DE ECONOMIST 133, NR. 1, 1985

UTILITY AND ECONOMICS* *

BY

ARIE KAPTEYN*

1 INTRODUCTION

'Utility' has been the central concept in economics for a long time. See, for example, Stigler (1950) for an excellent review of the history of the concept of utility in economics. There have been various attempts to undermine the position of the utility concept. The most famous is probably Samuelson's theory of revealed preference, which was an attempt to get rid of 'utility' for good.[1] Due to Houthakker (1950), we know now that revealed preference theory is basically equivalent to neoclassical choice theory. Thus, at present the utility concept is more important than ever in economic theory.

Still, economists have a somewhat ambiguous attitude towards utility, which is exemplified most clearly by the way in which they try to enhance their empirical knowledge of utility. As a rule, measurement takes place *indirectly;* asking people questions about their utility is considered to be fruitless. The only way in which economists allow themselves to learn about utility functions of economic agents is by observing their behaviour. This self-restraint with respect to the measurement of the central concept of their discipline has far-reaching consequences for the nature of economic theorizing. In this paper, I will discuss some of these consequences.

2 WHAT IS UTILITY?

'Utility' is roughly synonymous with 'satisfaction,' 'well-being,' 'welfare,' 'happiness,' 'pleasure,' *etc.* Generally, we can increase our utility by undertaking enjoyable activities or purchasing things we desire. In the words of Bentham (1823, p. 3): 'By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good or happiness ...' Of course, this description is rather too loose to serve as a definition. To Bentham and his contemporaries 'utility' was a *primitive term.* That is, they appeared to

* Professor of Econometrics, Tilburg University, the Netherlands.

* * This is a slightly adapted version of my inaugural address at Tilburg University. I thank Tom Wansbeek for his helpful comments.

1 Samuelson (1938, p. 62): 'I propose, therefore, that we start anew in direct attack upon the problem, dropping off the last vestiges of the utility analysis.'

agree on the meaning of 'utility,' without an exact definition being necessary.

In modern economic theory, this is not the common way the utility concept is introduced. Nowadays the *preference ordering* is taken as a primitive term. An individual's preference ordering ranks alternatives in order of preference. If two individuals express the same preference in each conceivable decision situation, their preference orderings are identical. If, however, given a choice between going to a soccer game and attending a concert of classical music, one individual chooses the soccer game and the other one prefers the concert, then their preference orderings are different.

One would expect that, in making decisions, an individual will try to enhance his utility. If I express a preference for a soccer game over a classical concert, then it seems reasonable to infer that my utility will be higher when I visit the game than when I attend the concert. If not, we would have some doubt about the utility concept used, or the concept of preference would be used in a somewhat perverse manner. In economic theory, such a perverse relationship between utility and preference cannot occur, because 'utility' is *defined* such that a preferred alternative always represents a higher utility level (or at least not a lower one) than the rejected alternative. 'Utility' is no longer a primitive term, but it is defined in terms of preferences.

An example may clarify this: microeconomic models of consumer behaviour describe how people choose from different consumption bundles. It is commonly assumed that consumers are able to rank all relevant alternatives in order of preference. A utility function, *i.e.* a relationship that tells us for any consumption bundle how much pleasure the consumer derives from it, has to satisfy the requirement that it agrees with the consumer's preference ordering, in the sense that each time the consumer prefers bundle A to bundle B, the utility function tells us that A gives more utility than B. Under rather plausible assumptions concerning the structure of a consumer's preference ordering, it can be shown that such a utility function exists (see, *e.g.*, Debreu, 1959). It is then said that the utility function *represents* the preference ordering.

For the simplest case, where only quantities of one good have to be compared, the utility function can be drawn in a simple diagram, as in Fig. 1, which depicts utility functions of income of three different individuals.[2]

3 ON ORDINAL AND CARDINAL MEASURABILITY

Looking at Fig. 1, we notice that each of the individuals prefer more income to less. In other words, each time they have a choice between two income levels, each individual chooses the higher income level. So, with respect to income the three individuals have identical preference orderings. 'More is better than less.' As a representation of a preference ordering the three utility functions are hence equivalent. In standard economic parlance, we say that 'utility' is an

2 In this paper, 'income' is defined throughout as annual after tax disposable income.

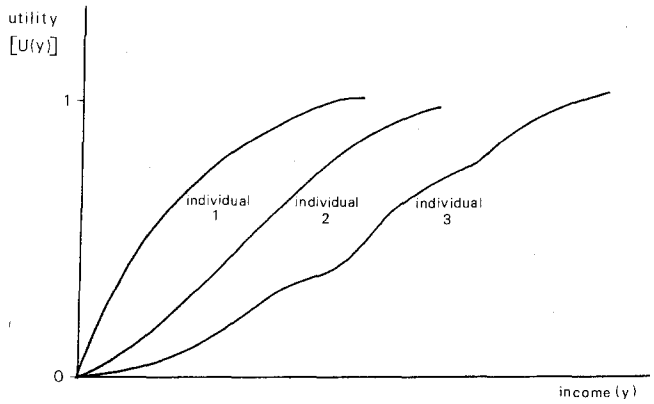

Figure 1 – Utility functions of income of three individuals

ordinal concept: The numbers on the vertical axis are arbitrary. The only requirement is that a higher utility level is associated with a higher number. Only the ordering of the numbers on the vertical axis counts, not their value.

