}$ really *be of the same sort* as the relationship between the individual magnitudes, x and r? Or may it, for instance, happen that a peculiar form of the income distribution could introduce a *bias* in the relationship between $\bar{x}$ and $\bar{r}$? More precisely: If the relationship between x and r were the same for all the individuals in the group, could it happen that this relationship was not the same as the resulting relationship between $\bar{x}$ and $\bar{r}$? Before we proceed to a discussion of the statistical determination of the utility curves it will be well to submit this question to closer analysis. In ~~an~~ order to do so I shall first prove a certain lemma on indirect means.

Lemma on the Indirect Mean of a Statistical Variable.

Let us consider two variables x and r that are connected by a functional relationship. This relationship may be defined in different ways. For instance, by expressing x as a function of r. Let this function be $x = h(r)$. The same relationship may, of course, also be described by expressing r as a function of x. Let this function be $r = k(x)$. Since the two functions $h(r)$ and $k(x)$ both express the same relationship there must evidently be an intimate connection between them. More precisely: Under certain very general assumptions it is possible to derive $k(x)$ when $h(r)$ is given, and vice versa. Therefore $k(x)$ is called the inverse function of $h(r)$, and similarly $h(r)$ is called the inverse function of $k(x)$. This sort of mutual relationship between functions, is denoted by a particular symbol: $h^{-1}(x)$ is used to express the inverse function of $h(r)$. That is, we put $k(x) = h^{-1}(x)$. The following are some simple examples of such inverse functions: If $h(r) = a + br$, that is, if $x = a + br$, where a and b are constants, we have $r = \frac{x-a}{b}$, that is, $h^{-1}(x) = \frac{x-a}{b}$. If $h(r) = r^2$, that is, if $x = r^2$, we have $r = \sqrt{x}$, that is, $h^{-1}(x) = \sqrt{x}$.

In these examples $h(r)$ is simply a low degree polynominal in r. In this case the inverse function can be expressed as a simple combination of multiplications and divisions of polynominals, extraction of square roots and the like. If $h(r)$ is a more complicated function, it may not be possible to express $h^{-1}(x)$ in such an elementary way, but that is unessential in this connection. There are other ways in which the inversion may be made, for instance, by development in a power series, by a simple graphical analysis and so on. The important thing in which we are here interested is not whether there exists a possibility of expressing $h^{-1}(x)$ in simple explicit form, but whether we have a possibility by some means or another to determine the shape of $h^{-1}(x)$ when the shape of $h(r)$ is given.

It may, of course, happen that $h^{-1}(x)$ is not a single valued function, even if $h(r)$ has this property. But also this difficulty may be gotten around by certain conventions as to which branch of the inverse of $h(r)$ shall be understood by the symbol $h^{-1}(x)$. A necessary and sufficient condition that the inverse of $h(r)$ shall be single valued in a certain x range is evidently that $h(r)$ is monotonic over the corresponding r range.

Now let there be given a frequency distribution $\varphi(r)$ of the values of r. That is to say, we assume that we have a certain number of elements, each element being characterized by a magnitude of r attached to it, and the elements being distributed with such a density that there are $\varphi(r)\, dr$ elements between r and $r + dr$. The magnitude

$$\bar{r} = \frac{\int r\varphi(r)dr}{\int \varphi(r)dr} \tag{3.8}$$

is then the mean of r. The integration in (3. 8) is to be extended to the whole r range for which there occur elements. Evidently it would not matter if we imagine the integration to be extended from $-\infty$ to $+\infty$ since the integrand $\varphi(r)$ will be zero wherever there are no elements. The magnitude

$$\bar{x} = \frac{\int x\varphi(r)dr}{\int \varphi(r)dr} = \frac{\int h(r)\varphi(r)dr}{\int \varphi(r)dr} \tag{3.9}$$

is the mean of x. More precisely x is the direct mean of x. It would also be possible to consider an indirect mean of x, that is a mean of x defined via the function $h(r)$. The procedure in constructing this indirect mean of x would be to take the *direct* mean of r and then ask: What magnitude of x corresponds to this mean of r? That is to say, we would define the indirect mean $\bar{\bar{x}}$ of x by

(3. 10) $\bar{\bar{x}} = h(\bar{r})$

Similarly we could define the indirect mean of r by

(3. 11) $\bar{\bar{r}} = h^{-1}(\bar{x})$

The question now arises: Is there a large difference between the direct and the indirect mean of x? The answer to this is given by the following formula[1]:

(3. 12) $$\bar{x} - h(\bar{r}) = \frac{1}{2} h''(\xi)\sigma^2$$

where $h''(\xi)$ designates the second derivative of $h(r)$ in a point ξ in the domain of integration, and σ^2 is the square of the standard deviation of the distribution $\varphi(r)$, that is,

(3. 13) $$\sigma^2 = \int (r - \bar{r})^2 \varphi(r) dr / \int \varphi(r) dr$$

The proof is easy. We simply have

$$(\bar{x} - h(\bar{r})) \cdot \int \varphi(r) dr = \int (h(r) - h(\bar{r}))\varphi(r) dr$$
$$= \int (r - \bar{r}) h'(\bar{r}) \varphi(r) dr + \frac{1}{2} \int (r - \bar{r})^2 h''(\xi_r) \varphi(r) dr$$

where ξ_r is a value between $\bar{r}$ and r. The first integral in the right member of the last formula is zero, and since $(r - \bar{r})^2$ is non-negative, the last integral is equal to $\frac{1}{2} h''(\xi)\sigma^2 \int \varphi(r) dr$ where ξ is a value of r in the total domain of integration. This proves (3. 12) since $\int \varphi(r) dr$ is different from zero.

From (3. 12) we see in particular that if $h(r)$ is a linear function, the direct mean and the indirect mean of x become equal. If $h(r) = \frac{1}{r}$, r assuming only positive values, $h''(r)$ is positive over the entire range of r, so that $\bar{x} > \bar{\bar{x}}$, which shows the familiar fact that the arithmetic mean of a positive variable is larger than the harmonic mean of this variable. Choosing $h(r) = \log r$ we see that the arithmetic mean is larger than the geometric mean, and so on.

From (3. 12) we further see that if the distribution $\varphi(r)$ is very concentrated (i. e. if σ is small), the difference between the direct and the indirect means become very small,

[1] This formula is only a special case of Hölder-Jensen's theory of convex functions. Rather than to refer the reader to this general theory I have preferred to give the present independent proof of the simple proposition here needed. It seems to me that the formula (3. 12) in all its simplicity exhibits something of the basic nature of the relationship between direct and indirect means. See also Darmois: Statistique mathématique. Paris 1928 p. 32.

regardless of the shape of the function $h(r)$.

* * *

Now let us apply this to the surface of consumption. Suppose that we have a great number of individuals (families), all of which have the same surface of consumption. Let $x = h(\alpha, r)$ be the function that expresses how the individual consumption (the family consumption) depends on α and r. Further let $\varphi(r)$ be the frequency distribution of the individuals (families) according to real income r, and let

(3. 14) $$\bar{r} = \int_0^\infty r\varphi(r) dr / \int_0^\infty \varphi(r) dr$$

be the mean income in the population, and let

(3. 15) $$\bar{x} = \int_0^\infty h(\alpha, r)\varphi(r) dr / \int_0^\infty \varphi(r) dr$$

be the mean consumption. If the individuals (families) behave according to our assumptions, the mean consumption $\bar{x}$ defined by (3. 15) is evidently equal to the total consumption in the group divided by the number of individuals (families).

Our first question is now: If the income distribution $\varphi(r)$ (and consequently also the mean income $\bar{r}$) is maintained unchanged, and the inverted price α varies, how will the average consumption $\bar{x}$ vary? Will it vary with α approximately in the same way as it would have varied if it had simply been computed in each price situation as the magnitude of the function $h(\alpha, \bar{r})$? It is indeed only when this condition is fulfilled that an observation of the co-variation between $\bar{x}$ and α will reveal the nature of the function $h(\alpha, r)$. The question here stated is the same as the question of whether the direct mean of x, namely, $\bar{x}$, will be approximately equal to the indirect mean $\bar{\bar{x}} = h(\alpha, \bar{r})$ for all the values of α considered. A condition under which this will be fulfilled is as we have seen in (3. 12) that the function $h(\alpha, r)$ considered as a function of r is nearly linear over that r range where the bulk of the individuals are found. The more concentrated the income distribution, the more closely will, of course, $h(\alpha, r)$ resemble a straight line over this range.

These conditions may not be fulfilled if we consider a large population as a whole, but they will generally be fairly well fulfilled if the population is sub-divided into income classes, Nos. 1, 2, . . ., and the mean income and mean consumption are considered separately within each class. This gives at the same time a possibility of determining empirically how the function $h(a, r)$ depends on r. The procedure in question would be to define the mean class incomes $\overline{r}_i$ and the mean class consumptions $\overline{x}_i$ by the formulae

$$\overline{r}_i = \frac{\int_{r_{i-1}}^{r_i} r\varphi(r)dr}{\int_{r_{i-1}}^{r_i} \varphi(r)dr}$$

(3. 16)

$$\overline{x}_i = \frac{\int_{r_{i-1}}^{r_i} h(a, r)\varphi(r)dr}{\int_{r_{i-1}}^{r_i} \varphi(r)dr}$$

and then consider the co-variation between a, $\overline{r}_i$ and $\overline{x}_i$. If the income classes are fairly narrow, this co-variation will be such that we have approximately $\overline{x}_i = h(a, \overline{r}_i)$. And this would hold good even if the income distribution $\varphi(r)$ should change, provided only that the income classes are narrow enough to make it plausible to assume that *within each class* the function $h(a, r)$ can be considered as approximately linear in r. This is the nature of the approximation involved in the application of the translation method to American data, described in section 7.

If only one class is considered and the necessary variation in the mean income is obtained by observing this class over a range of time, then the class should have as small an income spread as possible, so that the time-variation in the income distribution consists primarily in a change in the mean income, with most of the individuals always cluster tightly around the mean. This is the nature of the approximation involved in the application of the isoquant method to the Paris data, described in section 4.

Even if the above considered conditions are not exactly fulfilled, it would, of course, be possible to transform the relationship studied to an exact one. We could simply study the

relation between $\overline{x}$ and the *indirect* mean $\overline{\overline{r}} = h^{-1}(a, \overline{x})$ instead of the relation between $\overline{x}$ and $\overline{r}$. The function $h^{-1}(a, x)$ is the inverse function of $h(a, r)$ taken with respect to x, a being considered as a constant parameter under the inversion process. This relationship between $\overline{x}$ and $\overline{\overline{r}}$ gives, of course, a true picture of the function $h(a, r)$ since we always have exactly $\overline{x} = h(a, \overline{\overline{r}})$. Similarly we would get an exact relationship by considering the co-variation between the indirect mean $\overline{\overline{x}}$ and the direct mean $\overline{r}$. In fact, the indirect mean x is defined simply as $\overline{\overline{x}} = h(a, \overline{r})$. But this possibility of reducing the relationship between the mean income and the mean consumption to an exact one by considering the *indirect* instead of the direct mean for one of the variables involved, has more of a theoretical than a statistical interest, since in order to define the indirect mean $\overline{\overline{r}}$ or the indirect mean $\overline{\overline{x}}$, we have to know *a priori* the function $h(a, r)$ which it is just the object of the statistical analysis to determine.

4. THE ISOQUANT METHOD.

The simplest way of actually determining the function $w(r)$ is to construct one of the isoquants on the surface of consumption. If the data are sufficiently extensive we might construct several isoquants and thus check the result by seeing if the isoquants constructed are compatible in the sense that any one of them can be obtained from any other by multiplication with a stretch factor along the a axis. If this is so, we may select any of the isoquants as thus determined, change the notation along the a axis from a to w, and take the resulting curve as the marginal utility curve of money $w = w(r)$. This involves, of course, a selection of the unit of measurement of utility. If we later want to construct also the marginal utility curve for the commodity of comparison $u = u(x)$ and construct it in such a way that its ordinate becomes comparable to the ordinate of w, we have to remember that the unit of measurement of utility has already been chosen in the construction of the money utility curve. The scale to be used along the u axis is therefore not arbitrary any more. In order to find out what it should be we can simply select one particular observation point, say (a_1, x_1, r_1) and determine the u scale in such a way that $u(x_1) = w(r_1)/a_1$. If the $w(r)$ and the $u(x)$ curve are not to be used together, the unit of measurement along the ordinate axis may, of course, be chosen independently for the curve $w(r)$ and the curve $u(x)$.

If the data at disposal consist of time series for the quantities a, x and r for a given group of individuals, the isoquants may be constructed by the following time interpolation procedure. First one plots the three time series that represent the variation of a, x and r. The scale ought to be large enough to make graphical interpolation possible. If a, x and r are given by months, one could if necessary apply mechanical smoothing, for instance, a quarterly moving average. The time series a, x and r being plotted, one traces a straight horizontal line representing x = constant = x_1. The points of time where this

straight horizontal line cuts the time series x are marked off and also the points on the curves a and r that correspond to these points of time. Thus each of these points of time determine one a magnitude and one r magnitude. In other words, each of the points of time considered determine a point in an (a, r) diagram. The locus of points thus determined gives a numerical determination of one of the isoquants, namely, the isoquant corresponding to $x = x_1$.

It is, of course, also possible to adopt a certain analytical form of the functions $w(r)$ and $u(x)$ and fit them to the data by least squares or some other fitting principle. Generally it would, however, be advisable to use the time interpolation procedure as the principal basis of the analysis, since this method is independent of arbitrary assumptions regarding the shape of the analytical functions to be used in the fitting procedure. The result of the analytical fitting procedure should only be trusted insofar as the observation points determined by time interpolation group themselves fairly well around the isoquants determined by analytical fitting.

The isoquant method was the method that was applied to the previously mentioned Paris material. I cannot here go into detail about the preliminary work which consisted in comparing the different statistical series at disposal and sifting out those that could not be considered as adequate for the analysis. Nor can I give all the details regarding the numerical computations. I shall confine myself to enumerating the principal results.

The data were furnished by the Statistical Department of the "Union des Coopérateurs Parisiens." The following sets of data were used:

(1) Quantity of sugar sold per month in the stores belonging to the co-operative chain store system in the Paris region.
(2) The price of sugar in these stores.
(3) Total sales ("chiffre d'affaires") of the Cooperative Society per month (total for all commodities).
(4) The number of members at the beginning of the month.
(5) The price of living index computed by the statistical bureau of the Society.

All these data were monthly and they were used from June, 1920 to December, 1922, inclusive.

After having eliminated seasonal fluctuations from the total sales by the link relative method, the following ratios were taken as indices of our three variables[1] a, x and r: a was taken as equal to the ratio between the price of living index and the price of sugar; x was taken as the ratio between the quantity of sugar sold per month and the number of members (by interpolation the number of members was taken in the middle of the month), and r was taken as the ratio between total sales and the index of the price of living.

Table 1 gives the result of these computations.

Table 1.

Material: Paris Cooperative Societies.

	Inverted relative sugar price a	Quantity of sugar consumed x	Real (deflated) income r
1920 June	756	2710	376
July	749	1750	374
August	742	2976	385
September	794	2015	307
October	837	1115	273
November	1018	1164	252
December	1266	928	226
1921 January	1223	940	232
February	1162	710	226
March	1100	850	240
April	978	1223	253
May	975	1181	277
June	1081	1510	262
July	953	1971	269
August	944	1602	254
September	1040	1266	237
October	1228	1832	245
November	1208	1905	254
December	1196	2000	258
1922 January	1221	1652	255
February	1200	1728	268
March	1130	1870	281
April	1125	1710	273
May	1218	1780	259
June	1178	1962	243
July	1117	1749	252
August	1070	1823	252
September	1096	1778	263
October	1197	1900	253
November	1160	1945	259
December	1092	1955	273

[1] The notation used in my paper "Sur un Problème D'Economie Pure," is slightly different from the one here adopted.

From these data the isoquants corresponding to $x = 1550$, $x = 1750$, and $x = 1950$ were determined by time interpolation after having applied a three month moving average to each of the series a, x and r. The isoquants were also determined by an analytical method using as an expression for the money utility the following function:

$$w(r) = \frac{c}{\log r - \log a} \tag{4.1}$$

where c and a are constants; the constant a indicates the minimum of existence. This function is one of the simplest functions that satisfy a certain number of conditions which it seems plausible to impose on the money utility function. These conditions were discussed in my paper "Sur un Problème d'Economie Pure" previously referred to. The conditions in question exclude, for instance, all functions which are such that the money flexibility $\check{w}(r)$, defined by (2. 14) is of the form

$$\check{w}(r) = Re^{Q}$$

R and Q being rational functions of r, and they exclude a fortiori all functions $w(r)$ which are themselves of the form

$$w(r) = Re^{Q}$$

These conditions exclude amongst others the function

$$w(r) = \frac{c}{r - a}$$

that was suggested by Daniel Bernoulli. It seems to me that the place which Bernoulli's function has taken in the discussion of utility measurement and measurability is entirely unwarranted. Bernoulli's function has amongst others the property that it gives an absolute value of the money flexibility which is larger than unity for all incomes (above the minimum of existence), the absolute value of the flexibility only approaching unity as the income tends towards ∞. It is out of the question that any such function can give a good approximation to the law of variation of the money utility. The above mentioned conditions regarding the money utility function also excludes the function

$$w(r) = \frac{c}{(r - a)^2}$$

which has been suggested by Charles Jordan[1].

[1] The American Mathematical Monthly, Vol. XXXI, No. 4, (1924).

The isoquants determined in the Paris material by analytical fitting using the formula (4. 1) gave a result which checked fairly well with the result obtained by time interpolation. The points determined by time interpolation grouped themselves fairly well around the corresponding, analytically determined isoquants. The variation in money utility as determined by these methods is given in Table 2.

Table 2.

Paris Material, 1920—1922.

Real (deflated) income r	Marginal utility of money w	Absolute value of the flexibility of the money utility curve $-\check{w}$
(75%)	(284%)	(6,40)
(80%)	(201%)	(4,52)
85%	158%	3,55
90%	131%	2,96
95%	113%	2,55
100%	100%	2,25
105%	90,4%	2,03
110%	82,3%	1,85
115%	76,1%	1,71
120%	70,8%	1,59
125%	66,6%	1,50
130%	62,7%	1,41
135%	59,7%	1,34
140%	56,8%	1,28
(145%)	(54,3%)	(1,22)
(150%)	(52,3%)	(1,18)

(The figures in parenthesis are extrapolated outside the range of observation.)

5. THE QUANTITY VARIATION METHOD.

In the isoquant method the inverted relative price of the commodity of comparison a is the variable element through which the money utility curve is generated. Each price situation (with the selected constant magnitude of x) gives information only about one point on the money utility curve. Therefore, we only get as many points of this curve as there are different price situations (with the selected constant x) in the material. If we only have one such price situation we only get one point on the curve, that is to say, we do not get any information at all about the shape of the curve. The reason why the Paris material gave several points on the same money utility curve was that it consisted of time series, with considerable variation in the relative sugar price.

If the material consists of budget data, the situation is different. Each budget investigation will generally concern one particular place and time, and thus have reference only to one price situation. There may be available two or a few such budget investigations covering either places or times which exhibit different price situations, but there will generally not be available so many of these investigations that the price spread between them can be taken as a variable by which to generate points on the money utility curve. The question therefore arises: Instead of keeping x constant and observing the corresponding covariation between a and r, as we did in the isoquant method, can we keep a constant, observe the corresponding covariation between x and r, and use this as a means of determining the shape of the money utility curve? In other words: Can the money utility curve be determined by cutting the surface of consumption by an (x, r) plane instead of cutting it by an (a, r) plane? The answer is that it is not possible to determine the shape of the money utility curve by using one (x, r) plane, but it is possible to do it by means of two (or more) such planes. The method of constructing the money utility curve built on this principle we shall call the quantity variation method, because in this method the quantity consumed x is the variable by which the money utility curve is generated.

I shall first show that it is impossible to determine the shape of the money utility curve by one (x, r) plane. In fact, let us cut the surface of consumption with an (x, r) plane for the particular value $a = a_1$ and let us construct a two dimensional diagram, the ordinate of which measures marginal utility and the abscissa of which measures the quantity consumed x. On a secondary functional scale along the abscissa-axis we mark off the magnitude of the real income r which corresponds to x, according to the quantity-and-income curve associated with the value $a = a_1$.

Now without getting into the slightest discordance with the observational data at hand we may draw any curve in the diagram described and assume it to be the marginal utility curve $u(x)$ of the commodity of comparison. We only have to adopt as the marginal utility curve of money, that very same curve, the abscissa being however now given by the functional r scale and the ordinate by a_1 times the ordinate of $u(x)$. To a certain guess as to the shape of the $u(x)$ curve there consequently corresponds a certain shape of the $w(r)$ curve and vice versa. Guessing at the shape of one of these curves is just as legitimate as guessing at the shape of the other. The material does therefore, taken by itself, not furnish any definite information about the shape of either of the curves. Of course, if we make some specific assumption regarding the nature of one of the curves, for instance, assume that the money flexibility is constantly equal to unity in absolute value, or assume that the money utility curve has some other particular analytical form, it may be possible to determine the parameters of this function from a one-price material. This only shows how dangerous it is to apply a curve fitting procedure to the present problem. A money-utility curve determined in this way from a one-price material would have very little significance.

But if we have at our disposal data relating to two different (x, r) sections on the surface of consumption, then the function $w(r)$ as well as the function $u(x)$ can be statistically determined.

In fact let

(5. 1) $$r = r(a, x)$$

be the explicit expression for r obtained by solving the equilibrium equation (3. 4) for r. If a is kept constant then $r = r(a, x)$ is simply the relation between r and x which is statistically given by the quantity-and-income curve corresponding to the particular price situation a considered. We may therefore consider the function $r = r(a, x)$ as statistically determined, at least for a few particular values of a.

Introducing the expression (5. 1) in (3. 4) we get the equation

(5. 2) $$w(r(a, x)) = au(x)$$

From the way in which the function $r(a, x)$ is defined, it follows that (5. 2) is an identity. It holds good for any a and any x. Therefore, putting first $a = a_1$, and then $a = a_2$ and eliminating the function $u(x)$ we get

(5. 3) $$\log w(r(a_1, x)) - \log w(r(a_2, x)) = \log a_1 - \log a_2$$

and this equation holds good for any x.

For shortness we introduce the notation

(5. 4) $$r_1(x) = r(a_1, x) \qquad r_2(x) = r(a_2, x)$$

That is to say, $r_1(x)$ is the function that expresses how the real income depends on the quantity consumed of the commodity of comparison in the first budget material, and $r_2(x)$ is the function that expresses how the real income depends on the quantity consumed in the second budget material. Introducing this in (5. 3) and dividing through by $\log r_1(x) - \log r_2(x)$ we get

(5. 5) $$\frac{\log w(r_1(x)) - \log w(r_2(x))}{\log r_1(x) - \log r_2(x)} = \frac{\log a_1 - \log a_2}{\log r_1(x) - \log r_2(x)}$$

But the left number of (5. 5) is nothing else than an expression for the average money flexibility over the income interval from r_1 to r_2.

For an infinitesimal interval in the vicinity of the income point r, the money flexibility is namely by (2. 14) defined as

$$\check{w}(r) = \frac{d \log w(r)}{d \log r}$$

The expression

(5. 6) $$\check{w}(r_1, r_2) = \frac{\log w(r_1) - \log w(r_2)}{\log r_1 - \log r_2}$$

is therefore a plausible definition of the average money flexibility, but this is just the expression in the left member of (5. 5).

The only difference is that in the definition (5. 6) r_1 and r_2 are independent variables, while in the equation (5. 5) the variation in r_1 and r_2 is generated through the fact that both r_1 and r_2 depend on one and the same variable, namely, x. Introducing the flexibility $\breve{w}(r_1, r_2)$ (5. 5) takes the form

$$\breve{w}(r_1, r_2) = \frac{\log a_1 - \log a_2}{\log r_1(x) - \log r_2(x)} \tag{5. 7}$$

This formula holds good for a n y magnitude of x, and it has such a form that it can be directly applied to the observational budget material provided this material is such that it gives information about the functions $r_1(x)$ and $r_2(x)$ as well as information about the constants a_1 and a_2. To each (r_1, r_2) interval which may be determined from the two budget materials by attributing a certain magnitude to x, the formula (5. 7) gives a determination of the average money flexibility $\breve{w}(r_1, r_2)$.

If an income interval (r_1, r_2) thus determined is not too large, we may take this magnitude of the average flexibility $\breve{w}(r_1, r_2)$ as an approximate expression for the p o i n t flexibility $\breve{w}(r)$ in the logarithmic middle of the interval, that is for $r = \sqrt{r_1 r_2}$. In most practical cases the closeness of this approximation will be sufficient. As a rule, it will be of no use to work with any closer approximation because the accuracy obtained by putting $\breve{w}(r_1, r_2) = \breve{w}(r)$ as described is far closer than that determined by the o t h e r sources of error: the erratic variations or other disturbances in the material. Therefore if the income intervals (r_1, r_2) determined by letting x vary and noticing the magnitudes $r_1(x)$ and $r_2(x)$ which correspond to each magnitude of x, are not too large, we get through this variation of x a fairly good picture of the shape of the flexibility function $\breve{w}(r)$ for that part of the income range which is covered by the two budget materials used.

When the flexibility function $\breve{w}(r)$ is determined, it is easy to derive the money utility function $w(r)$. We have

$$\log w(r) = \log w(r_0) - \int_{s = r_0}^{r} (-\breve{w}(s)) d \log s \tag{5. 8}$$

It does not matter whether the log in (5. 8) is interpreted as lognat or $\log_{10}$ or the log in any other system. In an actual case the integration (5. 8) will, of course, have to be performed num-

erically. Since $\breve{w}(r)$ is a negative quantity we have for convenience expressed the formula (5. 8) in terms of $(-\breve{w}(r))$. This procedure of first determining the flexibility curve directly through the statistical material and then deriving the money utility curve itself indirectly through (5. 8) is the procedure that will, as a rule, be found the most convenient in practice when the quantity variation method is used.

In point of principle it would, however, have been possible to aim directly at the money utility curve. If the income intervals (r_1, r_2) determined from the variation of x are very large, this direct procedure will even have a certain advantage because it does not involve the approximation that consists in replacing $\breve{w}(r_1, r_2)$ by $\breve{w}(\sqrt{r_1 r_2})$. In order to explain this direct procedure let us revert to equation (5. 3). If, for shortness, we put

$$\delta = \log a_1 - \log_2 \tag{5. 9}$$

equation (5. 3) may be written

$$\log w(r_1(x)) - \log w(r_2(x)) = \delta \tag{5. 10}$$

Let us for a moment consider this equation as a functional equation by which the function $\log w(r)$ is defined. Looking at the equation (5. 10) from this point of view we see that it has a certain resemblance with a finite difference equation. In order to exhibit this similarity let us consider r_1 as the independent variable instead of x. Then r_2 becomes a certain function of r_1. Let us denote this function by $r_2 = R(r_1)$. The way in which the function $R(r_1)$ is defined may be visualized by imagining that we choose a certain magnitude of r_1, determine in the first budget material what x corresponds to this r_1, and then determine in the second budget material what income corresponds to this x. This income in the second budget material is just r_2, and from the process described it is clear that it becomes a function of r_1. Since r_1 is now considered as the independent variable we shall, for shortness, drop the subscript 1. If this is done, equation (5. 10) appears in the form

$$\log w(r) - \log w(R(r)) = \delta \tag{5. 11}$$

where r is now an independent variable and $R(r)$, as explained, is the function that expresses what income in the second budget material that shows the same commodity consumption as the income r in the first budget material.

If the function $R(r)$ had been of a particularly simple sort, namely,

$$R(r) = r + k \qquad (5.12)$$

where k is a constant, then (5. 11) would have been a true difference equation of the first order in the function $\log w(r)$. The solution of this equation would simply have been obtained by starting at some initial value of r, say r', attributing to $\log w(r')$ an arbitrary magnitude and then computing $\log w(r' + k)$ by subtracting the known magnitude of δ from $\log w(r')$, further computing $\log w(r' + 2k)$ by subtracting δ from $\log w(r' + k)$, and so on. This would have determined the function $\log w(r)$ in a set of equidistant points that lie k abscissa units apart, but between these points the function $\log w(r)$ would not have been determined by the functional equation. Or more precisely: Without infringing on the functional equation we could have chosen an arbitrary shape of the function $\log w(r)$ over an interval of length k. When this was done the value of $\log w(r)$ in any other point would have been uniquely determined through the functional equation. Of course, the indeterminateness here considered is only a theoretical indeterminateness. In practice it would be plausible to put up the further requirement that $\log w(r)$ should have a smooth course and this would practically do away with its arbitrary shape over the interval of length k, provided only that k is fairly small. In practice therefore the only arbitrariness that would be left is the arbitrary choice of the magnitude of $\log w(r)$ in the initial point r'.

