ECONOMIC THEORY

AND

OPERATIONS ANALYSIS

fourth edition

WILLIAM J. BAUMOL

Department of Economics
Princeton and New York Universities

1977

PRENTICE-HALL, INC., Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging in Publication Data

BAUMOL, WILLIAM J.
Economic theory and operations analysis.

(Prentice-Hall international series in management)
Includes bibliographies and index.
1. Microeconomics. 2. Economics, Mathematical.
3. Operations research. 4. Mathematical analysis.
I. Title
HB135.B38 1977 330'.01'84 76-46591
ISBN 0-13-227132-X

Printed in the United States of America

10 9 8 7 6

PRENTICE-HALL INTERNATIONAL, INC., London
PRENTICE-HALL OF AUSTRALIA, PTY. LIMITED, Sydney
PRENTICE-HALL OF CANADA, LTD., Toronto
PRENTICE-HALL OF INDIA PRIVATE LIMITED, New Delhi
PRENTICE-HALL OF JAPAN, INC., Tokyo
PRENTICE-HALL OF SOUTHEAST ASIA PTE. LTD., Singapore
WHITEHALL BOOKS LIMITED, Wellington, New Zealand

Demand Curves, Utility Surfaces, and Indifference Maps

9

1. Demand Curves

The demand curve is among those devices of economic theory which have found frequent employment in applied economics. In its traditional form, it sums up the response of consumer demand to alternative prices of a product—it can tell management what may be expected to happen to the demand for one of its outputs if the price of that item is changed.

This information is summarized in a graph (the demand curve itself) which shows how much will be demanded at every possible (hypothetical) price over the relevant range (Figure 1). For example, point D_0 on the demand curve DD' indicates that at price OP_0 the consumer, or group of consumers, for whom the curve is drawn will wish to purchase OX_0 units of the product.

Several features of the demand curve should be noted:

1. It is customary to represent the price level on the vertical axis and the quantity demanded on the horizontal axis.[1]

[1] This arbitrary convention would seem to be an inappropriate arrangement because in the present discussion we treat quantity demanded as the dependent variable and price as the independent variable. The origin of this practice is that this curve together with a supply curve has traditionally been used in the analysis of price determination in a competitive industry as described in Chapter 16, Section 3. However, even here the price cannot be considered a dependent variable since, in the supply-demand analysis, price and quantity are determined simultaneously.

2. The graph may refer to the demand of an individual consumer, or it may describe the aggregated demands of a group of consumers constituting a market. The market demand curve is obtained from the demand curves of the individuals composing it by adding up, for each price in turn, the quantities all the individual consumers demand at that price. That is, one adds the individual demand curves *horizontally* to obtain the market demand curves.

3. The demand curve assumes that there is no change in the values of other pertinent variables. Specifically, prices of other goods and consumer incomes are among the other things that are assumed to remain equal.

4. The graph depicts the situation at a single point in time, say 4:33 p.m. on June 12. Hence, all but one of the prices and quantities must be hypothetical—the curve must generally answer the "iffy" question: "If price were OP, how much would this (these) consumer(s) buy?"

5. The curve is generally assumed to have a negative slope. In economic terms, this is the plausible assertion that, other things being equal, more of the commodity would be demanded (OX_1 rather than OX_0) if the price were lower (OP_1 instead of OP_0). However, two exceptions should be mentioned: Cases of snob appeal and cases where consumers judge quality by price. Commodities like expensive jewelry may be purchased precisely because their price is high, and a fall in their price might reduce their snob

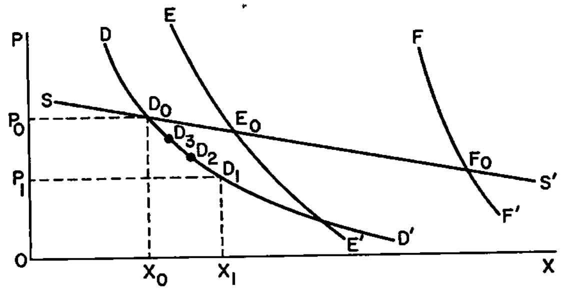

Figure 1

appeal and therefore, perhaps, their sale (although enough poorer consumers might then be induced to buy the items to make up for the loss of affluent customers). Similarly, when consumers have no ability to judge the quality of a good directly and use price as an indicator of quality, as they probably often do, a reduction in its price may cut into the demand for a good. The negative slope of the usual demand curve will be discussed again later in this chapter, along with another exceptional case that has been discussed widely in the theoretical literature.

It is important to recognize the *isotemporal* character of the demand curve—the fact that every one of its points represents one of the hypothetical possibilities available *for some given moment.* This property is a common feature of *all* the relationships used in static analysis, i.e., of virtually all the relationships in this book. This characteristic is troublesome both for empirical application and for comprehensibility of economic analysis by laymen. Yet, this attribute of our static relationships is not imposed as a mere caprice or as a means to introduce analytical complexity or elegance.

Rather, we use relationships that are isotemporal because most of our analysis concerns itself one way or another with *optimal choice.* Behavioral analysis in economic theory usually proceeds on the assumption that the decision-maker always arrives at decisions that are optimal in terms of his objectives; welfare analysis seeks to determine what decisions are optimal from the point of view of the public interest; etc. But we saw in Chapter 1 that, by definition, optimization analysis consists of the explicit or implicit comparison of the financial consequences of the choices available to the decision-maker. The need for isotemporal relationships follows at once. Suppose a seller is considering the selection of a price for 4:33 p.m. on June 12 and the price possibilities under consideration are \$12.99, \$14.99, and \$16.99. Obviously, the relevant consideration for the decision-maker is how much he will sell if he selects the first of these possibilities, how much he will sell instead if he selects the intermediate-price candidate, etc. But that is precisely the information that an isotemporal demand curve gives—it tells us how much would be sold *if some one or another of the possible* prices is the one that is selected *for the moment to which the decision pertains.* In sum, since optimization requires explicit comparison of the possibilities available at the time to which the decision applies, the relevant relationships (curves) *must* describe (and contrast) just those hypothetical possibilities for that time period.

2. *Shifting Demand Curves: Demand Functions*

Since the demand curve is defined to pertain only to a particular time, its shape and position are likely to change with the passage of time. At one moment DD' is the relevant demand curve, but at another instant the curve has the shape EE'. Such a change is described as a *shift* in the demand curve. This is contrasted with a *movement along* a demand curve, say from point D_0 to D_1.

A shift in a demand curve is normally accounted for by a change in the value of some of the variables which affect demand. For example, a rise in consumer income can lead to an upward shift in the demand curve from DD' to EE'. This means that at *any given price*, such as OP_1, the consumer(s) will demand more than before the shift. It should be noted, however, that if price happens to rise sufficiently at the same time, say to OP_0, consumers may end up buying less despite an outward shift in the demand curve. In such a case, the shift in demand is accompanied by an offsetting movement *along* the curve.

Besides income, many other variables can affect the position of the curve. A change in the amount of advertising, a change in price or quality, or the advertising approach of a competing product—even a change in the weather—can shift a demand curve. Some of the relevant variables may even be intangible and unquantifiable—for example, a change in consumer tastes can cause a shift in a demand curve—although we may prefer to go behind this phenomenon and seek the variables which account for the taste change.

To summarize, demand is a function of many variables such as price, advertising, and decisions relating to competing and complementary products. The relationship which describes this entire many-variable interconnection is called the *demand function*. By contrast, the demand curve deals only with two of these variables, price and quantity demanded, and ignores the others, or, rather, assumes that their values are held constant. Indeed, the distinction between a movement along and a shift in a demand curve may be described in terms of the variables involved. Any change in quantity demanded which results only from a variation in price is a movement along the curve, whereas change in the value of any other variable in the demand *function* is likely to shift the demand *curve*.

Several concluding observations about the distinction between shifts in and movements along the demand curve are relevant:

1. Phrases such as "a rise in demand" are ambiguous and should be avoided, since it is not clear whether they refer to a shift in or a movement along the curve.

2. The distinction is often important for applied economics. For example, the statement that a reduction in demand is deflationary is valid only if it refers to a downward shift in the demand curve, since a leftward movement (a decrease in quantity demanded) along a negatively sloping demand curve must, by definition, be concurrent with a rise in price. One can find cases where this has been misunderstood by legislators who thereupon have made nonsensical statements on inflation policy. Similarly, the sort of increase in demand which is most eagerly hoped for in a business firm will involve a shift in the demand curves for its products. In fact, it will normally result from autonomous changes in the values of the variables which are entirely outside the management's control. There is usually

some cost to the firm when the quantity of its products demanded increases as a result of a change in the firm's advertising expenditure, or in the incidental services which it provides to its customers, or in some other of its demand curve-raising activities. But an upward shift in demand which occurs because of a rise in national income or favorable weather comes to the company free.

3. The possibility that demand curves can frequently shift implies that a statistical investigation of the shape of such a curve requires the aid of relatively powerful methods. There is a serious difficulty in the obvious approach, which involves our taking price quantity data for a number of months and plotting them on a graph. For example, if in October OX_0 units were sold at (average) price OP_0 (point D_0 in Figure 1) whereas November and December sales were represented by points E_0 and F_0, respectively, this method would have us draw in the statistical "demand curve" SS', which, as we can see, really resembles *none* of the true demand curves (the DD' curve for October, the EE' curve for November, and the FF' December curve). The difficulty is that the true demand curve has shifted over this period. The naïve statistical method which has just been described does not even indicate this fact and it certainly offers us no means of correcting for it.

More sophisticated methods have been designed to deal with this so-called *identification* problem, and, more generally, with the econometrics of simultaneous equation estimation. However, it is desirable to defer discussion of these techniques until later.[2]

3. Elasticity: A Measure of Responsiveness

The most obvious piece of information we desire of a demand function (or from economic relationships of other varieties) is an indication of the effect on the "dependent" variable of a change in the value of one of the other variables. In the case of the demand curve, this involves measurement of the response in quantity demanded which can be expected to result from a given change in the price of the commodity.

The obvious measure of responsiveness is, of course, what we may call the marginal demand contribution of a price change, $\Delta x/\Delta p$, or the corresponding derivative, dx/dp, the change (fall) in quantity demanded caused by a unit change (rise) in price. It will be observed, incidentally, that this measure is the reciprocal of the slope of the demand curve $\Delta p/\Delta x$ (or dp/dx), so that the flatter the demand curve, the greater will be the value of this measure of responsiveness to price change. This peculiarity results from the oddity in the conventional drawing of the demand curve

[2]See the next chapter and its appendix.

which has already been noted—the fact that the value of the apparently dependent variable, quantity, is measured along the horizontal axis, and that of the independent variable, price, along the vertical axis. We might well get a better intuitive grasp of the degree of price responsiveness of a demand curve if the diagram were turned on its side.

In any event, the obvious measures of responsiveness, $\Delta x/\Delta p$ and dx/dp, are subject to a drawback which has led theorists to employ instead another measure—elasticity. The difficulty with, say, $\Delta x/\Delta p$ is that it deals with the absolute changes in quantity and price, which makes it difficult to compare the responsiveness of different commodities. Commodities are measured in different units—labor in hours, land in acres, and whiskey in fifths or quarts. There is no simple way of comparing a 20,000-quart increase in the quantity of Scotch demanded with a 3,000-acre rise in that of land. In economics it is difficult, because of the very nature of the animal, to impose uniform units on all the relevant magnitudes as is done in physics.

But the problem extends beyond the dissimilarity of units, because, even in the measurement of price change, the magnitudes are not readily comparable. Consider a 1-cent fall in the price of a package of bubble gum and an equal fall in the price of an automatic dishwasher. We might not be surprised to find bubble gum sales booming when habitués discover the bargain in this brand of the confection, but it is difficult to believe that a one-penny reduction in dishwasher prices would even be noticed. Though the measure $\Delta x/\Delta p$ would therefore almost certainly yield a much higher number in the case of chewing gum than in that of major household appliances (a much greater change in quantity demanded per penny price reduction), we would surely hesitate to conclude from this that the demand of the former was significantly more price-sensitive.

Theorists have concluded, from such considerations, that an appropriate measure of responsiveness of demand to price changes should employ percentage rather than absolute change figures. A 1 *per cent* (rather than a one-penny) fall in price then becomes the standard of comparison, so that the change in dishwasher price in our illustration is discounted as an insignificant price fall in comparison with that of the bubble gum.

Employing these percentage terms we have the definition

$$\text{price elasticity of demand for item } X = -\frac{\text{percentage change in quantity of } X \text{ demanded}}{\text{percentage change in the price of } X}.$$

The only peculiarity in the definition which remains to be explained is the presence of the minus sign before the fraction. This is inserted to make the elasticity number nonnegative. When the demand curve is negatively inclined, a rise in price (Δp positive) will lead to a fall in quantity (Δx

negative) so that in our elasticity fraction the numerator and denominator will be of opposite sign. Therefore the fraction will be a negative number, and a minus sign is needed to make the number positive. The insertion of this sign in the elasticity formula is, then, just a matter of linguistic convenience.

For our purposes it is necessary to define the elasticity measure somewhat more specifically. The percentage change in any quantity, x, is defined as 100 times the change in x, i.e., as $100\Delta x$, divided by x. For example, if quantity rises from 10 to 15, we have $\Delta x = 15 - 10 = 5$ and the percentage rise in x is $100\Delta x/x = 500/10 = 50$ per cent. Similarly, the percentage change in p is given by the expression $100\Delta p/p$. Therefore we have, by our definition of elasticity,

$$\text{price elasticity of demand} = -\frac{100\Delta x/x}{100\Delta p/p} = -\frac{\Delta x/x}{\Delta p/p}$$

(since we can divide both numerator and denominator by 100).