Although Fig. 1 refers to the simplest possible case where only quantities of one good (income) are compared, the same conclusion holds for situations in which one has to choose between bundles of goods. Consider, for example, the case of a man whose lunch consists of only two ingredients, water and bread. He is able to rank all water/bread combinations in order of preference. Assume that we are able to find a utility function which represents this preference ordering, *i.e.*, in each pair-wise comparison of water/bread bundles the preferred bundle has the highest utility. It is then very simple to find another utility function representing the same preference ordering, for example by multiplying all utility levels higher than 10 by 100 and by dividing all utility levels below 10 by 100. This transformation does not affect the ordering of utility levels.

Clearly, the preference ordering contains all the information we need to explain (or predict) consumer choice. This is because the preference ordering tells us, for each situation, which consumption bundle is preferred. Hence we can predict the consumer's choice in each situation. Since, in turn, an ordinal utility function provides us with complete information about the corresponding preference ordering, it also provides all information that is necessary to explain behaviour.

The conclusion that an ordinal utility function suffices to explain behaviour was first drawn by Fisher and Pareto at the turn of the century. Since then, ordinal utility has dominated economic theory (*cf.* Stigler, 1950).

Ordinal measurability is a rather weak property. Although someone's utility function can tell us that this individual prefers one liter of water and 5 slices of bread to half a liter of water and 6 slices of bread, it cannot tell us anything about utility *differences*. We do not know whether the utility difference

between 1 liter of water and 5 slices of bread and half a liter of water and 6 slices of bread exceeds the utility difference between half a liter of water and 6 slices of bread and half a liter of water and 5 slices of bread. It that were possible, utility would be said to be measurable on a *cardinal scale.*

Of course, the fact that an ordinally measurable utility concept suffices to describe economic choices[3] does not imply that cardinal measurement is impossible. Still, this appears to be the conclusion which most economists have drawn from Fisher's and Pareto's analyses.

The supposition that utility can only be measured on an ordinal scale has direct consequences for our capacity to learn something about utility. To illustrate this, let us reconsider Fig. 1. Apparently, it does not make sense to ask someone what utility level he attaches, for example, on a [0,1] scale, to an income of say, 75,000 guilders. The three graphs are equivalent representations of the same preference ordering. Yet they generate quite different answers to the above question. Also the reverse question, which income corresponds to a certain point on the utility scale, does not make sense. So, the measurement of utility functions by means of direct questioning (I will call this 'direct measurement') is impossible if utility can only be measured on an ordinal scale.

Indeed, the prevailing opinion among economists appears to be that direct measurement of utility is impossible. This is nicely illustrated by referee reports Bernard van Praag and I have received with respect to one of our papers.[4] One referee states: 'Unlike the authors, I am very dubious of the extent to which we can learn anything useful by asking people how happy they are, or what level of income will make them happy to a certain degree.' And the second referee adds: '... the authors – like everyone else – do not and cannot measure utility directly.'[5]

Now assume for a moment that indeed utility can only be measured on an ordinal scale. Still, out of mere stubbornness, we ask a respondent in a survey which income he associates with a utility level of, say, 0.5 on a [0,1] scale. What happens then? According to our assumptions, the respondent faces an impossible task. Psychologists may be able to predict what the respondent will do next; maybe he flatly refuses to answer the question, or he may express a low opinion about the interviewer's intelligence, or he may become aggressive towards the interviewer or simply throw him out.

I do not know whether Bernard van Praag was aware of these and other risks when he started asking people which income levels they associate with verbally

3 I ignore choices in uncertain or dynamic settings. In these cases a cardinal utility concept appears to be necessary. See, for instance, Von Neumann and Morgenstern (1944), and Koopmans (1972).

4 'A New Approach to the Construction of Family Equivalence Scales.'

5 The paper was eventually accepted for publication, but not by the journal for which these referees acted.

labelled utility levels like 'excellent,' 'good,' 'sufficient,' *etc.* (Van Praag, 1971). Since he started, these kind of questions have been posed to about 50,000 respondents in different countries and it appears that attempts at direct measurement of utility functions of income do not lead to strong emotions on the part of the interviewee nor to physical harm to the interviewer.