If $R(r)$ is not the simple linear function (5. 12), then the points in which the magnitudes of $\log w(r)$ are determined by the functional equation (5. 11) are not equidistant any more, but otherwise the procedure of solving the functional equation is very much the same: First we select an initial abscissa point r' and attribute to $\log w(r')$ an arbitrary magnitude. Then we go to the new abscissa point $r'' = R(r')$. That is to say, we find out what income r'' in the second budget material that shows the same commodity consumption as the income r' in the first budget material. From the functional equation (5. 11) we know that in this new abscissa point $\log w(r)$ shall be δ less than in the abscissa point r'. That is to say, we have $\log w(r'') = \log w(r') - \delta$. From r'' we go to the further abscissa point $r''' = R(r'')$. That

is to say, we now consider r'' as an income in the first budget material and ask what income r''' in the second budget material shows the same commodity consumption as the income r'' in the first budget material. The abscissa r''' being thus determined, the magnitude of the function $\log w(r)$ in this abscissa point is by the functional equation determined as $\log w(r''') = \log w(r'') - \delta = \log w(r') - 2\delta$. Thus we could continue and determine a series of points $r^{(n)}$. In any such point the magnitude of the function $\log w(r)$ would be equal to

$$\log w(r^{(n)}) = \log w(r') - (n - 1)\delta. \qquad (5.13)$$

Graphically, this procedure can be exhibited as in Fig. 3. The first thing to do is to plot in a ($\log x$, $\log r$) system the observation points on the quantity-and-income curve for the first budget material (i. e., the material where $a = a_1$). The points

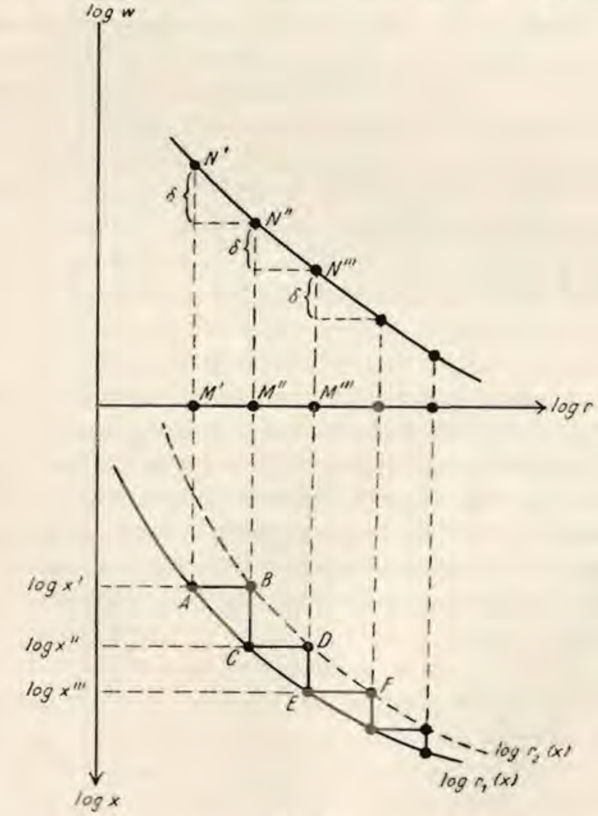

Fig. 3.

should be smoothed, preferably by freehand drawing, so as to obtain a continuous curve. No analytical formulas should be used in this smoothing process. Let the result be curve *ACE* . . . in Fig. 3. For convenience the log x axis in Fig. 3 is constructed vertically and downwards. Similarly the quantity-and-income curve for the second budget material should be drawn. In Fig. 3 it is *BDF* . . . Fig. 3 exhibits a situation which is plausible when $\alpha_1 > \alpha_2$.

When this is done, an initial point log x' on the log x axis should be chosen, and the zig-zag line *ABCDEF* . . . should be constructed. The segments *AB*, *CD* . . . are horizontal straight lines. And the segments *BC*, *DE* . . . are vertical straight lines. From the points determined by this zig-zag line the construction of the marginal utility curve for money can be made as follows.

The point A (where the quantity consumed is x') is taken as a starting point. We mark off the corresponding abscissa M' and choose an arbitrary initial ordinate $M'N'$ on the log w curve. The log w axis is constructed vertically and upwards. The point N' on this curve is thus fixed. When this is done we revert to the point A, and move from A to B, marking off the abscissa point M'' which corresponds to B. During the movement from A to B, x has been constantly equal to x'. Since A represents a point on the $r_1(x)$ curve and B a point on the $r_2(x)$ curve, the difference between the corresponding log w ordinates is by (5. 11) equal to the constant δ. The ordinate N'' on the log w curve is thus determined. Next we move from B to C. That is, we find out which x value on the $r_1(x)$ curve corresponds to the value x' on the $r_2(x)$ curve. This new x value is designated x''. Moving now from C to D, we get the abscissa point M'''. The drop in the log w ordinate when log r moves from M'' to M''' is still equal to the constant δ, which gives the point N''' on the log w curve. In this way we can continue and determine as many points on the log w curve as there are zig-zag turns between the log r_1 and the log r_2 curve.

We can also give a geometric interpretation of the flexibility determination built on equation (5. 7). We simply have to draw a series of horizontal lines $A'A''$, $B'B''$, . . . as indicated in Fig. 4. On the first of these lines we measure the horizontal distance between the point A' where the line cuts the log $r_1(x)$ curve and the point A'' where it cuts the log $r_2(x)$ curve. We

divide δ by this horizontal distance and plot this ratio as a vertical ordinate in the abscissa point corresponding to the midpoint between A' and A''. This gives one point on the flexibility curve — $\breve{w}(r)$. On this curve there is no arbitrary factor

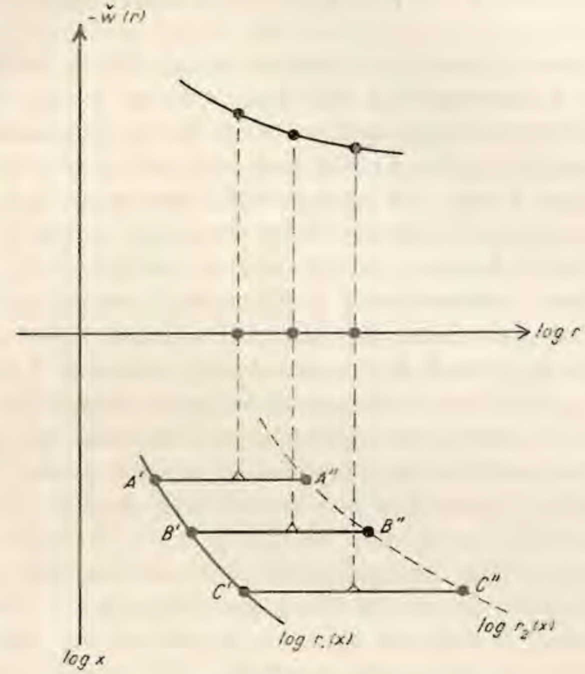

Fig. 4.

to choose, because the flexibility is a pure number uniquely determined from the statistical data. Similarly will the ratio between δ and the length $B'B''$, erected as an ordinate in the abscissa point midway between B' and B'' give a second point on the flexibility curve. And so on. This sort of construction can be carried through even though the two curves log $r_1(x)$ and log $r_2(x)$ should be such that no zig-zag turns fall within the range of observation.

6. THE TRANSLATION METHOD.

The quantity variation method developed in Section 5 is a method of constructing the money utility curve from the knowledge of two food quantity curves (or two other commodity curves) and the corresponding two magnitudes of α (supposed different). As a rule, the food quantity curves are not directly given in the budget material. What are given are the food expenditure curves, i. e. the curves exhibiting the relation between food expenditure ξ and nominal income ϱ. But the food quantity curves may be derived if we know, the food price p and the living price P in the first budget material relative to these magnitudes in the second budget material. By convention p and P in the second budget material may be put equal to unity, so that we may say that it is only p and P in the first budget material that are needed. If p and P in the first budget material are given, we simply have to divide in each point of the food-expenditure curve for the first budget material, the ξ coordinate by p and the ϱ coordinate by P. The curve thus obtained is the food quantity curve for the first budget material, that is, the curve exhibiting the relation between x and r in this material. If p and P in the second budget material are put equal to unity, the food quantity curve in this material is identical with the food expenditure curve. From the two food quantity curves thus obtained, the money utility curve may be constructed by the method of Section 5.

We shall now consider the situation which arises when there are no data available regarding the prices p and P. In this case the food quantity curves are not known, but the food expenditure curves are. The purpose of the discussion in the present Section is to show that the shapes of the given food expenditure curves contain information which makes it possible to draw conclusions regarding the prices p and P, and thus derive the data necessary to determine the money utility curve. For reasons which will presently become obvious, this method will be called the translation method of measuring utility.

Each food quantity curve represents a given magnitude of α, namely the magnitude of α that prevails in the place in question. Here and in the following I am speaking about the "places" to which the given budget materials refer. But the argument may, of course, be applied just as well to different points of time. Two places that have the same magnitude of α should also have the same food quantity curves, if the functions $w(r)$ and $u(x)$ are the same for these two places. And one of the assumptions underlying the present attempt at utility measurement is just that the functions $w(r)$ and $u(x)$ are the same for all the places considered. It might be that the interval covered by the food quantity curves will be different in the two places so that one of the curves will project beyond the other even though they refer to the same α. But over the interval where the two curves overlap they should be identical. It should be noticed that it is only the two food quantity curves that become identical in two places with the same α. The food expenditure curves are not necessarily identical because the two places have the same α. But if p and P, taken separately, are equal in the two places, then also the expenditure curves become identical.

From the way in which the food quantity curves are derived from the food expenditure curves it follows that if all the curves were drawn on a log chart, the unknown food quantity curve for each place would be identical with the known food expenditure curve for that place so far as the shape of the curve is concerned. The difference would only consist in a translation. In fact on the log chart the quantity curve is obtained simply by moving the expenditure curve a certain distance, log p, in the direction of the negative ξ axis and a certain distance, log P, in the direction of the negative ϱ axis. Such a translation can be looked upon as a translation in relation to an arbitrarily selected base place that remains fixed. The coordinates $(-\log P, -\log p)$ of the translation considered may be represented graphically as in Fig. 5.

Let the dotted lines in Fig. 5 represent the axes of the base place and the solid lines represent the axes of the translated place. The origin of this place, which originally coincided with the origin O of the base place is, after finished translation, located in some point R. This point R may be in either of the four quadrants around O. The relative positions of O and R give an expression for the magnitudes of $\log p$ and $\log P$. In this connection it is most convenient to consider the location of the point O as a point in the reference system of the translated place R. *The two quantities*, $\log p$ *and* $\log P$, *are simply the ordinate and the abscissa of the base place origin* O *considered as a point in the reference system of the translated place* R. In the example in Fig. 5, the situation is such that both $\log p$ and $\log P$ are positive.

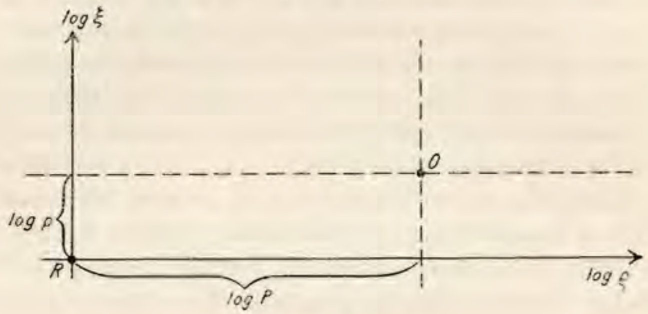

Fig. 5.

Thus we see that the problem before us can be reformulated in the following terms: The food expenditure curves (in log scales) for two or more budget places are known. By convention we may consider one of these curves as a food *quantity* curve. What are the components ($-\log P$, $-\log p$) of the translation which we must perform on the other expenditure curves in order to transform them into quantity curves? We now proceed to a discussion of this problem.

The first latitude principle.

If we have given several expenditure curves, each curve referring to a certain place (that is, a certain price situation), there are certain intrinsic properties of these curves which reveal something about the translation which the curves must undergo in order to be changed into quantity curves. One such thing is, for instance, that the curves (after finished translation) *must have no intersection points*. This simply follows from the fact that the quantity curves are contour lines of the

surface of consumption. This property gives valuable information, but it is, of course, in itself not sufficient to determine uniquely the translation which each of the expenditure curves must undergo. A further consideration of the properties of the surface of consumption permits, however, of determining other intrinsic properties in the food quantity curves, which makes it possible to determine the translations practically uniquely.

In order to exhibit these properties we shall use a graphical representation as in Fig. 6. In Fig. 6 R is the origin of the trans-

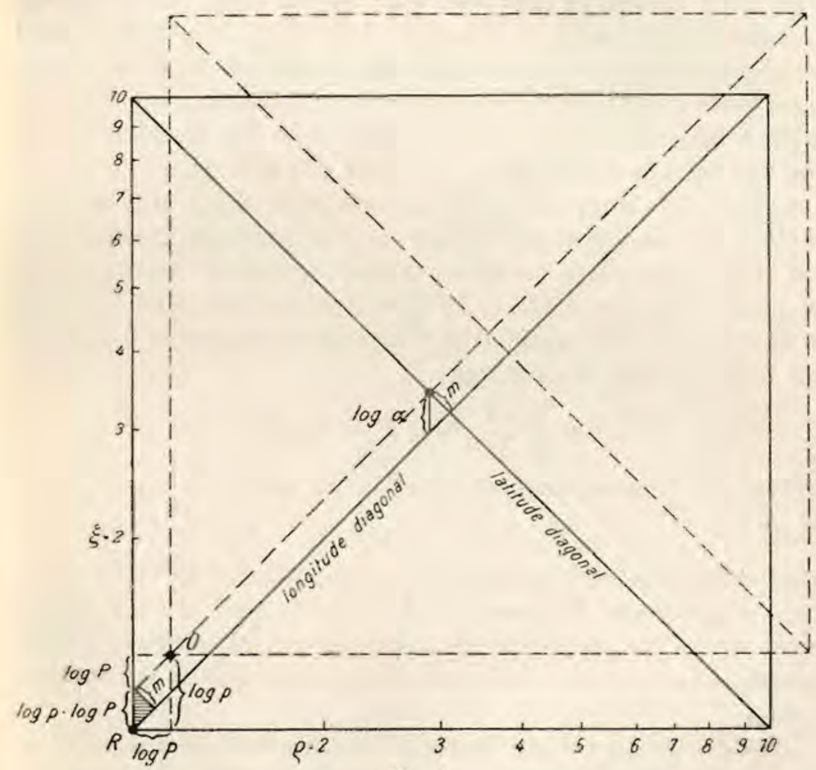

Fig. 6.

lated place (i. e. the point R represents the particular expenditure $\log \xi = 0$ and the particular income $\log \varrho = 0$ in this budget place). Instead of simply drawing one horizontal and one vertical axis we draw as indicated in Fig. 6 a whole quadrangle for the place considered and also draw the two diagonals of this quadrangle. These are the solid lines in Fig. 6. The diagonal going from the lower left hand corner to the upper right hand corner we shall call the longitude diagonal and the diagonal going from the upper left hand corner to the lower right hand corner we shall call the latitude diagonal. This form of the reference system will help in steering the charts when the translations are performed. The chart of the translated place, that is, the chart whose origin is the point R in Fig. 6, will for shortness be called the chart R.

In Fig. 6 we have indicated both the reference system of the chart R and the reference system of the base place chart. The latter system is indicated with dotted lines. After finished translation the longitude diagonal of the chart R will in general be displaced as compared with the longitude diagonal of the base place chart. This displacement can be measured along the latitude diagonal of the chart R. In Fig. 6 the displacement is marked m. The length of m we shall call the latitude of the chart R as compared with the base place chart. It is counted positive or negative accordingly as the longitude diagonal of the translated place is shown above or below the longitude diagonal of the base place. In the situation exhibited in Fig. 6 m is negative. By considering the shaded triangle in Fig. 6 we see that we have the relation

$$m = \frac{1}{\sqrt{2}} (\log P - \log p)$$

where $\sqrt{2}$ is taken positive. That is, we have

$$\log a = \sqrt{2}\, m \qquad (6.1)$$

This shows that the latitude m is proportional to the altitude difference between the base place and the translated place which we would get by representing the surface of consumption in a logarithmic system and with $\log a$ as the vertical axis.

In particular we see that any displacement of the chart R along its longitude diagonal means a simultaneous and proportional increase (or decrease) in both food price and living price, that is, a displacement such that the relation between food price and living price is kept constant. On the contrary, a displacement perpendicular to the longitude diagonal means a change in the ratio between food price and living price. More exactly stated, the change which takes place in this ratio is measured by the length m.

We thus see that there exists a simple and definite relation between the "water level" distance which the chart R has travelled during the translation and the "altitude" a connected with the final position of the chart. By virtue of this property we shall, after finished translation, have a situation like the one in Fig. 7.

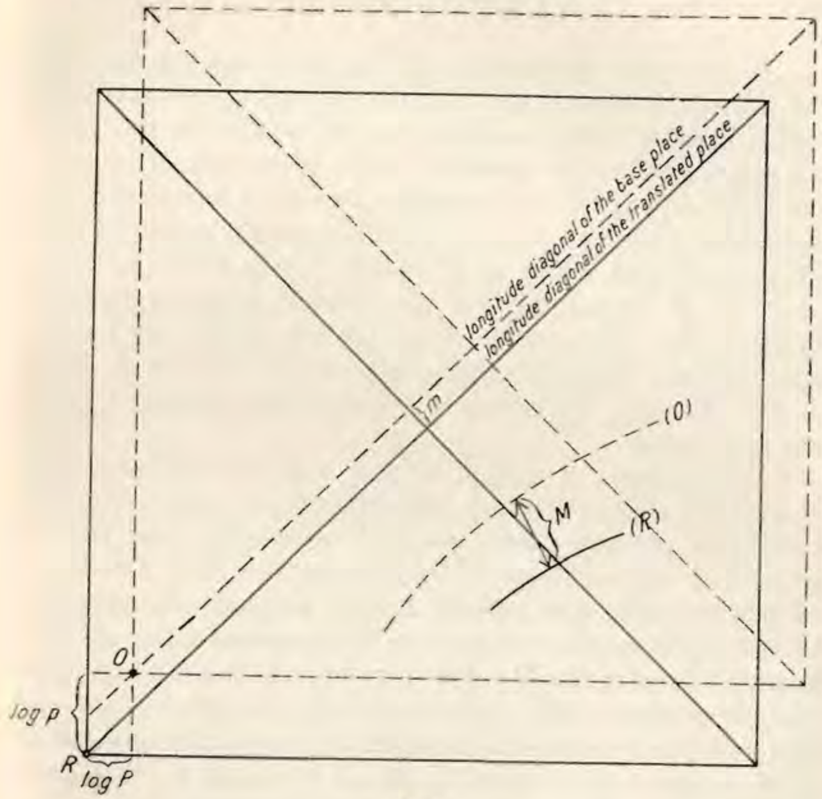

Fig. 7.

Fig. 7 is a picture of an actual situation found in the budget material studied. It shows the city Detroit (R in Fig. 7), as compared with the city San Francisco, the latter serving as the base place. According to the principle here considered, *if the longitude diagonal of the place R is below the longitude diagonal of the base place, then the food curve of the place R should also be below the food curve of the base place*, and vice versa. In fact, as we have seen, a negative m means a negative log a, that is, the relative food price is higher in the place R than in the base place. And a higher relative food price must give a lower food quantity curve and vice versa. This we shall call the first latitude principle.

The second latitude principle.

If the curves (O) and (R) in Fig. 7 are not too far apart, the altitude difference log a can as a first approximation be looked upon as being measured by how far the two curves are apart. Let us, for instance, define a measure of "how far the curves are apart" by drawing a free hand line roughly perpendicular to the two curves somewhere near the middle of them, for instance such a line as M in Fig. 7, and let us judge the length of M by the eye. For the moment, it is not necessary to define the "distance" between the two curves (O) and (R) more precisely than this. We shall refer to this rough notion of the "distance" by calling it the estimated distance between the two curves.

Introducing this notion, we can state the following rule: If we have three (or more) food expenditure curves, then the translation of these curves should be such that after finished translation the estimated distances between the curves are roughly speaking proportional to the perpendicular distances between their longitude diagonals. In other words, the m's shall be proportional to the M's. This principle we shall call the second latitude principle. The two first latitude principles in conjunction with the criterion that the two food quantity curves shall never intersect, gives already a means of making a rough guess at what the translation shall be. The procedure will be this: One of the places is selected as the base place. Then the other places are so arranged that the order of succession between the places becomes the same whether it is judged from their longitude diagonals or from the food curves themselves, and furthermore, such that none of the curves intersect any of the others. When this is done the translation is further perfected in such a way that the distances m between the longitude diagonals become roughly speaking proportional to the estimated distances M between the curves themselves.

The third latitude principle.

The translation can now be further corrected in the following way: Suppose we have three translated food curves (A), (B) and (C), as indicated in Fig. 8.

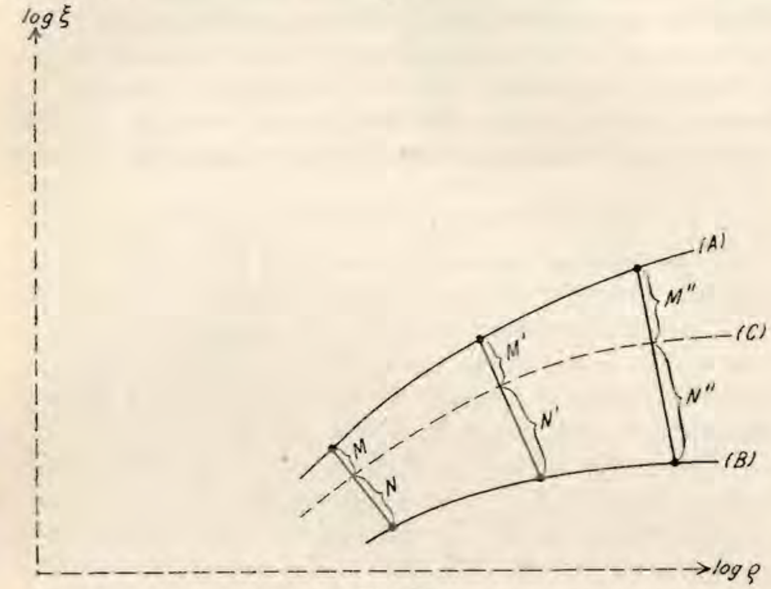

Fig. 8.

After the translation according to the first and second latitude principle has been made, we draw a series of normals to that one of the three curves that occupies a position between the two other curves. In Fig. 8 (C) is the central curve and there are indicated three normals to this curve. Let M and N, M' and N', M'' and N'' be the sections of these normals which are lying between (A) and (C), and between (C) and (B) respectively. Any of the ratios M/N, M'/N', and M''/N'' might be considered as an approximate expression for the ratio between the altitude of (A) and the altitude of (B), both altitudes measured from the place (C). That means that all the ratios M/N, M'/N', etc., should be approximately equal. And that is not all. The common magnitude of this ratio ought to be equal to the ratio between the latitude of A (measured from the longitude diagonal of the central place) and the latitude of B (measured from the central place). This principle for judging if the translations performed are correct, will be called the third latitude principle.

In practice, the food curves have turned out to be significantly different from straight lines under 45^0 angle. The principle here discussed is therefore a very powerful tool in determining the translations that the curves shall undergo. In most practical cases it will probably not be found necessary to carry the accuracy any further than this. In point of principle it is, however, possible to go still further, and even, theoretically, determine the translations exactly by the following principle.

The flexibility principle.

Let us consider the food curve for a given city (A), and compare it with the food curve for the base place (O). See Fig. 9. After finished translation we shall now consider both (A) and (O) as curves drawn in the same reference system, namely, the base place reference system. We shall further look upon the horizontal axis of this system as the axis for deflated income r (or rather for $\log r$) and the vertical axis as the axis for the food quantity x (or rather for $\log x$). The horizontal and vertical axes in the (A) system only have a meaning for the (A) curve, and for this curve they represent respectively nominal dollar income and nominal dollar expenditure for food.

Let m in Fig. 9 be the latitude of the chart R. The vertical line δ in Fig. 9 will then by (6. 1) be equal to $\log a$. Further consider any horizontal line A between the base place curve and the (A) curve. The ratio δ/A can then be taken as an expression for the average flexibility of the marginal utility curve of money over the income interval covered by the line A. We may express this fact by the formula

$$\check{w}(A) = \frac{\delta}{A} \qquad (6.\,2)$$

This simply follows from the theory of the quantity variation method. See in particular (5. 7). Formula (6. 2) gives a very convenient means of determining the money flexibility by readings from a graph such as Fig. 9. If we want to take account of the signs of the magnitudes δ and A that are read off from the graph, we must count the line A as the abscissa of its end point

on (A) minus the abscissa of its end point on (O). And correspondingly we must count δ as the ordinate of its end point on the longitude diagonal for (A) minus the ordinate of its end point on the longitude diagonal for (O). This will always give a negative value for $\check{w}(A)$ provided the first latitude principle holds good. Therefore, in practice we need not take account of

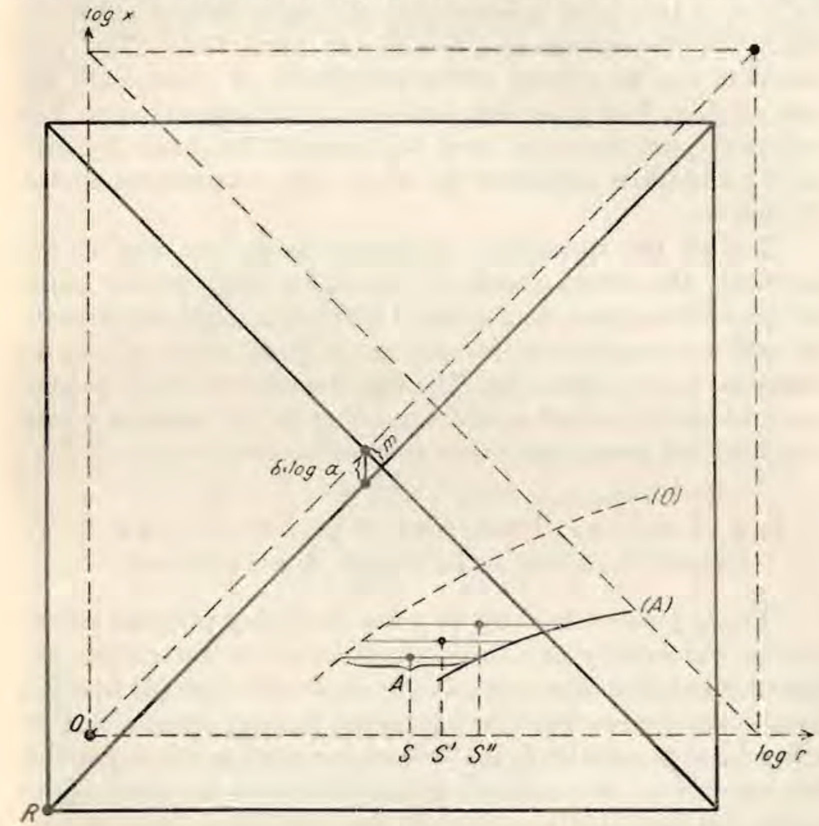

Fig. 9.

the sign but measure δ and A as pure lengths, and take the ratio between these pure lengths as an expression for $(-\check{w})$.

If the length A is not too large we may take the ratio δ/A as an expression for the point money flexibility in the midpoint S of A. Similarly, the magnitude of the money flexibility in the points S' and S'' of Fig. 9 would be approximately δ/A' and δ/A'' respectively, where A' and A'' stands for the horizontals between (O) and (A), the mid-point abscissae of which are respectively S' and S''. This procedure permits us to draw the money flexibility $\check{w}(r)$ as a function of r.

This being so, we see that if we have t w o food curves (A) and (B) besides the base place curve (O) we are able to draw t w o curves representing the money flexibility as a function of r. T h e s e t w o c u r v e s o u g h t t o b e i d e n t i c a l, and this gives another and still more refined condition which the translations of (A) and (B) must fulfill. This last condition can be utilized either graphically or numerically by least squares. The numerical procedure is considered below. The graphical procedure can most conveniently be made by free hand translations supported by slide rule computation of the flexibilities.

For all the translation work considered here and in the foregoing, the curves should be drawn on light tracing paper and placed on a glass tracing board with strong light underneath. By such an arrangement several curves (from 10 to 15) can be compared at the same time. The simultaneous free hand adjustment of such a set of curves according to the principles here discussed has proved to give a very satisfactory result.

A l e a s t s q u a r e m e t h o d b y w h i c h t h e f r e e h a n d t r a n s l a t i o n c a n b e r e f i n e d.

I now proceed to show how the flexibility principle can be utilized numerically as a final refinement of the translation. We shall assume that there are given two food curves (A) and (B) besides the food curve of the base place. Further we assume that a free hand translation is made so as to adjust as far as possible the position of the curves in accordance with the above principles. Let the situation after the free hand translation be, for instance, as in Fig. 10.

Then draw a curve (a) which is half way, horizontally measured, between (A) and the base curve. More precisely stated, the curve (a) is such that if we draw any horizontal line (that is, a line parallel to the log r axis) the section of this line which lies between (A) and (a) is equal to the section which lies between (a) and the base curve. Similarly, draw a curve (b) which is half way, horizontally measured, between the base curve and (B).