Moreover, since division by a fraction, $\Delta p/p$, is the same as multiplication by its reciprocal, $p/\Delta p$, we obtain the expression

(1) $$\text{price elasticity of demand} = -\frac{\Delta x}{x} \cdot \frac{p}{\Delta p} = -\frac{\Delta x}{\Delta p} \cdot \frac{p}{x}.$$

This expression, which will be used throughout the remainder of the elasticity discussion, helps now to describe two different elasticity concepts: *point elasticity* and *arc elasticity*. Arc elasticity is a measure of the *average* responsiveness to price change exhibited by a demand curve over some finite stretch of the curve such as D_0D_1 in Figure 1. One complication is inherent in the concept. In the elasticity formula (1), when price changes from P_0 to P_1 it is clear that $\Delta x = X_1 - X_0$, the change in quantity bought (Figure 1), and that $\Delta p = P_1 - P_0$. But what are the values of x and p? Since a range of values of x occurs along arc D_0D_1, no unique value of this variable is called for by the definition. It is customary for this purpose to use the average of the two end values of x, that is, to set $x = (X_1 + X_0)/2$, and to do the same for the percentage change in price. Hence the arc elasticity of demand is defined by the expression

$$-\frac{\Delta x}{\Delta p} \cdot \frac{p}{x} = -\frac{X_1 - X_0}{P_1 - P_0} \cdot \frac{(P_1 + P_0)/2}{(X_1 + X_0)/2}$$

so that, multiplying both numerator and denominator by 2, we have

(2) $$\text{arc (price) elasticity of demand} = -\frac{X_1 - X_0}{P_1 - P_0}\frac{P_1 + P_0}{X_1 + X_0}.$$

Point elasticity of demand is the corresponding concept for each particular point on the demand curve. But, at any such point there is no change in price ($\Delta p = 0$) or in quantity. We therefore define point elasticity in much the same way as the derivative concept in Chapter 4. That is, we take point elasticity to be the limit of the arc elasticity figure as the arc D_0D_1 is made smaller and smaller, first being cut down to D_0D_2, then to D_0D_3, etc. We thereby arrive at the definition

$$\text{(3)} \qquad \text{point price elasticity of demand} = -\frac{dx}{dp}\cdot\frac{p}{x},$$

where the derivative dx/dp has been substituted for $\Delta x/\Delta p$ in the elasticity definition (1).

Before leaving the question of definitions, it is well to point out that the elasticity concept can be (and has been) adapted to measure responsiveness in variables other than quantity and price. For example, we may measure the responsiveness of the supply, s, of some commodity to a change in interest rate, i, as

$$\text{interest elasticity of supply} = -\frac{\text{percentage change in supply}}{\text{percentage change in interest rate}}$$

$$= -\frac{ds}{di}\cdot\frac{i}{s}.$$

Similarly, when the price, p_j, of one commodity, j affects the quantity demanded, x_k, of another commodity k it is customary to define

$$\text{cross elasticity of demand} = \frac{dx_k}{dp_j}\cdot\frac{p_j}{x_k}.$$

The reader should try defining such concepts as the income elasticity of imports and the interest elasticity of investment.

4. Properties of the Elasticity Measure

The basic elasticity formula (1) permits us to see a relationship between an elasticity measure and the corresponding marginal measure of responsiveness, $\Delta x/\Delta p$, with which the elasticity discussion began. Elasticity is simply the marginal measure multiplied by the fraction $-p/x$.

This observation, in turn, helps us to see one of the peculiarities of the elasticity measure. Consider a straight-line demand curve like that in Figure 2. It is tempting to guess that the elasticity of such a demand

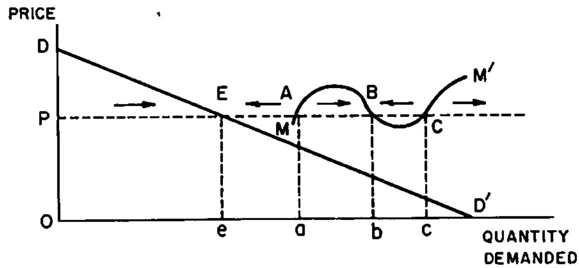

Figure 2

curve is the same throughout the length of the curve. Such constancy does hold for the marginal measure of price responsiveness, $\Delta x/\Delta p$, since it is the reciprocal of the slope of the demand curve which does not change along a straight line. There are two cases in which the elasticity measure also behaves in this way: If a demand curve is vertical (a fixed quantity demanded no matter what the price, so that $\Delta x/\Delta p = 0$), its elasticity is zero throughout, and at the other extreme, a horizontal demand curve has "infinite elasticity." But in any other straight-line case, such as DD' in Figure 2, elasticity is not constant. Indeed, it varies continuously from zero at point D' on the horizontal axis to any number as high as we like when we get close to the vertical axis (so that elasticity is said to approach infinity as we move toward point D)!

The reason for the variability in the elasticity of a straight line is readily seen from our last elasticity formula. We have just noted that the first fraction in this expression, $\Delta x/\Delta p$, retains the same value throughout the graph. But that is not true of the second fraction, p/x. At point D', we have $x = OD'$ and $p = 0$ so that $p/x = 0$, and hence the price elasticity of demand is zero also. As we move toward the left along the demand curve, the numerator of p/x increases while the denominator, x, approaches zero. Hence the value of the fraction grows larger and larger without limit and the same is consequently true of the price elasticity of demand.[3] We conclude that, except in the zero elastic vertical case and the infinite elastic horizontal case, elasticity of demand is certainly not constant along a straight-line demand curve. This complication is a price which we pay for using percentage figures instead of absolute figures in the elasticity measure.

[3] Note that I have avoided speaking of the elasticity being infinite at point D, where $x = 0$. Here the elasticity is not even defined because an attempt to evaluate the fraction p/x at that point forces us to commit the sin of dividing by zero. The reader who has forgotten why division by zero is immoral may recall that division is the reverse operation of multiplication. Hence, in seeking the quotient $c = a/b$ we look for a number, c, which when multiplied by b gives us the number a, i.e., for which $cb = a$. But if a is not zero, say $a = 5$, and b is zero, there is no such number because there is no c such that $c \times 0 = 5$.

However, even here there is an important compensation. Though the connection may at first not be obvious, the following theorem will lead us to a type of curve whose elasticity *is* constant throughout and which will offer us some useful insights.

Elasticity Proposition 1: Given any segment of the demand curve, a change in price within that segment will have no effect on the product px (price multiplied by quantity demanded) if and only if the elasticity of demand throughout the range is exactly equal to unity. More specifically, a change in price from p_0 to p_1 will yield $p_0x_0 = p_1x_1$ if and only if the elasticity of the arc D_0D_1 is unity, and each and every intermediate price change will also leave px unaffected if and only if point elasticity is unity at *every* point along this arc.[4]

The product px represents the amount which the consumer would spend and which the seller would therefore receive if quantity x were bought at price p. This theorem therefore states that if the price elasticity of his demand is unity, a fall in price will induce the consumer to increase his purchases by exactly the amount needed to keep his total outlay the same as it was initially. This is certainly plausible intuitively, for we may view a price reduction as having an expenditure-increasing effect (more demanded) and an expenditure-reducing effect (a lower price paid for each unit purchased). When the elasticity of demand is unity, the percentage fall in price is, by definition, exactly equal to the percentage rise in quantity demanded, and it is therefore believable that these two effects will then exactly offset one another, as the theorem asserts.

The theorem describes, implicitly, the type of demand curve along which elasticity of demand is constant. Specifically, it tells us that the elasticity will take the constant value unity along any curve characterized by the equation $px = K$ (any constant). Such a curve is called a *rectangular hyperbola* and has the shape of one of the curves depicted in Figure 3 (where different curves correspond to different values of K). That a

[4] The following argument demonstrates both Propositions 1 and 2 in terms of point elasticity. The arc elasticity proofs are just matters of tedious algebraic manipulation.

Proof (of Propositions 1 and 2): We want to determine the relationship between the elasticity, E, and the effect of a change in price, p, on total expenditure, $px = p \cdot f(p)$, where $x = f(p)$ is the equation of the demand curve. Then the effect of a change of a price on total expenditure is (by the formula for the derivative of a product)

$$\frac{d[pf(p)]}{dp} = p\frac{df}{dp} + f\frac{dp}{dp} = p\frac{dx}{dp} + x\frac{dp}{dp} = p\frac{dx}{dp} + x = x\frac{p}{x}\frac{dx}{dp} + x = -xE + x$$
$$= x(1 - E).$$

Hence, a change in price will leave xp constant, $d[p \cdot f(p)]/dp = 0$, if and only if $E = 1$. Similarly, $d[p \cdot f(p)] > 0$ if and only if $E < 1$ (demand inelastic), etc.

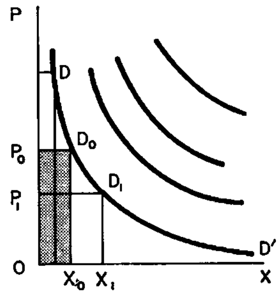

Figure 3

curve $px = K$ is of such a shape can be seen by noting that if our demand curve is DD', then, e.g., at price OP_0, consumer expenditure $OX_0 \times OP_0$ is represented by the area of the shaded rectangle $OX_0D_0P_0$ (= height OP_0 multiplied by width OX_0). Similarly, at price OP_1 expenditure is depicted by area $OX_1D_1P_1$. Since expenditure is constant along a demand curve of unit elasticity, it follows that all such rectangles must be equal in area. Hence the unit elastic demand curve must approach the axes of the diagram asymptotically, for as such a rectangle gets taller it must become narrower in order for its area to remain equal to that of its fellow expenditure rectangles. Moreover, such a demand curve must not touch either axis for at a point of intersection with an axis either p or x is zero so that px must equal zero rather than K. It can be shown, incidentally, that demand curves of constant elasticity 2 or $\frac{3}{4}$ or any other number are asymmetrical relative to the axes but roughly similar in shape to rectangular hyperbolas.

We may use a geometric argument to extend our first elasticity theorem as follows:

Elasticity Proposition 2: If a demand curve has elasticity less than unity (it is *inelastic*), a *rise* in price will *increase* consumer expenditure, px, and *vice versa*. If the curve has an elasticity greater than unity (it is *elastic*), a *fall* in price will increase consumer expenditure and *vice versa.*

A proof has already been provided in a footnote. The two elasticity theorems just given lie behind much of the use of the elasticity concept in applied economics. They are met, for example, in the analysis of problems of taxation, international trade, and pricing by private business. As a simple illustration, note that it will not ordinarily pay a firm to reduce the price of a product whose demand is inelastic, because this price reduction will tend to increase the number of units sold and hence the firm's total raw material, labor, and other costs, but, by Proposition 2, *it will also decrease the firm's revenue px*—clearly a losing proposition! As a second illustration, consider a country suffering from a "gold shortage." In such a case, popular writers often recommend that the country devalue its currency, thus making its products cheaper and hence leading foreigners to import more. There are a number of complications to be considered, but the one which is relevant for our purposes is the possibility that the elasticity of the foreigners' demand for that country's exports may be less than unity. Thus the country may find, after devaluing, that though it is shipping more abroad, it is actually earning less gold than before!

PROBLEMS

1. Given the definition of arc elasticity of demand as shown in Equation (2), prove that if total expenditure is constant along an arc, so that $P_1X_1 = P_2X_2$, then the arc elasticity must be unity.
2. If a firm's price elasticity of demand is greater than 1, can you say from this alone whether a fall in its price is profitable? Why?

5. *Utility Analysis of Demand*

Economic theory has long sought to go behind the obvious and observable demand phenomena which are summed up in the demand function in an attempt to explain these observations in terms of the structure of consumer desires. It seemed immediately apparent that there is some connection between demand and the *utility* of the commodity, i.e., the subjective benefit which the consumer obtains from its possession. But to classical economists this connection appeared to be limited largely to the fact that items totally without utility would not be demanded at all. To show that there is little or no connection with price, they called attention to the fact that water, which is essential to life and therefore to be considered of very great utility, commands only a very low and often no more than a zero price, whereas diamonds, whose utility was said to be less than that of water, are notoriously expensive.

This "diamond-water" paradox was explained by an analysis which was the focal point of the economic literature at the turn of the century. It was argued that the price of a commodity was determined not by its *total* but by its *marginal* utility. For this discussion it is convenient to evaluate the marginal utility of a commodity, X, in money terms (the amount of money the consumer is just willing to give up for another unit). The connection between price and marginal utility is that if to some rational consumer the marginal utility of some item, X, when he holds A units of X is more than its price, he can increase his welfare by purchasing some more units of X. This is so because, by definition, in these circumstances he receives more value than he gives up in such an exchange. Similarly, if the marginal utility of an Lth unit of the commodity is less than its price, the consumer can benefit by buying less than L units. He should, therefore, always buy such an amount of X that its marginal utility is equal to its price.[5]

The marginal utility theorists carried their analysis considerably further. For one thing, they argued, largely on introspective grounds, the more we possess of a commodity, the less we value an additional unit—the famous "law" of *diminishing marginal utility*. Partly, it was stated, this is so because we give priority to more highly valued uses—if we have only one piece of cake, we feed it to our child; if we have two, we divide it between husband and wife and a third we give to our mothers-in-law.

The marginal utility analysis of pricing and the diminishing marginal utility proposition can quickly dispose of the diamond-water paradox. The relative scarcity of diamonds results in their having a high *marginal* utility and, therefore, a high price, while the relative abundance of water means that its *marginal* utility and, consequently, its price will be low despite its high *total* utility.