Although the happy ending of this adventure may give us some confidence in the possibility of direct utility measurement, it is not the only piece of evidence. For example, it appears to be possible to discriminate between possible forms of utility functions. In other words, for the respondents in surveys in which direct utility measurement was undertaken, the three examples drawn in Fig. 1 are definitely *not* equivalent (see Van Herwaarden and Kapteyn, 1981).

4 INDIRECT MEASUREMENT OF UTILITY FUNCTIONS

In the remainder of the paper, I will consider the question of to what extent direct measurement of utility functions can help to solve some theoretical and practical problems in economics. To this end we first have to give some thought to the role of utility functions in empirical economic models.

In economic models of behaviour, it is invariably assumed that an individual behaves in such a way that his utility is maximized. In his attempts to maximize utility, he is hampered by one or more constraints. The standard example of this kind of modelling is the neoclassical model of consumption behaviour where the consumer faces the task of spending his income on consumption goods so as to maximize his utility. Under rather weak assumptions, the consumer's purchases are uniquely determined by his income and the prices he has to pay (together constituting the budget constraints) and the form of his utility function (representing his preferences). As said earlier, it suffices that the utility function is measurable on an ordinal scale.

Thus, in economic models an individual's behaviour is completely determined by two things: his preferences – represented by an ordinal utility function – and the constraints that limit his behaviour. If a researcher knows these constraints, he can employ observations on the individual's behaviour to draw conclusions about the individual's preferences.

This is the so-called revealed preference approach but I will usually denote it as *indirect utility measurement,* in contrast with direct utility measurement, as introduced above. The advantage of indirect utility measurement is that, in general, one can do without the assumption of cardinal measurability. Invoking Ockham's razor[6] this assumption has therefore been removed from economic theory.

The removal of the cardinality assumptions is not without cost, however.

6 'What can be done with fewer (assumptions) is done in vain with more.' Ascribed to William of Ockham (ca. 1285–1349). See Edwards (1967, p. 307).

The price that has to be paid consists of a sizeable number of other assumptions. To make that clear, let us once again consider the economic theory of consumer behaviour. To obtain reasonably accurate knowledge about the utility function of a consumer, we need a rather large number of observations on his behaviour in different situations. Specifically, we have to observe the consumption bundle he chooses at different combinations of prices and income. However, we cannot vary prices and income at will.[7] Hence a different approach is adopted. The utility function is assumed to have some, *a priori* specified, plausible functional form with some unknown parameters whose values have to be determined.

By and large, different families, or groups of families, are assumed to have identical utility functions. By observing families, or groups of them, for an extended period of time, sufficient variation in prices and income will occur so that one is able to estimate the unknown parameters. Once the demand system and hence the utility function has been estimated, future behaviour can be predicted.

What are, in view of this procedure, the additional assumptions that are required to measure utility functions indirectly? They include the following:

- The *a priori* specified functional form is correct.
- The utility functions of different families are indeed identical and they do not shift over time.
- Observed behaviour of families does indeed stem from the maximization of a utility function subject to constraints.
- The utility of families is not dependent upon the behaviour of other families.
- Families have at their disposal all relevant information required for an optimal decision. The information is certain.

This list of assumptions is easily extended. For example, in many empirical studies aggregation requires additional assumptions. On the other hand, some of these assumptions have been relaxed in certain studies.[8] However this may be, it is a fairly long list and each of the assumptions is questionable.

The most problematic aspect of this list of assumptions is not its length, but the virtual impossibility to test assumptions separately. By way of example, imagine that in the way described above we have estimated 'the' utility function of 'the' (representative) household in the Netherlands and the corresponding demand functions for consumption goods. Furthermore, imagine that we use the results to predict aggregate private consumption in the future.

7 Exceptions are to some extent prisons, psychiatric wards and animal laboratories where experiments are carried out occasionally. See, for instance, Battalio *et al.* (1973), Battalio, Kagel, Reynolds (1978), Lea (1982).

8 See, for instance, Phlips (1972, 1974), Manser (1976) and Pollak (1978), and the references therein for models where changing preferences are allowed. In Muellbauer (1977, 1980) one finds many references to models where utility functions depend on family composition.

Finally, suppose the predictions are inferior.[9] What should we conclude then? Were we mistaken in assuming that families try to maximize utility? Do different families have different utility functions? Do utility functions shift over time? Did we specify an incorrect functional form? Did we make mistakes in the aggregation of family demand functions to aggregate demand functions? Are there informational constraints that prevent families from finding a position with maximum utility? Disentangling these and other possibilities presents an arduous task.