Next, draw a number of v e r t i c a l lines between (a) and (b). In Fig. 10 there are indicated three such verticals, Nos. 1, 2, 3. The upper end points of these verticals are on (a) and the lower end points on (b). Consider particularly number 1 of these verticals. From the upper end point of the vertical number 1 draw a horizontal line between (A) and the base curve. From the lower end point of vertical number 1 draw a horizontal line

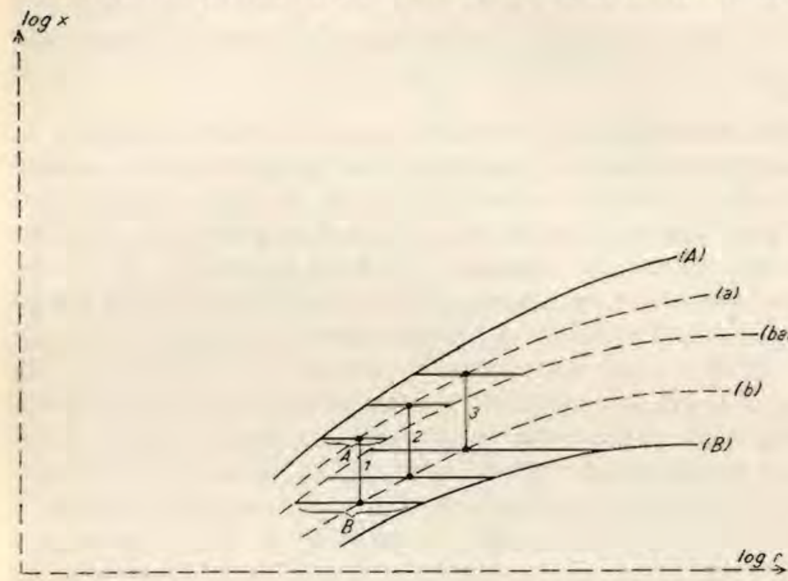

Fig. 10.

between the base curve and (B). The two horizontals thus drawn should be looked upon as forming a pair of horizontals connected through the vertical number 1. In the same way, draw the pair of horizontals connected through the vertical number 2, etc. In this way we obtain two series of corresponding quantities, namely, the lengths A and B of the upper and lower horizontals, respectively.

Now, consider the two quantities

$$A' = \frac{A}{\log p - \log P} \qquad B' = \frac{B}{\log q - \log Q}$$

where log p, log P, log q, log Q are the components of the free hand translation that h a s b e e n performed on (A) and (B) respectively in order to bring these curves to the position exhibited in Fig. 10. Since the money flexibility curves derived by pairing (A) with the base curve and by pairing (B) with the base curve ought to be identical, the two quantities A' and B' should be e q u a l, identically in $\log r$, that is, identically in all the abscissa points defined by the verticals 1, 2, 3 . . . of Fig. 10. As a matter of fact, they may not be rigorously equal everywhere, and part of this discrepancy might be looked upon as due to the fact that the translation has not been performed perfectly.

The idea therefore presents itself to extend the summation

$$f = \Sigma(A' - B')^2 \tag{6. 3}$$

over a set of abscissa points (that is over a set of verticals 1, 2, 3 . . .) and minimize this expression by considering it as a function of the four parameters, $\log p$, $\log P$, $\log q$, $\log Q$. Since we may assume that we know these four parameters approximately, namely, through the free hand translation, it will be most convenient to determine their final values by a set of corrections. The following procedure might be used.

First determine numerically the partial derivatives of f with respect to the four parameters, in the vicinity of the approximate magnitudes of the parameters which have been adopted. Then make a series of tentative displacements, all going in the negative direction of the gradient of f, and compute in each such point the magnitude of f as defined by (6. 3). Plott f as a function of the length of the displacement and determine graphically that length of the displacement which makes f the smallest. Consider the point thus determined as a corrected minimum point for f.

In practice the work might be arranged by the following scheme:

Observation number	A	B	A'	B'	$A' - B'$	$(A' - B')^2$
1						
2						
:						
Sum			$\Sigma A'$	$\Sigma B'$	$\Sigma(A' - B')$	$\Sigma(A' - B')^2$

In order to check the computation all the four sums indicated should be computed. When $f = \Sigma(A' - B')^2$ is computed, choose a number h_p which is small in comparison to the displacement og p that was used in the previously performed freehand trans-

lation. Now take the curve (A) in the position which it has after the free-hand translation, and displace it v e r t i c a l l y by the amount h_p. In other words, consider the curve $(\bar{A})$ which is obtained by letting the vertical displacement of the original expenditure curve for the locality (A) be equal to $\log p + h_p$ instead of $\log p$. But let the horizontal displacement of this curve still be equal to $\log P$, and let the curve (B) still have the displacements, $\log q$ and $\log Q$ respectively. In the situation thus obtained, repeat the computation of the magnitude defined by formula (6. 3). Let the magnitude (6. 3) in the new situation be f_p.

Next put the curve (A) back in the position determined by the freehand translation. And give it now a h o r i z o n t a l displacement by the amount h_P. In other words, consider the curve $(\bar{\bar{A}})$ whose vertical and horizontal displacements are $\log p$ and $\log P + h_P$ respectively. And match it with the curve (B) whose vertical and horizontal displacements are still $\log q$ and $\log Q$ respectively. Let the magnitude (6. 3) in this situation be equal to f_P.

Next put the curve (A) back to the position ($\log p$, $\log P$) and consider the position ($\log q + h_q$, $\log Q$) of the curve (B), h_q being a small vertical displacement. Let the magnitude (6. 3) in this situation be f_q. Finally, let f_Q be the magnitude of the expression (6. 3) when (A) is in the position ($\log p$, $\log P$) and (B) in the position ($\log q$, $\log Q + h_Q$), h_Q being a small horizontal displacement.

When the four magnitudes f_p, f_P, f_q, f_Q are thus determined, consider the composite displacement

$$\log p + H_p,\ \log P + H_P,\ \log q + H_q,\ \log Q + H_Q \tag{6. 4}$$

where

$$\begin{aligned} H_p &= h \cdot (f_p - f)/h_p \\ H_P &= h \cdot (f_P - f)/h_P \\ H_q &= h \cdot (f_q - f)/h_q \\ H_Q &= h \cdot (f_Q - f)/h_Q \end{aligned} \tag{6. 5}$$

and h is a small negative quantity. The quantity f defined by (6. 3) should be computed for a set of different magnitudes of h (and with the fixed magnitudes of f_p, f_P . . . h_p, h_P . . .), the corresponding one dimensional curve with abscissa h and ordinate f plotted, and that magnitude of h selected which on this curve gives the least f. The translation point (6. 4) for this magnitude of h may be considered as defining a corrected translation. If necessary this corrected point may be taken as a new starting point, the whole process repeated and thus a still more correct minimum point for f determined.

It should be noticed that the above application of the least square principle does not involve the use of any preconceived formula with parameters to be determined. Furthermore the process can of course be applied not only to the present translation problem but to any least square minimalization problem in several variable.

Summary of the technique involved in the translation method.

(1) Reduce the income and food expenditure figures to a member basis, as explained in Section 7. Plot the food-expenditure curve for each place using logarithmic scale both on the income axis and the food axis.

(2) Smooth each of these curves independently. Graphical smoothing can generally be used to greatest advantage. Avoid analytical smoothing based on specific formulae with parameters to be determined.

(3) Copy the smoothed curves on light tracing paper, one curve on each sheet. Mark the reference system on each sheet by drawing not only the two axes but a whole square with its two diagonals, as shown in Fig. 6. This will help in steering the charts under the free hand translation.

(4) Choose one of the places as a base place. Trace the reference system of this place (including the two diagonals) with heavy lines. Mark the base place origin with a heavy dot. Fix this chart on a glass drawing board with strong light underneath.

(5) Take the other places and try to fit them on to the base place by free hand translation vertically and horizontally, using the three latitude principles and the criterion of no intersection, as explained above. An attempt may also be made to improve on the adjustment by a free hand use of the flexibility principle. As new places are taken into account adjust if necessary the translations of the places that were treated first. Something between 10 and 15 places can be adjusted simultaneously

this way, if so many budget places are available. On the sheet of each place that is finally adjusted, mark the point where the base place origin shows through the tracing paper. The coordinates of this point measured in the reference system of the place considered, are the coordinates of the log of the food price and the log of the living price in the place in question.

(6) When all adjustments are made, the curves should be fixed together so as to be kept in a constant position. The system of curves thus obtained may be called the composite food curve chart. It is utilized for the construction of money flexibility curves. Each set of two places gives one money flexibility curve, or a part of such a curve. The flexibility curves should have a log scale along the horizontal axis, that is, along the axis of r, but should have an ordinary scale along the vertical axis, that is along the axis of the flexibility. Each such flexibility curve is derived by measuring on the final composite food curve chart the horizontal distance A between two food curves and also measuring the vertical distance δ between the corresponding two longitude diagonals as indicated in Fig. 9 and formula (6. 2). The ratio δ/A gives the absolute magnitude of the flexibility. The measurement of δ and A on the composite chart can be made in centimeters, inches, or any other conventional unit. Tentative flexibility computations of this kind may be used already during the translation (see under "The flexibility principle" above).

(7) When the flexibility curves are constructed, put a new sheet of tracing paper over the composite chart and copy the base place reference system and all the food curves which have been used and which are showing on the composite chart. This new chart gives a picture of the surface of consumption, the food curves being the contour lines in (log x, log r) coordinates.

(8) Since the parts of the money flexibility curve obtained from the various pairs of places used, will as a rule not coincide exactly, a compromise must be made, preferably by graphical smoothing. If an analytical smoothing is wanted, the formula

$$(6.6) \qquad -\breve{w}(r) = \frac{\gamma}{\log r - c}$$

might be tried where c and γ are constants determined by

$$(6.7) \qquad c = \frac{\sigma_{(-\breve{w})\log r}}{\sigma_{(-\breve{w})}}$$

$$\gamma = M_{(-\breve{w})\log r} - cM_{(-\breve{w})} \tag{6. 8}$$

In (6. 7) and (6. 8) $(-\breve{w})$ denote the ordinate and $\log r$ the abscissa of the plotted observation points. $M_{(-\breve{w})\log r}$ and $\sigma_{(-\breve{w})\log r}$ is respectively the mean and the standard deviation of the product $(-\breve{w})\log r$ taken over all the observation points. And $M_{(-\breve{w})}$ and $\sigma_{(-\breve{w})}$ is respectively the mean and standard deviation of $(-\breve{w})$. The log in formulae (6. 6) and (6. 7) may be taken either as natural log or as the log in any other system.

From the graphically smoothed curve of the flexibility the marginal money utility itself may be determined by numerical integration, using the formula (5. 8). If the flexibility curve has been smoothed by formula (6. 6), the determination of the marginal utility w itself does not necessitate any numerical integration. In this case we simply have

$$w(r) = \frac{b}{(\log r - c)^{\gamma}} \tag{6. 9}$$

where c and γ are determined by formulae (6. 7) and (6. 8), b is an arbitrary constant, and log stands for the same sort of log as was used in (6. 6).

7. APPLICATION OF THE TRANSLATION METHOD TO THE U. S. BUDGET MATERIAL FOR 1918—19.

The United States Bureau of Labor Statistics made a Budget Study in 1918—19 including ninety-two cities. The results are published in the Bureau of Labor Statistics Bulletin No. 357. For the present study the following thirteen cities were selected: New York City; Detroit, Michigan; Houston, Texas; San Francisco, Calif.; Minneapolis, Minn.; Boston, Mass.; Buffalo, N. Y.; New Orleans, La.; Portland, Me.; Baltimore, Md.; Savannah, Ga.; Chicago, Ill.; Cleveland, Ohio.

For each of these cities the food expenditure curve was plotted using logarithmic scale along both the income axis and the food expenditure axis. The income is not given directly in the tables published by the Bureau of Labor Statistics, but the total expenditure is given and also the surplus or deficit so that the income can easily be computed.

The data immediately given in the tables are the family expenditure for food, ξ_F in each of seven groups of the family income ϱ_F. For each group is also given the magnitude of the coefficient of equivalent adult males E. For the present purpose both the income figures and the food consumption figures were reduced from a family basis to an individual basis. This was done by dividing both the family expenditure for food and the family income with the coefficient E. The ratios thus obtained $\xi = \xi_F/E$ and $\varrho = \varrho_F/E$ were taken as expressing the member expenditure for food and the member income respectively.

For each city were plotted both the curve showing the relation between family expenditure and family income, and the curve showing the relation between ξ and ϱ. The family expenditure curves showed a much greater tendency towards linearity than the member expenditure curve. This fact has a very simple explanation. It simply shows the proportionality effect of the family size as measured by E. The nature of this effect is best exhibited by an extreme example. Suppose that the whole population consisted of a great number of identical members, each of whom had a given income, say $500 a year, and spent a given sum, say $200 a year, for food. Suppose that from this population we constructed a great number of families by grouping together a certain number of these identical members. The number of members in a family might be determined by random sampling. If we would investigate the relation between family food expenditure ξ_F and family income ϱ_F in the set of families thus constructed we would find a perfect linear correlation. And the correlation would have been nearly perfect even if there had been some spread between ξ and ϱ. This fact would, however, have no significance whatsoever for the problem of utility measurement. The important thing in the utility measurement is to know the relation between ξ and ϱ, not the relation between ξ_F and ϱ_F.

For each of the thirteen cities studied, the member expenditure curve was smoothed graphically. In the smoothing process each of the member expenditure curves was treated independently, no comparison being made between the forms of the various curves.

From the set of thirteen expenditure curves thus obtained, the following nine curves were selected for a further analysis: San Francisco, Calif.; Minneapolis, Minn.; Houston, Texas; New York City; Savannah, Ga.; New Orleans, La.; Buffalo, N. Y.; Boston, Mass.; Detroit, Mich.

Each of these food expenditure curves was traced on a separate sheet of light tracing paper. The reference system was drawn as indicated in Fig. 6. San Francisco was chosen as a base place and a free-hand translation of the other eight curves was performed, using the principles explained in Section 6. This translation gave the following result (Table 3 p. 61) regarding the food price and the living price in the cities considered.

The situation obtained after finished translation is pictured in Fig. 11. The curves thus obtained have the significance of food-q u a n t i t y curves.

Each of the eight quantity curves Nos. 1, 2, ... 8 thus obtained was then paired with the base place curve 0. Since the curves 3 and 8 turned out to be identical (the only difference being that curve 8 projected a little further to the right than

Table 3.
Year: 1918—19

No.	City	Food price p	Living price P	Inverted food price $\alpha = \frac{P}{p}$
0	San Francisco, Calif.	1.00	1.00	1.00
1	Minneapolis, Minn.	1.00	0.86	0.86
2	Houston, Texas	1.10	0.99	0.90
3	New York City	1.56	1.24	0.80
4	Savannah, Ga.	0.66	0.72	1.09
5	New Orleans, La.	0.78	0.89	1.14
6	Buffalo, N. Y.	1.24	1.04	0.84
7	Boston, Mass.	1.63	1.20	0.74
8	Detroit, Mich.	1.35	1.07	0.80

the curve 3) there was in this way obtained 7 pairs of curves. From each of these pairs a number of points on the flexibility

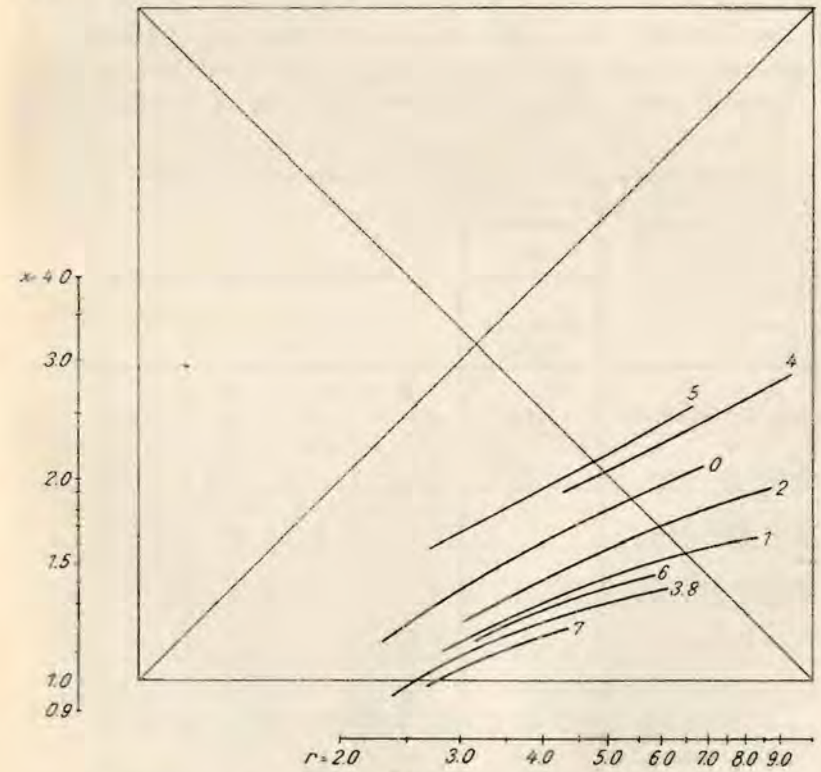

Fig. 11.
Contour Lines of the Surface of Consumption when α is „altitude".

curve of money were determined. Each such point was determined by measuring in centimeters the two lenghts δ and A, forming the ratio δ/A and considering this ratio as the magnitude of the flexibility ($-\check{w}$) in the middle of the real income interval covered by the length A. The number of points on the flexibility curve which were thus determined from each pair of food quantity curves ranged from 1 to 3. In all seventeen points on the flexibility curve were determined. They are given in Tab. 4.

Plotting these seventeen points in a diagram with log r along the horizontal axis and $-\check{w}$ along the vertical axis the chart of Fig. 12 was obtained.

The heavy line in Fig. 12 is a free-hand smoothed curve. Light lines in the chart connect observation points which correspond to one and the same pair of food quantity curves. It is remarkable that not only does the group of the seventeen observation points taken as a whole indicate distinctly the course of the flexibility curve, but each of the point sets relating to a given pair of food quantity curves shows the same characteristic features in the variation of the absolute value of the flexibility,

Table 4.

Pairs of food quantity curves o San Francisco matched with:	log $a = \delta$ measured in centimeters (Compare Fig. 9)	Individual points measured				
1 Minneapolis, Minn.	0.85	$r =$	3.3	4.3	5.7	
		$-\check{w} =$	.425	.34	.213	
2 Houston, Texas ..	0.55	$r =$	2.88	4.28	6.2	
		$-\check{w} =$	.523	.365	.275	
3 New York City ..	1.15	$r =$	2.8	3.35	4.05	
		$-\check{w} =$	.575	.48	.329	
4 Savannah, Ga. ...	— 0.55	$r =$			4.7	5.7
		$-\check{w} =$			0.407	.305
5 New Orleans, La..	— 0.77	$r =$			3.1	5.5
		$-\check{w} =$			.428	0.35
6 Buffalo, N. Y. ...	0.9	$r =$		2.62	3.75	4.35
		$-\check{w} =$		.528	.36	.29
7 Boston, Mass.	1.65	$r =$		3.25		
		$-\check{w} =$		.55		
8 Detroit, Mich.....	Same as 3, New York City					

namely a monotonic decrease from about 0.6 in the beginning to 0.3 at the end of the real income interval covered.

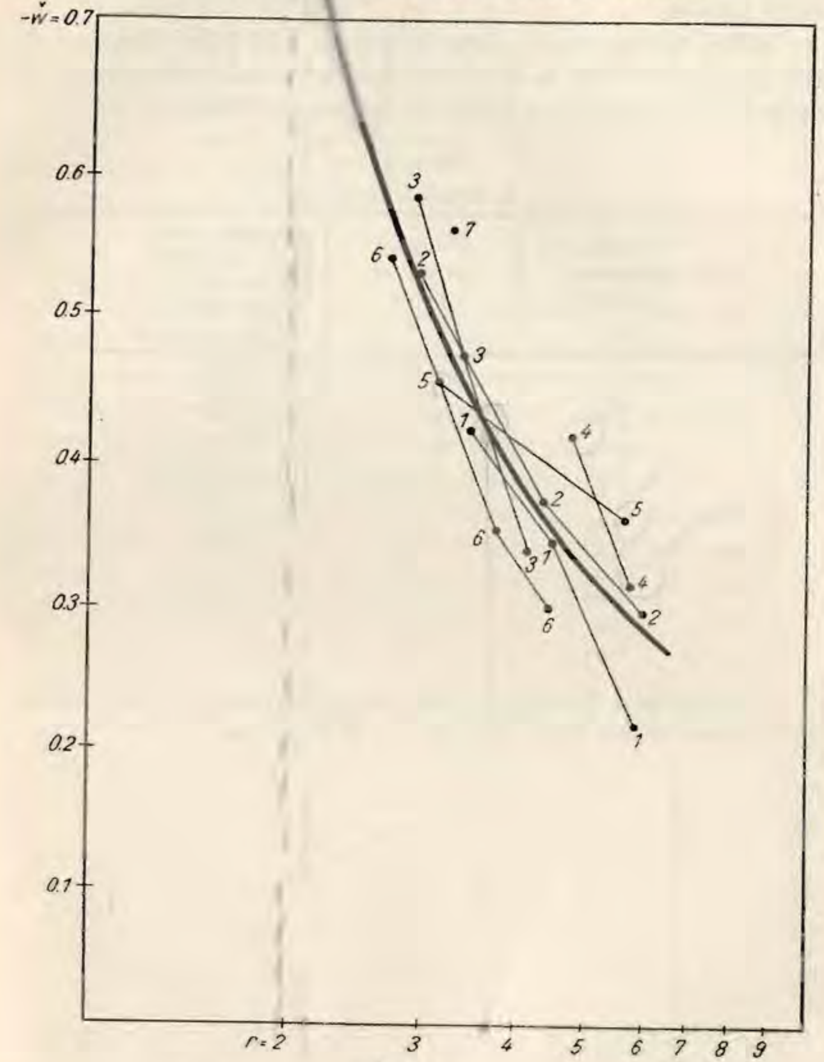

Fig. 12.
Flexibility of the Marginal Utility of Money
w = marginal utility of money; r = real (deflated) income;
$$-\check{w} = -\frac{d \log w}{d \log r} = \text{absolute value of the money flexibility.}$$

It should be remembered that the curve of Fig. 12 is the curve of the f l e x i b i l i t y of the marginal utility of money.

Since the marginal utility curve itself is the integral of the flexibility curve, its relative accuracy will, of course, be still much higher.

After the flexibility curve in Fig. 12 had been constructed, the marginal utility w as a function of r was determined by numerical integration. The result is given in Table 5.

Table 5.
U. S. Material, 1918—19

Real (deflated) income r	Marginal utility of money w	Absolute value of the flexibility of the money utility curve $= -\breve{w}$
2.40	10.00	.617
2.62	9.50	.559
2.90	9.03	.510
3.17	8.63	.467
3.48	8.30	.428
3.80	7.98	.396
4.16	7.72	.362
4.55	7.48	.333
5.00	7.26	.312
5.40	7.07	.294
5.91	6.89	.278
6.50	6.72	.261

A graphical picture of the marginal money utility curve as determined by Table 5 is given in Fig. 13.

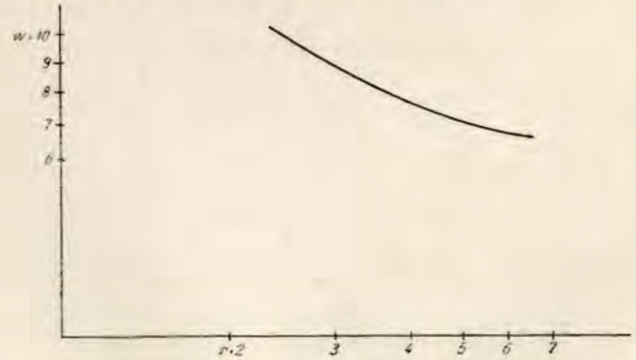

Fig. 13
Marginal Utility of Money as a Function of the Income
The curve is determined by numerical integration of the freehand smoothed curve of the flexibility
w = marginal utility of money
r = real (deflated) income

According to the numerical results here obtained, **the absolute value of the money flexibility is less than unity over the entire real income interval considered.** And the spread of the observation points (see Fig. 12) is so small that there can be no doubt about the significance of the fact that the flexibility is less than unity. This confirms the hypothesis which I made in 1926 in Statsökonomisk Tidsskrift.

The fact that we have obtained practically the same flexibility determination by pairing *different* cities seems to me to give a high degree of significance to the whole investigation. Even the most hard boiled anti-utilitist must at least admit, I believe, that we are here in presence of an empirical fact in mass behavior that is of considerable importance and demands an explanation. Conceivably he may, perhaps, still refuse to explain this uniformity in terms of utility equilibrium. But this refusal would, it seems, be of little weight as long as no other plausible explanation is offered. And I believe we shall most likely have to wait very long for another interpretation that can compete with the utility interpretation in simplicity and generality.

8. THE GENERAL FLEXIBILITY EQUATION.

The methods of Sections 5 and 6 are built on the fact that two observation points (a_1, x_1, r_1) and (a_2, x_2, r_2) which show the s a m e food quantity, that is, $x_1 = x_2$, and d i f f e r e n t relative prices, that is $a_1 \neq a_2$, may be utilized to obtain one observation of the money flexibility. The question then naturally arises: If the material at our disposal is in such a form that we do not get observations showing rigorously the same food quantity would it not be possible even then to obtain an approximate expression for the money flexibility, for instance, by utilizing in each flexibility determination three or more observation points instead of two, and through this additional information partly overcome the lack of observation-pairs with the same food quantity? This problem will be discussed in the present Section.

For convenience we shall now write the equilibrium equation in the following logarithmic form

$$\log w(r) - \log u(x) = \log a \qquad (8.\ 1)$$

We assume that there is given a set of observation points on the surface of consumption: (a_1, x_1, r_1), (a_2, x_2, r_2), ... etc. For each of these points the equilibrium equation (8. 1) must be fulfilled. If we select any two of the observation points, for instance the points Nos. (1) and (2), and subtract the equilibrium equation in the second point from the equation in the first point, we get

$$(\log r_1 - \log r_2) \cdot \breve{w}(r_1, r_2) - (\log x_1 - \log x_2) \cdot \breve{u}(x_1, x_2) = \log a_1 - \log a_2 \qquad (8.\ 2)$$

where $\breve{w}(r_1, r_2)$ defined by (5. 6) is the average money flexibility over the income interval (r_1, r_2), and

$$\breve{u}(x_1, x_2) = \frac{\log u(x_1) - \log u(x_2)}{\log x_1 - \log x_2} \qquad (8.\ 3)$$

is the average food flexibility over the quantity interval (x_1, x_2).

Equation (8. 2) will be called t h e g e n e r a l f l e x i b i l i t y e q u a t i o n. By putting $x_1 = x_2$ in this equation

we get immediately the formula (5. 7) that served as the basis for the quantity variation method. Similarly, if we have two complete budget curves Nos. (1) and (2), and put $r_1 = r_2$ in (8. 2), we get

$$\breve{u}(x_1, x_2) = -\frac{\log a_1 - \log a_2}{\log x_1(r) - \log x_2(r)} \qquad (8.\ 4)$$

where r is the common magnitude of r_1 and r_2, and $x_1(r)$, $x_2(r)$ are the two functions that represent how the quantity consumed varies with the income r on the first and second budget curve respectively. By attributing to r in (8. 4) different magnitudes, we get different points on the food flexibility curve. And from the food flexibility thus determined, the food utility curve may of course be derived by integration, in the same way as the money utility curve was derived from the money flexibility curve by (5. 8). This method of determining the food flexibility and the food utility may be called the i n c o m e - v a r i a t i o n m e t h o d.

In order to study the situation when the quantities x and the income r in the material at hand are all different, we shall introduce the following notation

$$\begin{aligned} \breve{w}_{12} &= \breve{w}(r_1, r_2) \\ \breve{u}_{12} &= \breve{u}(x_1, x_2) \\ a_{12} &= \log a_1 - \log a_2 \\ x_{12} &= \log x_1 - \log x_2 \\ r_{12} &= \log r_1 - \log r_2 \end{aligned} \qquad (8.\ 5)$$

and similarly if 1 and 2 are replaced by any other subscripts: 3, 4, ... etc. Obviously $x_{21} = -x_{12}$, $r_{21} = -r_{21}$ and if $x_1 = x_2$ we have $x_{12} = 0$, if $r_1 = r_2$ we have $r_{12} = 0$, and so on.

With this notation the general flexibility equation can be written

$$r_{12}\breve{w}_{12} - x_{12}\breve{u}_{12} = a_{12} \qquad (8.\ 6)$$

The average money flexibility $\breve{w}(r_1, r_2)$ as defined by (5. 6) may be looked upon as the first divided difference of the function $\log w(r)$ with respect to $\log r$. Similarly we may define the second order divided difference of $\log w(r)$ by [1])

[1] The second order divided differences are usually defined without the factor 2 introduced in the right member of (8. 7). I have used the definition (8. 7) in order to make the second difference directly comparable to the second derivative defined by (8. 10).

$$(8.\ 7)\quad \check{\check{w}}(r_1, r_2, r_3) = \frac{\check{w}(r_1, r_2) - \check{w}(r_2, r_3)}{\frac{\log r_1 + \log r_2}{2} - \frac{\log r_2 + \log r_3}{2}} = 2\,\frac{\check{w}(r_1, r_2) - \check{w}(r_2, r_3)}{\log r_1 - \log r_3}.$$

Denoting for shortness this second order difference by $\check{\check{w}}_{123}$, we can write (8. 7) in the form

$$(8.\ 8)\qquad \check{\check{w}}_{123} = 2\,\frac{\check{w}_{12} - \check{w}_{23}}{r_{13}}$$

Similarly we may consider the second order divided difference of the function $\log u(x)$, namely,

$$(8.\ 9)\qquad \check{\check{u}}_{123} = 2\,\frac{u_{12} - u_{23}}{x_{13}}$$

We shall call $\check{\check{w}}_{123}$ the average money a c c e l e r a t i o n taken over the income points (r_1, r_2, r_3). Similarly $\check{\check{u}}_{123}$ will be called the average food acceleration taken over the quantity points (x_1, x_2, x_3).