This law of diminishing marginal utility was also used as an explanation of the negative slope which is alleged to characterize most simple demand curves. The argument is that if the marginal utility of a commodity falls when the consumer purchases more of the item, he can only be induced to buy more of a good by a fall in its price.

Another important function of the law of diminishing marginal utility arises out of the need for *second-order* equilibrium conditions. It will be recalled (Section 5 of Chapter 4) that a marginal equation such as "price equals marginal utility" is not enough to guarantee that the consumer is getting the maximum possible utility for his money. There may be several purchase levels at which the equation holds. For example, referring back to Figure 2, we see that if the marginal utility curve has the peculiar shape of curve MM' and price is OP, then there are three purchase levels Oa, Ob, and Oc at which marginal utility equals price. However, these are not all optimal purchase levels. In fact, two of these, Oa and Oc, are extremely *dis*advantageous to the consumer! If, for example, the consumer increases his purchase quantity from Oa (direction of an arrow), he enters a region where marginal utility exceeds price, and it will pay him to increase the amount he buys even more. Only when he gets to the true equilibrium point B (where marginal utility is diminishing—the curve of MM' has a negative slope) does it pay him to stop increasing his purchases. Similarly,

[5] More formally, if x is the amount of X purchased, and if $u(x)$ is the total utility of the purchase (measured in dollars), the consumer presumably seeks to maximize the difference between this total utility and his expenditure, px, i.e., he seeks to maximize $u(x) - px$. Differentiating with respect to x and setting the result equal to zero, we obtain $du/dx - p = 0$, i.e., $p = du/dx$, the marginal utility of x units of good X.

from quantity Oc it pays the consumer either to increase or decrease his purchases—not to stay at Oc (direction of the arrows). In sum, even if price equals marginal utility but marginal utility is *increasing* (points A and C), the consumer is at a point of *minimum*, not maximum, net gain. The "price equals marginal utility" condition only assures us that the consumer is on neither the uphill nor the downhill side of a total utility hill, but this means that he may be either at the top of the hill or the bottom of the valley (see Figure 4 of Chapter 4). Only if marginal utility is diminishing (as at point B) do we know that he must be at a point of *maximum* net gain.[6] From B it pays him to move neither to the right nor to the left (see arrows at point b). Finally, if the law of diminishing marginal utility is valid, the entire marginal utility curve will have a negative slope (curve DD' in Figure 2). There will then be only one point, E, where marginal utility is equal to price, and it will always pay the consumer to move toward the corresponding purchase level, Oe (arrows). The law of diminishing marginal utility thus guarantees that there will be only one possible equilibrium level, Oe, and that it will possess an element of stability—consumers will always be motivated to move toward that point.

At the beginning of this section we employed a monetary measure of marginal utility to make our comparison between the price of a commodity and its marginal utility. The marginal utility of X in money terms was defined as the maximum amount of money which a consumer is willing to pay for an additional unit of X. But the marginal utility theorists were generally dissatisfied with such a measure, for when money becomes scarcer, they maintained, its subjective marginal value will increase, like that of any commodity. Hence, an attempt to measure the marginal utility of X by asking the person how much *money* an additional unit is worth to him is like calculating length with a rubber ruler which stretches as we measure. Marginal utility must, according to this view, be measured in its own, subjective, units—we may call them utils. Some noted economists believed that subjective introspective experiments can be conducted successfully and that marginal utility, measured in utils rather than some directly observable unit (like money), can be known to diminish. That is, by thinking about our own feelings about additions to our holdings of, say, packages of spaghetti, we can come to be sure that additional packages are

[6] This is of course the second-order condition—the requirement that if we are maximizing, the second derivative of the maximand must be negative. In the current case (see the preceding footnote) the maximand is $u(x) - px$ whose first derivative with respect to x is $mu_x - p$, where we write mu_x for marginal utility of x. The second derivative, then, is dmu_x/dx, which is required to be negative by the second-order conditions. That is, marginal utility must be declining, as the text asserts.

worth less and less to us in these absolute units (which corresponds to no objective experience that any of us has ever had). This view can be referred to as the *neoclassical cardinal utility* position.[7]

6. Indifference Maps: Ordinal and Cardinal Utility

Many theorists, who classify themselves as *ordinalists*, believe that measurement of subjective utility on an absolute scale is neither possible nor necessary. They question the validity of the introspective data of neoclassical cardinal utility and maintain that all consumer behavior can be described in terms of preferences, or rankings, in which the consumer need only state which of two collections of goods he prefers, without reporting on the magnitude of any numerical index of the strength of this preference.

The geometric device employed to represent this sort of ordinal preference information is the indifference map (Figure 4a). In this diagram

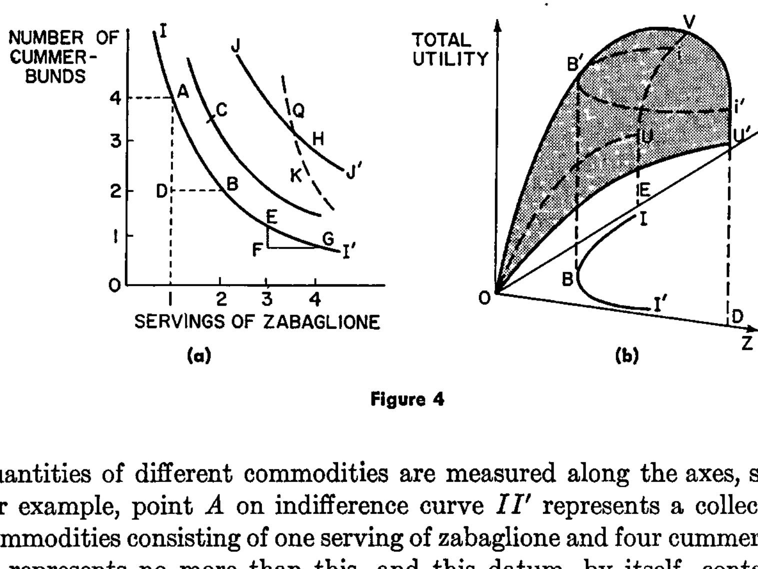

Figure 4

quantities of different commodities are measured along the axes, so that, for example, point A on indifference curve II' represents a collection of commodities consisting of one serving of zabaglione and four cummerbunds. It represents no more than this, and this datum, by itself, contains no information about the consumer in question. In particular, *it does not mean*

[7] That view is briefly discussed again in Chapter 17, where it is contrasted with von Neumann-Morgenstern cardinal utility, an entirely different sort of construct despite the similarity in nomenclature. Neoclassical cardinalism is also mentioned in the next section, where it is contrasted with the ordinalist position.

that he is indifferent between the four cummerbunds and the serving of Italian dessert. We note also that every possible combination of these two items can be represented by a point in this diagram.

We may now define an indifference curve as the locus of points each of which represents a collection of commodities such that the consumer is indifferent among any of these combinations. For example, the presence of point B on curve II' means that the consumer is indifferent between *collections* A and B, that is, between the combination of four cummerbunds and one serving of zabaglione (point A) and the combination which consists of two units of each of these items (point B). The indifference map consists of the infinite set of indifference curves such as II' and JJ' (there is, by assumption, one through every point in the diagram) of which only a few can be shown in any actual drawing.

If, for reasons which will be discussed presently, we go along with the assumption that the consumer prefers combinations represented by points on higher indifference curves (e.g., he prefers collection C to A), the indifference map provides us with a complete and simple report on the consumer's ordering of all possible combinations of the two items, for if two combinations are represented by points on the same indifference curve, the consumer is indifferent between them, and in any other case he prefers that collection which is represented by a point on a higher indifference curve.

Let us digress briefly to see how the indifference map is related to the neoclassical cardinal utility representation of the consumer's tastes, leaving until Sections 14–19 a discussion of utility functions in an ordinal analysis. The three-dimensional Figure 4b shows the same consumer's utility surface, which is constructed as follows: Lay Figure 4a on a horizontal surface to constitute the floor of the diagram. Any point, such as B, on this floor again represents a collection of these two items. Now suppose we have somehow found out the number of utils which this collection, B, can yield to the consumer. We erect over point B a flagpole BB', whose length is equal to the number of utils. Similarly, such a flagpole is erected above every point on the floor of the diagram representing the utility of every possible combination of the two items. For example, DU' is the utility of the collection of OD servings of zabaglione (and zero cummerbunds), whereas EU is the utility of OE cummerbunds. If we now stretch a canvas over the top of the collection of flagpoles, this canvas is the consumer's *cardinal* utility surface, $OUVU'$ (shaded surface).

Since all combinations of consumer goods represented by points on an indifference curve II' have equal utility, the flagpoles above such a curve must all be of equal height, i.e., the portion of the utility surface which lies directly above an indifference curve (such as IBI') must all be of a single height (curve $iB'i'$). In other words, the consumer's indifference curves are the contour lines (iso-utility lines) of his utility surface. They are the loci of commodity combinations of equal utility, just as the contour lines on an ordinary geographic map are loci of combinations of latitude and longitude of equal height above sea level.[8]

However, to an ordinalist there is one important respect in which this geographic analogy does not hold. A contour line on an ordinary map is labeled by a number which indicates the height of its points above sea level. But an indifference curve bears no number to indicate the corresponding height of the utility surface—no cardinal utility number is attached to the curve. Hence indifference curves do not contain cardinal utility information—they only record preferences—the order in which the consumer ranks the various commodity combinations. From utility information we can deduce preferences; the consumer prefers the item whose utility is highest—but the converse is not true: The statement that the consumer prefers A to B gives us no numerical utility magnitudes.

7. Properties of Indifference Curves

The slope of an indifference curve has a significant economic interpretation. For example, in Figure 4a we see that the arc AB has the slope AD/DB. But in moving from point A to B the consumer loses AD (2) cummerbunds and gains DB (1) serving of zabaglione. Since A and B are indifferent, it must mean that the DB unit gain in his zabaglione holdings just compensates him for his AD unit cummerbund loss. Thus the absolute (i.e., positive) value of the slope, $AD/DB = \frac{2}{1} = 2$, indicates that it takes one serving of zabaglione to supply heart balm to the consumer for the loss of two cummerbunds. This absolute value of the slope, called the consumer's *marginal rate of substitution of zabaglione for cummerbunds*, therefore represents the number of units of the latter whose loss can be made up by a unit gain in the former. It is the consumer's psychological rate of exchange between the two commodities.

We can also show that this slope (which in the rest of this chapter is taken to mean its *absolute value*) is equal to the fraction (marginal utility of zabaglione/marginal utility of cummerbunds),[9] that is, the marginal

[8] If we describe the utility surface by means of a function $u = f(x_1, \cdots, x_n)$, where x_i is the quantity of good i consumed, then the equation of an indifference curve is $f(x_1, \cdots, x_n) = k$ (constant) with different indifference curves corresponding to different values of k.

[9] *Proof:* If arc AB is sufficiently small, the utility loss involved in giving up AD units of cummerbunds is the marginal utility of such a unit (mu_c) multiplied by AD, the number of units involved, i.e., the loss in giving up $AD = (AD) \times (mu_c)$. Similarly, the utility gain involved in acquiring DB units of z is $(DB) \times (mu_z)$, where mu_z represents the marginal utility of zabaglione. Since the gain and the loss just offset one

rate of substitution of Z for C equals

$$\text{slope of } II' \left(= \frac{\Delta c}{\Delta z}\right) = \frac{\text{marginal utility of } Z}{\text{marginal utility of } C}$$

where z and c represent, respectively, the quantities of zabaglione and cummerbunds.

Two features of this result bear some discussion:

1. In the equation

$$\frac{\Delta c}{\Delta z} = \frac{\text{marginal utility of } Z}{\text{marginal utility of } C}$$

it is noteworthy that c appears in the *numerator* of the left-hand fraction but in the *denominator* of the fraction on the right-hand side of the equation and that the reverse holds for z. This inverse relationship between Δc and the marginal utility of c is easily explained. $\Delta c = AD$ units of C is the amount of C which the consumer is willing to give up for $\Delta z = DB$ units of Z. But the more valuable C is to him (the greater the marginal utility of C), obviously the less the consumer will be willing to give up in exchange for Δz; i.e., the smaller will be Δc; hence the inverse relationship.

2. A second thing to be noted is that marginal *utility* seems to have sneaked back into the analysis despite the ordinal nature of the indifference map. However, its return is not as serious as it may appear from the point of view of the ordinalist. Only the *ratio* of two marginal utilities ever

---

another (points A and B are indifferent), we have $AD \times mu_c = DB \times mu_z$. Dividing both sides of the equation by $mu_c \times DB$ we obtain the required result:

$$mu_z/mu_c = AD/DB = \text{the slope of } II'.$$

Alternatively, one can obtain the equation of the text from the expression for the utility function, $u = f(c, z)$, and the formula for total differentiation (Chapter 4, Section 7). Since along an indifference curve total utility must be constant, we must have $du = 0$ or

$$du = \frac{\partial u}{\partial c}\,dc + \frac{\partial u}{\partial z}\,dz = 0.$$

Bringing the first term over to the right and dividing through by $dz\ \partial u/\partial c$ we obtain at once

$$\frac{dc}{dz} = -\frac{\partial u/\partial z}{\partial u/\partial c},$$

which is the relationship in the text.

occurs in indifference analysis. In such a ratio we measure the marginal utility of one commodity not in terms of utils but in terms of the other commodity. We ask how much of C an additional unit of Z is worth (the marginal rate of substitution of Z for C). Thus we are, in effect, back to measuring marginal utility in terms of money, or some other commodity, and that is perfectly satisfactory to the ordinalist.