Now assume that we are able to measure a consumer's utility function directly, that is, without resort to observations about his behaviour. Such measurement could be the result of extensive interviews, or the outcome of psychological experiments.[10] *In this case the assumptions can be tested separately.* We can first investigate what the form of the utility function is. Next, we investigate whether different consumers have identical utility functions. If not, we can try to find out what causes the differences. In the third place, the measured utility functions can then be used to see if consumers do maximize utility or whether they follow different rules of behaviour.

Thus we see that if it is possible to measure utility directly, independent of observations about behaviour, research into economic behaviour can be split up into a number of elements that can be investigated separately. This makes research into economic behaviour simpler, although not simple.

The three elements distinguished here, measurement of utility functions, explanation of differences between individuals, tests of behavioural hypotheses, constitute a research program to which I have devoted a substantial part of my time, in collaboration with others. A major part of this work has been done as part of the so-called Leyden Income Evaluation Project.

It is worth noting that the research program outlined here fits in with the predominant paradigm in economics. But the direct measurement of utility functions leads to a research practice that is quite different from existing practice. In the remainder of the paper, this will be illustrated by means of the second part of the research program: the explanation of differences in utility functions among consumers.

5 PREFERENCE FORMATION

I will denote theories dealing with the explanation of differences in utility functions among individuals, or families, as theories of preference forma-

9 A rather plausible assumption, considering experiences with the prediction of aggregate consumption in the Netherlands over the past twenty years. See Van der Leeuw (1984).

10 As experiments dealing with decisions under uncertainty. See Grether (1978) or Schoemaker (1982) for reviews.

tion.[11] It is customary to distinguish two aspects to preference formation, *viz. habit formation* and *preference interdependence.* Habit formation is the phenomenon by which my behaviour in the *past* (or results of that behaviour) influences my *present* preferences. Preference interdependence denotes the phenomenon by which behaviour of *others* influences *my* preferences. In the first instance, I like to have a house with a garden because I am used to having one. In the second instance, I like to have a house with a garden because most of my friends have one.

To incorporate preference formation into demand systems creates a number of thorny problems. As observed by Duesenberry (1949, p. 17): 'Ordinarily we try to measure preference parameters (or functions of them) by market behavior, since we cannot observe the preferences directly. With shifting parameters we should be carrying indirect measurement a step further. We would not only have to measure the preference parameters but the parameters of the relation governing shifts in the preference.' To find the correct specifications to describe these shifts then becomes a risky enterprise, especially since we do not know the correct form of the utility function nor the correct behavioural rule (utility maximization or something else), *etc.* As a result, most economists have abstained from incorporating preference formation into models of demand. Habit formation is encountered occasionally in empirical work.[12] Empirical studies of preference interdependence in demand systems are at present virtually nonexistent.[13]

Some economists simply deny that preferences may change, *e.g.* Stigler and Becker (1977, p. 76) who state that '. . . tastes (do) neither change capriciously nor differ importantly between people . . . one does not argue over tastes for the same reason that one does not argue over the Rocky Mountains – both are there, will be there next year, too, and are the same to all men.' In their paper, they explain a number of instances of apparent shifts in preference, *e.g.* addiction, advertising and fashion, by assuming that in all these cases it is an individual's efficiency as a producer of pleasure that changes, but not his preferences. It is hard to imagine that their paper would have been written had the authors had available directly measured utility functions and had these utility functions been different for different individuals.

In sum, indirect (revealed preference) measurement of utility functions makes it harder to model preference formation adequately. On the other hand this makes it easier to maintain that preferences are constant and the same for everyone, irrespective of empirical evidence.

11 For, utility functions represent preferences. So, if preferences differ, utility functions differ, and *vice versa.*

12 *E.g.,* Houthakker and Taylor (1970), Philips (1972, 1974), Manser (1976), Darrough, Pollak and Wales (1983).

13 An exception is Kapteyn, Van de Geer, Van de Stadt, Wansbeek (1984).

6 ARE PREFERENCES CONSTANT AND THE SAME FOR EVERYONE?

To answer this question, let us consider some examples. The first example is from Duncan (1975). In two surveys of housewives in the Greater Detroit area, held in 1955 and 1971, the respondents were asked to express their satisfaction or dissatisfaction with their standard of living. Although real median income in the 1971 survey was 42% higher than in the 1955 survey, the distribution of responses in both years was virtually the same. Within each survey, however, satisfaction with one's standard of living correlates significantly, and positively, with income.

Can we reconcile these outcomes with constant and identical preferences? If preferences are constant and the same for everyone, one would expect that people with a higher income are more satisfied with it than people with a lower income. This expectation is confirmed in both years. But one would also expect the 1971 respondents to be more satisfied, on average, than the 1955 respondents. This expectation is not confirmed at all. Duncan's findings can be explained better by the hypothesis that preferences are *relative*. That is, people evaluate their income or standard of living by comparing it to the income or standard of living of others.