In conjunction with these average accelerations we shall also consider the p o i n t accelerations

$$(8.\ 10)\qquad \check{\check{w}}(r) = \frac{d\check{w}(r)}{d \log r} = \frac{d^2 \log w(r)}{(d \log r)^2}$$

and

$$(8.\ 11)\qquad \check{\check{u}}(x) = \frac{d\check{u}(x)}{d \log x} = \frac{d \log u(x)}{(d \log x)^2}$$

From the definition of the point accelerations it follows that these accelerations are expressions for h o w f a s t t h e f l e x i b i l i t i e s c h a n g e as we change the abscissa (income r or quantity x) in question. The average accelerations admit of a similar interpretation. It is indeed a well known proposition in the theory of divided differences, that if $\check{\check{w}}_{123}$ is the second order divided difference of $\log w(r)$ taken over the point triple (r_1, r_2, r_3), and if the point acceleration $\check{\check{w}}(r)$ is continuous in the interval between the smallest and the largest of the numbers r_1, r_2 and r_3, then there exists at least one value r_0 of r in this interval such that

$$(8.\ 12)\qquad \check{\check{w}}_{123} = \check{\check{w}}(r_0)$$

A similar proposition holds, of course, for $\check{\check{u}}_{123}$.

From (8. 12) and (8. 8) we infer that

$$(8.\ 13)\qquad \check{w}_{23} = \check{w}_{12} + {}^1\!/_2 r_{31} \check{\check{w}}(r_0)$$

Similarly we have

$$(8.\ 14)\qquad \check{u}_{23} = \check{u}_{12} + {}^1\!/_2 x_{31} \check{\check{u}}(x_0)$$

where x_0 is a magnitude of x in the interval between the smallest and the largest of the numbers x_1, x_2 and x_3.

Now let us consider the two pairs of observation points (1, 2) and (2, 3). From these two pairs of observations we deduce that $\check{w}_{12}$ and $\check{u}_{12}$ must satisfy the following two equations

$$(8.\ 15)\qquad \begin{aligned} r_{12}\check{w}_{12} - x_{12}\check{u}_{12} &= a_{12} \\ r_{23}\check{w}_{12} - x_{23}\check{u}_{12} &= a_{23} - R \end{aligned}$$

where

$$(8.\ 16)\qquad R = {}^1\!/_2[r_{23} r_{31} \check{\check{w}}(r_0) - x_{23} x_{31} \check{\check{u}}(x_0)]$$

Solving the two equations (8. 15) for $\check{w}_{12}$ and $\check{u}_{12}$ we get

$$(8.\ 17)\qquad \check{w}_{12} = \frac{a_{12}x_{23} - a_{23}x_{12} + x_{12}R}{r_{12}x_{23} - r_{23}\,x_{12}}$$

$$(8.\ 18)\qquad \check{u}_{12} = -\frac{r_{12}a_{23} - r_{23}a_{12} - r_{12}R.}{r_{12}x_{23} - r_{23}x_{12}}$$

If $x_1 = x_2$, (8. 17) reduces to $\check{w}_{12} = \frac{a_{12}}{r_{12}}$. And if $r_1 = r_2$ (8. 18) reduces to $\check{u}_{12} = -\frac{a_{12}}{x_{12}}$. Thus we get again back to the formulae on which the quantity-variation method and the income-variation method respectively was built.

If x_{12} is not rigorously equal to zero, the unknown remainder term R will occur in the numerator of (8. 17). However, the term $x_{12}R$ is of the third power in the quantities a_{ij}, x_{ij} and r_{ij} while the other terms in the numerator are only of the second power. Therefore if the three points (1), (2) and (3) are not too far apart on the surface of consumption, the term $x_{12}R$ in the numerator will be comparatively small, and may, as an approximation be dropped. We thus get

$$(8.\ 19)\qquad \check{w}_{12} = \frac{a_{12}x_{23} - a_{23}x_{12}}{r_{12}x_{23} - r_{23}x_{12}}$$

and similarly

$$(8.\ 20)\qquad \check{u}_{12} = -\frac{r_{12}a_{23} - r_{23}a_{12}}{r_{12}x_{23} - r_{23}x_{12}}$$

The last two formulae are expressed in terms of directly observable magnitudes only.

From the point of view of the relative significance of the remainder term R, the formulae (8. 19) and (8. 20) are all the more exact the closer the points (1), (2) and (3). However, the points must not be lying too close together, because then the formula (8. 19) or (8. 20) will give a result which is nearly of the form $\frac{0}{0}$, and such an expression is very sensitive for e r r a t i c influences. The absolute value of the relative error committed by computing $\breve{w}_{12}$ by the approximation formula (8. 19) instead of by the correct formula (8. 17) is less than or equal to the absolute value of

$$(8.21) \qquad z = \frac{1}{2} \frac{x_{12}(|r_{23}r_{31}| + |x_{23}x_{31}|)}{a_{12}x_{23} - a_{23}x_{12}} M$$

where M is an upper bound for the absolute values of $\breve{\breve{w}}(r)$ and $\breve{\breve{u}}(x)$ in the intervals in question. This means that the correct magnitude of $\breve{w}_{12}$ must be lying between $(1 + z)$ and $(1 - z)$ times the approximate magnitude of $\breve{w}_{12}$ found by (8. 19).

We can therefore formulate the following criterion: If nothing is known a priori regarding the money acceleration or the food acceleration, the three points (1), (2) and (3) which are used in the approximation formula (8. 19) for the money flexibility, s h o u l d b e c h o s e n s o a s t o m a k e t h e a b s o l u t e v a l u e o f

$$(8.22) \qquad \varphi_w = \frac{|r_{23}r_{31}| + |x_{23}x_{31}|}{a_{12}x_{23} - a_{23}x_{12}} x_{12}$$

a s s m a l l a s p o s s i b l e. In particular we see that if it is possible to choose the points in such a way that $x_1 = x_2$ but $a_1 \neq a_2$, we get the ideal case $\varphi_w = 0$. This is another way of exhibiting the significance of the isoquants and the quantity variation method.

If the object is to determine the food flexibility, the points should be chosen so as to make the absolute value of

$$(8.23) \qquad \varphi_u = \frac{|r_{23}r_{31}| + |x_{23}x_{31}|}{r_{12}a_{23} - r_{23}a_{12}} r_{12}$$

as small as possible.

There are several ways in which it may be worthwhile to attempt a practical application of the preceding formulae. The following are some tentative suggestions. If a system of scatter-points in $(r, \breve{w})$ coordinates, respectively $(x, \breve{u})$ coordinates, are determined by (8. 19) or (8. 20), it may be attempted to perform a least square smoothing with (8. 22) respectively (8. 23) (or certain simple functions of these quatities) as inverse weights.

If the prices are not known so that (8. 19) cannot be applied as it stands, the following procedure may be tried: Introduce into (8. 19) $x_{12} = \xi_{12} - p_{12}$, $r_{12} = \varrho_{12} - P_{12}$, $a_{12} = P_{12} - p_{12}$ where $\xi_{12} = \log \xi_1 - \log \xi_2$ and similarly for ϱ, p and P. By certain assumptions about the variation of the unknowns p and P taken in conjunction with the fact that all the magnitudes of the flexibilities determined by (8. 19) shall refer to the s a m e money flexibility curve it may be possible (for instance by a least square minimalization) to determine approximately the unknown p's and P's and thereby determine approximately the magnitudes of the money flexibility. This would be a sort of modification of the translation principle to the case where only a set of discrete observations are available. Whether such a procedure will give a worthwhile result can of course only be tested by a thoroughgoing statistical analysis.

9. MONEY UTILITY AND THE PROBLEM OF INDEX NUMBERS.

The connection between the money utility and the index number problem may be looked upon from two different viewpoints: In the first place, the statistical technique by which we have in the preceding Sections attempted to construct utility curves is to some extent based on the use of index numbers. P and in certain cases also p are index numbers. The question may arise: Do index numbers constructed in the usual way by combining directly observed prices and quantities into a more or less mechanical formula, give an adequate expression for the notions P and p that are involved in the utility study? We shall see that it is desirable somewhat to modify the orthodox conception of index numbers when the notion is applied in the utility study. In the second place, adopting a more general viewpoint, we could ask: Can the concept of money utility quantitatively defined help to elaborate a definition of the notion of a price index number which is more significant even for general purposes. We shall here consider these two points in turn.

If index numbers are constructed in the usual way by actually observing prices and quantities and combining these data through one of the classical index number formulae, it is of great importance that the kinds and qualities of goods that enter into the index are the same at all the points of time, or all the places, compared. If the kinds or qualities of the goods that enter into the index differ at the different times or places, the index will not give a correct expression for that thing which it is intended to express, namely, a change in prices alone. One of the important factors that determine whether a good or a certain grade of a good shall be selected for incorporation in a price index or not, is therefore whether it is possible to define the good or the grade exactly, so that the same thing will be understood by it in the different points of time or different places considered.

This selection will, however, tend to

produce a bias, so that the index number constructed on this basis does not become quite representative for the notion of general prices. The selection in question will in fact tend to dampen the price fluctuations. The fluctuations as measured by an index number constructed on the basis of such a selection will show less price spread than it ought to, and this damping effect is particularly strong in geographical index numbers. Let us, for instance, take the construction of a geographical food price index. If the index should be restricted to contain only those foodstuffs that exist in exactly the same qualities in the various places, the index would contain virtually only a few standard articles like granulated sugar, wheat flour of certain grades, etc. But this kind of standard articles are exactly those for which one would expect the least geographical spread in prices. The spread which exists for this kind of goods will, as a rule, only be that which is due to transportation costs or, may be, tariff walls. But this by no means accounts for all the differences there are in prices of foodstuffs. In particular this is very far from representing all the price difference which is of significance in the utility study. A great part of the foodstuffs and often some of the most important items on the list are of a more or less locally characterized sort. They do not have their exact counterpart in the food budgets of other places, and has therefore a more or less locally determined price. When we want to compare "the expensiveness" of food in general, at different places it would be utterly misleading to disregard these articles. And in particular it would be misleading to do so in a utility study. But disregarding these articles is really what we do when we require that the items entering into the food price index shall be strictly "comparable."

I have recently had an opportunity to see a good example of the fact that such directly constructed food price indexes give a too low spread. While Professor Fisher and I were still working along the line of Fisher's method and my isoquant method, we tried, of course, various sources for geographical price indexes within the United States. Amongst others we received from Professor Paul H. Douglas, through his courtesy, a series of geographical living price and food price indexes which he had constructed. The spread shown in these data was much below what one would guess at as a plausible spread, and furthermore the way in which the spread was distributed over the country did not at all appear reasonable [1]). It was therefore found that these price data could not be used for the purpose of utility measurement.

If the food price and the living price indexes are determined by the translation method, we dodge these difficulties arising from the difference in kind and grade of goods in different places, and therefore also avoid the damping effect produced by selecting only "comparable" commodities. Since the damping effect is of such a basic importance in the utility study, I believe that the index numbers of Table 3 are more significant for the present purpose than a set of directly determined index numbers would be.

This does not mean, of course, that the figures in Table 3 are necessarily more significant for general purposes. It may be that the translation method, some time in the future when it has been tested by more extensive numerical work, will prove to be useful also as a means of determining geographical index numbers for general purposes. As a matter af fact, I think it will. But so far, the results given in Table 3 should be considered rather as a set of parameters that has served in the construction of the utility curve, exhibited in Table 5 than as a set of index numbers with an independent and general significance. Of course, I believe that anybody familiar with the geographical price variations in the United States will agree that, in the general course, the figures of Table 3 are plausible, also as an expression, for the generally accepted notion of food prices and living prices. In particular the expensiveness of the great east coast cities as exhibited in Table 3 is, I think, a very reasonable feature. But in detail there may be corrections to make if the index numbers were to be used for general purposes.

I now proceed to the more general aspect of the connection between the notion of money utility and the problem of index numbers. The main idea back of our attempt to construct an index number of "the price of living" is, I think, this: We have two points of time (or place), Nos. (1) and (2). And we want to

[1] Dr. Royal Meeker, who has a most thoroughgoing knowledge of the food price and living price situation over a great part of the States, agrees with me in this point of view.

know by how much we must multiply the income in (1) in order to get an income in (2) that will make an individual just as well off. This formulation of the problem is more or less implicitly involved, I believe, in all the actual attempts of constructing index numbers of "the price of living." Let ϱ_1 be a given nominal income in the first point of time (or place), and ϱ_2 the corresponding nominal income in the second point of time (or place), i. e. the income that would make the individual just as well off as he is with the income ϱ_1 in the first point of time (or place). In the orthodox formulation of the index number problem it is implicitly assumed that the income ϱ_2 in question can be found simply by multiplying ϱ_1 by a certain factor P_{21} which is independent of ϱ_1. In other words it is assumed that we have

$$(9.1) \qquad \varrho_2 = P_{21} \cdot \varrho_1$$

where P_{21} is a constant, independent of ϱ_1. And this constant P_{21} is just taken as an expression for the "living price" at the point of time (or place) No. (2) as compared with the "living price" at the point of time (or place) No. (1). It seems to me that the whole index number problem appears in a more fruitful setting when we drop the orthodox assumption that P_{21} is a constant, and simply formulate the notion of "living price" by saying that ϱ_2 is some function of ϱ_1

$$(9.2) \qquad \varrho_2 = R_{21}(\varrho_1)$$

The nature of this function would then express the price of living in the point (2) as compared with the point (1). If this point of view is adopted, the notion of the price of living in one point as compared with the price of living in another point, does not appear as a number any more but as the shape of a curve, namely the curve that gives a graphical picture of the function R_{21}. This curve we shall call the reduction curve between the two points considered. It expresses the rule by which ϱ_2 is reduced to a scale where it becomes comparable with ϱ_1.

The orthodox assumption can be formulated by saying that the reduction curve shall be a straight line through origin, such as, for instance, the line OL in Fig. 14. A straight line through origin is indeed the graphical expression for a relationship of the form (9.1) where P_{21} is independent of ϱ_1.

And the slope (the angular coefficient) of this straight line would just be an expression for the orthodox notion of "living price." Instead of a reduction c u r v e we would then simply have a reduction n u m b e r. In the present more general formulation of the problem the curve connecting ϱ_2 and ϱ_1 may be a curve with a varying slope, as for instance the curve *OMK* in Fig. 14. Any set of two points of time (or place) compared would be represented by such a curve. It would only be the curve itself that would be able to give the complete expression for the comparison between the two points of time (or place).

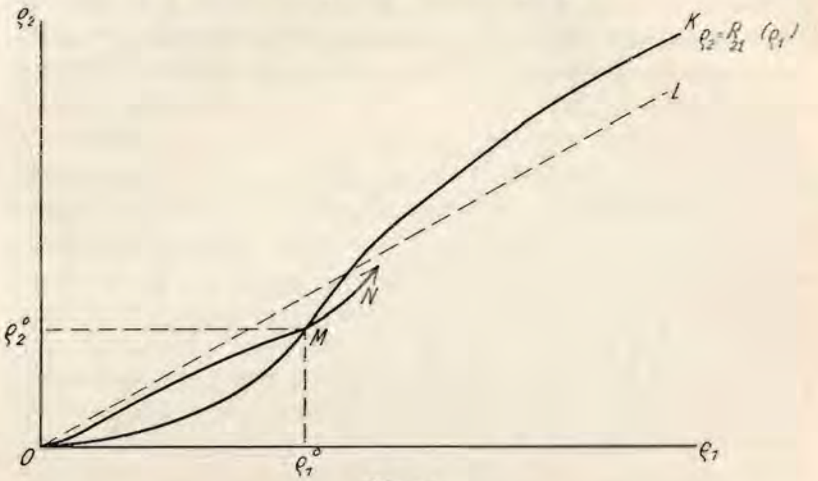

Fig. 14.

If we adopt the present generalized point of view, the notion of "living price" cannot be expressed as a n u m b e r , unless we attach this number to a definite p o i n t on the reduction curve. If the curve is monotonically increasing, this fixation of a point on the curve may be uniquely made by fixing a certain magnitude of ϱ_1, for instance the magnitude ϱ_1^0. Let *M* be the corresponding point on the reduction curve and let ϱ_2^0 be the corresponding income in the point of time (or place) No. (2). There are two numbers which can be attached to the point *M*, and which, each in their sense gives an expression for the notion of "living price." First we have the slope of the straight line from origin to *M*. This slope is equal to the value of the ratio

$$P_{21}(\varrho_1) = \frac{R_{21}(\varrho_1)}{\varrho_1} \qquad (9.\ 3)$$

when we insert ϱ_1^0 for ϱ_1. In the second place we have the slope of the tangent to the reduction curve, in the point *M*, that is to say the magnitude of the derivative

$$R'_{21}(\varrho_1) = \frac{dR_{21}(\varrho_1)}{d\varrho_1} \qquad (9.\ 4)$$

in the point $\varrho_1 = \varrho_1^0$. The ratio defined by (9. 3) we shall call the r e l a t i v e living price between (2) and (1) for the income ϱ_1, and the derivative (9. 4) we shall call the m a r g i n a l living price between (2) and (1) for the income ϱ_1.

The relative living price $P_{21}(\varrho_1)$ expresses by how much we must multiply ϱ_1 in order to get the corresponding ϱ_2. In other words we have

$$\varrho_2 = P_{21}(\varrho_1) \cdot \varrho_1 \qquad (9.\ 5)$$

The relative living price is therefore the notion that comes nearest to the orthodox concept of a "living price." The difference is only that the multiplicator P_{21} is now considered as a function of ϱ_1.

Prima facie the above formulation of the index number problem may seem a little strange, but on reflection I believe it will be found to be a rather plausible formulation. The principle of considering P_{21} as a function of ϱ_1, is really adopted de facto, I think, although the theoretical consequences have not been drawn. For instance: The Ford inquiry has the object to investigate prices in certain European cities as compared with prices in Detroit, in order to find a basis for a comparison of the workmens wages in these places. And the United States Foreign office has long ago been facing a similar problem in connection with the fixation of the salaries of their ambassadors, and legation employees around the world. Is it to be expected that the "living price" ratio between, say Berlin and Rome should be exactly the same when it is defined by a comparison between workmens wages as when it is defined by a comparison between salaries for ambassadors? Certainly not. What we do when we say that P_{21} is a function of ϱ_1 is only to take account of one aspect of this variation in the living price concept. Of course there may also be other factors besides the size of the income which tend to produce a variation in the ratio considered. This simply shows that even the assumption that P_{21} is a function of ϱ_1 is not general enough to completely cover the real situation. But at least the assumption that P_{21} is a function of ϱ_1 is, I think, nearer to reality than the simple assumption that it is a constant. A similar argument would hold good it seems even if we did not consider general living prices but other sorts of prices that are frequently made the object of index number construction: wholesale commodity prices, general food prices, etc.

In terms of the notions here developed the problem of living price determination is the problem of defining some principle by which, to every income ϱ_1 in (1) we can associate a certain income ϱ_2 in (2). If the mode of living is not very much a like in the two situations, it appears as exceedingly difficult to make this association on the basis of direct observation of prices and quantities. This is true already in the simple case when P_{21} is assumed constant, and it would be still more true in the case where we consider P_{21} as a function of ϱ_1. In particular it seems that these difficulties would be virtually unsurmountable when the comparisons made are geographical rather than temporal.

The question therefore arises: Does there exist any economic parameter which is a pure number, defined only by prices and quantities in (1) or only by prices and quantities in (2) (so that no direct price or quantity comparison between (1) and (2) is necessary), and which further is such that it offers a possibility of creating a plausible association between ϱ_1 and ϱ_2?

If we had available a significant money utility curve for (1) and also such a curve for (2), then it seems that the money flexibility would be a parameter of the kind we need. In fact, if we want to keep our definition of real income in absolute terms, independent of geographical and temporal variations in the nature of the goods consumed, it seems rather plausible, or more than that: is seems more or less inevitable, to attach the real income definition to some basic feature of human behavior which we know depends very closely upon how high up in the scale of living the individual finds himself. Such a feature of the behavior is, however, the way in which the individual reacts for an increase in income. And a quantitative measure of this behavior is just the money flexibility.

The practical carrying through of this principle would be as follows: Let us assume that the money flexibility functions $\check{w}_1(\varrho_1)$ and $\check{w}_2(\varrho_2)$ for (1) and (2) are determined by some method

or another. And let us assume that both these functions are monotonic functions, so that a given $\check{w}_1$ will determine uniquely a certain ϱ_1, and a given $\check{w}_2$ will determine uniquely a certain ϱ_2. If this is so, we can determine an association between ϱ_1 and ϱ_2 by the following process. First select a certain magnitude of ϱ_1. Then determine what flexibility $\check{w}_1$ in (1) corresponds to this ϱ_1. Then go to that point on the flexibility curve for (2) where the flexibility is equal to $\check{w}_1$. And finally read off what income ϱ_2 in (2) corresponds to this flexibility. In other words, if we denote the inverse function of $\check{w}_2(\varrho_2)$ by

(9. 6) $$\varrho_2 = R_2(\check{w}_2)$$

then the function R_{21} defined by (9. 2) would, by the above association process, become equal to

(9. 7) $$\varrho_2 = R_{21}(\varrho_1) = R_2(\check{w}_1(\varrho_1))$$

By virtue of our assumption regarding the monotony of the

Table 6.

(I) Absolute value of the money flexibility $-\check{w}$	(II) German income measured in Reichsmarks ϱ_1	(III) American income measured in Dollars ϱ_2	(IV) Living price in the U. S. as compared with Germany. Actual currency basis P_{21}	(V) Living price in the U. S. as compared with Germany. Gold basis
2.0	500	200	0.400	1.68
1.9	580	234	0.404	1.69
1.8	615	250	0.407	1.71
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:

two functions $\check{w}_1(\varrho_1)$ and $\check{w}_2(\varrho_2)$, (9. 7) will define ϱ_2 uniquely as a function of ϱ_1. The law of association (9. 7) can be exhibited in tabular form as indicated in Tab. 6.

As an example we have in Tab. 6 taken a comparison between Germany and the United States. In column (I) of Table 6 are listed different magnitudes of the money flexibility. In column (II) are listed the incomes, in Reichsmark, that, according to the German money flexibility curve, correspond to the given magnitude of the flexibility, and in column (III) are listed the incomes, in Dollars, that, according to the American flexibility curve, correspond to the given magnitudes of the flexibility.

In column (IV) are given the ratio $P_{21} = \varrho_2/\varrho_1$ between the numbers in columns (III) and (II). This ratio is the living price as defined by equation (9. 3). The essential point in this connection is that the ratio P_{21} changes with ϱ_1. If this ratio is multiplied by the gold content of the Dollar and divided to the gold content of the Reichsmark, we get the living price comparison expressed on a gold basis. The latter ratio is given in column (V). Similarly a reduction can be made to an exchange rate basis, regardless of wether the countries have gold redemption or not. Table 6 is not built on actual observation, it is only intended as an example.

As a modification of the above procedure it may be found useful not to establish the association between $\breve{w}_2$ and $\breve{w}_1$ simply by putting $\breve{w}_2$ equal to $\breve{w}_1$ but by putting $\breve{w}_2$ equal to, say a constant plus $\breve{w}_1$, or equal to a constant times $\breve{w}_1$, or equal to some other simple function of $\breve{w}_1$. By such a process it may be possible to overcome the difficulty which arises when there

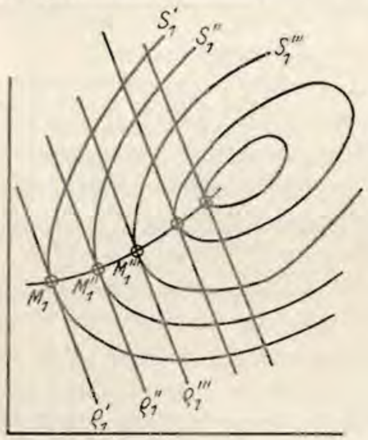

Fig. 15 a.
Indifferences surfaces in (1)

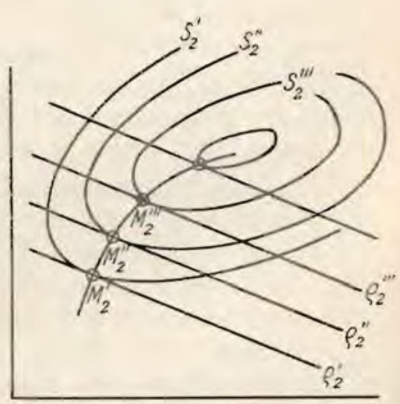

Fig. 15 b.
Indifference surfaces in (2)

exists some systematic difference between the typical want constitution for (1) and the typical want constitution for (2), that introduces a bias in the flexibility comparison.

This whole analysis can easily be generalized to the case where we do not assume that the nominal money utility is of the form (1. 2) but adopt the more general assumption (1. 1). In this case the association principle between ϱ_1 and ϱ_2 can best be expressed in terms of the notion of indifference

surfaces. We assume that the nature of the want constitution of a typical individual in the point of time (or place) (1) is given by a set of indifference surfaces, such as those expressed in Fig. 15 a. Similarly let Fig. 15 b represent the set of indifference surfaces for the typical individual in (2). The axes are quantity axes for the commodities. For simplicity we have assumed only two commodities. The want constitution in (1) and the want constitution in (2) need not be the same in every respect. We even do not need to assume that there is the same kind of commodities in (1) and (2), but we must assume that there exist some principle by which any given indifference surface in (2) is associated with a certain indifference surface in (1), and vice versa. One such principle, namely, the flexibility principle, is discussed below. Let S'_1 and S'_2, S''_1 and S''_2, S'''_1 and S'''_2 . . . etc. be sets of corresponding indifference surfaces in (1) and (2).

The price situation (that is: all the individual prices) are also assumed given, both in (1) and (2). Consequently we may determine what nominal incomes ϱ'_1, ϱ''_1, ϱ'''_1 . . . etc. the individual need in (1) in order that his equilibrium of exchange shall be realized in a point which lies respectively on the surfaces S'_1, S''_1, S'''_1 . . . etc. These incomes will be those that are represented by budget planes tangent to the respective surfaces S_1. (See Fig. 15 a.) Similarly we can find out what incomes ϱ'_2, ϱ''_2, ϱ'''_2 . . . etc. that are associated with the surfaces S'_2, S''_2, S'''_2 . . . etc. For simplicity we may assume that the association between the incomes ϱ_1 and the surfaces S_1, is unique, and similarly for the association between the incomes ϱ_2 and the surfaces S_2.

If an individual in (1) increases his income he will move along the curve M'_1, M''_1, M'''_1 . . . Under this variation we may define the notion of money flexibility, so that a certain magnitude of the flexibility is attached to each point on the curve M'_1, M''_1, M'''_1 . . . Similarly there may be attached a certain magnitude of the money flexibility to each point on the curve M'_2, M''_2, M'''_2 . . . This gives a means of establishing a unique association between the surfaces S_1 and S_2 by utilizing corresponding values of the flexibility, as above explained.

Since there exists a unique association between the indifference surfaces S_1 and S_2, the above procedure will define an unique association between the incomes ϱ_1, and ϱ_2, that is to say, it will define ϱ_2 as a function of ϱ_1, and thus define the notion of living price[1]).

If the present notion of living price as a function, not as a number is adopted, the question of subjecting the index to fulfill certain tests, must, of course, be formulated in another way than when the index is a number.

[1] The idea of utilizing the notion of utility in order to define an index number has been considered by various writers; for instance, Knut Wicksell, F. Y. Edgeworth, and others. See also Dr. Royal Meeker's article „Cost of Living" in Encyclopedia Britannica. It seems, however, that none of these writers have utilized the utility notion in the way suggested in the above discussion.

10. MONEY UTILITY AND THE SUPPLY CURVE OF LABOR.

Although the question of the labor supply curve and its determining factors have been the object of much discussion, it does not seem that the more complicated aspects of this problem have yet been cleared up. In the present Section I shall utilize the notion of money utility in an attempt to throw some light on this question.

Briefly the supply curve of labor may be defined as the curve that exhibits how the number of hours worked per day y depends on the wage rate per hour q. This definition is, however, only preliminary. The various notions involved in the present statement need to be specified more exactly. One point that needs specification is the question as to what extent the individual wage earner under the wage system of our days based upon collective bargaining has a possibility to vary the amount of labor supplied by him. The following analysis is not hinged upon such a possibility. Quite regardless of whether or not it is possible for each of the workers to vary his supply of labor it is certain that the way in which hours of work and wage per hour is valued *on the average* in the group of workers included in the collective wage agreement, plays an important role in the fixation of the collective wage agreement itself. The labor supply function may always be looked upon as an expression for this average or typical valuation. And if there exists a possibility for each worker to vary his supply of labor, the labor supply function even has an individual significance.

Next we have to consider the meaning of "hours worked". When I say that y represents "hours worked" this must be understood as a simplified term used for shortness of expression. In an actual case y could, of course, not be measured simply by the clock. We would have to make some correction for the varying intensity of the work. We may even have to express y in terms of piece work, or express y by a more or less complicated formula involving both actual hours and piecework, etc. The main point is that y should be expressed as something per which the total wage is computed, so that it becomes possible to define the notion of the wage rate q. These qualifications being made, we shall in the following for shortness continue to call the units in which y is measured "hours."