In indifference curve analysis it is customary (at least implicitly) to make these assumptions about the psychology of the consumer:

Assumption 1 (nonsatiety): The consumer is not oversupplied with either commodity, i.e., he prefers to have more of C and/or Z.

Assumption 2 (transitivity): If A, B, and D are any three commodity combinations and if A is indifferent with B and B is indifferent with D, then the consumer is also indifferent between A and D. This condition simply requires that the consumer's tastes possess a conceptually simple type of consistency.

Assumption 3 (diminishing marginal rate of substitution): Consider two collections represented by points along the same indifference curve (e.g., A and E in Figure 4a). Then if at one of these points, E, the consumer has a relatively small supply of one commodity, C, and a relatively large supply of the other, then at E the marginal utility of the relatively scarcer C will be large in comparison to that of Z, i.e., the consumer will there be willing to give up only a relatively small amount of C in exchange for an additional unit of Z. Thus, in Figure 4a, at point A the consumer is willing to give up AD units of C for an additional unit of Z. But at point E, where C is scarcer, he is only willing to pay the smaller number EF units of C for the same increment in his holdings of Z.

These assumptions permit us to deduce four properties of indifference curves which normally characterize their drawings:

PROPERTY A *(by Assumption 1)*. An indifference curve which lies above and to the right of another represents preferred combinations of commodities.

Proof: Consider the indifference curves II' and JJ' in Figure 4a, and combination B on II' and Q on JJ'. Since point Q is above and to the right of point B, it involves more of both commodites C and Z. Hence, by Assumption 1, the consumer must prefer Q to B, and therefore he must prefer every point on JJ' (all of which are indifferent with Q) to any point on II'.

PROPERTY B. Indifference curves have a negative slope (by Assumption 1).

Proof: Start, e.g., at point A in Figure 4a and move from it to the right so that the consumer holds more of commodity Z as a result. By Assumption 1 the consumer must prefer this new point (he cannot be indifferent between it and A) unless at the same time it involves his having less of the other commodity, C. In other words, if he is to be indifferent between the new point and A, it must lie below A as well as to its right, as does point B.

PROPERTY C. Indifference curves can never meet or intersect, so that only one indifference curve will pass through any one point in the map (by Assumptions 1 and 2).

Proof: Suppose on the contrary that two indifference curves, JJ' and the dashed curve, were to intersect at point Q. Pick point K on the dashed indifference curve and point H on JJ' where H lies above and to the right of K. By Property A (Assumption 1) H must be preferred to K. But H is indifferent with Q, and Q is, in turn, indifferent with K. Hence, by Assumption 2, H must be indifferent with K. Since H cannot be both indifferent with and preferred to K, the intersection of the two curves which led to this self-contradictory result cannot possibly occur.

PROPERTY D. The absolute slope of an indifference curve diminishes toward the right (the curve is flatter at point E than it is at point A) so that the curve is said to be *convex to the origin* (by contrast with SS' in Figure 7, which is said to be concave to the origin). This theorem is a direct consequence of Assumption 3, which states that the marginal rate of substitution of Z for C [which, it will be remembered, is represented by the slope of the curve (neglecting minus signs)] is smaller at E than at A (Figure 4a).

8. Violation of the Premises about Indifference Curves. Satiation and Lexicographical Orderings

While the shapes that have just been described are frequently assumed to hold and are extremely convenient analytically for reasons that will be indicated presently, they are necessarily valid *only* on the psychological assumptions listed at the beginning of the section, i.e., nonsatiety, transitivity, and diminishing marginal rate of substitution. Any or all of these conditions can be violated in reality and there is nothing necessarily pathological about such violations, as will now be shown.

Assume that our consumer ultimately does become sated with zabaglione—after the refrigerator and the freezer are filled with it the householder regards further quantities of the dessert with apprehension and perhaps with hostility. Suppose Z^* is the maximal desired quantity of Z (and similarly, let C^* be the satiation quantity of cummerbunds). What happens to the shape of the indifference curves beyond these quantities?

In Figure 5a we see that rectangle OZ^*SC^* is the region of nonsatiation:

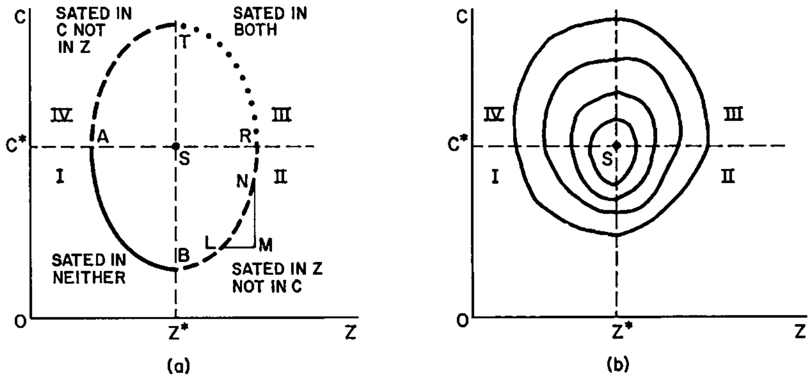

Figure 5

Any point in that region (which has been labeled region I) represents a combination of the two goods which leaves the consumer wanting more of either or both. In that region we see a normally shaped indifference curve—the solid locus AB. However, at any point in region II to the right of Z^* but below C^* (e.g., point L), the consumer still wants more of C but now he desires *less* of Z. Hence if he gets still more Z (the move from L to M) and yet remains indifferent, he must be compensated for the (repugnant) rise in quantity of Z by a desired *rise* in C. Thus the indifference curve in this region must have a *positive* slope. The reader should verify that the same argument holds for region IV in which there is too much C but more Z is still desired. However, in region III, where the consumer has more than he wants of either item, the indifference curve will again acquire its negative slope since there, to compensate him for an *addition* in his unwanted holding of Z, he must be *relieved* of some of his unwanted C. That is, to leave him indifferent a rise in his Z holdings must be accompanied by a fall in his C, and vice versa.

Figure 5b suggests more clearly what is going on by showing a set of several indifference curves. It reveals them to be closed contours, one inside the other, converging to the saturation point, S (sometimes called the "bliss point"), at which the consumer possesses exactly the maximal amounts he wants of each of the two commodities. The indifference curves can be taken as the contour lines of a smooth utility hill with a single maximum point, S, with the surrounding indifference curves representing decreasingly desirable possibilities as they move further from the bliss point.

Thus we see

1. The conventional indifference curves really are only segments of the complete indifference curves—those portions lying in region I, the region of nonsatiation (scarcity). This is, of course, the relevant region for most economic analysis since budgetary limitations do keep consumers from complete satiation in *every* commodity (even the wealthiest of absolute rulers has not been able to afford all the military equipment he wanted for his armies).
2. In other regions the negative slope of the indifference curve need not hold. Moreover, the curve need not be convex to the origin (region III).[10]
3. At points B and T in Figure 5a (zabaglione satiation) the indifference curve is horizontal (a small change in Z neither adds to nor subtracts from his welfare, and so no change in quantity of C is needed to compensate him for such a change). Similarly, at points R and A (cummerbund satiation) the indifference curves are vertical.

It is also possible to think of plausible cases in which the nonintersection property of indifference curves will be violated (intransitivities). This will occur, for example, where the consumer cannot distinguish small differences (1.0003 and 1.0004 ounces of Z look the same to him and seem to assuage his hunger equally). In that case neighboring indifference curves will be equally preferable though one is a tiny bit higher than the other. In this case one says that indifference curves are "thick," that is, they encompass a narrow area rather than a locus (curve) of zero thickness.

Finally, we note that underlying the entire discussion is a premise rarely questioned in elementary texts—the assumption that such curves *exist*. But even that is not necessarily true. It is easy to describe an interesting preference relationship for which no such curves exist. The standard

---

[10] Intuitively, as the quantities of Z rise and C fall as we move from point T toward R in Figure 5a, further additions to the holdings of Z become increasingly unbearable, while further decreases in the excess holdings of C become less urgent. Hence, to get the consumer to accept further increases in his Z he must be compensated by ever-larger declines in his C (increasing marginal rate of substitution).

Actually, this sort of shape of indifference curve can occur also in region I, where it characterizes the behavior of an addict or a collector (the more of either commodity he has, the more urgently he wants even more of it). If addiction to zabaglione were to characterize its consumption, as we move toward point Z^* additional units of this item will become very valuable and additional C comparatively worthless, i.e., the consumer will be willing to trade a small addition in Z for a great loss in C.

counterexample, which is of interest in itself, is called a *lexicographical ordering*, i.e., it uses a ranking criterion analogous to that used in ordering words in a dictionary. Suppose a government of a very poor country with a mild climate considers two objectives: More food (x_1) and more clothing (x_2). Since starvation and malnutrition are serious problems and neither cold nor modesty (!) are considered pressing issues, the government prefers *any* increase in food output *no matter what happens to clothing production.* Then no increase in clothing output can make up for a unit decline in food output. However, the government does favor more clothing output for its decorative and amusement value, provided no food need be given up to get it. Thus, if we start out with 8 units of food and 8 of clothing (point A in Figure 6), the government will prefer any point involving more food than

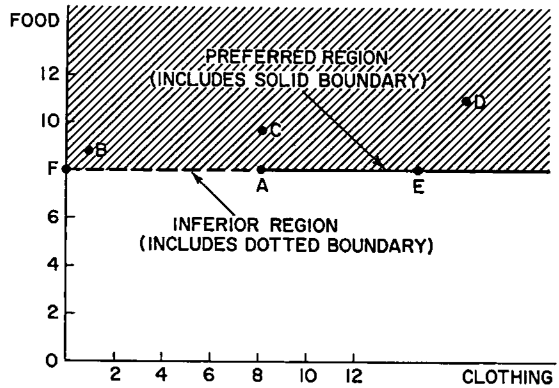

Figure 6

A regardless of the associated quantity of clothing (e.g., points B, C, or D). It will also prefer any point on AE to the right of A (more clothing with no less food). However, it will consider inferior to A any point below A or on the line segment FA to the left of A. Thus, every point in the diagram other than A is either preferable to A or less desirable than A. *There can be no second point that is indifferent to A*, so that no indifference *curve* through A (or through any other point in the diagram) is possible, just as our discussion was intended to show.

PROBLEMS

1. Show the pertinent indifference curve and equilibrium point for a wealthy ruler who has all the zabaglione he can possibly want but wishes he could afford more military equipment.
2. Explain why point F in Figure 6 is considered inferior to A but B, which is very close to F, is superior to A.
3. Explain the analogy between the lexicographical ordering as described in the text and the ordering of words in a dictionary.

9. Price Lines: Consumer Income and Prices

By itself, an indifference map cannot possibly predict consumer behavior because it leaves out two vital types of information—the income of the consumer and the prices of the commodities. The indifference curves do not ask the consumer which combination he believes will give him the most for his money. It is merely a hypothetical ranking of various commodity combinations—perhaps castles in Spain against yachts in Portugal —taking no account of what the consumer can afford.

Price and income information is supplied in an indifference diagram by another curve which is called the *price line* or, sometimes, the *budget line*. Since the axes of the diagram present only quantities of commodities rather than amounts of money, dollar prices and incomes cannot be shown directly. Instead, the price line does the next best thing and indicates what amounts of the commodities a given amount of money can buy.

For example (Figure 7), suppose \$50 spent exclusively on commodity Z will, at its current price, buy OP' units of that commodity, whereas the same amount spent entirely on C will purchase exactly OP units of that item. Suppose, moreover, that every point such as A on line PP' represents a combination of the two commodities which sells for \$50 (e.g., \$10 worth of Z plus \$40 of C). Then line PP' is a *price* or *budget* line. Such a line is defined as the locus of all combinations of commodities which cost some fixed amount of money (e.g., our illustrative \$50).

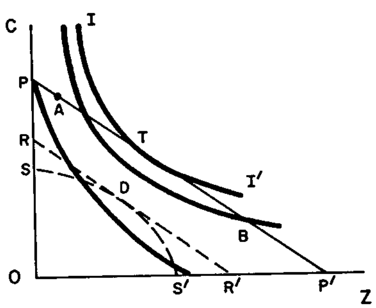

Figure 7

If the prices of both commodities are fixed, that is, they do not vary with the amounts of the goods which are purchased, the price line will possess the following properties:

1. It will be a straight line.
2. It will have a negative slope.
3. Its slope will be equal to the *negative inverse* of the ratio of the prices of the two commodities, i.e., we will have $\Delta c/\Delta z = -p_z/p_c$, where p_z and p_c are the unit prices of Z and C, respectively.
4. Suppose two price lines involve the same commodity prices but represent the expenditure of different amounts of money (say \$50 for PP' and \$30 for RR'). Then the two lines will be parallel.

The equation of the price line is, in the fixed price case, given by a simple expression. If the consumer buys z units of commodity Z, his total expenditure on this item will be $p_z z$ (the price per unit, p_z, multiplied by the number of units purchased). Similarly, expenditure on the other commodity is given by $p_c c$ so that total expenditure is given by

$$p_z z + p_c c = m, \tag{4}$$

where m is the amount of money spent (our illustrative $50 for line PP') and is therefore constant along a price line. This, then, is the equation of a price line.[11]

The four properties can readily be generalized to take account of price variability. There are two possibilities: Either that buying in quantity will make the commodities scarce relative to the quantities demanded and so raise their prices to the purchaser (as wages go up when the demand for labor increases), or, on the other hand, that he will be offered discounts if he buys in larger quantities (special today: one elephant, $200, or two for $325). The former possibility, which can be interpreted as a case of diminishing returns to an increased number of dollars spent by a large purchaser on a given commodity, will yield a curved price line which, like SS', is concave to the origin (Figure 7). The reason is that as one moves toward the axes from an interior point such as D, a greater proportion of this large consumer's fixed amount of money, m, is spent on one of the commodities; thus near S almost all of it is spent on commodity C. This raises the price of C against the consumer so that his m dollars will buy only OS—which is less than the OR units he could obtain for m dollars if the price of C were fixed at the level it is at point D.