The idea that preferences are relative is also confirmed by a well-known study of Easterlin (1974), who uses self-ratings of happiness by individuals in various countries – rich and poor, eastern and western. *Within* each country, the happiness ratings correlated positively with income, but *across* countries there is no discernible relationship between national income per capita and the mean happiness rating of a country.

As one might expect, the relativity of preferences, or evaluations, is not restricted to the income domain. For instance, Davis (1966) finds that self-confidence of American students – and especially their career plans – depend heavily on their performance relative to their fellow students. This happens despite the fact that in the American system the best student at one university might very well have been the worst student at another university.[14]

The number of these examples can be increased almost at will. In sociology, *Relative Deprivation* theory (RD theory, for short) explains phenomena like these.[15] The extent to which an individual is relatively deprived with respect to a certain dimension depends on whom he compares himself with and whether these others are doing better or worse on this dimension than the individual himself. Suppose, the relevant dimension for me at this moment is 'playing chess' and I compare my own abilities with those of the faculty members at the department of econometrics of Tilburg University. As far as chess ability is concerned, I will be rather deprived. If, instead of comparing myself with my

14 See also Bassis (1971) for a related analysis and Kapteyn and Wansbeek (1982) for an interpretation.

15 See, for instance, Stouffer *et al.* (1949), Merton and Kitt (1950), Davis (1959), Runciman (1966).

colleagues in the department, I compare myself with the other members of my family, my deprivation is really not so dramatic. However, this tends to take a turn for the worse, with my children getting older, and getting better at the game.

One is not necessarily deprived relative to others. One can also compare to one's own situation in the past. Comparison of my present chess playing ability to my abilities in the past reinforces my deprivation with respect to this dimension.

7 CAN EVALUATIONS BE ABSOLUTE?

Before returning to the more restricted area of preference formation, let me briefly discuss a related theory from experimental psychology. This so-called *Adaptation Level* theory (AL theory, for short) has been developed by Helson (1964). The central notion in this theory – adaptation level – is defined by Helson as follows '... adaptation level is defined as a weighted geometric mean of all stimuli impinging upon the organism from without and all stimuli affecting behavior from within.'[16] Of course, this is a rather vague description, which has to be operationalized within a concrete context. Therefore, I briefly present some examples.

The first example deals with an experiment carried out by Helson and Kozaki.[17] Four groups of five experimental subjects were shown random patterns of 10, 12, 14, 16 and 18 dots respectively. Each pattern was visible during 3/10ths of a second. Before this, one group was shown a random pattern of 4 dots, the second one a pattern of 13 dots and the third group a pattern of 32 dots. The fourth group (the control group) was not shown a dot pattern in advance. During the experiment, the experimental subjects had to estimate how many dots they were shown each time. It turned out that the first group, which had been shown 4 dots before the experiment, consistently gave the highest estimates whereas the third group, which had been shown 32 dots in advance, gave the lowest estimates. The remaining two groups gave estimates in between. Here, the adaptation level is defined as a weighted geometric average of the random dot pattern shown to them before the experiment. By this definition, the first group has the lowest adaptation level, and hence provides the highest estimates in the experiment. The third group has the highest adaptation level and therefore gives the lowest estimate of the number of dots shown to them.

A somewhat related experiment is due to Ross and Thibaut (1974). Experimental subjects were shown 19 slides, each for one-fourth of a second. The slides showed 16×16 matrices containing zeros and ones. The subjects were told that each matrix represented the judgment of a jury of 16 laymen who had evaluated paintings with respect to 16 different dimensions, like color, com-

16 Helson (1964, p. 59).

17 *Cf.* Helson and Kozaki (1968). The experiment is also described in Helson (1971).

position, perspective, *etc.* A 'one' represents a good evaluation and a 'zero,' represents a bad evaluation. Thus, the more ones a slide shows, the better was the corresponding painting evaluated. For half of the subjects, the nineteen slides were presented in increasing order of the number of ones. For the other half of the subjects, the presentation was in decreasing order of the number of ones. After that, both groups were shown one more slide with exactly as many ones as zeros. The subjects were asked to rate this last picture on a scale from 1 (an extremely bad painting) to 20 (an outstanding painting). It turns out that the first group, which saw slides in order of an increasing number of ones, gave a significantly lower rating to this last painting than the second group.

The common elements of RD theory and AL theory are obvious. Evaluations, opinions and perceptions are nonconstant, but are formed under the influence of previous experiences. These previous experiences then serve as a standard of comparison. In view of the large variety of phenomena that is covered by AL theory and RD theory, it would be quite a surprise if utility functions were immutable. And indeed they are not! Approximately eight years ago, I formulated a theory of preference formation which has major communalities with AL and RD theory, albeit that at the time I had never heard of AL theory.[18]

8 A THEORY OF PREFERENCE FORMATION

I will briefly sketch the basic features of my theory and then pay some attention to empirical evidence.