As the number of hours worked on a given day increases, the marginal disutility of labor v increases. We may consider v as a function of y.

$$v = v(y) \tag{10.1}$$

and picture it by a graph such as Fig. 16. There will be a certain minimum work per day y_0 which the individual will want to do just for the pleasure of the work itself. In other words, y_0 is the point where the labor disutility curve passes from negative to positive. If the individual has to work more than y_0 hours per day there will be attached an actual disutility to the labor. If he is prevented from working so much as y_0 hours per

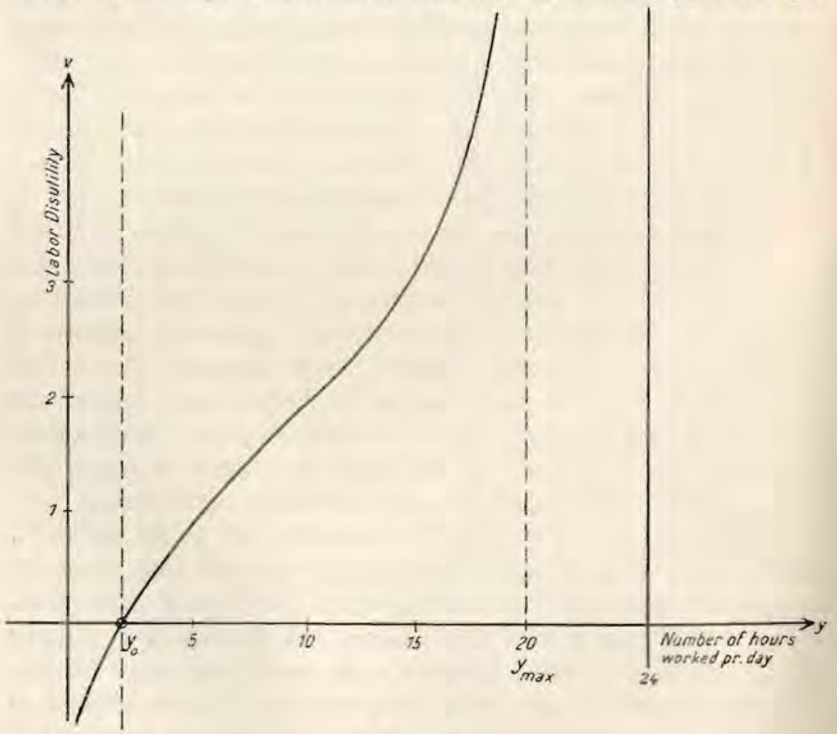

Fig. 16.
Disutility Curve for Labor.

day, labor has a negative disutility, that is, the exertion of labor actually appears as a "good" with a positive utility. For the following analysis it is only the part of the curve that is lying to the right of the point y_0 that we need to take account of. There is also another point on the abscissa axis of Fig. 16 that is of importance, namely, the physiological maximum of work per day y_{max}. Beyond this point the individual cannot go without breaking down. If y is actually measured by the clock, the point y_{max} will, of course, lie somewhere before the ultimate limit $y = 24$ hours. In practice the limit will not be an absolutely fixed magnitude that holds good for every day. It only holds good as an average. For one or a few days work might perhaps be kept on continously. In order to get a perfectly realistic picture of the exhaustion process we would have to follow the individual from day to day and, so to speak, study his breakdown historically. However, the details of this process are not essential for the purpose which we have here in mind. Here it will be sufficient to take account of the exhaustion process only so far as it can be done by introducing the parameter y_{max}.

In any point on the labor disutility curve we may consider the labor-flexibility defined by

$$\breve{v}(y) = \frac{dv(y)}{dy} \cdot \frac{y}{v(y)} \tag{10.2}$$

The nature of the labor disutility curve will, of course, be one of the essential factors that determine how the labor supplied varies with the wage rate. But it is by no means the only factor underlying this relationship. There are in particular two other factors that play an important role: the income composition and the money utility curve. I shall first discuss at some length the nature of the income composition. In particular I shall describe it by certain characteristic "flexibilities" that will be of importance for the following discussion. It seems to me that this is a rather vital but generally neglected part of the labor supply discussion. Next I shall derive a set of general formulae expressing how the labor supply elasticities can be expressed in terms of the income composition flexibilities, the labor disutility flexibility, the money flexibility and certain other factors. Finally I shall utilize the formulae obtained for a closer discussion of certain important special cases.

By the income composition I mean the way in which the remuneration coming from the work on the various days enters as elements into the total income budget of the individual. In practice the income from work on a given day will always be looked upon in its connection with income from work on other days and possibly also in connection with income from other sources. Even the most shortsighted person will think of more than one day when he disposes of his income. In order to take a full account of this fact we should have to go into a systematic study of saving. This will not be attempted here. We shall only take account of the "carry over" of income from one period to another, so far as it can be done by assuming that the income period, that is to say, the period to which the money utility function $w(r)$ refers, is longer than the period, namely a day, to which the labor disutility function $v(y)$ refers. For shortness of expression we shall call the income period a "year", but there will be nothing in the following argument that is based on the notion of year in the calendar sense. We simply assume that the income year includes a certain number n of working days.

We shall number the working days in a year 1, 2 . . . n and denote the number of hours worked on these days y_1, y_2 . . . y_n. The wage rates that are effective on these days we shall denote q_1, q_2 . . . q_n. In general we shall not assume that the same wage rate obtains throughout the year.

The simplest way in which the total yearly spendable nominal income ϱ can depend on the daily wages is through straightforward summations with no additional income from other sources. In this case we have

$$\varrho = y_1q_1 + y_2q_2 + \ldots + y_nq_n \tag{10.3}$$

If there is a certain yearly income σ from a source other than the work considered we have

$$\varrho = y_1q_1 + y_2q_2 + \ldots + y_nq_n + \sigma \tag{10.4}$$

If taxes are levied, say with a percentage t on income from the work considered, with a percentage τ on income from other sources, and a per capita tax T, then the total yearly spendable income will be

$$\varrho = (y_1q_1 + y_2q_2 + \ldots + y_nq_n)(1 - t) + \sigma(1 - \tau) - T \tag{10.5}$$

The formulae (10. 3), (10. 4) and (10. 5) represent examples of

income composition. Many other and more complicated situations are of course possible. In order to insure generality we shall simply assume that the total yearly spendable income is some function

$$\varrho = \varrho(y_1, y_2 \ldots y_n, q_1, q_2 \ldots q_n) \tag{10.6}$$

of the variables indicated. In order to characterize the nature of this function we introduce its partial derivatives which we denote as follows:

$$\varrho'_i = \varrho'_i(y_1 \ldots y_n, q_1 \ldots q_n) = \frac{\partial\varrho(y_1 \ldots y_n, q_1 \ldots q_n)}{\partial y_i} \tag{10.7}$$

$$\varrho'_{(j)} = \varrho'_{(j)}(y_1 \ldots y_n, q_1 \ldots q_n) = \frac{\partial\varrho(y_1 \ldots y_n, q_1 \ldots q_n)}{\partial q_j} \tag{10.8}$$

Further we shall introduce the partial flexibilities

$$\varrho_i = \varrho_i(y_1 \ldots y_n, q_1 \ldots q_n) = \frac{\partial\varrho(y_1 \ldots y_n, q_1 \ldots q_n)}{\partial y_i} \cdot \frac{y_i}{\varrho(y_1 \ldots y_n, q_1 \ldots q_n)} \tag{10.9}$$

$$\varrho_{(j)} = \varrho_{(j)}(y_1 \ldots y_n, q_1 \ldots q_n) = \frac{\partial\varrho(y_1 \ldots y_n, q_1 \ldots q_n)}{\partial q_j} \cdot \frac{q_j}{\varrho(y_1 \ldots y_n, q_1 \ldots q_n)} \tag{10.10}$$

And finally the "accelerations"

$$\varrho_{ik} = \frac{\partial\varrho'_i(y_1 \ldots y_n, q_1 \ldots q_n)}{\partial y_k} \cdot \frac{y_k}{\varrho'_i(y_1 \ldots y_n, q_1 \ldots q_n)} \tag{10.11}$$

$$\varrho_{i(k)} = \frac{\partial\varrho'_i(y_1 \ldots y_n, q_1 \ldots q_n)}{\partial q_k} \cdot \frac{q_k}{\varrho'_i(y_1 \ldots y_n, q_1 \ldots q_n)} \tag{10.12}$$

If the income composition is of the special form (10. 5) and σ, t, τ and T are independent of the y's and q's, we simply have

$$\varrho'_i = q_i(1 - t) \tag{10.13}$$

$$\varrho'_{(j)} = y_j(1 - t) \tag{10.14}$$

$$\varrho_i = \frac{q_iy_i(1 - t)}{\varrho} \tag{10.15}$$

$$\varrho_{(j)} = \frac{q_jy_j(1 - t)}{\varrho} \tag{10.16}$$

If either the tax t or τ is progressive, or if we have other complications in the composition of the spendable income, the simple formulae (10. 13) to (10. 16) do not hold good any more.

If we know the flexibilities ϱ_i and $\varrho_{(i)}$, we can express the percentage change in ϱ which will be produced by given small percentage changes in the y's or in the q's. If y_1 is increased (or decreased) with the percentage $\frac{\delta y_1}{y_1}$, y_2 with the percentage $\frac{\delta y_2}{y_2}$ ... etc., then the resulting percentage change in ϱ will be

$$(10.\ 17)\qquad \frac{\delta\varrho}{\varrho} = \varrho_1 \cdot \frac{\delta y_1}{y_1} + \varrho_2 \cdot \frac{\delta y_2}{y_2} + \ldots + \varrho_n \cdot \frac{\delta y_n}{y_n}$$

Similarly if q_1 is increased (or decreased) with the percentage $\frac{\delta q_1}{q_1}$, q_2 with the percentage $\frac{\delta q_2}{q_2}$ ... etc., the resulting percentage change in ϱ will be

$$(10.\ 18)\qquad \frac{\delta\varrho}{\varrho} = \varrho_{(1)} \cdot \frac{\delta q_1}{q_1} + \varrho_{(2)} \cdot \frac{\delta q_2}{q_2} + \ldots + \varrho_{(n)} \cdot \frac{\delta q_n}{q_n}$$

The above formulae are built on the assumption that the y's and q's for each day may vary independently. However, sometimes there will be things at work that tend to keep the wage rate q the same over a certain period (long time wage contracts, etc.). And there may also be things that tend to keep y at the same level over a certain period. We shall therefore introduce a set of range flexibilities that expresses with what percentage the yearly, spendable income will change when the number of hours worked per day is changed in one and the same proposition over a certain sequence of days, say over the days (1, 2 ... ν), while all other things remain unchanged. In other words, we assume that

$$(10.\ 19)\qquad \frac{\delta y_1}{y_1} = \frac{\delta y_2}{y_2} = \ldots = \frac{\delta y_\nu}{y_\nu}$$

and that the common ratio expressed by (10. 19) is different from zero while the rest of the δy and all the δq are equal to zero. Under this assumption we form the ratio

$$(10.\ 20)\qquad \varrho_{1,2\cdots\nu} = \frac{\delta\varrho}{\varrho} \Big/ \frac{\delta y_k}{y_k}$$

where k is any of the numbers 1, 2 ... ν. On account of (10. 19) if does not matter which one of the numbers 1, 2 ... ν we put k equal to. The ratio (10. 20) is the flexibility of the income with respect to a range-variation in hours worked. It should be

noticed that (10. 19) does not involve the assumption that we actually have $y_1 = y_2 = \ldots = y_\nu$. But in many cases where it is plausible to assume (10. 19) the situation will be such that the further condition $y_1 = y_2 = \ldots = y_\nu$ is also fulfilled.

Similarly we may assume that

$$(10.\ 21)\qquad \frac{\delta q_1}{q_1} = \frac{\delta q_2}{q_2} = \ldots = \frac{\delta q_\nu}{q_\nu}$$

and that the rest of the increments δq and all the increments δy are zero. And under this assumption we may form the ratio

$$(10.\ 22)\qquad \varrho_{(1,2\cdots\nu)} = \frac{\delta\varrho}{\varrho} \Big/ \frac{\delta q_k}{q_k}$$

where k is any of the numbers 1, 2 ... ν. This ratio is the flexibility of the income with respect to a range-variation in the wage rate.

In particular if the range is the entire year we use the notation

$$(10.\ 23)\qquad \varrho_{\cdot} = \varrho_{1,2\cdots n}$$

$$(10.\ 24)\qquad \varrho_{(\cdot)} = \varrho_{(1,2\cdots n)}$$

The range flexibility for a certain range is equal to the sum of the daily flexibilities in the range. That is to say we have

$$(10.\ 25)\qquad \varrho_{1,2\cdots\nu} = \varrho_1 + \varrho_2 + \ldots + \varrho_\nu$$

and

$$(10.\ 26)\qquad \varrho_{(1,2\cdots\nu)} = \varrho_{(1)} + \varrho_{(2)} + \ldots + \varrho_{(\nu)}$$

(10. 25) simply follows by inserting (10. 19) in (10. 17) and using the definition (10. 20). (10. 26) is proved in a similar way.

Just as we have considered the daily flexibilities and the range flexibilities of the function ϱ, we may consider the daily flexibilities and the range flexibilities of the function ϱ'_i defined by (10. 7). The daily flexibilities of ϱ'_i are already defined by (10. 11) and (10. 12). The range flexibility of ϱ'_i with respect to a variation in the y's and the q's respectively we denote $\varrho_{i;1,2\cdots\nu}$ and $\varrho_{i(1,2\cdots\nu)}$. These range flexibilities are equal to

$$(10.\ 27)\qquad \varrho_{i;1,2\cdots\nu} = \varrho_{i1} + \varrho_{i2} + \ldots + \varrho_{i\nu}$$

$$(10.\ 28)\qquad \varrho_{i(1,2\cdots\nu)} = \varrho_{i(1)} + \varrho_{i(2)} + \ldots + \varrho_{i(\nu)}$$

where the ϱ_{ik} and the $\varrho_{i(k)}$ are the magnitudes defined by (10. 11) and (10. 12). In particular when the range is the whole year we use the notation

(10. 29) $$\varrho_{i_*} = \varrho_{i;1\cdot2\cdots n}$$

(10. 30) $$\varrho_{i(*)} = \varrho_{i(1\cdot2\cdots n)}$$

In the following we shall consider the yearly price of living P as a function, amongst others, of the wage rates, that is we put

(10. 31) $$P = P(q_1, q_2 \ldots q_n)$$

The nature of this function we shall express by the daily flexibilities

(10. 32) $$P_{(j)} = \frac{\partial P(q_1 \ldots q_n)}{\partial q_j} \cdot \frac{q_j}{P(q_1 \ldots q_n)}$$

and the range flexibilities

(10. 33) $$P_{(1\cdot2\cdots\nu)} = P_{(1)} + P_{(2)} + \ldots + P_{(\nu)}$$

In particular when the range is the whole year we use the notation

(10. 34) $$P_{(*)} = P_{(1\cdot2\cdots n)}$$

The above procedure of considering the wage rate or the hours worked separately for each day or for certain sub intervals of the year helps to make the usual static conception of the labor supply mechanism more realistic. But it cannot, of course, give a complete picture of the actual happenings in a continuously changing market. It is at best only a substitute for a truly dynamic analysis of the shiftings in the labor market.

A third factor which determines the shape of the labor supply curve is the money utility function $w(r)$. There are in particular two characteristic abscissa points in this function, which will play a rôle in the present analysis, namely, the real income r_0 which represents the physical minimum of existence, and the real income r_1 which represents the point where the absolute value of the money flexibility $-\breve{w}(r)$ passes from above unity to below unity. These two income points may be called the minimum-of-existence-point and the hyperbola-point. The latter point is a point where the money utility curve is tangent to an equilateral hyperbola.

We now have the necessary tools to proceed to a closer analysis of how the three data: the labor-disutility curve, the income composition, and the money-utility curve determine the supply of labor.

Between the money utility and the labor disutility there exists for each day an equilibrium equation similar to the one we have studied for the money utility and the utility of the commodity of comparison. However, the very general assumptions we have made about the form of the income composition makes it necessary on one point to modify the equation. Instead of the commodity price p which occurs in the denominator of a in the right member of equation (3. 4) we cannot simply introduce the wage rate q. We must instead introduce the partial derivative of the net spendable yearly income ϱ with respect to the number of hours worked on the particular day considered. In other words, if the day considered is the day No. i, we have to introduce the derivative ϱ'_i defined by (10. 7).

In fact, suppose that the hours worked on the i-th day, namely y_i, is increased by the small amount δy_i, while all other things are kept constant. This will involve a small labor disutility equal to $v(y_i)\delta y_i$. And the yearly spendable income ϱ will be increased by the small amount $\varrho'_i\delta y_i$. But increasing ϱ by the small amount $\varrho'_i\delta y_i$ means creating a utility equal to $\omega(\varrho, P)\varrho'_i\delta y_i$ where $\omega(\varrho, P)$ is the money utility measured per dollar, P being as before the living price. Since the disutility suffered and the utility gained ought to be equal, we must have

(10. 35) $$\omega(\varrho, P)\varrho'_i = v(y_i)$$

The equilibrium equation for the day No. i must consequently be

(10. 36) $$w\left(\frac{\varrho}{P}\right) = \frac{P}{\varrho'_i}\, v(y_i).$$

The last equation is perfectly similar to the equation of the surface of consumption. We only have the symbols y_i, ϱ'_i and v instead of x, p and u.

Introducing the explicit expression for the functions ϱ, P and ϱ'_i in (10. 36) we see that the equilibrium equation for the day No. i can be written

(10. 37) $$w\left(\frac{\varrho(y_1 \ldots y_n, q_1 \ldots q_n)}{P(q_1 \ldots q_n)}\right) = \frac{P(q_1 \ldots q_n)}{\varrho'_i(y_1 \ldots y_n, q_1 \ldots q_n)}\, v(y_i)$$

If we assume that the nature of the functions $w(r)$, $\varrho(y_1 \ldots y_n, q_1 \ldots q_n)$, $P(q_1 \ldots q_n)$ and $v(y)$ are known, then (10. 37) involves only the two sets of variables $y_1 \ldots y_n$ and $q_1 \ldots q_n$. For each day we would have an equilibrium equation such as (10. 37).

This gives a system of n equations that can be looked upon as defining implicitly each y_i as a function of the n wage rates $q_1, q_2 \dots q_n$. These functions we shall denote

$$(10.38) \qquad y_i = y_i(q_1, q_2 \dots q_n)$$

The functions $y_i(q_1 \dots q_n)$ $(i = 1, 2 \dots n)$ thus determined are the supply-functions for labor. The function $y_i(q_1 \dots q_n)$ expresses how the quantity of labor supplied on the i-th day depends on the wage rates $q_1, q_2 \dots q_n$ that are paid on the n days of the year.

In order to characterize the supply functions (10. 38) we shall introduce the partial elasticities of labor supply:

$$(10.39) \quad y_{i(j)} = y_{i(j)}(q_1, q_2 \dots q_n) = \frac{\partial y_i(q_1 \dots q_n)}{\partial q_j} \cdot \frac{q_j}{y_i(q_1 \dots q_n)}$$

The elasticity (10. 39) is an elasticity with respect to a one-day wage variation. It measures the change which takes place in the labor supplied on the day No. i when the wage-rate on the day No. j is changed and all other wage rates remain unchanged. More precisely, it is the ratio between the percentage change in the labor supplied on the day No. i and the corresponding percentage change in the wage rate on the day No. j (these changes being assumed small). If we know the partial elasticities $y_{i(j)}$ we can find out by which percentage $\frac{\delta y_i}{y_i}$ the labor supplied on the day No. i will increase (or decrease) if the wage rates on the days $1, 2 \dots n$ are subject to the small percentage changes $\frac{\delta q_1}{q_1}, \frac{\delta q_2}{q_2} \dots \frac{\delta q_n}{q_n}$. This labor supply change is given by the formula

$$(10.40) \qquad \frac{\delta y_i}{y_i} = y_{i(1)} \frac{\delta q_1}{q_1} + y_{i(2)} \frac{\delta q_2}{q_2} + \dots + y_{i(n)} \frac{\delta q_n}{q_n}$$

Besides the elasticity with respect to a one-day variation in the wage rate defined by (10. 39) we shall also consider the range elasticity of a given day's labor supply with respect to a wage change that is made effective for a range of several days. More precisely expressed: We assume (10. 21) and then define the range elasticity of the i-th days labor supply as

$$(10.41) \qquad y_{i(1,2 \dots \nu)} = \frac{\delta y_i}{y_i} \Big/ \frac{\delta q_k}{q_k}$$

where k is any of the numbers $1, 2 \dots \nu$. That is to say, we assume

that the wage rates on the days $1, 2 \dots \nu$ are all increased (or decreased) with a certain percentage, the same for all these days, while the wage rates on the other days remain unchanged. Then we ask: By how many per cent will this cause the labor supply on the particular day No. i to increase (or decrease)? This percentage is given by (10. 40). And the ratio between this percentage (positive or negative) increase in the labor supply on the day No. i and the corresponding percentage increase in the wage rate for the days No. $1, 2 \dots \nu$, is what we have called the labor supply elasticity for the day No. i with respect to a wage-change over the range $(1, 2 \dots \nu)$. In practice it is this kind of elasticity that is of the greatest importance.

By virtue of the general proposition on the cumulation of daily elasticities into range elasticities we have

$$(10.42) \qquad y_{i(1,2 \dots \nu)} = y_{i(1)} + y_{i(2)} + \dots + y_{i(\nu)}$$

In particular, if the range is the whole year we use the notation

$$(10.43) \qquad y_{i(\cdot)} = y_{i(1,2 \dots n)}$$

From the above discussion it is apparent that the present theoretical problem of the labor supply can be looked upon as the problem of studying the functions (10. 38) considered as a set of functions defined implicitly by the system of equations we obtain from (10. 37) by putting successively $i = 1, 2 \dots n$. In such a study the elasticities form a very valuable tool. I now proceed to show how we can, through the implicit definition of the labor supply function, express the labor supply elasticities in terms of the money flexibility, and some of the other flexibilities defined above.

Let j be a fixed number in the sequence $1, 2 \dots n$, and let us differentiate the equation (10. 37) partially with respect to q_j. Noticing that the derivative of $w(r)$ with respect to r is $\frac{\breve{w} \cdot w}{\varrho / P}$, the derivative of ϱ with respect to y_k is $\frac{\varrho_k \cdot \varrho}{y_k}$, etc., we get

$$(10.44) \quad \frac{\breve{w} \cdot w}{\varrho/P} \, \frac{P\left[\Sigma_k \frac{\varrho_k \cdot \varrho}{y_k} \cdot \frac{y_{k(j)} y_k}{q_j} + \frac{\varrho_{(j)} \varrho}{q_j}\right] - \varrho \frac{P_{(j)} P}{q_j}}{P^2} =$$

$$= \frac{P \breve{v} \cdot v}{\varrho'_i \, y_i} \cdot \frac{y_{i(j)} y_i}{q_j} + v \, \frac{\varrho'_i \frac{P_{(j)} P}{q_j} - P \left\{\Sigma_k \varrho_{ik} \frac{\varrho'_i}{y_k} \cdot \frac{y_{k(j)} y_k}{q_j} + \frac{\varrho_{i(j)} \varrho'_i}{q_j}\right\}}{\varrho'^2_i}$$

Rearranging the terms in this expression and introducing for convenience the symbol $e_{ik} = \begin{Bmatrix} 1 & (i=k) \\ 0 & (i \neq k) \end{Bmatrix}$, we get

$$(10.45) \quad \sum_{k=1}^{n} (e_{ik}\breve{v} - \varrho_k \breve{w} - \varrho_{ik})\, y_{k(j)} = \varrho_{i(j)} - (-\breve{w})(\varrho_{(j)} - P_{(j)}) - P_{(j)}$$

$$\begin{pmatrix} i = 1, 2 \ldots n \\ j = 1, 2 \ldots n \end{pmatrix}$$

The last equation holds good for any i and any j. It is therefore a system of linear equations by which the elasticities $y_{k(j)}$ can be determined. It would not be difficult to indicate the solution explicitly in determinant form. But this is not necessary for our purpose.

Since the coefficient of $y_{k(j)}$ in (10. 45) is independent of j we can by extending a summation over j to both sides of (10. 45) immediately deduce the following system of equations for the range-elasticities

$$(10.46) \quad \sum_{k=1}^{n} (e_{ik}\breve{v} - \varrho_k \breve{w} - \varrho_{ik})\, y_{k(1,2\ldots\nu)} =$$
$$= \varrho_{i(1,2\ldots\nu)} - (-\breve{w})(\varrho_{(1,2\ldots\nu)} - P_{(1,2\ldots\nu)}) - P_{(1,2\ldots\nu)}$$

In particular, if the range considered is the entire year, we have

$$(10.47) \quad \sum_{k=1}^{n} (e_{ik}\breve{v} - \varrho_k \breve{w} - \varrho_{ik})\, y_{k(*)} =$$
$$= \varrho_{i(*)} - (-\breve{w})(\varrho_{(*)} - P_{(*)}) - P_{(*)}$$

For certain questions discussed in the following it is important to know under what conditions the labor supplied on the various days of the year will be equal. In this regard we have the following proposition: In order that the quantities of labor supplied on the various days of the year shall be equal, it is necessary and sufficient that the derivatives ϱ'_i are equal. This simply follows from equation (10. 37) by noticing that in this equation ϱ'_i and $v(y_i)$ are the only magnitudes that depend on i. In order that the $v(y_i)$ shall be equal it is therefcre necessary and sufficient that the ϱ'_i are equal. And since the function $v(y)$ in monotonic, this also becomes the necessary and sufficient condition that the y_i shall be equal. The conditions that all the ϱ'_i shall be equal can be interpreted by saying that the

remuneration from the work considered shall contribute to the total income in such a way that there is no difference between the effect on total yearly spendable income produced by an additional hour of work done in one part of the year and an additional hour of work done in another part of the year.

The above proposition refers to the magnitudes of the ϱ'_i in the daily equilibriumpoints. We may also ask another question: Does the function (10. 7) depend on i when the variables occuring in it are allowed to vary more freely? It is of no practical interest to consider the case where (10. 7) is independent of i when all the variables are allowed to vary independently. In the case (10. 13) for instance such an independence would only be possible if $t = 1$, that is if all the labor income is taxed away. But it may be of interest to consider the case where (10. 7) is independent of i identically in all those independent variables that are left when certain conditions are imposed on the $2n$ variables occuring in (10. 7). We shall in particular consider the following case: We assume that the wage rates paid on the various days are equal and require that the function (10. 7) shall be independent of i identically in the variable q that expresses the common magnitude of the q_i and also identically in the n variables y_i. In this case we shall say that (10. 7) is semi-identically independent if i, or shorter that the income composition (10. 6) is semi-identical; (10. 13) is an example of such a case. If the rate of interest enters into the problem, or if taxes are levied with different percentages in different parts of the year the income composition is not semi-identical.

If the wage rates are equal on the various days and if the income composition is semi-identical, then the quantities of labor supplied on the various days must be equal. In fact, if the wage rates are equal and the income composition semi-identical, ϱ'_i in (10. 37) is independent of i, which by the above proposition entails that y_i is also independent of i. Obviously this holds good no matter what the common magnitude of the wage rates is.

If the wage rates are equal and the income composition semi-identical, the whole problem of the labor supply therefore appears in the following simple form: y is a certain function of q

(10. 48) $$y = y(q)$$

which is implicitly defined by an equilibrium equation of the form

(10. 49) $$w\left(\frac{\varrho(y,q)}{P(q)}\right) = \frac{P(q)}{\varrho'(y,q)}\, v(y)$$

where $\varrho(y,q)$, $\varrho'(y,q)$ and $P(q)$ are the functions (10. 6), (10. 7) and (10. 31) respectively when all the y's and all the q's are equal and the income composition is semi-identical. In this case the labor supply elasticity may simply be defined as

(10. 50) $$\check{y} = \check{y}(q) = \frac{dy(q)}{dq}\cdot\frac{q}{y(q)}$$

This definition is identical with the definition (10. 43) in the present case. Further the flexibilities $\varrho_\bullet$, $\varrho_{(\bullet)}$ and $P_{(\bullet)}$ are now equal to respectively

(10. 51) $$\varrho_y = \frac{\partial\varrho(y,q)}{\partial y}\cdot\frac{y}{\varrho(y,q)}$$

(10. 52) $$\varrho_q = \frac{\partial\varrho(y,q)}{\partial q}\cdot\frac{q}{\varrho(y,q)}$$

and

(10. 53) $$P_q = \frac{d\,P(q)}{dq}\cdot\frac{q}{P(q)}$$

And the "accelerations" $\varrho_{i\bullet}$ and $\varrho_{i(\bullet)}$ are respectively equal to

(10. 54) $$\varrho_{yy} = \frac{\partial\varrho'(y,q)}{\partial y}\cdot\frac{y}{\varrho'(y,q)}$$

and

(10. 55) $$\varrho_{yq} = \frac{\partial\varrho'(y,q)}{\partial q}\cdot\frac{q}{\varrho'(y,q)}$$

We may determine $\check{y}$ by putting $y_{k(\bullet)} = \check{y} =$ independent of k in (10. 47) which gives

(10. 56) $$\check{y} = \frac{\varrho_{yq} - (-\check{w})(\varrho_q - P_q) - P_q}{\check{v} + (-\check{w})\,\varrho_y - \varrho_{yy}}$$

This formula can of course also be developed directly by differentiating (10. 49) with respect to q, on the assumption that y is a function of q.