For a completely analogous reason, quantity discounts (increasing returns to increased expenditure on any one item) will result in a budget line which, like II', is convex to the origin.

There remains one point to discuss about price lines. What do they tell us about the consumer's income and the prices of the various products? First, to deal with the information on consumer income which is conveyed by a price line, it is convenient to define the multicommodity analogue of a price line [Equation (4)]—the algebraic budget relationship for all of

---

[11] The four properties of the price line are readily derived from this equation. Dividing both sides by p_c and rearranging terms the equation becomes

$$c = -\frac{p_z}{p_c} z + \frac{m}{p_c}.$$

If we now change our notation, writing y for c, x for z, a for $-p_z/p_c$, and b for m/p_c, this becomes the standard *linear* equation of Chapter 2, $y = ax + b$, with (*negative*) *slope* $a = -p_z/p_c$. The four price-line properties follow directly from this result, as the reader should verify.

Note also that m/p_c = the total amount of C the consumer could purchase if he were to spend all of his income on that item.

the (say, 1,257) different commodities which the consumer buys or considers buying:

$$p_1x_1 + p_2x_2 + \cdots + p_{1257}x_{1257} = m,$$

where, e.g., x_2 is the quantity of commodity number 2 purchased and p_2 is its unit price. In such a multicommodity budget equation it is convenient to consider savings to be one of the 1,257 goods which he buys or can buy for his money. On this interpretation, the consumer has no choice but to spend all his money (either on savings or on some other commodity) and the only relevant price line is the one which uses up all of the funds which he has available to him. This price line, then, specifies the consumer's *real income* (or wealth). It tells us just what combinations of commodities he can afford to buy, given prices and his money income.

So much for the income information supplied by a price line. Let us now see what the price line tells us about prices.

Property 3 states that the slope of such a line tells us the *ratio* of the prices of the commodities. If the slope is -2, we know that the price of Z must be twice the price of a unit of C (note again the inverse relationship, $-p_z/p_c = \Delta c/\Delta z$).

To summarize, the price line specifies the real purchasing power which is available to the consumer and the ratio of the prices of the two commodities. But since monetary quantities are not shown anywhere on the diagram, it is impossible for the price line by itself to specify either the level of the consumer's liquid assets or the money price of any commodity.[12]

10. Equilibrium of the Consumer

The consumer who wants to get the most for his money will want to land on as high an indifference curve as his purchasing power permits—the highest indifference curve which can be reached from his budget line. This optimal purchase combination is given by the point of tangency, T, between the price line and indifference curve II' (Figure 7), for it is clear, by inspection of the diagram, that any other point on the price line, such as B, will be intersected by an indifference curve which lies below II'. In this way, the indifference map together with the price line permit us to predict the demand pattern of the "rational" consumer—the consumer who spends his money efficiently in the pursuit of his needs and interests. We say that T is a *point of equilibrium* because once the consumer arrives at

---

[12] However, if we know any one of these values, the others follow at once. For example, if the price of C is known to be \$10 and the price line shows Z to be twice as expensive as C, then the price of Z must obviously be \$20. Similarly, since his money buys OP units of C at \$10 per unit, his expendable money must be OP times \$10.

the decision to purchase the combination of commodities represented by that point, he has no motivation to revise his purchase plans.

The tangency condition of equilibrium immediately yields another equilibrium condition. At their point of tangency the slope of the price line and that of the indifference curve must, by definition, be equal. But we know that the (absolute value of the) slope of the budget line is equal to the (inverse) ratio of the two prices, whereas the slope of the indifference curve is equal to the (inverse) ratio of the two marginal utilities or to the marginal rate of substitution of Z for C. Therefore, in equilibrium we must have

$$\frac{p_z}{p_c} = \frac{mu_z}{mu_c} = \text{marginal rate of substitution of } Z \text{ for } C.$$

This is the marginal condition of equilibrium of the consumer. It resembles the neoclassical equilibrium condition that price must equal the marginal utility of a commodity, but states, instead, that the *ratio* of the marginal utilities of two commodities must equal the *ratio* of their prices.

The logic of this condition is easily demonstrated. Suppose the condition is violated so that, e.g., the first of these fractions is greater than the second. Then, multiplying both sides by the presumably positive number mu_c/p_z, we obtain the inequality $mu_c/p_c > mu_z/p_z$. But if item C costs, e.g., $p_c = \$5$ per unit, we can for \$1 obtain $1/5 = 1/p_c$ units of this item, and $(1/5)mu_c = (1/p_c)mu_c$ therefore represents the utility which can be obtained spending an additional dollar on C. The last inequality therefore states that the consumer can acquire more utility out of an additional dollar spent on C than from another dollar spent on Z. If this is so, he cannot possibly be getting the most for his money—he can get more by reallocating his funds, spending less on Z and more on C. This is illustrated in Figure 7, where we note that at point B the absolute value of the slope of the indifference curve is less than that of the price line (the indifference curve is flatter) so that we have $mu_z/mu_c < p_z/p_c$ and so, as before, $mu_z/p_z < mu_c/p_c$. It therefore pays the consumer to plan to buy less of Z and more of C, i.e., for him to move upward and to the left along the price line from B toward the point of tangency T. We see then that B violates our equilibrium condition and *that it does so in a way which motivates the consumer to move toward the equilibrium point* T. Thus with curves of the usual shape (as in the diagram), the equilibrium point possesses an element of stability. From any other point on the price line the consumer is motivated to move in the direction of the point of tangency.

Indeed, the shape we have assumed for the indifference curves plays an important role in our tangency solution. If any one of the four properties of indifference curves (listed in Section 7, above) were violated, consumer equilibrium would not occur at a point of tangency. Thus, if Property A were violated so that the consumer wished, say, to be on the lowest attainable indifference curve, his optimal point would be P rather than T, i.e., he would end up spending his money exclusively on one commodity. If Property B were violated so that the slope of the indifference curves was not negative, there could be no point of tangency with the negatively sloping price line. If Property C (nonintersectability of indifference curves) were violated, a number of points of tangency might occur (Figure 8a), and if the indifference curves were concave to the origin, in violation of

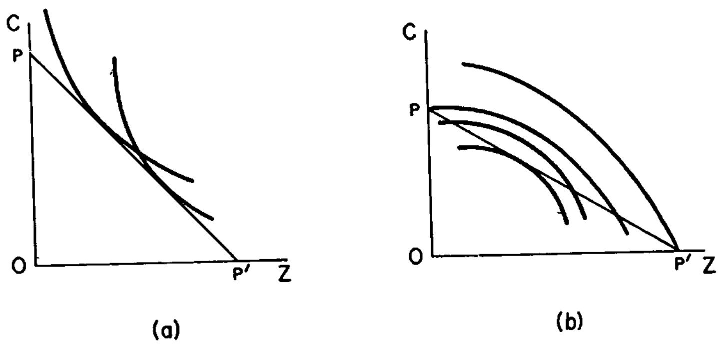

Figure 8

Property D, the point of tangency would yield the lowest attainable indifference curve, whereas the highest indifference curve would lie at one of the end points of the price line (P' in Figure 8b), so the rational consumer would again end up spending all of his money on just one commodity! Note that at the point of tangency in Figure 8b the consumer is at the point of *minimum* utility on his price line.[13]

PROBLEM

Show that if indifference curves are positively sloped the optimal point is likely to occur at a corner (an end point of the price line). Interpret this case in terms of Figure 5.

11. *Responses to Price and Income Changes*

If the income of the consumer increases, his budget line will retain its slope (relative prices remain unchanged) but that line will then shift

[13] What has gone wrong here is that while at the tangency point the first-order conditions for a maximum are satisfied (MRS equal to the ratio of prices), the second-order conditions are those required for a minimum rather than a maximum (i.e., it is as though, in a one-variable function, the second derivative were positive).

upward (he can get more goods for his increased money supply). In other words, income changes cause parallel shifts in the budget line, and a set of parallel budget lines (Figure 9a) shows how the consumer's possible purchases will vary with changes in his income. On each such line we can find the equilibrium point of tangency (points T_1, T_2, etc.). The curve OW, which is the locus of all such points, shows how the consumer's purchases of the two commodities will vary when his income changes. Such a curve is called an *income-consumption* curve or, sometimes, an Engel curve (named after an early student of the effects of income changes on consumer expenditure patterns).

Normally, consumers may be expected to increase their purchases of commodities as their incomes rise. But sometimes, if an item is of low quality, demand for it will drop as the consumer's financial position improves and more desirable commodities are substituted for it. Such an item is called an *inferior good*. Plausible examples of inferior goods are recapped automobile tires, poorly made clothing, poor cuts of meat, etc., any of which the consumer may be buying only because he can afford no better. In Figure 9a commodity Z is taken to be an inferior good. This is shown by the relative positions of points T_3 and T_4 (the negatively sloping segment of OW). The latter point lies to the left of the former (it represents a lower quantity of Z) despite the fact that it (T_4) is on a higher budget line and therefore involves a higher income for the consumer.

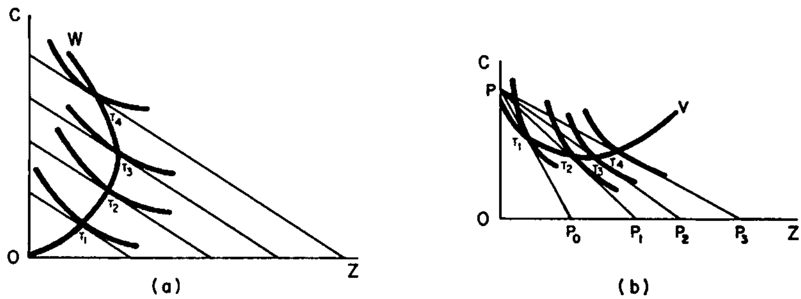

Figure 9

Next, we can investigate the effects on the consumer's purchases of changes in the price of one of the commodities. Suppose the price of Z falls, other things remaining equal. This means that the buyer can get more of this commodity for his money (e.g., OP_1 instead of OP_0 in Figure 9b), though he can only obtain the same amount of C as before (quantity of OP). We see, then, that a fall in the price of the item on the horizontal axis leads the price line to flatten out by swinging to the right. Figure 9b represents a number of such price lines and the corresponding equilibrium tangency points. Curve PV, the locus of these points of tangency, shows how changes in the price of Z affect the purchases of *both* commodities. PV is called a *price-consumption curve* or sometimes, particularly in international trade theory, an *offer curve*.

It will be noted that the *income-consumption* curve, OW, begins at the origin (point O) because with zero income the consumer can buy none of either commodity. By contrast, the *price*-consumption curve, PV, characteristically begins at point P, the pivot point of the swinging price line in Figure 9b. The reason is that, as the price line approaches the vertical axis (the price of Z increases further and further), the consumer finds that he gets less and less of Z for his money. Eventually, when its price goes high enough, the consumer will be forced out of buying Z altogether and he will therefore spend all his money on the remaining commodity, C, i.e., he will buy OP units of C and no Z (point P).

The offer curve construction can readily be translated into an ordinary demand curve for the consumer[14] if one of the commodities represented in the diagram is M, the money held by the consumer (Figure 10a). By this device money values are inserted into the indifference map. As before, let PV be the offer curve so that if the consumer buys zero units of the commodity he will have \$30 for himself (point P). Now consider point A on

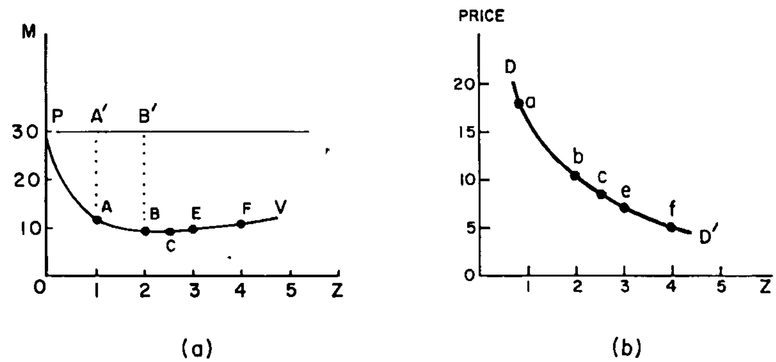

Figure 10

the offer curve which represents the consumer possessing $z = 1$ unit of commodity Z and $m = \$12$. Since in moving from P to A he acquired 1 unit of the good but gave up $18 = 30 - 12$ dollars, the price per unit at A must be \$18. Thus, point A states that the consumer will buy 1 unit of the commodity if its price is \$18. This information is recorded by point a in Figure 10b. Similarly, point B on his offer curve involves the buyer's spending $20 = 30 - 10$ dollars on *two* units of the good so the price *per unit* must be $\$20/2 = \$10$. Hence (point b in Figure 10b) the offer curve

[14] It should be noted, however, that even if the consumer has preference patterns for which an indifference map exists (cf. Section 8) it need not follow that a corresponding demand function exists unless the (ordinal) utility function is differentiable.

tells us that he is prepared to buy two units if the price is \$10. Points c, e, and f in Figure 10b are derived similarly. These are clearly points on the consumer's demand curve since they indicate how many units he is prepared to buy at different prices. DD', the locus of all such points, is the demand curve for this consumer.