The central notion of the theory is the so-called *perceived distribution.* Let us, for didactic reasons, only consider preferences with respect to consumer goods and, moreover, restrict our attention to just one good. If we let 'income' be this one good, then the perceived distribution is a perceived income distribution. I will try to clarify this concept by means of a series of graphs.

The solid line in Fig. 2 represents, for some hypothetical society, the income distribution. That is, for an arbitrary income measured along the horizontal axis (*e.g.* the point y) the corresponding point at the vertical axis (point A) indicates which proportion of the population has an income less than or equal to this income.

This income distribution may be observable for the Central Bureau of Statistics, but for individuals in society the distribution is rather irrelevant, as they do not observe all other individuals and their incomes. An individual observes only a subset of the population, and this subset is not chosen randomly. The dashed line in Fig. 2 represents the income distribution that is perceived by someone who knows mainly people with low incomes. For ease of language, I shall say that this individual has a 'poor reference group.' The

18 See Kapteyn (1977). I got acquainted with AL theory through discussions with Richard Easterlin. Hans Werner has been so kind as to provide me with a number of recent references to the literature.

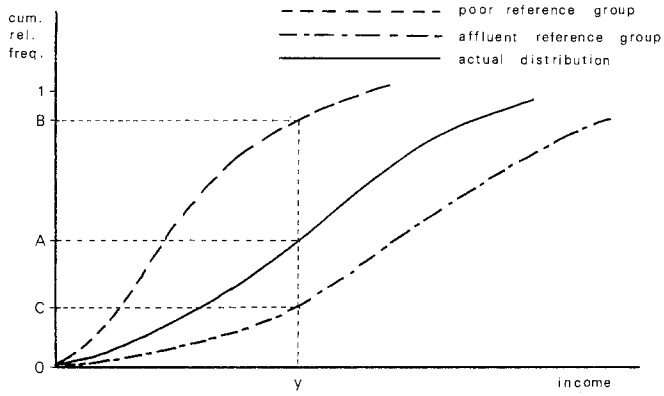

Figure 2 – Actual and perceived income distribution in a hypothetical society

proportion of people in his reference group with an income less than or equal to y is equal to OB. Let us, finally, consider the third curve in Fig. 2, which is located to the right. This is the perceived income distribution of someone with an affluent reference group. In this reference group only a proportion OC of all people have an income less than or equal to y.

Assuming that the notion of a 'perceived distribution' is more or less clear, I will next complicate this notion a little bit. First of all, the definition will be broadened to also incorporate an individual's own income. In the perception of incomes, one's own income plays a role, and probably an important one. In the second place, I want the definition of a perceived distribution to encompass *dynamic* aspects. Fig. 3 illustrates what I mean by that.

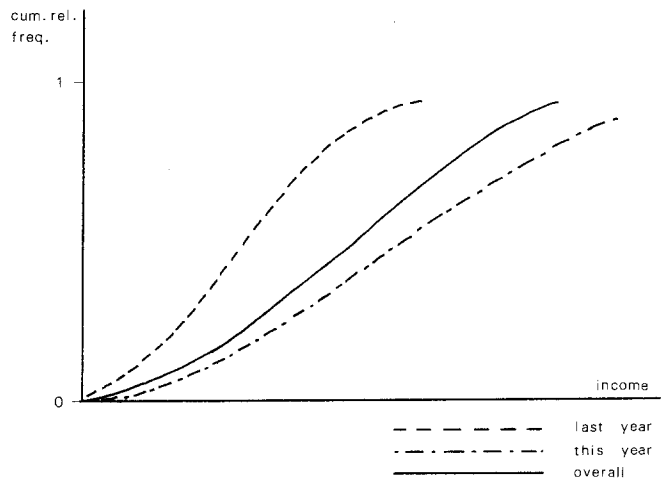

Figure 3 – The overall perceived income distribution as a convex combination of contemporaneous perceived distributions

In Fig. 3 we consider an individual who has, during his life, perceived two income distributions, *viz.* one during last year (the curve located to the left) and one during this year (the curve located to the right). I call these distributions contemporaneous distributions in the two separate years. I define the *overall perceived distribution* (today) as a convex combination of the two contemporaneous distributions. The weights in the convex combination represent the relative importance of the two years in the formation of the overall perception. The magnitude of the weights is an empirical matter. It is assumed in Fig. 3 that the individual under consideration gives a higher weight to this year than to the previous year. This appears to be a reasonable assumption, as one will have somewhat less vivid memories of the previous year.