As an example of an income composition satisfying the conditions under which (10. 56) is valid we may consider the one we would get from (10. 5) when $q_1 = q_2 = \ldots = q_n$ and

the tax parameters t, τ and T are constants, independent of the other variables involved. (One or more of the tax parameters being possibly equal to zero.) In this case we get:

(10. 57) $$\varrho_{ik} = 0 \qquad \varrho_{i(k)} = e_{ik} = \begin{cases} 1 & (i = k) \\ 0 & (i \neq k) \end{cases}$$

From (10. 56) or by solving the equation (10. 47) under the assumption (10. 57) we now get

(10. 58) $$\check{y} = \frac{1 - (-\check{w})\left[\dfrac{nyq(1-t)}{\varrho} - P_q\right] - P_q}{\check{v} + (-\check{w})\dfrac{nyq(1-t)}{\varrho}}$$

In particular if there are no taxes and the price of living is not appreciably influenced by the wage rate in the work considered, we get

(10. 59) $$\check{y} = \frac{1 - (-\check{w})\left(1 - \dfrac{\sigma}{\varrho}\right)}{\check{v} + (-\check{w})\left(1 - \dfrac{\sigma}{\varrho}\right)}$$

And under the still more special assumption that the worker has no income from outside sources, that is $\sigma = 0$, we get

(10. 60) $$\check{y} = \frac{1 - (-\check{w})}{\check{v} + (-\check{w})}$$

The equations (10. 58), (10. 59) and (10. 60) may, of course, also be developed by first computing the coefficients ϱ_y, ϱ_q, ϱ_{yy} and ϱ_{yq} directly on the assumption that ϱ is of the form (10. 5) with all the y's and all the q's equal and t, τ and T constant, and then introducing the values obtained into (10. 56).

We first procede to a closer discussion of the simple case where (10. 60) holds good. In (10. 60) both $\check{v}$ and $(-\check{w})$ are positive, $\check{y}$ has therefore the same sign as $1 - (-\check{w})$. In the present very simple case we consequently have the proposition: *An increase in the wage rate will increase the labor supply if we are in a point in the income scale where the absolute value of the money flexibility is less than unity.* Conversely: If we are in an income point where the absolute value of the money flexibility is larger than unity, an increase in the wage rate will decrease the labor supply. Since $\check{y}$ is nothing else than the labor supply elasticity the formula (10. 60) gives of course not only information about the direction in which the quantity of labor supplied changes with a change in the wage rate, but also information about the *velocity* with which this change takes place.

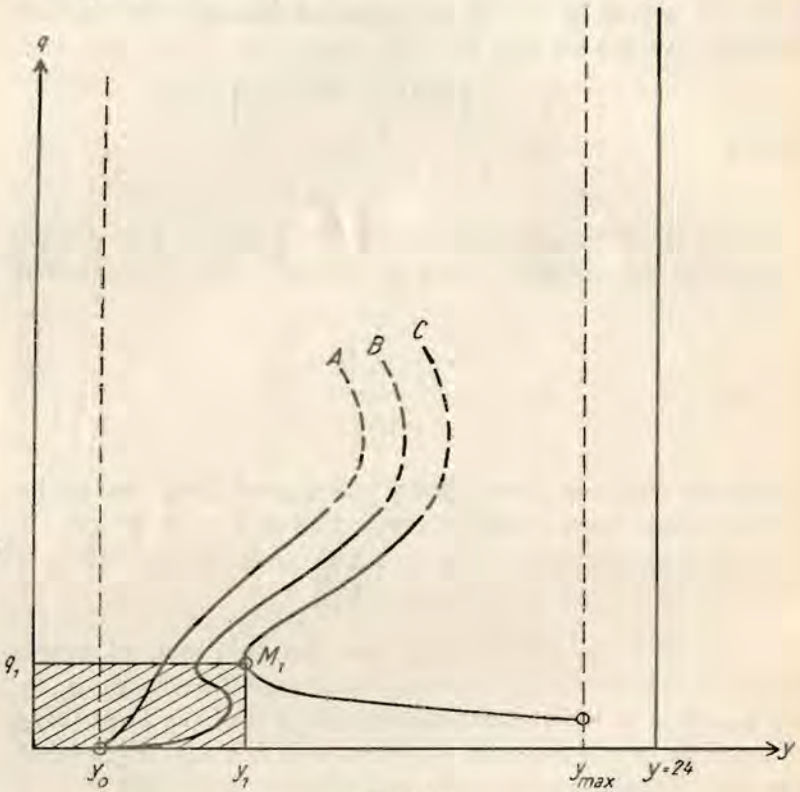

Fig. 17.
Labor supply curves.

In view of the fact that the absolute value of the money flexibility for small incomes is very high, much above unity, and from thereon decrases, as income increases, passing unity for a certain real income r_1, we see that the lower part of the labor supply curve, in the present case must have a shape as shown by curve C in Fig. 17. That is to say, over the very lowest part of the wage rate scale, the labor supply curve will be decreasing. It will even approximately have the shape of an equilateral hyperbola:

The hours worked must virtually be increased in the inverse proportion of the one in which the wage rate is further decreased. This is the "sweat-shop" situation, of which we had examples in Europe in the beginning of the Industrial Revolution. Even today such situations may become established when the worker has no outside source of income. This last condition is, of course, essential for a situation like the one exhibited in the lower branch of curve C in Fig. 17. As the wage rate increases the quantity of labor supplied will decrease until a certain point, which in the curve C of Fig. 17 is marked M_1, and from where the quantity of labor supplied again will start *increasing*, as the wage rate increases further. This point is characterized by the fact that the area of the shaded rectangle in Fig. 17 is equal to that income (counted pr. *day*) for which the absolute value of the money flexibility passes unity. Since we assume at present $P =$ constant, we may put $P = 1$, so that real and nominal income now become the same. A rectangle as the shaded one in Fig. 17 represents, of course, *income* (pr. day), since its area is wage rate times hours worked. I am inclined to believe that the portion of the curve immediately below M_1 represents a rather common situation at least in many countries in Europe.

It may be that still higher up the labor supply curve turns back again as indicated by the upper dotted branch of the curve C in Fig. 17. The necessary and sufficient condition that this shall be so is that the absolute value of the money flexibility for very large incomes again starts increasing and goes up above unity. No statistical information about this end of the money utility curve is at present available. And it even seems doubtful whether it will be possible to conceive of the money utility for the very high part of the income scale in the same way as for the lower part. Over the lower part, money utility has a definite meaning in terms of consumption, while this is not so for the very high part of the scale. Some further remarks on this point are made in section 12.

In so far as it is plausible to assume that the typical working man in our days depends on his wage *as* nearly the only source of income, the relationship exhibited in the curve C of Fig. 17 throws an interesting light on the collective bargaining fight. So long as the equilibrium occurs in a point on the lower branch of the curve (where the flexibility of the money utility is larger than unity) this fight must be particularly violent from the laborers' point of view. Here the laborers must try either to hold status quo or simultaneously to obtain both a wage increase and shorter hours. On the upper branch this does not hold good.

It is not improbable that one of the reasons why the American labor market has been more peaceful than the European resides in the fact revealed by the statistical analysis of the money utility, namely that the average laborer in America has reached an income interval where the absolute value of the flexibility of the money utility is less than unity. That is to say, the collective bargaining fight in America should take place in some point on the upper branch of the curve C in Fig. 17, while in Europe it is still going on on the lower branch of the curve.

The points here discussed have an immediate bearing on an argument put up by Professor Knight. Knight contends that the labor supply curve must a l w a y s be sloping downwards. And he tries to prove it by the following argument: If there is an increase in the wage rate, and the amount of labor supplied is maintained constant (or even increased), then the money utility will diminish. But, when the money utility diminishes the individual will increase his purchase of the various goods which he consumes. One of these goods is leisure, so consequently he will increase the amount of leisure purchased. That is, he will diminish the quantity of labor supplied. This argument is however wrong. If the money utility decreases, then it is quite true that the consumption of a good, w h o s e p r i c e h a s n o t c h a n g e d , will be increased. But if the price of the good has changed simultaneously with the money utility decrease, then the question of whether its consumption shall increase or decrease is still an open one. More precisely expressed, it will depend on whether the decrease in money utility has taken place in a stronger or a weaker proportion than that in which the price of the commodity has increased. If there takes place a decrease in money utility as a consequence of an increase in the wage rate then the price of leisure is certainly n o t constant. More precisely, it has changed exactly like the wage rate, because the price of leisure is nothing else than the wage rate. The question is therefore whether the decrease in money utility takes place in a stronger proportion than the increase in wage rate.

But the percentage with which the wage rate increases is, under the present simplified assumptions, equal to the percentage with which income would increase if number of hours worked were maintained unchanged. So the question is if the decrease in money utility takes place in a stronger or weaker proportion than that in which the income increases. In other words, the question is whether the money flexibility is larger than or less than unity in absolute value. Thus, by making the necessary corrections to Knight's argument, we get back to the criterion which we originally deduced from (10. 60)[1].

We now proceed to a study of the next simplest case, namely, the one where the laborer has some source of income other than the work considered, but where the other simple assumptions specified above are fulfilled. In this case we have $\varrho = nqy + \sigma$ where σ is the income from outside sources. We may also in this case assume $P = 1$, so that ϱ becomes equal to r and σ equal to real outside income s. For convenience in making comparison with the independent variable in $\check{w}(r)$, which is r, we shall here consider the real income r instead of the nominal income ϱ; and the real outside income s instead of σ.

The expression for the labor-supply-elasticity can then be written

$$\check{y} = \frac{\frac{r}{r-s} - (-\check{w})}{\frac{r}{r-s}\check{v} + (-\check{w})} \qquad (10.\ 61)$$

In this case the direction in which the quantity of labor supplied will change when there occurs a change in the wage rate, does not depend on how $(-\check{w})$ compares with unity, b u t o n h o w $(-\check{w})$ c o m p a r e s w i t h t h e i n v e r s e p r o p o r t i o n $\frac{r}{r-s}$ w h i c h i n c o m e f r o m t h e p a r t i c u l a r k i n d o f

[1] The error in Knight's argument has already been pointed out by Professor Robbins (Economica, 1930). However, Professor Robbins makes his argument depend on a distinction between "effort price" and nominal price, which, it seems to me, is irrelevant for the problem in hand. ¶ Professor Pigou has an argument similar to Knight's. Pigou contends that an income tax will always make the laborer work more. This is correct for a capita tax or a tax levied on other income than that from work. But it is wrong if the tax is levied with a fixed percentage on the labor income. Such a tax has the same effect as a decline in the wage rate. The effect therefore depends on the money flexibility.

work considered, bears to total income. If the outside income is very large, it is most likely that the fraction $\frac{r}{r-s}$ in any equilibrium point will be larger than the money flexibility. Therefore the supply-curve of labor will probably be rising in any point. That is, we will have a curve like A in Fig. 17 (no attention being here paid to the upper dotted part). But if the income from other sources is comparatively unimportant, it may happen that there are points where the supply curve of labor is sloping down. It is even possible from (10. 61) to deduce the necessary and sufficient condition that such points shall exist, and furthermore if they exist to show where they are. In fact, let us construct a graph of the function

$$(10.62) \qquad -\check{W}(r) = \frac{r}{r-s}$$

This function is nothing else than the money flexibility we would get by adopting as the money utility function the Bernoullian function $\frac{c}{r-s}$ where however s does not necessarily represent the physical minimum of existence r_0 but some quantity that may be either equal to or smaller or larger than r_0; c is an arbitrary constant.

The function (10. 62) we shall call the Bernoullian flexibility, when the outside income is s. The graph of this function is exhibited in Fig. 18. The absolute value of the Bernoullian flexibility is infinite for $r = s$, and monotonically decreasing to the value $+1$, which is reached for $r = \infty$. The absolute value of the actual money flexibility $-\check{w}(r)$ is monotonically decreasing from $+\infty$ at $r = r_0$ and through the value $+1$ which is passed at $r = r_1$. If we draw a graph of the actual flexibility curve on the same chart where the Bernoullian flexibility curve was plotted, it may evidently happen that these two curves have certain intersection points. Over some parts of the income range the actual money flexibility may be the largest, over other parts the Bernoullian flexibility may be the largest. Those parts of the income interval where the actual money flexibility is less than the Bernoullian flexibility in absolute value correspond to the rising parts of the labor supply curve, and vice versa. (It is only

that part of the income scale where $r \geqq s$ we need to take account of.) According to the shape of the actual flexibility curve we may therefore have various possible more or less complicated

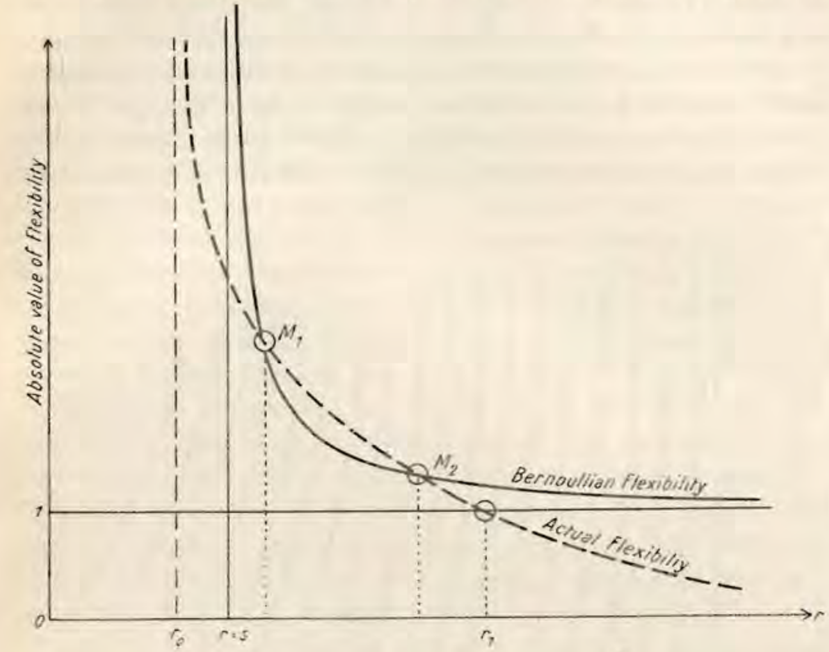

Fig. 18.

Comparison between the actual money flexibility — $\check{w}(r)$ and the Bernoullian flexibility — $\check{W}(r)$.

forms of the labor supply curve, where rising and falling branches alternate. In Fig. 18 the actual flexibility is less than the Bernoullian flexibility before the point M_1 and after the point M_2. Between M_1 and M_2 we have the opposite situation. Therefore if the actual money flexibility varies as indicated in Fig. 18 we get a labor supply curve as B in Fig. 17. Moreover, the nature of the labor supply curve will change in a characteristic way as the outside income s changes. We may visualize the effect of a change in s by imagining that the asymptote to the Bernoullian flexibility, namely the vertical $r = s$ in Fig. 18, is moved to the right or to the left. In particular we see that if the asymptote is moved to $s = r_1$, the Bernoullian flexibility must always be larger than the actual flexibility in absolute values (for $r > s$) and consequently the labor-supply curve monotonically increasing as A in Fig. 17. On the other extreme if the asymptote is moved to $s = 0$ which means that there is no outside income, the Bernoullian flexibility in Fig. 18 will be represented by the horizontal line with ordinate $+ 1$. In this case, the supply curve of labor has the shape indicated by C in Fig. 17.

By the following procedure we get a graphical representation which exhibits both the effect produced by a change in the outside income s, and the effect produced by a change in the shape of the actual money flexibility curve. First we attribute

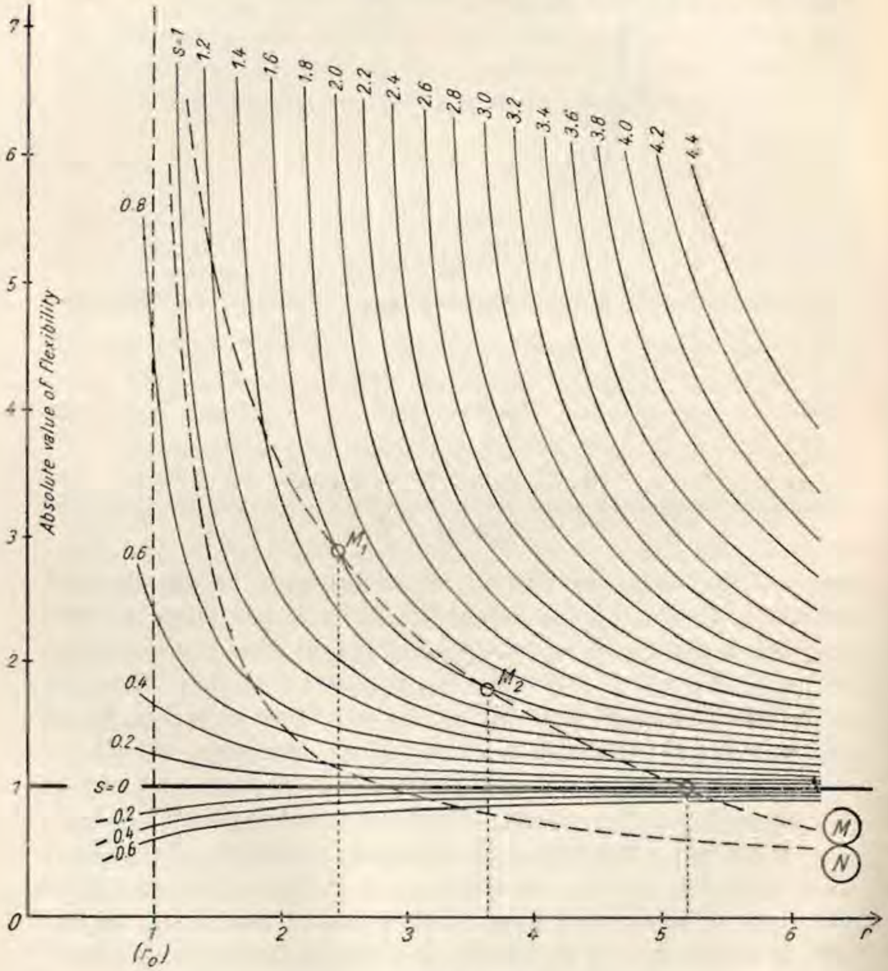

Fig. 19.
Family of Bernoullian Flexibility curves.

to s a set of different magnitudes, and for each s we draw a graph of the function $\frac{r}{r-s}$. Thus we get a family of curves as exhibited in Fig. 19. If, in the same system of axis, we draw a graph of the actual money flexibility (counted positive), we get a nomogram from which we can immediately read off where the labor supply curve will be rising and where it will be falling, under various assumptions regarding the magnitude of the income from outside sources. Suppose, for instance, that the money flexibility curve has a shape like the curve M, that is the upper dotted curve in Fig. 19. In Fig. 19 we have chosen the unit of measurement for the income such that $r = 1$ represents the minimum of existence. This is expressed by the fact that the actual money flexibility curve shoots up to infinity as r approaches 1. From Fig. 19 we see immediately that the money flexibility, as represented by the curve M, will be above the Bernoullian flexibility, corresponding to the outside income $s = 0$, so long as $r < 5.2$ and below it for $r > 5.2$. The Bernoullian flexibility when $s = 0$, is indeed represented by the horizontal line with ordinate equal to 1. Consequently we now have the case where the lower part of the labor supply curve only has one swing, namely a swing as shown by the lower part of C in Fig. 17. If the outside income is $s = 1.8$ the curve M in Fig. 19 is situated entirely below the Bernoullian flexibility curve. Therefore, we now have a labor supply curve where the lower part is monotonically increasing as A of Fig. 17. And if the outside income is $s = 1.6$, the flexibility curve M will be below the Bernoullian flexibility curve for small incomes (that is, total incomes only a little larger than the outside income), and also below the Bernoullian flexibility curve for very high incomes, and in an interval in the middle (between M_1 and M_2 in Fig. 19) it will be above. This is therefore the case represented by B in Fig. 17.

The lower dotted curve N in Fig. 19 is a graph of the flexibility

(10. 63) $$-\check{w}(r) = \frac{1}{\text{lognat } r}$$

This is the flexibility we get if the money utility function is of the form (4. 1) with the minimum of existence $r_0 = 1$. It is unessential what sort of logs we use in the definition (4. 1).

The flexibility will always be given by (10. 63). If the actual flexibility is of the form (10. 63) then the lower part of the labor supply curve can never have two swings, no matter what the outside income is. In fact, if the money flexibility is of the form (10. 63) we have

$$\frac{r}{r-s} - (-\check{w}(r)) = \frac{r}{(r-s)\ \text{lognat}\ r}\left(\text{lognat}\ r - \left(1 - \frac{s}{r}\right)\right)$$

Since $r > s$ and $r > 1$ the labor supply curve is therefore rising or falling in a given point accordingly as the function

$$\text{(10. 64)} \qquad \text{lognat}\ r - \left(1 - \frac{s}{r}\right)$$

is positive or negative in this point. That is to say there can be a turn in the labor supply curve only each time when (10. 64) passes zero. But the derivative of (10. 64) with respect to $\frac{1}{r}$, namely $s - r$, is negative, since $r > s$, so that the function (10. 64) is constantly increasing with increasing r, and can therefore at most have one zero point in the range considered. More precisely: If $s > 1$ the labor supply curve will, in the case here considered, have no turn but be monotonically increasing. And if $s < 1$ it will have one turn, the lower part being falling and the higher part rising. In fact, (10. 64) is equal to $(s - 1)$ for $r = 1$. That is to say, if $s > 1$, (10. 64) is positive for $r = 1$, and constantly increasing for higher r, so (10. 64) must be positive everywhere in the range considered. On the other hand (10. 64) is equal to $+\infty$ for $r = +\infty$. If $s < 1$, (10. 64) will therefore be monotonically increasing from a negative magnitude at $r = 1$ to a positive magnitude at $r = +\infty$, and must consequently pass zero exactly once in the range.

If the information about the shape of the labor supply curve which is obtained by an inspection of Fig. 19, is not sufficient, the labor supply curve itself can be constructed in the following way. We shall for convenience now let r denote the yearly income d i v i d e d by the number of working days in the year n. In other words r represents now a daily income, however it is not the income on a particular day, but the average income per day computed for a whole year. Similarly, we shall let s represent the outside income per day computed for the year. In the simple case at present considered we have

$$\text{(10. 65)} \qquad r = yq + s$$

and since we assume for convenience $P = 1$, the equilibrium equation becomes

$$\text{(10. 66)} \qquad w(r) = \frac{v(y)}{q}$$

We shall write the last two equations in the forms

$$\text{(10. 67)} \qquad q = \frac{v(y)}{w(r)}$$

$$\text{(10. 68)} \qquad q = \frac{r-s}{y}$$

Eliminating q from these two equations we get

$$\text{(10. 69)} \qquad y \cdot v(y) = (r - s) \cdot w(r)$$

If s is given, the last equation can be used for plotting the labor supply curve, using r as a variable parameter. First we plot the graph of the function

$$\text{(10. 70)} \qquad V(y) = y \cdot v(y)$$

Then, attributing a certain value to r, we read off from the graph of (10. 70) what abscissa point, i. e. what y, that corresponds to the ordinate $(r - s)w(r)$. This y being determined, the corresponding q is found from (10. 68). By changing r we may in this way determine a series of points on the labor supply curve corresponding to the given magnitude of s considered.

If the whole family of labor supply curves corresponding to a set of different magnitudes of s are wanted, it is, however, easier to proceed as follows. First we plot in a (y, q) system the family of curves that are obtained from the labor disutility curve by multiplying the ordinate of this curve with the s t r e t c h f a c t o r $\frac{1}{w(r)}$. For each value of r we get a different stretch factor and the magnitude of the stretch factor is simply read off from the graph of the money utility curve. The equation of such a stretched disutility curve is (10. 67). Each of the stretched disutility curves should be marked with a number, namely, the corresponding magnitude of r. See Fig. 20. The stretched disutility curves are analogous to the demand curves defined by the surface of consumption. In the case of the demand curve we could assume that the magnitude of the income was not influenced by how the income was used, that is to say, a change in the consumption of the commodity of comparison would not in itself affect the size of the total income and would therefore not affect the money utility. In the case of the stretched disutility curve we do something of the same sort: we now vary y and q, while we keep r constant. In other words, we find out what would have happened if r had not been influenced by y and q.

When the stretched disutility curves are constructed, we construct in the same (y, q) system the family of equilateral hyperbolas

$$q = \frac{r}{y} \qquad (10.71)$$

and mark each of these hyperbolas with a number, namely the corresponding magnitude of r. These curves whe shall call the labor-income curves. See Fig. 20.

This being done we select again a value of r and now mark off the intersection point between the stretched disutility curve and the labor-income curve that correspond to the selected magnitude of r. Then we select another magnitude of r and mark off the intersection point between the stretched disutility curve and the labor-income curve corresponding to *this* magnitude of r. And so forth. In this way we get a whole series of intersection points. *The curve through these points is the labor supply curve* which obtains when no outside income exists. This is the labor supply curve marked 0 in Fig. 20. Notice for instance how this curve passes through the intersection point between the stretched disutility curve $r = 4$ and the labor-income-curve $r = 4$.

Now let us perform the same operation over again, this time however pairing each stretched disutility curve with the equilateral hyperbola that carries a number which is s less than the number carried by the stretched disutility curve. If this is done, we get the labor supply curve which obtains when the outside income is s. Thus we get a whole family of labor supply curves. Each of these curves should be marked with a number, namely the corresponding magnitude of s. In Fig. 20 there are drawn four such labor supply curves, namely those corresponding to $s = -1$, 0, 1,5 and 3 (the dotted lines in Fig. 20). In order to become familiar with the principle on which Fig. 20

is constructed one should follow one of the labour supply curves, for instance the one corresponding to $s = 3$ (the labor supply curve to the left), and vizualize that each point on this curve is the intersection point between a certain labor income curve and a certain stretched disutility curve, the stretched disutility curve always (along the labor supply curve considered) bearing a number which is $s = 3$ higher than the corresponding

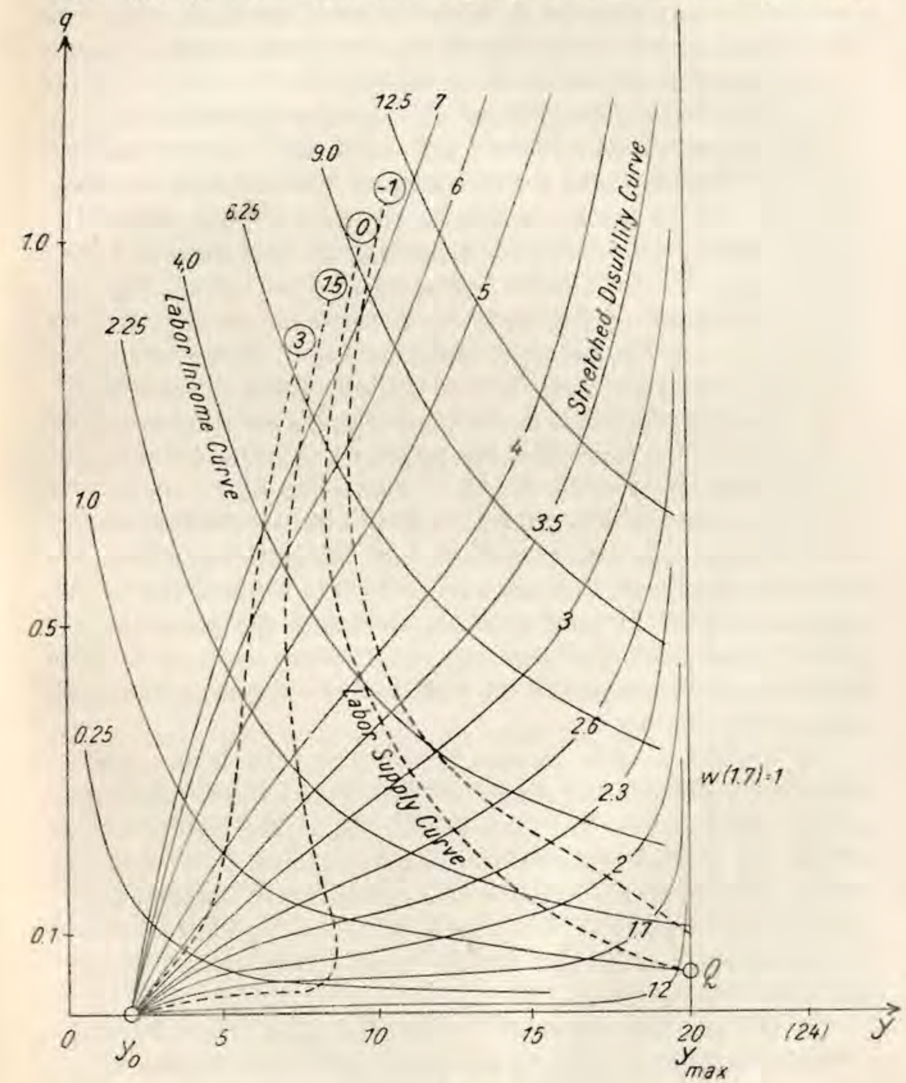

Fig. 20.

labor income curve. The labor supply curve considered passes for instance through the intersection point between the labor income curve 1 and the stretched disutility curve 4, through the intersection point between the labor income curve 4 and the stretched disutility curve 7 and so on. For the next labor supply curve in fig. 20 there is a difference of 1.5 between the number on the stretched disutility curve and the number on the corresponding labor income curve. Between the labor income curves that are actually drawn in fig. 20 one should of course imagine a whole system of intermediate curves. These may be interpolate by the eye. Similarly for the stretched disutility curves.