It is noteworthy that the *offer* curve gives us information about the elasticity of the *demand* curve. For example, inspection of PV tells us that to the left of point c the demand curve DD' must be elastic. To see why this is so, note that the unit price at point B (\$10) is lower than that at A (\$18) but that total consumer expenditure on the commodity at B ($BB' =$ \$20) is greater than that at A ($AA' =$ \$18). Thus a fall in price has produced a rise in total expenditure (the price-consumption curve has a negative slope). By elasticity Proposition 2 in Section 4 of this chapter, expenditure will rise when price falls only if the demand curve is elastic. The reader should have no difficulty showing that DD' is unit elastic at horizontal point c and inelastic to the right of point c.

12. *Income and Substitution Effects: The Slutsky Theorem*

It is customary to analyze somewhat further the effect on purchases of a change in the price of one of the commodities. The effect, for example, of a fall in the price of Z is classified into categories: the income and the substitution effect. Its lower price makes Z a better buy relative to C than it was before, and, as will be shown presently, that consequence by itself would always induce the consumer to increase his purchase of Z (the Slutsky theorem). This price-ratio portion of the effect of a price change on purchases is called the *substitution* effect. Purchases of Z will be substituted for those of C because Z is *relatively* more price-attractive than it was initially.

But the fall in price of Z also affects the purchases of both commodities in another way—it increases the purchasing power of the consumer's income. This will, in turn, always increase the purchases of *both* commodities provided that neither of them is an inferior good, the demand for which is reduced by an increase in real income. The income effect, then, is the effect *on the consumer's purchases* of the rise in real income which results from a fall in the price of commodity Z. Note that the income effect refers to the resulting change in his purchases and *not* to the change in his real income.

To summarize, a fall in the price of any commodity, X, will affect the consumer's demand for X. This effect may be subdivided into two parts: the substitution effect, which always increases the demand for X, and the income effect, which will increase the demand for X unless X is an inferior good. Thus, ignoring this exceptional possibility, *the demand curve for X must have a negative slope*, i.e., a fall in the price of X must increase the demand for that commodity. Even if X is an inferior good its demand curve will still have a negative slope unless the income effect is stronger than the substitution effect, for, as will soon be shown, the substitution effect of a lower price of X is always a rise in the demand for X. In addition, in practice, the income effect for most consumers' goods is likely to be small because a buyer's outlay on any one commodity constitutes a relatively small proportion of his budget, so a fall in the price of that item alone will not increase his real income significantly.[15]

Two different graphic depictions of the income and substitution effects have been employed in the literature. In Figure 11a and 11b let PP' and PP'' be two price lines involving different prices of commodity Z, and let A and B represent the (tangency) equilibrium points on the two price lines. The total effect of the price change on the amount of Z purchased is, therefore, ab. Our object is to divide ab into two parts—the income effect and the substitution effect. For this purpose the change in position of the price line is divided artificially into two parts: a parallel shift (a change in real income with no change in relative prices) and a pivot or twisting (change in slope) of the price line (a change in relative prices with no change in real income). To accomplish this division we employ an imaginary price line RR' in Figure 11a (or SS' in Figure 11b) which is parallel to one of the price lines (they have the same relative prices) and is, in some sense, at the same real income level as the other. Here is where the ambiguity in interpretation occurs (the source of difference between the two diagrams).

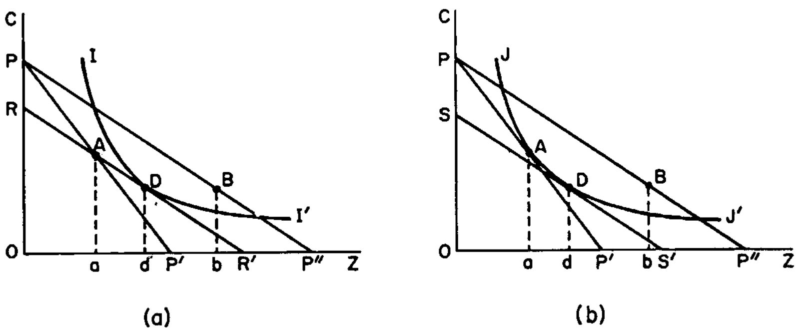

Figure 11

[15] But, at least as a remote possibility, we see that a very inferior good for which the income effect is very high provides another possible case of a positively sloping demand curve. The other two cases which were mentioned in Section 1 of this chapter (snob appeal and quality judged by price) do not show up in the usual indifference map analysis because each of these involves the consumer's preference structure being changed by the price change. He values platinum collar stays or a brand of frozen chop suey more highly when its price rises. In other words, the consumer's indifference curves shift when there is a swing in the price line—a possibility which has not been considered in the text.

When do two price lines, which are not identical, represent the same real income? One highly persuasive solution is to say that this occurs when they both yield the same satisfaction to the consumer, i.e., they are both tangent to the same indifference curve, as are PP' and SS' in Figure 11b (they are tangent, at points A and D, respectively, to indifference curve JJ'). There is another solution, which is perhaps less satisfying intuitively but which is very useful and which we will need presently. This is to say that RR' (Figure 11a) yields the same income as PP' if RR' passes through point A so that it just gives the consumer enough money to buy combination A, the combination he would buy if PP' were in fact the prevailing price line. In this case, the point of equilibrium, D, on the imaginary price line RR' lies on an indifference curve II', which is not tangent to PP'. Indeed, since line RR' in Figure 11a is higher than SS' in 11b, the indifference curve II' to which RR' is tangent must lie above indifference curve JJ' in 11b, which is tangent to both SS' and the original price line, PP'.

The income and substitution effects can now be read off from the diagrams. The substitution effect is ad, the change in purchase of Z which results from the twisting of the imaginary price line, whereas the income effect is db, the effect of the parallel shift in the price line.

In this two-commodity analysis. Figure 11b can be used to show that when the price of z falls (the price line flattens out), the substitution effect must lead to a rise in the demand for Z, for SS' and PP' are both tangent to the same indifference curve. But since SS' is the flatter price line, its point of tangency, D, must occur to the right of A, the point of tangency of PP' (because the slope of an indifference curve gets smaller toward the right). Hence, with the lower relative price of Z (RR' or SS') the demand for $Z(d)$ will be greater than the demand for Z when the price line is PP'. Unfortunately this argument is not valid when there are more than two commodities so that the consumer's preferences cannot be summed up in a two-dimensional indifference map. Presently, more general proofs of this result, the *Slutsky theorem*, will be presented (Chapter 13, Section 8).

13. The Role of the Income Effect

The reader may well wonder what the fuss is all about—why the mere classification of the effects of a price change into two portions—the substitution and the income effect—should have elicited so much attention in the literature. The essence of the matter is that Eugen Slutsky and, after him, J. R. Hicks and R. D. G. Allen discovered independently that such a price effect has two portions one of which, the substitution effect, is predictable in sign and in many of its other characteristics. For example, we have just seen that the substitution effect of a rise in the price of X on the quantity of X purchased will always be negative (the Slutsky theorem).

But these authors noted that there is another portion of the overall effect (the income-effect portion) whose behavior is unpredictable in general—and it must therefore be stripped away so that the systematic and predictable behavior of the substitution effect can be revealed. This discovery can be likened to a filter which eliminates the static from the transmission of sound so that the underlying message can be made out.

This, then, is the unexalted role of the income effect—it is discussed primarily in order to permit us to remove it. One should not be misled by the subclassification of possible values for the income effect—the statement that for "normal" goods it has one sign and for "inferior" goods it has another. This is only a little more than the use of nomenclature to put a better appearance on our ignorance. What this last subclassification asserts, in essence, is that the income effect can go either way and that we can think of realistic examples of both cases. Hence, we can make firm predictions about demand reactions to price changes only if this undependable portion of the price effect has been removed.

Much of more sophisticated consumer theory proceeds accordingly, discussing matters in terms of "net" concepts—after removal of the income effects—rather than the corresponding gross concepts which correspond more closely to the observable data but whose behavior patterns vary in a manner that defies generalization.

14. Complements and Substitutes

The distinction between substitute and complementary commodities is easily grasped intuitively. Vodka and gin are substitute commodities—they serve the same general purposes, and if we have more of one, we will tend to want less of the other. On the other hand, bread and butter or gin and vermouth are complements—for many consumers they are better together and hence an increase in the availability of one tends to stimulate the demand for the other. But how does one measure these relationships? One straightforward approach makes use of the cross elasticity of demand—the effect of a change in the price of X_1 on the demand for X_2. If goods are substitutes, we expect the cross elasticity to be positive, and we expect the reverse if they are complements, for if X_1 and X_2 are substitutes, a rise in p_1 will decrease x_1 (the quantity of X_1 demanded), and as a result the consumer will seek more of the substitute. Consequently, the rise in p_1 will lead to an increase in x_2, and so the cross elasticity, $(dx_2/dp_1)p_1/x_2$, will be positive. The reverse will be true of complements, for the rise in p_1 will decrease x_1 and hence decrease x_2.

Or will it? The answer is that it always will[16] *unless* the ambiguous

[16] For proof of this statement see Chapter 14, Section 9, Proposition 10.

income effect messes matters up once again. For example, suppose the consumer is a relatively impecunious martini drinker. Then a rise in the price of gin may lead him to increase his purchases of cheap vermouth, which he will use instead of the imported variety. It is clearly the income effect which stimulates his consumption of the inferior good, cheap vermouth, as the rise in price of gin reduces the consumer's real income. Thus, though the goods are complements, their cross elasticity will in this case be positive. Similarly, a rise in the price of hamburger may lead a poor family to decrease its demand for steaks even though they are substitutes.

There is worse to come: Because of the asymmetry of the income effect the cross elasticity of demand for good 1 with respect to the price of good 2 may be positive and yet the elasticity of demand for good 2 with respect to the price of good 1 may be negative, for one good may play a small part in the consumer's budget and so a rise in its price will have a negligible income effect while the reverse may be true of the other good.

To make it easier to describe these cases economists use the following classifications:

1. Good 1 is a *gross substitute* for 2 if its cross elasticity of demand with respect to p_2 is positive,
2. It is a *gross complement* if that cross elasticity is negative,
3. It is a *net substitute* if the cross elasticity is positive *after the income effect is removed,*
4. It is a *net complement* if the cross elasticity is negative after removal of the income effect.

15. Compensated Demand Curves

The elimination of the income effect has also been carried out for demand curves, and so much of recent analysis has been carried out in terms of a *compensated demand curve,* i.e., the demand curve after adjustment to remove income effects. This curve describes the result of the conceptual experiment described in the following steps:

a. Start from some initial price-quantity combination.

b. Consider some alternative price, e.g., a price higher than the initial one.

c. Adjust the consumer's income so as to leave him with the real purchasing power he possessed initially; e.g., if price is increased, he must be compensated by an increase in income sufficient to permit him to purchase the initial quantity combination, should he choose to do so.

d. Now examine the effect of the price change on his purchases after the compensation for income effect (step c.).

Graphically the process looks as shown in Figure 12. Here DD' is the ordinary (uncompensated) demand curve. Suppose the initial price-quantity combination is p_a, x_a and we consider the consumer's behavior at the alternative price, p_b. Without compensation his purchases would be reduced to x_b. But to compensate him for the erosion of his purchasing power stemming from the price rise, the consumer is provided a (conceptual) infusion of income. If X is not an inferior good, this means his purchases will not fall quite as much as if he had received no compensation, i.e., instead of going from x_a to x_b they will decrease only to x_c. This means that AC is a *compensated demand curve* through point A. Thus, the compensated demand curve for a rise in price will generally lie to right of the ordinary demand curve (except at the initial point), provided the commodity in question is not an inferior good.

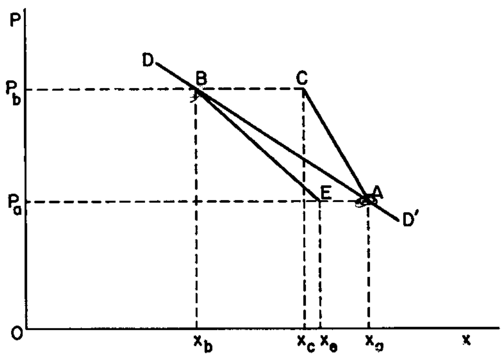

Figure 12

We note at once that there is not just one single compensated demand curve—indeed, there will be a different one for every initial price-quantity combination, i.e., for each initial point (like A) on the ordinary demand curve DD'. Moreover, there are also compensated demand curves for price *decreases*, and these (like curve BE) will usually lie to the left of the ordinary demand curve. For if price falls, in the absence of compensation the consumer's real income will rise. Thus, after enough income has been taken away to offset this gain, if the good is not inferior he will buy less than he would have otherwise. For example, with the lower price, p_a, substituted for p_b the uncompensated quantity demanded will rise to x_a, but with the (negative) compensation to offset the rise in real income accompanying the price fall, quantity demanded will rise only to x_e.

16. Ordinal Utility Functions: Monotonic Transformations

Even in an ordinalist analysis it is often convenient to conduct the calculations in terms of utility functions rather than an indifference map.

By approaching the consumer's decision as a matter of utility maximization subject to a budget constraint one is enabled to use all of the mathematical apparatus of constrained maximization, the powerful Lagrangian techniques, their extension in Kuhn-Tucker methods, etc.