Of course, most people live longer than two years, but the principle remains the same. In subsequent periods an individual perceives various contemporaneous income distributions and today's overall perception is a convex combination of all contemporaneous distributions. This is illustrated in a slightly different fashion in Fig. 4. The curve located most to the left now presents the overall perceived income distribution in the previous year, *i.e.*, a convex combination of all contemporaneous income distributions perceived by the individual up to and including last year. The right hand curve represents this year's contemporaneous perceived income distribution. The solid line is today's overall perceived income distribution. It is, in turn, a convex combination of the two other curves. Notice that the solid line is closer to the left-hand curve than to the right-hand one. This reflects the presumption that all previous years combined receive a greater weight than this year's contemporanous distribution. Whether or not this presumption holds true can be verified empirically.

To keep my story reasonably simple, I will not consider further complications, such as the fact that distributions of income usually refer to families

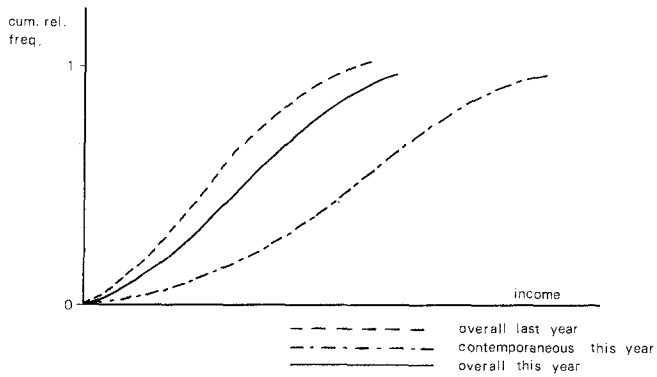

Figure 4 – The overall perceived income distribution this year as a convex combination of the overall perceived distribution of last year and the contemporaneous distribution of this year

rather than individuals. Nor will I consider an extension of the notion of a perceived distribution to more dimensions, although that would not create any new conceptual problem. Furthermore, I will refer to 'overall perceived distributions' as 'perceived distributions' for short.

Thus, for didactic reasons, I restrict myself to the distribution of incomes. My theory of preference formation now states that someone's evaluation of some income y on a zero-one scale is *equal to the value of the perceived distribution function corresponding to that income y*. For example: if according to my perceived distribution function 70% of all incomes are below 60,000.00 guilders, then I evaluate an income equal to 60,000.00 guilders by 0.7. If 40% of the incomes are lower than 35,000.00 guilders, then I evaluate an income of 35,000.00 guilders by 0.4, *etc.*

Let us now return to the Detroit housewives. Despite an average real increase of their incomes by 42%, their average satisfaction with their standard of living had not increased, although within each survey satisfaction increased significantly with income. These outcomes would have been predicted by my theory. For, the incomes have increased by 42% but the *distribution* has probably not changed very much and presumably the perceived distributions have not changed either.

Easterlin's findings also follow from the theory, provided at least that in perceived income distributions incomes in foreign countries play a negligible role. Under these circumstances per capita national income is basically irrelevant for one's satisfaction, but one's position in the income distribution does matter.

Recall, finally, the experimental subjects who were shown jury reports on paintings. The subjects who were shown slides with an increasing number of ones end up having a perceived distribution which is located farther to the right than the subjects who were shown slides with a decreasing number of ones, provided that more recently shown slides have more influence on a perceived distribution than the earlier ones. If we then next show the subjects a slide with 50% ones, this slide looks less favourable compared to the first distribution than compared to the second distribution. This explains the differences in evaluation.

It is not difficult to supplement this kind of qualitative evidence with new examples. I prefer, however, to briefly sketch some quantitative results that were obtained recently (*cf.* Van de Stadt *et al.*, 1985). The results are based on an analysis of the first two waves of the so-called income evaluation panel of the Central Bureau of Statistics, which started in 1980. Evaluations of income are measured for each respondent in the panel by means of Van Praag's income evaluation question. The result of this measurement is known as the respondent's *welfare function of income* (WFI). An individual's WFI describes his evaluations of income levels on a zero-one scale.[19] It follows from my

19 See Van Praag (1968) for its theoretical basis.

theory of preference formation that an individual's measured WFI has to be identical to his perceived income distribution. Using the information from the panel, and with some econometric skill, one can quantify how for each individual in the panel the overall perceived income distribution has changed from 1980 to 1981.

Fig. 5 illustrates some of the results. It shows for an arbitrarily selected individual in the panel the overall perceived income distribution in 1980, the overall perceived income distribution in 1981 and the contemporaneous distribution in 1981. The figure shows that the 1981-contemporaneous distribution has considerably less influence on the overall perceived distribution in 1981 than the overall perceived distribution in 1980. To be a little more precise, according to our estimates the most recent contemporaneous distribution receives a weight of approximately 20% and all preceding contemporaneous distributions combined receive a weight equal to 80%.