The labor disutility curve on which the constructions in Fig. 20 are based is the curve shown in Fig. 16. This curve is itself one of the curves in the family of stretched disutility curves, namely the curve corresponding to that income r for which $w(r) = 1$. This curve is the curve $r = 1.7$ in Fig. 20. The vertical scale of Fig. 20 is much smaller than the vertical scale of Fig. 16. The money utility curve used in the construction of Fig. 20 is the one obtained by integrating the flexibility curve M in Fig. 19. Some of the labor supply curves constructed on the basis of the curve M in Fig. 19 should, as we have seen, have a double turn. Such a curve is $s = 1.5$ in Fig. 20.

Let us analyze Fig. 20 by starting in the situation where there is no outside income. In this case the labor supply curve has a one-turn-shape. It is the curve $s = 0$ in Fig. 20. The curve ends abruptly in the point Q whose abscissa is the physiological maximum of work per day y_{max} and whose ordinate is the wage rate which is equal to the minimum of existence r_0 (measured per day) divided by y_{max}.

If a little outside income is available, the labor supply curve will still have the same one-turn-shape and the vanishing point Q will still have the same abscissa y_{max}. But the ordinate of Q will be a little less than before namely equal to $(r_0 - s)/y_{max}$. As the outside income increases this ordinate will decrease until it has become zero. This will happen when $s = r_0$, that is, when the individual has enough outside income to take care of his minimum of existence.

If the outside income is further increased, the vanishing point will suddenly jump to the point y_0 on the abscissa axis. This expresses the fact that if the minimum of existence is assured, the individual will not work more than y_0 so long as the wage rate is zero. And if the wage rate now starts rising from zero, the quantity of labor supplied will increase. The same will hold good for all higher s: All the labor supply curves for these higher magnitudes of s will start rising from the point y_0 on the abscissa axis. We can formulate this in the following proposition: I f t h e m i n i m u m o f e x i s t e n c e i s c o v e r e d b y a n o u t s i d e i n c o m e , t h e v e r y l o w e s t p a r t o f t h e l a b o r s u p p l y c u r v e i s r i s i n g . A n d t h e c u r v e s t a r t s f r o m t h e p o i n t y_0 o n t h e a b s c i s s a a x i s .

This fact can be seen not only from Fig. 20, but also by comparing the curve M of Fig. 19 with the Bernoullian flexibility curve that corresponds to any s larger than unity. The Bernoullian flexibility has a pole in $r = s$. And if $s > 1$, the actual flexibility is finite for $r = s$. Therefore, for values of r that are just a little more than s, the Bernoullian flexibility must be larger than the actual flexibility (in absolute value), and consequently the labor supply curve rising.

Whether the labor supply curve, when s is larger than the minimum of existence, shall go on rising monotonically as r increases from the point $r = s$, or whether it shall turn back (as indicated in B of Fig. 17 and by the curve $s = 1.5$ in Fig. 20), that we cannot tell without knowing more about the rate with which the money flexibility declines. There is, however, one thing we can say in general. If the supply curve does turn back, it must do so in some place in the income interval which is situated before the point r_1 where the money flexibility passes below unity. After that point there will never be any chance that the money flexibility shall catch up with the fraction $\frac{r}{r-s}$. (So long as we do not pay any attention to the very highest income range where we have not yet any definite information about the money flexibility.) And we can also say that if the labor supply curve does turn back once in the income interval before r_1, then it must later on turn forward again. More generally: for each backturn there must be a forward turn, because in the u p p e r portion of that income range for which information about the money flexibility is available, the labor supply curve must be increasing (this is indicated in Fig. 17 where the upper parts of all the three solid lines are rising).

The above discussion gives also an answer to the question of how a dole will effect the willingness to work, so far as this willingness can be expressed by the shape of the labor supply curve: a dole will push the labor supply curve to the left and possibly make the vanishing point jump from Q to y_0. Of course the preceding discussion does not treat the other and perhaps more important questions connected with the dole, how it will affect the whole mental and moral status of the workers etc. Capita taxes may be looked upon as "negative doles". The analysis of their effect is therefore also contained in the preceding discussion. In Fig. 19 are drawn three Bernoullian flexibility curves corresponding to negative magnitudes of s. And in Fig. 20 is drawn the labor supply curve corresponding to $s = -1$.

Fig. 20 gives information not only about the shape of the labor supply curve under various situations with respect to outside income, but it also gives, through the family of labor-income curves, information about the size of the income that corresponds to each point on the labor supply curve. We simply have to read off the number marked on the hyperbola that intersects the labor supply curve in the point considered. This number indicates how much income is derived from the work considered. And adding this to the outside income (which is constant along one and the same labor supply curve and equal to the number marked on this curve) we get the total income. We could also, if we want to, read off the total income directly, namely by noticing the number marked on the stretched disutility curve that intersects the labor supply curve in the point considered.

We can also utilize the families of curves in Fig. 20 for various other purposes. We can, for example, ask this question: If the outside income and the total income are both given, in what sort of a market situation must our individual then be? The answer is given by finding the intersection point between the stretched disutility curve corresponding to the given total income, and the labor-income curve corresponding to the given difference between total and outside income. This point determines both q and y. In other words it determines the wage rate that must obtain in the market in order that our individual shall adjust his work so as to reach the given total and outside

income. And it also determines the number of hours which he then will work per day.

The relationship between the various families of curves exhibited in Fig. 20 is probably the strategic point on which it would be most fruitful to concentrate the effort, if, some day, the problem of the labor supply curve should be attacked directly through a statistical investigation similar to those that have been carried through on the surface of consumption.

A further study of the more complicated case where the wage rate affects the living price must take the formula (10. 56) or (10. 58) as a starting point. And in the still more complicated cases the system (10. 45) must be used.

11. MONEY UTILITY AND THE INCOME TAX.

Should the income tax be made progressive? And if so, what should the exact rate of progression be? These are important questions which have been discussed to great lenght in the last decades, both from the practical and the theoretical angle. Most of the theoretical considerations in the matter have been in some way or another connected with the idea of money utility. In the present Section I shall point out the interrelationships between the most important of the theories propounded in this field and give a systematic discussion of the conditions under which a progressive tax may be deduced. These conditions seem sometimes to be confused. I shall also point out what conclusions can be drawn from the actual results regarding the measurement of money flexibility, given in the previous Sections. These measurements will, I believe, attach to the present discussion at least a little of that definiteness which, on account of incomplete factual information, has been lacking in the earlier discussions of the matter.

In the present discussion I shall leave aside all practical considerations relating to the effect of the income tax on the *production* activity in the society or relating to the "services" rendered to the citizen by the state (the "benefit theory"). I shall confine myself to the analysis of the sacrifice which the tax puts on the individual. In other words, I shall only try to find out what the tax rate should be if it were determined uniquely on the basis of the individual sacrifice aspect of the problem. Even with this limitation the problem does not appear as a problem whose solution can be derived from an economic-theoretical analysis of money utility alone. *The problem involves two steps, of which only the last is a purely economic-theoretical question.* The first step is the statement of the *justice-definition*. The second step is the application of this definition in an actual computation of what the tax-rate ought to be. And it is only this latter part of the problem that belongs entirely to the

domain of economic theory. To be more specific: The object of income tax graduation is to distribute the tax burden "justly." But "justly," "justice," etc., are only words. Before we can proceed to an actual computation based on money utility, the meaning of these words must be defined quantitatively *and in terms of the money utility curve*. For instance: Is it "just" that the absolute amount of total utility given up, should be the same for everybody? Or should rather the relative amount of total utility given up be the same? These are only two examples of justice-definitions. There exist many others, some of which we shall consider below. And each of these definitions entails a different consequence with respect to the determination of the tax-rate. The question of knowing which one or which ones of these different justice-definitions is most appropriate as a basis for tax distribution can not be finally settled by the economic theorist. To a great extent this is a question which the social philosopher, the moralist, and the man with common sense must settle, or at least have a considerable influence upon. Of course, economic theory can help somewhat also in the solution of this definition problem, for instance by showing what will be the consequences if we try to realize this or that abstract ideal. But alone it cannot tell which ideal should be chosen. When the question of the justice definition *is* settled, however, then the rest is a problem entirely for the economic-theorist and statistician.

This being so, the fact that we have found some statistical evidence of the monotonic decrease of the absolute value of the money flexibility from very high magnitudes for small incomes, to magnitudes less than one for higher incomes, *does not yet involve any conclusion as to whether the income tax should be progressive or not.* For one sort of justice definition this flexibility shape may entail a progressive tax, and for another justice definition it may entail a proportional, or even a regressive tax. One of the main objects of the present Section is just to discuss these various cases more closely.

In the discussion in the present Section the following notions and symbols will be used. We shall assume that the price of living is constant, and equal to unity so that the nominal income and the real income is the same. It is, of course, the notion of real income that is the important one in this connection. We shall denote the income by the symbol r which we have already adopted as the symbol of real income. The amount of the tax we shall denote

$$(\text{II. 1}) \qquad s = s(r)$$

so that $r - s$ is the spendable income left after the tax is paid. The tax rate we shall denote

$$(\text{II. 2}) \qquad \Phi = \Phi(r) = \frac{s}{r}$$

In order to take care of the case where the tax rate Φ changes with the size of the income, we have indicated that Φ is a function $\Phi(r)$ of r. If s is so small that it may be considered as infinitesimal in connection with r, we denote the tax rate by the small letter φ, thus

$$(\text{II. 3}) \qquad \varphi = \varphi(r) = \frac{s}{r}$$

To characterize the nature of the functions Φ and φ, we introduce the tax-progressivities

$$(\text{II. 4}) \qquad \check{\Phi} = \check{\Phi}(r) = \frac{d\Phi(r)}{dr} \cdot \frac{r}{\Phi}$$

$$(\text{II. 5}) \qquad \check{\varphi} = \check{\varphi}(r) = \frac{d\varphi(r)}{dr} \cdot \frac{r}{\varphi}$$

The marginal money utility and the marginal money flexibility will, as before, be denoted $w(r)$ and $\check{w}(r)$. In addition to these notions we shall also need the corresponding notions of total money utility and total money flexibility defined as follows

$$(\text{II. 6}) \qquad W = W(r) = \int_{\bar{r}}^{r} w(z)\, dz$$

$$(\text{II. 7}) \qquad \check{W} = \check{W}(r) = \frac{dW(r)}{dr} \cdot \frac{r}{W}$$

where $\bar{r}$ is a conventionally determined lower limit of the total utility integration Since the marginal money-utility becomes infinite as r approaches the minimum of existence r_0, we shall assume $\bar{r} > r_0$.

Furthermore we shall consider the rectangle-utility[1]

$$(\text{II. 8}) \qquad Q = Q(r) = rw$$

[1] Q is sometimes called "the effective utility": This term does not seem to be a good one. To most people it will convey, I believe, some sort of a marginal idea.

this is the area of the rectangle whose base line is the total income, and whose height is the marginal money utility corresponding to this income. For the rectangle-flexibility

$$(\text{II. 9}) \qquad \check{Q} = \check{Q}(r) = \frac{dQ(r)}{dr} \cdot \frac{r}{Q}$$

we have the formula

$$(\text{II. 10}) \qquad \check{Q} = 1 - (-\check{w})$$

By taking the flexibility of (II. 6) we see that we have

$$(\text{II. 11}) \qquad \check{W} = \frac{rw}{W} = \frac{Q}{W}$$

Therefore, the second order flexibility of W, namely[1]

$$(\text{II. 12}) \qquad \check{\check{W}} = \check{\check{W}}(r) = \frac{d\check{W}}{dr} \cdot \frac{r}{\check{W}}$$

is equal to

$$(\text{II. 13}) \qquad \check{\check{W}} = \check{Q} - \check{W} = 1 - (-\check{w}) - \frac{rw}{W}$$

For the second order flexibility of the marginal money utility we use the notation

$$(\text{II. 14}) \qquad \check{\check{w}} = \frac{d\check{w}}{dr} \cdot \frac{r}{\check{w}}$$

The inverse function of $w = w(r)$ we shall denote

$$(\text{II. 15}) \qquad r = \mathfrak{r}(w)$$

and the inverse function of $W = W(r)$ we shall denote

$$(\text{II. 16}) \qquad r = R(W)$$

The flexibility of the function (II. 15) namely

$$(\text{II. 17}) \qquad \check{\mathfrak{r}} = \check{\mathfrak{r}}(w) = \frac{d\,\mathfrak{r}(w)}{dw} \cdot \frac{w}{\mathfrak{r}(w)}$$

is equal to

$$(\text{II. 18}) \qquad \check{\mathfrak{r}}(w) = \frac{1}{\check{w}(r)}$$

where the connection between the arguments is $w = w(r)$. Similarly the flexibility of the function (II. 16), namely

$$(\text{II. 19}) \qquad \check{R} = \check{R}(W) = \frac{dR(W)}{dW} \cdot \frac{W}{R(W)}$$

[1] The second order flexibilities here considered are not exactly analogous to the second order divided differences considered in Section 8.

is equal to

(II. 20) $$\check{R}(W) = \frac{1}{\check{W}(r)}$$

where $W = W(r)$.

It is often useful to have at disposal certain general formulae regarding the flexibilities of a sum or a product of two functions, the flexibility of a function of a function, etc. Such formulae will be particularly useful in the discussion of the present Section. I shall therefore list some of these formulae here. The proofs are very simple, using the classical rules of derivation, so I don't think it will be necessary to develop the proofs here. The operation of taking the flexibility of a function we shall in general denote by the symbol F l e x. Thus, if $u(x)$ and $v(x)$ are two functions of the variable x, we have Flex $u(x) = \frac{du(x)}{dx} \cdot \frac{x}{u(x)}$ and similarly for $v(x)$. If it is necessary to indicate the variable with respect to which the flexibility is taken, we use the notation $\text{Flex}_{(x)}\, u(x)$. This notation being adopted we have the following rules:

(II. 21) Flex $c = 0$ where c is a constant $\neq 0$.

(II. 22) Flex $x = 1$ and Flex $\frac{1}{x} = -1$ where x is the variable with respect to which the flexibility is taken

(II. 23) Flex $cu(x)$ = Flex $u(x)$ where c is a constant $\neq 0$.

(II. 24) Flex $(u(x) \cdot v(x))$ = Flex $u(x)$ + Flex $v(x)$

(II. 25) Flex $\frac{u(x)}{v(x)}$ = Flex $u(x)$ — Flex $v(x)$

(II. 26) $$\text{Flex}\,(u(x) + v(x)) = \frac{u(x)\ \text{Flex}\ u(x) + v(x)\ \text{Flex}\ v(x)}{u(x) + v(x)}$$

(II. 27) $\text{Flex}_{(x)}\, G(g(x)) = \text{Flex}_{(g)}\, G(g) \cdot \text{Flex}_{(x)}\, g(x)$, where G is a function of g, and g again a function of x

It will be noticed that with the flexibilities it is the formulae for products and quotients that are the simplest, while with the ordinary derivatives it is the formulae for sums and differences that are simplest.

I now proceed to a discussion of a series of more or less plausible justice-definitions and of the consequences which these definitions entail for the determination of the income-tax.

(I) T h e a r e a - d i f f e r e n c e p r i n c i p l e (the principle of "equal sacrifice"). Let us consider an individual who knows that at two different occasions he will find himself with the incomes r_1 and r_2 respectively. And let us imagine that he has the choice of giving up as a tax e i t h e r the sum s_1 in the first situation o r the sum s_2 in the second situation. Evidently he will choose to pay his tax in the first or second situation accordingly as $W(r_1) - W(r_1 - s_1)$ is less than or larger than $W(r_2) - W(r_2 - s_2)$. Therefore if we should determine s_1 and s_2 in such a way that it becomes indifferent to the individual whether he pays s_1 in the first situation or s_2 in the second situation, s_1 and s_2 must be such that

(II. 28) $$\int_{r_2 - s_2}^{r_1} w(z)\,dz = \int_{r_1 - s_1}^{r_2} w(z)\,dz$$

Now let us consider a whole series of situations with different incomes r_i $(i = 1, 2, 3 \ldots)$. And let us imagine that to each of these situations is associated a tax of the amount s_i $(i = 1, 2, 3 \ldots)$, respectively. The individual is under the obligation that he must pay o n e of these taxes, but he can himself choose that one of the situations where he wants to do it. If we should determine the tax amounts s_i in such a way that it becomes indifferent to the individual w h e r e he pays the tax, then the s_i must be determined in such a way that

(II. 29) $$\int_{r_i - s_i}^{r_i} w(z)\,dz = c$$

where c is a constant independent of i. If the money utility curve $w(r)$ is given and the constant c is given, there is by (II. 29) associated a certain s_i to every r_i. Moreover, we see that s_i depends on i only by the fact that it depends on the magnitude of r_i. We can express this by saying that the tax amount s is a function $s = s(r)$ of the income r, and that this function is defined by the equation

(II. 30) $$\int_{r-s}^{r} w(z)\,dz = c$$

The last equation we can also write

(II. 31) $$\int_{(1-\Phi)r}^{r} w(z)\,dz = c$$

By (II. 31) the tax rate Φ is defined as a function of r. (II. 31) can also be written

$$W((1 - \Phi)r) = W(r) - c \tag{II. 32}$$

The principle expressed by (II. 30), (II. 31) or (II. 32) we shall call the a r e a - d i f f e r e n c e p r i n c i p l e, because it expresses the fact that the difference between the area under the money utility curve before and after the levy of the tax is a constant independent of r.

So far we have not considered different individuals. From a formal point of view, the direct comparison between the utilities referring to different individuals is not absolutely necessary in the present analysis. It is sufficient, and, from the point of view of theoretical invulnerableness, even desirable to define the relation between s and r by considering the same individual in different situations instead of different individuals. However, in the practical application the relation between s and r thus determined would have to be applied also to cases where the different magnitudes of r stand for the incomes of different individuals. In speaking about the principle expressed by (II. 30) we may therefore use the elliptic expression that' 'everybody gives up the same amount of utility". But we should remember that theoretically the construction of the relationship between s and r does not involve a direct intra-individual utility comparison.

The area-difference principle expressed by (II. 30) is one way of formulating the justice definition. I don't think that it is a particularly plausible formulation. The main theoretical objection against it would be that it is built on an analysis of the behavior of an individual who only has the obligation to pay a tax in o n e of the many income situations considered. We shall presently see that if he must pay a tax in e v e r y o n e of the situations, which is, of course, the more realistic assumption, then we arrive at a quite different principle. But although the area-difference principle does not appear theoretically as a very well founded principle, we must include it in our analysis because it has been used by several eminent economists. Foremost among these are E m i l S a x[1].

If the area-difference principle is adopted, the explicit expression for the tax-rate Φ as a function of the income r becomes

$$\Phi(r) = 1 - \frac{R(W(r) - c)}{r} \tag{II. 33}$$

[1] "Die Progressivsteuer." Zeitschrift für Volkswirtschaft, etc. 1892, and "Die Wertungstheorie der Steuer." Ibid. 1924.

where R is the inverse function defined by (II. 16). If we take the flexibility of (II. 33) with respect to r, using the general flexibility relations (II. 26) and (II. 27) we get

$$\check{\Phi} = \frac{r-s}{s}\left(1 - \frac{Q(r)}{Q(r-s)}\right) \tag{II. 34}$$

If we develop (II. 33) as a power series in the constant c, we get

$$\Phi = \frac{c}{rw} + \ldots \tag{II. 35}$$

The first term of this expansion is the expression for the tax-rate in the case where the tax is very small (strictly speaking, infinitesimal). This tax rate we have denoted φ, so that we have in the present case

$$\varphi = \frac{c}{rw} \tag{II. 36}$$

If we take the flexibility of (II. 36), using (II. 24) and (II. 25) we get

$$\check{\varphi} = (-\check{w}) - 1 \tag{II. 37}$$

The formula for the infinitesimal tax rate (II. 36) we could also have developed by the following direct argument: If the individual with the income r pays the infinitesimal tax s, he gives up an amount of utility equal to $s.w(r)$. Putting this equal to a constant c, independent of r, we get $sw = c$, that is $\frac{s}{r} = \frac{c}{rw}$, which is (II. 36).

From the preceding formulae we can immediately read off the conditions that the tax shall be progressive. The tax will be progressive, proportional or regressive, accordingly as the progressivity $\check{\Phi}$, or in the case of an infinitesimal tax $\check{\varphi}$, is positive, zero or negative. From (II. 37) we therefore see that the necessary and sufficient condition that an infinitesimal tax shall be progressive in the income point r (when the area-difference principle is adopted) is that the money flexibility in this income point is larger than unity. If the infinitesimal tax shall be progressive everywhere in a certain income range, it is consequently necessary and sufficient that the money flexibility is larger than unity in absolute value everywhere in this range. From (II. 34) is seen that this is also a s u f f i c i e n t condition that the finite tax shall be progressive everywhere in the range. If the money flexibility is larger than unity in absolute value, we see indeed from (II. 10) that Q is constantly decreasing, so that $Q(r-s)$ is larger than $Q(r)$ and consequently $\check{\Phi}$ determined by (II. 34) positive. It should be noticed that this condition regarding the finite tax is not necessary. If the total level of the tax is high (that is, the level of s is high), it may happen that $Q(r-s)$ is larger than $Q(r)$ f o r a n y r i n t h e r a n g e, although the range contains small subintervals where the function $Q(r)$ is increasing. However, if we request that we shall have progressivity, no matter what the general level of s is, then, of course, the condition considered becomes both necessary and sufficient. In the following we will frequently encounter this situation where the condition considered is necessary and sufficient so far as the infinitesimal tax is concerned, but only sufficient so far as a finite tax is concerned.

If the statistical evidence regarding the shape of the money utility curve which we have obtained, is significant, the area-difference principle will lead to a heavily progressive tax for small incomes and a regressive tax for medium incomes. I feel that this implausible result is due to the fallacy of the area-difference principle rather than to errors in the statistical results.

The preceding formulae give, of course, not only information about whether or not the area-difference principle leads to a progressive tax, but gives also a means of actually constructing the tax rate function $\Phi(r)$ when the money utility function $w(r)$ is known.

(II) T h e a r e a - r a t i o p r i n c i p l e (the principle of "proportional sacrifice"). Instead of deciding that everybody shall give up the same absolute amount of total utility as we did in the area-difference principle, we may decide that everybody shall give up the same p e r c e n t a g e of the total utility which he had before the tax was levied. In order not to run into difficulties with infinite total utilities we have to define the total utility in proportion to which the amount of lost utility is measured, by introducing a conventional lower limit of integration $\bar{r} > r_0$ as was done in formula (II. 6). This convention being adopted, the principle here considered may be expressed by the formula

$$\int_{r-s}^{r} w(z)\,dz = c \int_{\bar{r}}^{r} w(z)\,dz \tag{II. 38}$$

where c is a constant independent of r. We may write the same equation in the form

$$W((1-\Phi)r) = (1-c)W(r) \tag{II. 39}$$

The principle of justice definition expressed by these formulae, we shall call the area-ratio principle. This principle formed the basis of the income tax theory of the Dutch School of economists, in particular represented by N. G. P i e r s o n[1], C o r t v a n d e r L i n d e n[2], and A. G. S t u a r t[3].

Solving the equation (II. 39) we get the following explicit expression for Φ

$$\Phi(r) = 1 - \frac{R((1-c)\,W(r))}{r} \tag{II. 40}$$

If we take the flexibility of (II. 40) with respect to r, using the general flexibility relations (II.26) and (II.27) we get

$$\check{\Phi} = \frac{r-s}{s}\left(1 - \frac{\check{W}(r)}{\check{W}(r-s)}\right) \tag{II. 41}$$

If we develop (II. 40) as a power series in the constant c, we get by (II. 20)

$$\Phi = \frac{c}{\check{W}(r)} + \dots \tag{II. 42}$$

So that the expression for the infinitesimal tax now becomes

$$\varphi = \frac{cW}{rw} \tag{II. 43}$$

And the flexibility of φ, becomes equal to

$$\check{\varphi} = \frac{rw}{W} + (-\check{w}) - 1 \tag{II. 44}$$

The formula (II. 43) for the infinitesimal tax rate could also have been developed by the following direct reasoning: If the tax is infinitesimal the utility given up is φrw. Putting his equal to a constant c times W, we get (II. 43).

From the preceding formulae we can deduce the following conditions: In order that the infinitesimal tax shall be progressive everywhere in a certain income range, (when the area-ratio principle is adopted), it is necessary and sufficient that in

[1] Grondbeginselen der Staatshuishoudkunde (2d ed, 1886).
[2] Die Theorie der Belastningen (1885).
[3] Bijdrage tot der Theorie der Progressive Inkomstenbelastning (1889).

every point in this range the money flexibility (counted positive) plus the ratio between the rectangle-utility and the total utility, is larger than unity.

If this condition is fulfilled we see from (II. 13) that $\breve{W}$ is monotonically decreasing over the range considered, so that $\breve{\Phi}$, according to (II. 41) must be positive. The condition considered is therefore also a sufficient condition that the finite tax is progressive in every point in the range.

Comparing (II. 37) with (II. 44) we see that the condition now considered is easier to fulfill than the progressivity condition in the case of the area-difference principle. Consequently: *If the area-difference principle gives a progressive tax, the area ratio principle must a fortiori do so.* And the rate of increase of the tax rate is larger in case of the area-ratio principle. Furthermore we see that the difference between the two conditions become all the smaller, the closer the conventional lower limit of integration $\bar{r}$ is to the physical minimum of existence. W increases indeed as $\bar{r}$ decreases. And if $\bar{r} \to r_0$ it will in most cases be plausible to assume that W tends towards infinity. In this case the first term in (II. 44) drops out so that (II. 37) and (II. 44) become identical. We can interpret this by saying that if $\bar{r} \to r_0$, *the area-ratio principle is transformed into the area-difference principle.* The situation can also be visualized intuitively by noticing that if $\bar{r} \to r_0$, so that the integral in the right member of (II. 38) tends towards infinity, this integral becomes independent of the upper limit r. That is to say the entire right member in (II. 38) becomes independent of r, which means that we have come back to the justice formulation (II. 30).

Pierson and Cort van der Linden believed that the progressivity condition according to the justice-principle they had adopted (the area-ratio principle) only was that the money utility was decreasing. This error was corrected by Cohen-Stuart, who found the correct progressivity criteria in the case where the area-ratio principle is adopted. The rest of his analysis has now little interest since it is, to a large extent, built on the Bernoullian utility function, which according to all now available evidence is very far from representing the true course of the money utility.

In the search for other possible justice definitions, a natural thing to do would be to see what the situation becomes when the above difference and ratio principles are applied, not to total but to *marginal* money utility. There does not seem to be any well founded theoretical consideration that would lead up to such a justice definition, but the consequences which may be derived from these marginal principles seem to have at least so much heuristic value as those derived from the corresponding total-utility principle. Moreover, the marginal principles have already been introduced into the literature, so it will be well to state exactly what their nature is.

(III) *The marginal-difference principle.* This principle states that the absolute amount by which the marginal money-utility is increased as a consequence of the tax levy, shall be the same for all the tax payers. In other words, we shall have

$$w(r - s) = w(r) + c \qquad \text{(II. 45)}$$

Solving equation (II. 45) we get

$$\Phi(r) = 1 - \frac{\mathfrak{r}(w(r) + c)}{r} \qquad \text{(II. 46)}$$

where $\mathfrak{r}(w)$ is the inverse function defined by (II. 15). We consequently have

$$\breve{\Phi} = \frac{r - s}{s}\left(1 - \frac{\breve{w}(r) \cdot w(r)}{\breve{w}(r - s) \cdot w(r - s)}\right) \qquad \text{(II. 47)}$$

If we develop (II. 46) as a power series in c we get

$$\Phi = \frac{c}{(-\breve{w})\, w} + \dots \qquad \text{(II. 48)}$$

For the infinitesimal tax rate we therefore have

$$\varphi = \frac{c}{(-\breve{w}) \cdot w} \qquad \text{(II. 49)}$$

and consequently

$$\breve{\varphi} = (-\breve{\breve{w}}) + (-\breve{w}) \qquad \text{(II. 50)}$$

where $\breve{\breve{w}}$ is the second order marginal money flexibility defined by (II. 14). From (II. 50) we deduce: The condition that the infinitesimal tax shall be progressive (when the marginal difference principle is adopted) is that the marginal money flexibility (counted algebraically with its proper sign) plus the second order marginal money flexibility shall be negative. This is at the same time the condition that the product $(-\breve{w})w$ shall be decreasing. If this condition is fulfilled we see from (II. 47) that the finite tax must also be progressive.

Since all our statistical evidence shows that both $(-\breve{w})$ and w are decreasing, the product $(-\breve{w})w$ must a fortiori be decreasing. We have therefore good reason to say that: The marginal-difference principle leads to a progressive tax practically everywhere over that part of the income range for which observations of the money utility are at present available.

The condition here considered can evidently also be formulated by saying that the logarithmic derivative

$$-\frac{d\,w(r)}{d \log r} \tag{II. 51}$$

shall be decreasing.

(IV) The marginal-ratio principle. According to this principle the tax shall be determined in such a way that the marginal money utility is increased in the same proportion for all taxpayers. In other words we shall have

$$w(r-s) = (1+c)w(r) \tag{II. 52}$$

where c is a constant.

This principle has formed the basis of the theory of K. Schönheyder[1]. This theory is especially interesting because it is built on the following two assumptions regarding the shape of the money utility curve: (1) The money flexibility (counted positive) is decreasing over the whole income range; (2) The money-flexibility (counted positive) is less than unity, except for the poor. Both these assumptions are in perfect agreement with the statistical evidence now available. Schönheyder's assumptions are particularly remarkable because they are contrary to what most economic theorists have believed to be true. Even as late as in 1920 Edgeworth[2] wrote that it was strongly to be presumed that "the satisfaction as dependent on income increases at a rate which diminishes more rapidly than does the rate of increase pertaining to the simple function proposed by

[1] Statsökonomisk Tidsskrift, Oslo (1907). Less clearly the same principle has been stated by Robert Meyer in his book. Die Principien der gerechten Besteuerung (1884). See in particular p. 312 and p. 332. See also Oskar Jæger: Finanslære, Oslo 1930, p. 269. Meyers statements are rather vague. His expressions p. 330 »Intensitätsdifferenz«, »kleinere Schritte« etc. suggest that he is here thinking of the marginal difference principle.