But what utility function can an ordinalist use, since he does not believe that absolute psychic utility can be measured? The answer is that for any given indifference map there is ordinarily[17] an infinity of utility functions any one of which will do just as well as any other. Given two indifference curves A and B with the latter preferable to the former, we can say arbitrarily that any output combination on the former offers 7 utils and the latter 11 utils or instead we can say that A provides 3 utils and B provides 59 utils. As long as points on the preferred indifference curve are assigned the higher utility numbers they provide all the information on preferences that the ordinalist needs for his calculations. We know that above any indifference curve the utility surface is level and that the surface gets higher as we move to higher indifference curves, but no more. The actual height of the utility surface above any indifference curve is left completely unspecified, so that any of an infinite number of utility surfaces is usually consistent with any given indifference map—that is, all of the surfaces in the set will give us the same indifference map. Thus, to an ordinalist the surface in Figure 4b represents only one of the infinite number of utility surfaces consistent with the given indifference map.

The switch from one such acceptable utility function to another is said to involve a *monotonic transformation*. A *transformation* may be described as the replacement of one set of numbers by another. A transformation is monotonic if a higher number in the first set is always replaced by a higher number in the second set. Table 1 illustrates the relationship.

TABLE 1

Goods Collection	a_1	a_2	a_3	a_4	a_5
First set of utility numbers	3	7	9	11	15
Second set of utility numbers	1	2.6	55	59	60
Third set of utility numbers	3	7	10	8	15

We see that the replacement of the first set of numbers by the second is, indeed, a monotonic transformation since whenever the numbers in the first row increase, the numbers in the second also increase. On the other

[17] There are some cases that may be considered pathological in which a peculiar indifference map precludes the existence of *any* utility surface consistent with it. This is the so-called integrability problem—an indifference map which permits no utility function is called *nonintegrable*.

hand, the replacement of the first set of numbers by the third is not monotonic, for the 9-util entry in row 1 is replaced by (transformed into) a 10-util entry in row 3, while the 11-util entry is transformed into an 8-util figure. Thus, the figures in row 1 assert that goods combination a_4 is preferable to a_3, whereas the utility figures in row 3 indicate that the preferences are reversed.

If we use $u = f(x_1, \cdots, x_n)$ to describe any one of the infinity of utility functions for a given indifference map, then any monotonic transformation of this function, i.e., any other acceptable utility function, is written

$$u^* = g(u), \qquad dg/du > 0,$$

where g is *any* function of u whose value increases whenever u increases. But that is just what is meant by monotonicity, and it is also precisely what is meant by the $dg/du > 0$. It is important to note that *any monotonic transformation of a utility surface will leave the indifference curves unchanged.*[18]

We will see next that the class of utility functions acceptable to an ordinalist is characterized by a property called *quasi-concavity*, which is extremely useful analytically. But in order to define the concept we must first deal with a preliminary matter.

17. *Interior Points on a Line Segment*[19]

Consider any two points A and B with coordinates (y_a, x_a) and (y_b, x_b), respectively (Figure 13). Now connect A and B by a line segment. The remainder of the chapter will make heavy use of the formula for the coordinates of any given point on this line segment which is given by the rule that the coordinates of any interior point on the line segment AB

---

[18] It is easy to prove that the indifference curves will be unaffected by the choice between u and u^* as a utility function. The shape of the indifference curve through any point in the indifference map for commodities 1 and 2 is given by the slope of that curve which we know is given by $-mu_1/mu_2 = -(\partial u/\partial x_1)/(\partial u/\partial x_2)$. The issue is what happens to that ratio when the utility function u is replaced by its monotone transform $u^* = g(u)$. But by the chain rule of differentiation (Chapter 4, Section 2, Rule 8) we know that $\partial u^*/\partial x_i = (dg/du)\ \partial u/\partial x_i$. Therefore, with the utility function u^* we deduce that the slope of the indifference curve becomes

$$-\frac{du^*/dx_1}{du^*/dx_2} = -\frac{(dg/du)\ \partial u/\partial x_1}{(dg/du)\ \partial u/\partial x_2} = -\frac{\partial u/\partial x_1}{\partial u/\partial x_2}.$$

Therefore, both utility functions necessarily yield the same number for the slope of the indifference curve through any point in the indifference map.

[19] The remainder of this chapter is made up of relatively advanced material.

will be a weighted sum of the corresponding coordinates of A and B, with a fixed weight, k, applied to *each and every* coordinate of A, and the weight $1 - k$ assigned to *each and every* coordinate of B, and where k is some number between zero and unity. More specifically, we have (in the two-dimensional case) the rather tedious but important

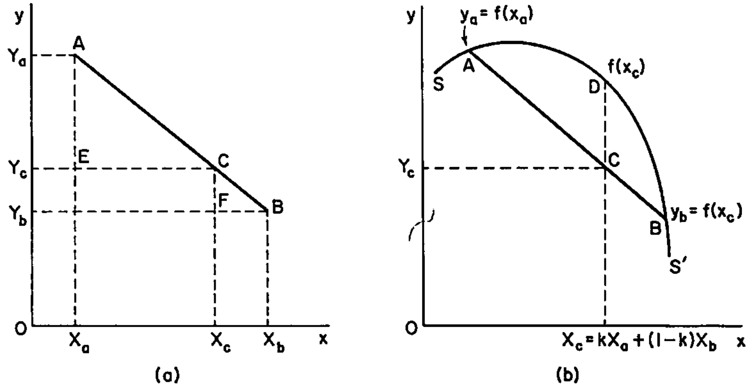

Figure 13

Proposition 3: Let point C with coordinates (y_c, x_c) be any point on line segment AB. Write x_c as a weighted average of x_a and x_b, so that $x_c = kx_a + (1 - k)x_b$, where $0 < k < 1$. (x_c is then called a *convex combination* of x_a and x_b.) Then, the y coordinates of points C, A, and B will satisfy the corresponding equation with the same value of k, i.e., we will have[20] $y_c = ky_a + (1 - k)y_b$. Moreover, $k/(1 - k)$ equals the ratio $(x_b - x_c)/(x_c - x_a)$.

It should be noted that the same result holds in n-dimensional space: Let A and B be two points with respective coordinates $(x_{1a}, \cdots, x_{na})$ and

[20] *Proof:* The triangles AEC and CFB are similar. Hence

(i) $$(x_b - x_c)/(x_c - x_a) = FB/EC = FC/EA = (y_b - y_c)/(y_c - y_a).$$

But substituting for x_c the expression $kx_a + (1 - k)x_b$, as given in the text, the first of the preceding fractions becomes

$$(x_b - x_c)/(x_c - x_a) = [(x_b - kx_a - (1 - k)x_b]/[kx_a + (1 - k)x_b - x_a]$$
$$= \frac{k(x_b - x_a)}{(1 - k)(x_b - x_a)} = k/(1 - k).$$

Hence by (i) we must also have

$$(y_b - y_c)/(y_c - y_a) = k/(1 - k) \text{ or } (y_b - y_c) - (ky_b - ky_c) = ky_c - ky_a,$$

which gives us our result: $y_c = ky_a + (1 - k)y_b$.

$(x_{1b}, \cdots, x_{nb})$ and if C is any point in the interior of the line segment connecting A and B, then there exists a number k such that $0 < k < 1$ and such that if x_{ic} is the ith coordinate of C, then $x_{ic} = kx_{ia} + (1 - k)x_{ib}$ for each and every i.

18. Concave and Strictly Concave Functions

Section 5 of Chapter 7 offered intuitive definitions of *concave* and *convex functions.*[21] Using the customary frame of reference—the shape of the surface as viewed from the floor of the diagram—we can envision a concave function as one having the general shape of an inverted bowl, while a convex function is shaped like an upright bowl. When a function is concave, if we take any two points on the surface of the bowl and connect them by a line segment, it is clear intuitively that every interior point on that connecting line segment will lie beneath the surface of the bowl. This characteristic is used by mathematicians to define a concave function.

Specifically, the theorem on the formula for interior points of a connecting line segment (Proposition 3 of the preceding section) is used to define concavity. The general notion is illustrated in Figure 13b in which SS' is the graph of a concave function $y = f(x)$. A and B are any two points on the surface and C is any point on the connecting line segment, and since the curve is concave, C lies below point D on SS' directly above x_c, the x coordinate of C. Now we have

(i) the y coordinate of point D, $y_d = f(x_c) = f[kx_a + (1 - k)x_b]$ by the formula for x_c from the preceding section;

(ii) by Proposition 3 the y coordinate of point $C = y_c = ky_a + (1 - k)y_b = kf(x_a) + (1 - k)f(x_b)$.

Therefore, we have

Definition: The function $y = f(x)$ is *strictly concave* if for any two values of x, call them x_a and x_b, every point $C = (y_c, x_c)$ on the line connecting (y_a, x_a) and (y_b, x_b) lies *below* the corresponding point $D = (y_d, x_c)$ on the graph of the function, i.e., [by (i) and (ii)] if

$$y_c = kf(x_a) + (1 - k)f(x_b) < f[kx_a + (1 - k)x_b] = y_d.$$

Definition: The function $y = f(x)$ is *concave* if for any two values of x, call them x_a and x_b, every point $C = (y_c, x_c)$ on the line connecting points

[21] The reader may well want to review the discussion and the distinction between the concept of a convex set (region) and that of a convex function.

$A = (y_a, x_a)$ and $B = (y_b, x_b)$ lies *on or below* the corresponding point $D = (y_d, x_c)$ on the graph of the function, that is, if the $<$ in the previous definition is replaced by $\leq$.

These concepts are readily extended to the $(n + 1)$-variable case $y = f(x_1, \cdots, x_n)$, where we merely write $y_a = f(x_{1a}, \cdots, x_{na})$, etc., and leave all other elements of the definitions of concavity and strict concavity completely unchanged.

Now strict concavity is the natural extension of the second-order maximum condition requiring a negative second derivative of the function being maximized. In effect, strict concavity requires that if the function is differentiable, its second derivatives be negative along *any* cross section, i.e., in any direction in the $(n + 1)$-dimensional (!) graph representing the function. Looked at another way, obviously in seeking to maximize we want the relevant function to be shaped like a hill or an inverted cup, and that is just what we mean by (strict) concavity.

Thus, if we are to analyze consumer behavior in terms of utility maximization, it would be convenient for the utility function to be concave, with the second-order conditions for maximization thereby satisfied. Unfortunately, ordinal utility analysis cannot accept such an assumption. For, as we have seen, given any indifference map, there is an infinity of utility functions corresponding to it, any one of which is as acceptable as any other. Furthermore, as we will confirm next, among those that are acceptable there will be some utility functions that are concave and some that are not. This is shown clearly by Figures 14a and 14b, both of which have the same indifference curves for combinations of x_1 and x_2, yet the first of which has a utility surface that is concave and the latter of which

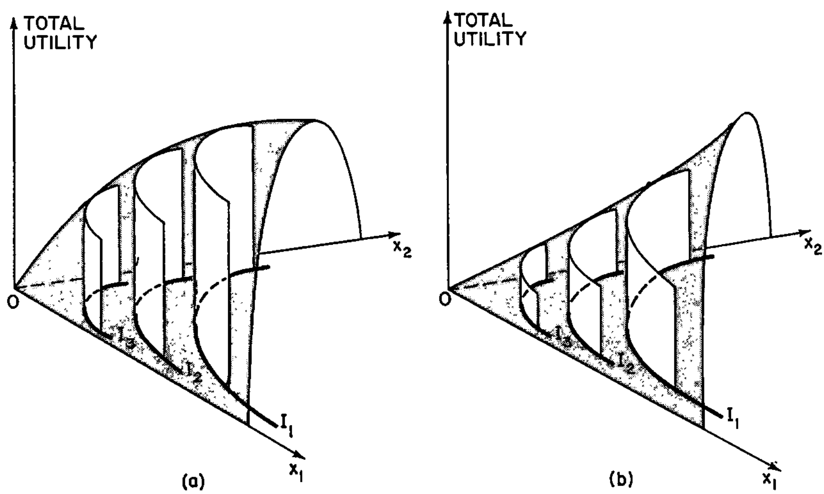

Figure 14

does not. The one utility surface is clearly a monotonic transformation of the other, and hence neither is more valid than the other as a representation of the indifference map. Thus, we simply cannot take an ordinalist point of view and yet require a utility function to be concave. Concavity of the utility function essentially has no implications about preferences in an ordinalist world because it is irrelevant for the shapes of the indifference curves. It is therefore just undefinable in any operational or observable terms. Some substitute concept has to be found by the ordinalist to serve the purposes of the second-order conditions. This substitute is a weaker condition, quasi-concavity, to which we now turn.

19. *Quasi-Concave Utility Functions*

Intuitively, a function is taken to be strictly quasi-concave if it follows one of two behavior patterns: (a) It is *monotonic* throughout; that is, in any direction in which its graph slopes uphill it does so "forever," i.e., for all values of its variables, and in any direction in which it slopes downhill, it also never reverses direction; (b) alternatively, if the function is not monotonic, it will have one single maximum with no other bumps or dents. Where the *second* alternative holds we see that the quasi-concave function does, indeed, resemble a concave relationship, as is illustrated in Figure 14a. But where the first alternative applies, as we will see (Figure 14b) that the shape of a quasi-concave surface may depart significantly from that of one that is concave. The formal definition of quasi-concavity bears some resemblance to that of concavity:

Definition: A function $y = f(x_1, \cdots, x_n)$ is *quasi-concave,* if given any two sets of values[22] of the x's $\mathbf{x}_a = (x_{1a}, \cdots, x_{na})$ and $\mathbf{x}_b = (x_{1b}, \cdots, x_{nb})$, where, say,

$$f(\mathbf{x}_a) \leq f(\mathbf{x}_b),$$

then

$$f(\mathbf{x}_c) \geq f(\mathbf{x}_a)$$

for $\mathbf{x}_c$ any point on the line segment connecting[23] $\mathbf{x}_a$ and $\mathbf{x}_b$.

That is, the function is quasi-concave if given any two points A and B, on its surface, then the height of any intermediate point, C, on a cross section through A and B is at least as great as the lower of the two points A and B. Similarly, we have

---

22 Note that here we introduce the vector notation $\mathbf{x} = (x_1, \cdots, x_n)$ so that $f(\mathbf{x})$ represents $f(x_1, \cdots, x_n)$.

23 [so that there exists a k value $0 < k < 1$ such that for any $i = 1, \cdots, n$, $x_{ic} = kx_{ia} + (1 - k)x_{ib}$]

Definition: A function $f(\mathbf{x})$ is *strictly quasi-concave* if for any two values $\mathbf{x}_a$ and $\mathbf{x}_b$ and any point $\mathbf{x}_c$ on the line segment connecting them, $f(\mathbf{x}_c)$ is greater than at least one of $f(\mathbf{x}_a)$ and $f(\mathbf{x}_b)$, i.e., if either

$$f(\mathbf{x}_c) > f(\mathbf{x}_a) \quad \text{or} \quad f(\mathbf{x}_c) > f(\mathbf{x}_b).$$

It is easy to prove

Proposition 4: Every function which satisfies the definition of concavity automatically satisfies that of quasi-concavity, and, similarly, every strictly concave function is automatically strictly quasi-concave.

Though we omit a formal proof, the reason for this result is not difficult to see. By definition, interpreting $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ as before, a concave function is one for which $f(\mathbf{x}_c)$ is at least as great as a weighted average of $f(\mathbf{x}_a)$ and $f(\mathbf{x}_b)$, i.e., for which $f(\mathbf{x}_c) \geq kf(\mathbf{x}_a) + (1 - k)f(\mathbf{x}_b)$. But then $f(\mathbf{x}_c)$ must obviously be at least as great as the smaller of the two items in the average, which is what quasi-concavity requires.

But while every concave function is therefore automatically quasi-concave, the converse is not true.

Proposition 5: A function which is quasi-concave need not be concave.

That is just what we mean by saying that quasi-concavity is a weaker condition than concavity. There are many functions that are quasi-concave but not concave. An example is all that is needed to prove the proposition. The function $y = x^2$ is clearly not concave for its graph "goes" upward toward the right at an increasing rate somewhat like the surface in Figure 14b, and so the line segment connecting any two points on its graph lies above its graph, not below it as concavity requires. However, $y = x^2$ *is* quasi-concave, for take any two values of x, $x_a < x_b$. For any x_c between them we may write $x_c = x_a + \delta$, $\delta > 0$. Then $y(x_c) = (x_a + \delta)^2 > x_a^2 = y(x_a)$, as is required for strict quasi-concavity.

Thus we have shown that $y = x^2$ is a function that is (strictly) quasi-concave but not concave (or strictly concave).

Next we show

Proposition 6: A quasi-concave function cannot have two (local) maxima.

Proof by contradiction: Suppose the contrary, that $y = f(\mathbf{x})$ is quasi-concave and yet possesses two (separated) local maxima $\mathbf{x}_a$ and $\mathbf{x}_b$. By definition every local maximum point is surrounded by points of lower altitude; therefore there must be a point $\mathbf{x}_c$ on the line segment joining $\mathbf{x}_a$ and $\mathbf{x}_b$ such that $f(\mathbf{x}_c) < f(\mathbf{x}_a)$ and $f(\mathbf{x}_c) < f(\mathbf{x}_b)$. But this contradicts the premise that $f(\mathbf{x})$ is quasi-concave.

This means that if a quasi-concave function has more than one maximum point they must be contiguous with the top of the graph, forming a level plateau.[24] This completes our characterization of the quasi-concave functions themselves. They may have a maximum point, but for a function that is strictly quasi-concave, never more than one, and if they have none, they will be monotonic. They need not be concave and include utility surfaces like that in Figures 14b as well as that in Figure 14a.

Next, we prove a proposition which shows that the quasi-concavity of utility functions is compatible with an ordinalist analysis, which, it will be recalled, treats two utility functions to be interchangeable if one is a monotonic transform of the other.

Proposition 7: Any function $y^* = g(y) = g[f(x)]$ obtained by a monotonic transformation from a quasi-concave (strictly quasi-concave) function $y = f(\mathbf{x})$ must itself also be quasi-concave (strictly quasi-concave).

Proof: By definition of quasi-concavity with $\mathbf{x}_c$, $\mathbf{x}_a$, and $\mathbf{x}_b$ defined as before, $y_c = f(\mathbf{x}_c) \geq y_a = f(\mathbf{x}_a)$. Then by the monotonicity of the transformation $y_c^* = g(y_c) \geq g(y_a) = y_a^*$, which proves the quasi-concavity of y^*, for it shows that y_c^* must also equal or exceed the smaller of y_a^* and y_b^*.

Thus, quasi-concavity is not a characteristic which evaporates as one utility function is replaced by another obtained from the former by a monotone transformation.

Finally, we come to another proposition that reveals the reason for adoption of the premise of strict quasi-concavity for utility functions, for though this premise is weaker than that of strict concavity, it is nevertheless sufficient, if used along with the premise of nonsatiety, to guarantee that indifference curves are convex to the origin. And once that property of indifference curves is satisfied all of the usual analysis of consumer behavior proceeds without difficulty.[25] Thus we conclude our discussion with

---

[24] But if the function is *strictly* quasi-concave, even two such points with equal values of $f(\mathbf{x})$ are impossible. Indeed, we have the following more general proposition: A function that is *strictly* quasi-concave cannot have two *global* maxima $\mathbf{x}_a$ and $\mathbf{x}_b$.

Proof: If both points are global maxima, we must have $f(\mathbf{x}_a) = f(\mathbf{x}_b)$. Hence, if $\mathbf{x}_c$ is any point on the line segment joining $\mathbf{x}_a$ and $\mathbf{x}_b$, we must have by strict quasi-concavity $f(\mathbf{x}_c) > f(\mathbf{x}_a) = f(\mathbf{x}_b)$. But this obviously contradicts the assertion that at points $\mathbf{x}_a$ and $\mathbf{x}_b$ the function $f(\mathbf{x})$ attains a global maximum.

[25] We also need for this purpose the negative slope of the indifference curves and the property that curves farther from the origin are always preferred, but these, as we have seen, follow from the nonsatiety assumption.

Proposition 8: If the consumer is not sated in any commodity (he prefers more of any good or any combination of goods, holding all other quantities constant) and his utility function is *strictly* quasi-concave, then any of his indifference curves (surfaces) will be convex to the origin.

Geometric demonstration: Given a strictly quasi-concave utility function (surface $OSTU$ in Figure 15a), select any two points A and B of equal

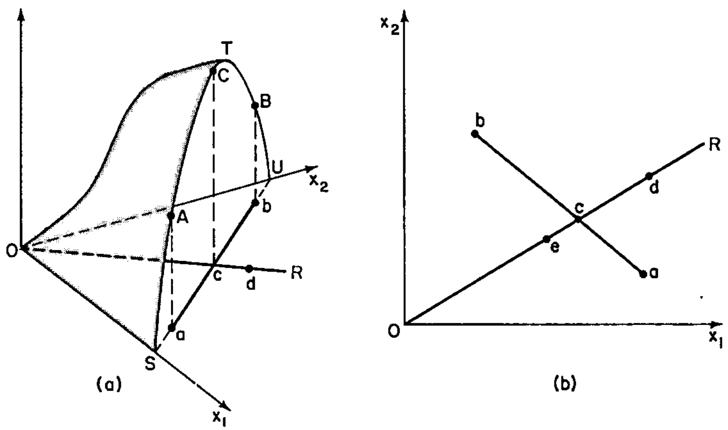

Figure 15

height (utility) on its surface so that their projections a and b on the floor of the diagram (the x_1, x_2 plane) lie on the same indifference curve. Now connect a and b by a line segment and draw any ray (line segment), OR through the origin. To find another point on the (unknown) indifference curve which connects indifferent points a and b we ask where the ray, OR, can intersect that curve. First, we know that the unknown intersection point cannot be point c where OR crosses ab for by the definition of strict quasi-concavity the utility of point c must be greater than the utility of at least one of points a and b and hence c cannot be indifferent to points a and b. Moreover, that point of indifference on OR cannot be a point like d above and to the right of ab since d offers more of both goods x_1 and x_2 than does point c, so that by nonsatiety d must be preferred to c, which is in turn preferred to a and b. So, if there is any point on OR which is indifferent to a and b, it must be a point such as e, which lies below and to the left of ab, but that is exactly what we mean by convexity to the origin of the indifference curve through points a, e, and b.

20. *Elementary Mathematics of Demand Analysis*

As we have noted, much of standard demand analysis is based on the formulation which takes the consumer to maximize a utility function

$$u = f(x_1, \cdots, x_n)$$

subject to a budget constraint

$$p_1x_1 + p_2x_2 + \cdots + p_nx_n = m,$$

where x_i is the quantity of commodity i purchased by the consumer, p_i is its price, and m is the total amount of money available to him. Using a Lagrangian approach to the problem (cf. Chapter 4, Section 8) we obtain the expression

$$u_\lambda = f(x_1, \cdots, x_n) + \lambda(m - p_1x_1 - p_2x_2 - \cdots - p_nx_n),$$

which we maximize by setting each of its partial derivatives equal to zero:

$$\frac{\partial u_\lambda}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda p_1 = 0$$

$$\vdots$$

$$\frac{\partial u_\lambda}{\partial x_n} = \frac{\partial f}{\partial x_n} - \lambda p_n = 0$$

$$\frac{\partial u_\lambda}{\partial \lambda} = m - p_1x_1 - \cdots - p_nx_n = 0.$$

Here, by definition, $\partial f/\partial x_i$ is the marginal utility of i. Dividing the preceding equation corresponding to commodity i by the equation which refers to commodity j, we have

$$\frac{\partial f}{\partial x_i}\bigg/\frac{\partial f}{\partial x_j} = \frac{p_i}{p_j}. \tag{5}$$

This is the equilibrium condition which was derived in a less formal manner earlier in the chapter. It states that in equilibrium the ratio of the prices of the two commodities must be equal to the ratio of their marginal utilities, i.e., to their marginal rate of substitution.

Let us see finally how the utility function can be used to derive a specific demand relationship.

Example: Suppose a consumer has \$90 available to be divided between commodities A and B, and suppose the unit price of B is fixed at 20 cents. What will be his demand equation for A if his utility function is $u = \log x_a + 2 \log x_b$?

Answer: We are given $p_b = 0.2$ and $m = 90$. By direct differentiation of the utility function we obtain $du/dx_a = 1/x_a$ and $du/dx_b = 2/x_b$. Substituting these into equilibrium condition (5) we have

$$\frac{1}{x_a}\Big/\frac{2}{x_b} = \frac{p_a}{0.2}$$

or

$$x_b = 10p_a x_a.$$

Substitution of this value of x_b, $p_b = 0.2$ and $m = 90$ into the budget $m = p_a x_a + p_b x_b$ yields

$$90 = p_a x_a + 0.2(10p_a x_a) = p_a x_a + 2p_a x_a$$

or

$$p_a x_a = 30.$$

This is our desired demand equation, which, incidentally, happens to be a rectangular hyperbola.

To summarize, given a utility function, income, and the prices of all other commodities, we obtain the demand for the remaining commodity by direct substitution into the equilibrium condition (5) and the budget constraint.

PROBLEMS

Find the demand function for commodity A, given

1. $p_b = \$3$, $m = \$20$, $u = 4x_a x_b$.
2. $p_b = \$12$, $m = \$246$, $u = e^7 x_a x_b$.
3. $p_b = \$8$, $m = \$100$, $u = 2x_a - 3x_a^2 + x_b - 4x_b^2 + 782$.

REFERENCES

American Economic Association, George J. Stigler and Kenneth E. Boulding (eds.), *Readings in Price Theory*, Richard D. Irwin, Inc., Homewood, Ill., 1952, Articles 1–4.

Henderson, James M., and Richard E. Quandt, *Microeconomic Theory*, 3rd edition, McGraw-Hill Book Company, New York, 1971, Chapter 2, Sections 1–7.

Hicks, J. R., *Value and Capital*, 2nd edition, Oxford University Press, New York, 1946, Part I and Mathematical Appendix to these chapters. (Rather difficult reading.)

———, *A Revision of Demand Theory*, Oxford University Press, New York, 1956.

Malinvaud, Edmond, *Lectures on Microeconomic Theory*, North-Holland Publishing Company, Amsterdam, 1972, Chapter 2 (rather difficult).

Marshall, Alfred, *Principles of Economics*, 8th edition, Macmillan & Co., Ltd., London, 1922, Book III and pp. 838–840.

Morgenstern, Oskar, "Demand Theory Reconsidered," *Quarterly Journal of Economics*, Vol. XLII, February 1948.

Samuelson, Paul A., *Foundations of Economic Analysis*, Harvard University Press, Cambridge, Mass., 1947, Chapters V and VI. (Highly mathematical.)

Walsh, V. C., *Introduction to Contemporary Microeconomics*, McGraw-Hill Book Company, New York, 1970, Chapters 1 and 3 (axiomatic treatment of utility theory).