[2] Economic Journal, p. 399 (1920).

Bernoulli ... (and) the presumption ... is now commonly, though not universally accepted." In other words, it was "commonly though not universally accepted" that the money flexibility (counted positive) was larger than unity. (In the original Bernoullian utility function no deduction was made for the minimum of existence, so that the absolute value of the money flexibility according to this function is constantly equal to unity. It is this fact that Edgeworth refers to.)

From (II. 52) we deduce

$$\Phi = 1 - \frac{\mathrm{r}((1+c)\,w(r))}{r} \tag{II. 53}$$

and consequently

$$\breve{\Phi} = \frac{r-s}{s}\left(1 - \frac{\breve{w}(r)}{\breve{w}(r-s)}\right) \tag{II. 54}$$

Developing (II. 53) as a power series in c we get

$$\Phi = \frac{c}{(-\breve{w})} + \ldots \tag{II. 55}$$

so that

$$\varphi = \frac{c}{(-\breve{w})} \tag{II. 56}$$

The last formula can also be deduced directly by noticing that if the tax is infinitesimal, the percentage with which the marginal money utility will increase as a consequence of the tax is $(-\breve{w})$ times the percentage with which the income diminishes. This is nothing else than the definition of the marginal money flexibility. Putting the percentage increase in the marginal money utility, namely $(-\breve{w})\varphi$, equal to a constant c, we get (II. 56).

From (II. 56) we immediately deduce

$$\breve{\varphi} = (-\breve{\breve{w}}) \tag{II. 57}$$

This shows that the necessary and sufficient condition that the infinitesimal tax shall be progressive in any income point in a certain range (when the marginal-ratio principle is adopted) is that the money flexibility (counted positive) is decreasing, everywhere in the range. From (II. 54) is seen that this is also a sufficient condition that the finite tax shall be progressive.

Schönheyder found this condition correctly. But he made the mistake of believing that when the area-ratio principle is adopted, the condition for a progressive tax is that the marginal money flexibility (counted positive) shall be *increasing*. As we have seen the condition in case of the area-ratio principle is still a condition regarding the *size*, not regarding the direction of change of the marginal money flexibility. The condition is indeed, according to (II. 44), that the absolute value of the marginal money-flexibility added to the ratio between rectangle utility and total utility, shall be larger than unity.

Since all the statistical results regarding the money-flexibility indicate that its absolute value is monotonically decreasing (at least so far up in the income range as the evidence goes), the consequence of the marginal-ratio principle should be a progressive tax.

(V) *The principle of complete levelling* (the "minimum sacrifice" principle). All the principles considered so far have been more or less of a heuristic sort. None of them have been built on a basic analysis of the behavior of the taxpayer under the condition which he actually has to face, namely, the condition that a tax will be levied in *each* of the income situations in which he will find himself. If this more realistic assumption is adopted, we are led to a "justice principle" of an entirely different sort.

Let us first suppose that the individual knows that at two different occasions he will be in the income situations r_1 and r_2 respectively. And let us suppose that he has to pay a tax, the total amount of which is *given* and equal to s, while the installments in which the tax shall be paid is a matter of choice for the individual. In other words, he shall pay s_1 in the first situation and s_2 in the second situation, such that

$$s_1 + s_2 = s \qquad \text{(II. 58)}$$

But otherwise s_1 and s_2 are arbitrary. This principle, it seems to me, is as true an expression for "justice" in taxation as we can get it by a simple theoretical scheme: The individual does not choose the total amount of the tax, *but he chooses the distribution of it*.

How will he now determine the two magnitudes s_1 and s_2? Evidently he will do it in such a way as to maximize the expression

$$\int_{\bar{r}}^{r_1 - s_1} w(z)\, dz + \int_{\bar{r}}^{r_2 - s_2} w(z)\, dz \qquad \text{(II. 59)}$$

In other words, the problem is to determine s_1 and s_2 so as to make (II. 59) as large as possible with the side relation (II. 58) fulfilled. This leads to the condition

$$w(r_1 - s_1) = w(r_2 - s_2) \qquad \text{(II. 60)}$$

More generally, if there are a series of income-situations r_i ($i = 1, 2, 3 \ldots$) each with a certain tax s_i ($i = 1, 2, 3 \ldots$) attached to it, we get the condition

$$w(r_i - s_i) = c \qquad \text{(II. 61)}$$

where c is a constant independent of i. This defines s_i as a function of r_i. And since r_i does not depend on i otherwise than through r_i, we see that we can consider r as a continuous variable, and s as a function of r defined by

$$w(r - s) = c \qquad \text{(II. 62)}$$

Since w is a monotonic function, (II. 62) is equivalent with

$$r - s = \mathfrak{r}(c) = \text{constant} \qquad \text{(II. 63)}$$

In other words, according to this principle, the tax shall be put on so as to leave the same amount of all incomes. *All incomes are simply levelled down to the uniform magnitude* $\mathfrak{r}(c)$. We may call this the principle of complete levelling. This principle has been studied in particular by F. Y. *Edgeworth*[1] and T. N. *Carver*[2]. But it seems that the principle is rather general in the influence which it has exerted on the thought in this field. Consciously or unconsciously I think it is back of a great part of the argument, and particularly I think it is this principle, if any, that is back of the commonsense conviction that the income tax ought to be progressive. Of course, everybody will admit that in practice, the principle ought to be modified because of the disastrous effect which its strict adoption would have on the production activity in society when this activity is organized on a private capitalistic basis. But that is another question. If one only takes account of the subjective utility side of the problem and assumes that the situation can be fairly well represented by the utility function $w(r)$ of a single variable, then the conclusion according to the principle here discussed will be a com-

[1] The Pure Theory of Taxation. Part III. Economic Journal, p. 550 (1897).

[2] The Minimum Sacrifice Theory of Taxation. Political Science Quarterly (1904).

plete levelling of incomes. Of course, this will involve a negative tax for the smallest income.

It may be objected against the levelling principle that it introduces into the taxation a social and political point of view which goes far beyond the pure fiscal considerations. This it may be claimed, is expressed by the fact that the levelling principle does not satisfy the "leave them as you find them" criterion. As I see it such a criterion as "leave them as you find them" has no meaning at all. There would be only one way of rigorously living up to this criterion namely to levy no tax at all. As soon as any tax at all is levied, the taxpayers are not, strictly speaking, left as we find them. And as soon as we try to modify the criterion so as to make it possible to change the situation in *some* respects, we are immediately facing the question of *which* features of the situation that shall be left unchanged. In other words, we are again back in a discussion of the various possible principles mentioned above. And when we get back into that situation, it seems that, so far as the logical argument is concerned, the levelling principle. holds a position which is at least as strong as that held by any of the other principles.

In practice, it would obviously be necessary to make some modification in the leveling principle. Not only on account of the capitalistic production argument mentioned above, but also because the assumption that $w(r)$ is a function only of r, is too narrow. In practice we will have to take account of the effect of the *habitual income* for instance. (This notion is defined more fully in Section 12.) And it may well be that if this and possibly other modifications are taken into account it will turn out that some combination of the principles (I)-(IV) offer a workable and convenient approximation to the correct solution even if the levelling principle is adopted as the logical basis of the theory.

It is quite obvious without any further analysis that the levelling principle will lead to a progressive tax, provided only that the money utility is decreasing, which is a fact that cannot seriously be doubted. (Strictly speaking, we only need to assume that the money utility function is monotonic so that its inverse function is uniquely determined.) However, we want more than a criterion for whether or not the tax should be progressive.

We also want an exact expression for the rate of the progression. This can easily be obtained from (11. 63). The explicit expression for the tax rate as determined from this equation is

$$\Phi(r) = 1 - \frac{\mathfrak{r}(c)}{r} \qquad (11.\ 64)$$

And the tax progressivity becomes

$$\check{\Phi} = \frac{\mathfrak{r}(c)}{r - \mathfrak{r}(c)} = \frac{r - s}{s} \qquad (11.\ 65)$$

In other words, the tax rate is simply represented by an equilateral hyperbola with the horizontal $\Phi = 1$ and the Φ axis as asymptotes. The tax rate is monotonically increasing (for positive r). It passes from negative to positive for $r = \mathfrak{r}(c)$, and approaches unity as r increases. In the case of the levelling principle it has no meaning to speak of an infinitesimal tax in the same sense as when one of the other principles is adopted.

The distinction between the various "justice" principles discussed above has not always been kept clear. Even such an eminent student of the progressive income tax as R. A. Seligman comments on the difference between the various principles in a surprisingly superficial way. He says, for instance, that the difference between Emil Sax's theory and the theory of the Dutch School is "une simple différence de mots"[1]. And he objects against Edgeworth's distinction between "equal sacrifice" and "proportional sacrifice" (which in this connection is just the same as the distinction between Sax's theory and the Dutch theory) on purely terminological grounds. Seligman does not appear to have seen anything else in Edgeworth's distinction than just a matter of nomenclature. He does not seem to be aware of the fact that the two principles in question differ with respect to a well defined property that entails a definite difference in the consequences which can be deduced regarding the progressivity of the tax.

(VI) *The general marginal principle.* The modifications that it is necessary to apply to the levelling principle in practice can be conceived of in different ways. Both the marginal difference principle and the marginal ratio principle can be looked upon as expressing such a modification. Possibly

[1] I quote from p. 229 of the French edition of Seligman's book "Progressive Taxation".

there may be other modifications worth while a closer study. The question therefore arises: Can we formulate a general principle that includes all conceivable modifications of the levelling principle, and at the same time offers a useful means of classifying these modifications and exhibiting those features in the adopted justice definition on which the resulting tax progressivity depends? The most natural procedure in formulating such a principle would be, it seems, to focus the attention on the nature of the relationship that is postulated between the marginal money utility before and after the tax is levied. This relationship can be expressed by a function. Let $w(r)$ and $w(r - s)$ be respectively the marginal money utility before and after the levy of the tax. Then $w(r - s)$ will be a certain function of $w(r)$. We denote this function $G(w)$ so that we have

(II. 66) $$w(r - s) = G(w(r))$$

Assuming a particular form of the function $G(w)$ means adopting a particular sort of justice-definition. $G(w)$ is simply a precise way in which to state the justice definition. We may therefore call $G(w)$ the justice-function. We shall characterize the nature of this function by introducing its flexibility

(II. 67) $$\check{G}(w) = \frac{dG(w)}{dw} \cdot \frac{w}{G(w)}$$

If we require that the tax shall be distributed in such a way that the amount left of a large income is never smaller than the amount left of a smaller income, and if we further assume that $w(r)$ is a monotonically decreasing function, then $G(w)$ will be a single valued and monotonically non-decreasing function. In this case there is a unique correspondence between the tax rate function $\Phi(r)$ and the function $G(w)$. On the one hand we now have

(II. 68) $$G(w) = w\,([1 - \Phi(\mathfrak{r}(w))]\mathfrak{r}(w))$$

The first letter w in the right member of (II. 68) stands for the function sign expressing how the ordinate depends on the abscissa in the marginal money utility curve. The second and third letter w in the right member of (II. 68) is a variable, namely, the argument of $G(w)$.

On the other hand we have

(II. 69) $$\Phi(r) = 1 - \frac{\mathfrak{r}(G(w(r)))}{r}$$

(II. 68) expresses the function $G(w)$ in terms of the function $\Phi(r)$. And (II. 69) expresses $\Phi(r)$ in terms of $G(w)$.

The function $G(w)$ is a perfectly general expression for the nature of the justice definition. For instance: The marginal difference principle, the marginal ratio principle and the principle of complete levelling can be looked upon as being obtained from (II. 66) by putting respectively $G(w) = w + c$, $G(w) = (1 + c)w$ and $G(w) = c$. And the formula (II. 69) is a general expression for the way in which the tax rate depends on the justice definition. There are three functions entering into this problem: The marginal money utility function $w(r)$, the justice-function $G(w)$ and the tax-rate function $\Phi(r)$. If the two first functions are known, then $\Phi(r)$ may be determined, namely, by (II. 69). But the knowledge of $w(r)$ alone is not sufficient to determine $\Phi(r)$. This is another way of expressing the fact we have already insisted upon, namely, that our statistically determined money utility curve in itself neither proves nor disproves the "justice" of a progressive income tax, it will do so only when a particular form of $G(w)$ is assumed.

Taking the flexibility of (II. 69) we get

(II. 70) $$\check{\Phi} = \frac{r - s}{s}\left(1 - \frac{\check{w}(r)}{\check{w}(r - s)}\,\check{G}(w(r))\right)$$

In the case of the marginal difference, marginal ratio and levelling principle respectively, we have $\check{G} = \frac{w}{w+c}$, $\check{G} = 1$ and $\check{G} = 0$. Introducing this in (II. 70) we derive (II. 47), (II. 54) and (II. 65) as three special cases of (II. 70).

In the case where the tax s is infinitesimal we have $w(r - s) = w(r)\,[1 - \check{w}(r) \cdot \varphi]$. Inserting this in (II. 66) we get

(II. 71) $$\varphi = \frac{G(w) - w}{(-\check{w})w}$$

In the case where the marginal difference or the marginal ratio principle is adopted, we have respectively $G(w) = w + c$ and $G(w) = (1 + c)w$. Introducing this in (II. 71) we get (II. 49) and (II. 56) as special cases.

From (II. 71) we get

$$\breve{\varphi} = \frac{G(\breve{G} - 1)}{G - w}\breve{w} - \breve{\breve{w}} \qquad \text{(II. 72)}$$

This gives the general progressivity condition. As special cases of (II. 72) we deduce (II. 50) and (II. 57).

Can the formulation of the "justice" problem which we have here adopted, namely, the formulation (II. 69) or the equivalent formulation (II. 70), be of any practical use? I think it can. Not because I believe there is any great chance that we shall ever be able to determine the exact nature of the function $G(w)$ on a priori grounds, but because I think (II. 69) or (II. 70) offer a natural and plausible formulation of the interpolation problem inherent in the graduation of the income tax rate. Suppose, for instance, that it has been decided by political or social considerations, by common sense judgement, or in some other way, that the three incomes \$3,000, \$50,000 and \$1,000,000 shall pay respectively 1%, 7% and 50% income tax. If this question is settled, it seems that it would be a plausible procedure to determine the intermediate percentages by interpolation in the following way.

If the marginal money utility is known through statistical observation (or for the higher income range through the interview method mentioned in Section 12), then the above numerical data will by (II.68) determine uniquely three points on the curve $G(w)$. Through these three points we may then interpolate the function $G(w)$ by some more or less empirical method, for instance by assuming $G(w)$ to be a second degree parabola. (This assumption includes amongst others the marginal difference, marginal ratio and levelling principles as special cases.) And when $G(w)$ is determined in this way, the tax-rate function $\Phi(r)$ is determined by (II. 69). It is true that this procedure involves some arbitrariness, namely, in the interpolation form adopted for $G(w)$. But this arbitrariness is much less harmful than the arbitrariness we would have introduced by interpolating the tax rate function $\Phi(r)$ directly. What we do when we interpolate $G(w)$ is, so to speak, to utilize our knowledge of the money utility curve to transfer the arbitrariness in the interpolation process to a less vital part of the system.

The various principles of justice here discussed are summarized in table 7.

Table 7.
Principles of "Justice".

Name of principle	Author	"Justice" definition	Condition under which the principle in question entails a progressive tax. (The numbers in parenthesis give reference to formulae in the text).
Area-difference principle ("Equal sacrifice")	Emil Sax	$\int_{r-s}^{r} w(z)dz = c$	$(-\breve{w}) > 1$ (II. 37)
Area-ratio principle ("Proportional sacrifice")	N. G. Pierson Cort van der Linden A. G. Cohen-Stuart	$\int_{r-s}^{r} w(z)dz = c\int_{\overline{r}}^{r} w(z)dz$	$\frac{rw}{W} + (-\breve{w}) > 1$ (II. 44)
Marginal-difference Principle		$w(r-s) = w(r) + c$	$(-\breve{w}) + (-\breve{\breve{w}}) > 0$ (II. 50)
Marginal-ratio Principle	K. Schönheyder Robert Meyer	$w(r-s) = (1+c)w(r)$	$(-\breve{\breve{w}}) > 0$, (i. e. $(-\breve{w})$ decreasing) (II. 57)
Levelling-principle ("Minimum sacrifice")	F. Y. Edgeworth T. N. Carver	$w(r-s) = c$	$(-\breve{w}) > 0$ (i. e. w decreasing)
General Marginal Principle		$w(r-s) = G(w(r))$	$G(\breve{G}-1)\breve{w} > (G-w)\breve{\breve{w}}$ (provided $G > w$) (II. 72)

12. FURTHER PROBLEMS.

In the present paper I have on purpose avoided all the more or less philosophical questions connected with the notion of utility as a quantity. The statistical results themselves as determined through the methods employed, have been taken as a sufficiently exact definition of the utility notion involved. There are several important aspects of the problem which are thus left out of the picture. These I intend to take up in the above mentioned paper to be published in the transactions of the Norwegian Academy of Science.

Then there is the problem of a dynamification of the whole theory. As already mentioned, a first approach to this aspect of the problem is contained in my paper "Statikk og Dynamikk i den Økonomiske Teori"[1], and a further elaboration will be given in the Norwegian Academy of Science paper referred to. Then again there is the problem of *structural* changes. An analyses of such changes is somewhat connected with but not identical with the dynamic analysis. The kind of thing I have here in mind may best be illustrated by an example: Suppose we have a spiral spring. If we attach a weight to the end of the spring, we will notice a lengthening of the spring. And by experiment we may find a definite law connecting the size of the weight and the extent of the lengthening. This law would hold good for weights that are not too heavy *and not left on the spring for too long a time*. If we leave a very heavy weight on the spring for a long period of time, we would, when we got back, find that the original law connecting weight and lengthening had changed. In economics we have many phenomena of a similar sort. For instance: If a person gets more income, the representative point on his money utility curve will move along the curve downwards and to the right. But if the point is left in this position for a considerable time, this will have the effect of *shifting the whole curve upwards*, so that the money utility as represented by the

[1] Nationalökonomisk Tidsskrift. Kbhvn. 1929. pp. 321—379.

ordinate of the point in question again starts increasing, even though the new income remains constant. There are several ways in which we could attempt to draw this phenomenon into the analysis. One way would be to consider the nominal money utility w as a function not only of the actual income ϱ and the actual price of living P, but also of the *habitual income* $\bar{\varrho}$ and the habitual price of living $\bar{P}$. By the habitual income I mean the average income which the individual has had for the last years. In practice, it may perhaps be defined as a simple moving average for the last 5 or 10 years.

We may even connect the notion of habitual income with the notion of components in time-series, and define the first order habitual income as the average income during the last "40 month" cycle, and further define a second order habitual income as the average income during the last "10 year" cycle, and so on. Similarly for the habitual price of living. Which one, if any, of these notions that will be fruitful as a tool of further analysis can only be decided by future investigations.

If the notion of habitual income is adopted, we would have to write the nominal money utility in the form

$$\omega(\varrho, P, \bar{\varrho}, \bar{P}) \tag{12.1}$$

This function would still satisfy a proportionality equation in ϱ and P analogous to the equation (2. 13), namely

$$\lambda\omega(\lambda\varrho, \lambda P, \bar{\varrho}, \bar{P}) = \omega(\varrho, P, \bar{\varrho}, \bar{P}) \tag{12.2}$$

But in $\bar{\varrho}$ and $\bar{P}$ the proportionality equation would be of a little different form. The factor λ outside the function sign would be lacking because the money utility w is not a notion that is measured per unit of $\bar{\varrho}$, w is still measured per unit of ϱ, so we would have

$$\omega(\varrho, P, \lambda\bar{\varrho}, \lambda\bar{P}) = \omega(\varrho, P, \bar{\varrho}, \bar{P}) \tag{12.3}$$

Therefore, if we let

$$w(r, \bar{r}) = \omega(r, 1, \bar{r}, 1) \tag{12.4}$$

be the special function that expresses how the money utility depends on actual income r and habitual income $\bar{r}$ when both the actual living price and the habitual living price are equal to unity, then we would have

$$\omega(\varrho, P, \bar{\varrho}, \bar{P}) = \frac{1}{P}\, w\left(\frac{\varrho}{P}, \frac{\bar{\varrho}}{\bar{P}}\right) \tag{12.5}$$

The problem is therefore reduced to a problem regarding the function of two variables $w(r, \bar{r})$.

Similarly we could consider the marginal utility of the commodity of comparison u as a function of actual consumption x and habitual consumption $\bar{x}$, and we would have an equilibrium equation of the form

$$w(r, \bar{r}) = a. u(x, \bar{x}) \tag{12. 6}$$

where $a = P/p$ is the inverted relative commodity price as before. Since we now have added two dimensions to the problem, the actual statistical determination of the relationship would, of course, offer new and interesting difficulties.

Even in the static theory, without structural complications there is a problem that needs a further study than the one presented in the preceding Sections; namely, the actual numerical determination of the money flexibility for higher incomes. In such a study it will be essential I think to introduce what may be called group-utilities. We may, for instance consider the expenditure group (A) consisting only of current expenditure for food, clothing, shelter, and so on. In short, all those things that enter into the type of budgets that formed the basis of the statistical study in Section 7.

Next we would have an expenditure group (B) including all the expenditures entering into (A), and in addition certain expenditures connected with wants that may cumulate in intensity over a long period of years, and then, if the means are available, be satisfied in a relatively short period of time. Typical for this type of wants is the want for travel (for pleasure or for study). This sort of expenditure plays an important rôle in the life of the middle class, particularly the intellectual part of the middle class. Such things as "expenditure" for saving might also be classified under the group (B). In short (B) will include a large number of items which are not of a current sort, and which therefore cannot be given a significant analysis unless by considering the interrelationship between income and expenditure budgets for a long period of years. But still the items entering into consideration here are only such items that have a definite connection with an ultimate consumption purpose.

In a third group (C) we would have to list not only all the above expenditures but also expenditures that have no definite

relation to the actual consumption of the individual, but are nevertheless of great concern to him. Typical in this category is Mr. Henry Ford's expenditures. The fraction of his total expenditure which has to do with his personal consumption is infinitesimal. When he thinks about his various income and expenditure items it is not because he relates them to his own consumption but because these items mean success or failure to a certain activity with which he has identified his whole life and ambition. He pursues this activity not because it is a means of "earning his living", but because the activity itself has fascinated him. Maybe it is the mere satisfaction of seeing things grow bigger, maybe the feeling of power, or perhaps some idealistic idea of the "social usefulness" of such an undertaking. The underlying motive is unessential in this connection. The main thing is that the way in which he looks upon his income and expenditure has nothing to do with his personal consumption, at least if we take the word consumption in its usual sense. Mr. Ford's activity is only an extreme example in our analysis. There are many other types resembling it: The activities of nearly every great business man or speculator has some features of this sort.

When we get to the higher income groups it is therefore essential to specify what sort of money-utility we have in mind. Is it the sort of money utility to which we arrive when we limit the use of money to expenditures of the type (A), or shall we allow all expenditures of the type (B), or even all expenditures of the type (C)? Of course, the groups, (A), (B) and (C) are only rough groups, put up for convenience. Other and more specialized groups may be of interest in particular connections. If we specialize the groups more and more, we finally get down to the individual commodity utilities μ, defined in Section 2, i. e. the individual commodity utilities measured per dollar's worth.

Conceivably the notion of money utility may perhaps be defined in exact terms even if we think of income and expenditure of the type (C). But the magnitude of the money utility as thus defined would vary so fundamentally from one moment to another, according to the particular business circumstances under which the individual finds himself, that it would, at best, be very difficult to fit this notion into a systematic theoretical analysis. The situation is much simpler with a money utility of the type (B). And for a money utility of the type (A) the theoretical difficulties we have to face when we consider large incomes do not seem to be any greater than those we have already faced for small incomes. But so far as the actual measurements is concerned the higher income groups offer some difficulties, also for the money utility notions (A) and (B). The main difficulty lies in a lack of reliable statistical data. For the higher incomes we shall therefore, I believe, have to revert to some sort of interview-method in order to obtain actual information about the money flexibility. I developed such a method in 1922 and have since tried it out occasionally on friends. A few results are given in Statsökonomisk Tidsskrift, Oslo, 1926, p. 332. But so far, the investigations along these lines have not been carried through in a systematic and extensive way. I have worked enough with this method, however, to become most hopeful that it will furnish a very valuable supplement to and check on the statistical methods developed in the previous Sections. I even think that in a sense the results obtained by the interview method may be looked upon as more reliable than anything else, because they are obtained by going more directly to the base of the individual judgments. The method has further the great advantage that is does not require any great amount of work and may therefore, if properly organized, easily be carried through on a great scale over large geographical areas and large classes of the population.

In short, the method consists in questioning in a certain systematic way persons of different social classes and different income ranges as to what sort of exchange transactions they would agree to if their income for some reason or another should change. By an intelligent and systematic formulation of the questions the magnitude of the money flexibility can be determined within very narrow limits of error.

The questions can be varied in many different ways. In each case they will, of course, have to be adopted to the intellectual and cultural status of the person questioned. For instance, the questions must be accompanied by much longer and more detailed explanations when the person questioned is a man in the street than when he is, say, a professional economist. The latter will, of course, understand much more quickly what it

is all about. In the case of the man in the street it may perhaps also be necessary to tell a little about the purpose and the deeper meaning of the questions. Otherwise the questions may look a little queer to him. The following is a brief outline of a set of questions I recently used on a friend of mine, a college professor living in the U. S. with an income of $6000 a year, wife and no children.

"You know the standard of living which you can keep with your present income. Try to realize this standard clearly in your mind. Only think of that part of the income which you usually consume. Disregard saving. And disregard also great occasional expenditures such as a travel or the like. Only think of the current consumptive use of the income in that year when the income flows. Now imagine that your income next year will be 50% larger than this year. Year after next it will again drop down to its usual level. The increased income next year must be consumed in that year. Otherwise it is lost. There does not exist any possibility of saving. (To be tactful one may operate with percentage changes. This is sufficient for the flexibility determination. But if the person does not object, it is better to use absolute changes. The questions are then easier to formulate. And in the results obtained one will know the absolute size of the income involved.)

This being so, you will probably want to transfer some of next year's income to this year. In other words, you would probably want to use some of your increased income now instead of crowding all the increased consumption on next year. Imagine that it is impossible for you on regular conditions to obtain a loan in a bank. You negotiate only with me. I am the only one who can help you out of the situation. I will put at your disposal $500 this year. This, however, I won't do without a heavy premium. I want more money back from you next year than I give you now. Do you accept the deal if I ask $1,000 back next year, that is a premium of 100%? (Answer: No.) Do you accept if I only ask $510 back next year? (Yes.) $800? (No.) $ 550?(Yes.), etc."

It is easy to understand when one is approaching the indifference point. The time taken to think the question over then becomes longer. With a little patience one can easily determine an upper limit that will be decidedly refused, and a lower limit that will be decidedly accepted. The average between this upper and lower limit can then be taken as the magnitude of the amount of next year's income which will be exchange-indifferent with the amount offered as an increase to this year's income. In the actual case here considered it turned out to be \$600. And the ratio between the amount received this year and the amount given up next year will express approximately the ratio between the two money utilities, provided the amounts involved in the transaction are not too large. Since 600/500 = 1.2, the average money flexibility (counted positive) over the income range from \$6000 to \$9000 for the man questioned in this case is approximately equal to

$$(-\breve{w}) = \frac{\log 1.2}{\log 1.5} = \frac{0.079}{0.176} = 0.448$$

The whole questioning can now be repeated under the assumption that the income next year is smaller than the income this year. We may even try out questions with a whole set of different alternatives for next year's income.

Whatever the nature of the questions we have the following general formula for the determination of the money flexibility:

$$(-\breve{w}) = \frac{\log (\delta_2/\delta_1)}{\log (r_2/r_1)} \tag{12. 7}$$

where r_2 and r_1 are respectively the large and the small of the two incomes that occur in the questions, and δ_2 and δ_1 are respectively the large and the small of the two amounts that are exchange-indifferent. Approximately (12. 7) can be considered as the p o i n t flexibility in the income point $r = \sqrt{r_1 r_2}$.

Formulating in this way a series of questions on assumptions that range from next year's income being virtually a bare minimum of existence and up to next year's income being very high, a w h o l e f l e x i b i l i t y c u r v e m a y b e c o n s t r u c t e d.

All the flexibility curves I have determined in this way have shown the same characteristic features as those exhibited by the statistically determined flexibility curves, namely a monotonic decrease in the absolute value of the flexibility from magnitudes higher than one for small incomes to magnitudes less than one for larger incomes.

ERRATA.

In the printing some of the accents ˇ have dropped out:

Page	Formula	
69	(8. 18)	u should have one accent
69	(8. 20)	u " " " "
79	Tab. 6	w " " " "
103	Fig. 18	w " " " "
		W " " " "
125	(11. 50)	φ " " " "
		the last w should have two accents
135	(11. 50)	" " w " " " "
135	(11. 72)	the last w in the second line should have two accents.

Also add the letter v after the three dots in formulae (10. 20) and (10. 27) on pages 88 and 89, and in the line preceding formula (10. 27).