Since, according to the theory, the overall perceived distribution and the WFI are identical, one can also read Fig. 5 as follows. An individual's 1981 WFI is a convex combination of his 1980 WFI and the 1981-contemporaneous perceived distribution, where the most recent contemporaneous distribution causes an 'innovation' of the WFI with 20%.

The contemporaneous perceived income distribution consists of two essential elements: one's own income and the incomes of all others in the social reference group. Our research indicates that one's own income gets a weight which is approximately two times higher than all other incomes combined. It can be shown that as a result an increase of one's own income by 10% has the same beneficial effects on the satisfaction level as a *de*crease by 10% of the incomes of all others in the social reference group.

All evidence we have considered so far indicates that the preference formation theory provides a good explanation of differences in the evaluation of income among individuals.

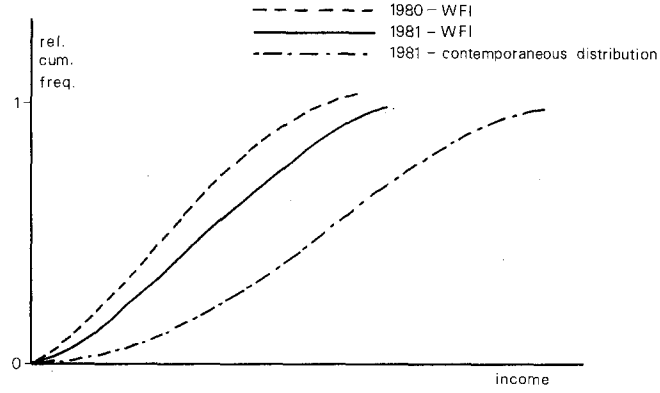

Figure 5 – The 1981-WFI as a convex combination of the 1981-contemporaneous distribution and the 1980-WFI

9 SOME CONSEQUENCES FOR ECONOMICS

Let me return to the theme I started out with. By using a directly measurable cardinal utility concept, Van Praag's WFI, it appears possible to quantify and test a theory of preference formation. Although I have only mentioned investigations with respect to incomes, earlier work has shown that similar analyses can be carried out with respect to somewhat narrower expenditure categories (*cf.* Kapteyn *et al.*, 1980). The aforementioned research with respect to RD and AL theory makes it clear, moroever, that the preference formation theory is part of a quite general mechanism that can also be observed outside economics.

What are the consequences of our results for economics? The answer is 'none,' if we assume that the utility concepts we have used in the empirical tests of the preference formation theory are not related at all to the indirectly measured utility concepts that underlie models of economic behaviour. It has to be admitted that the utility functions that have been measured directly hitherto are usually one-dimensional, or two-dimensional at most.[20] Furthermore, there have been very few attempts to predict behavior on the basis of these directly measured utility functions.[21] Thus, at the very least there is a difference in emphasis between our own research and research into neoclassical models of behaviour. The main reason for this is that we want to learn more about utility functions before their impact on behaviour is investigated. In the last instance, it is a matter of *belief* to claim that models that deal with directly measured utility functions, *e.g.* my preference formation theory, are also valid for the indirectly measured utility functions that underlie models of behaviour. I find it hard to imagine, however, that verbal evaluations and indirectly observed preferences would be entirely unrelated. For example, when I observe that someone's opinion about the income he needs to maintain a decent standard of living depends heavily on the incomes in his social reference group, then I find it quite implausible that his preferences regarding how to spend his income are not related to the consumption pattern in his reference group.

A related example deals with the measurement of the cost of living associated with different family compositions. Both the direct and the indirect approach aim at finding the compensating money amounts (family allowances) which would leave families that differ only with respect to their composition equally well off. If the two approaches would yield different compensating amounts, it would imply that direct and indirect measurement of utility give different results. In that case I would not rest until an explanation for the discrepancy was found. Hitherto, such a discrepancy has not occurred.

20 An example of a directly measured two-dimensional utility function can be found in Dagenais (1977).

21 An exception is Kapteyn *et al.* (1979).

To repeat, it is ultimately a matter of belief, but the belief is not unfounded. So let us repeat the question of what consequences our findings have for economics, now assuming that directly and indirectly measured utility functions are closely related concepts. I will mention a number of them.

In the first place, the preference formation theory makes it clear that there is no such thing as a 'true' functional form of the utility function. Whenever perceived distribution functions change, utility functions change with them. Attempts to find the true form of utility functions, or to approximate the true form by flexible specifications (*cf. e.g.*, Christensen, Jorgenson, Lau, 1975, or Jorgenson and Lau, 1979) become futile. There is no true functional form.

A related, somewhat surprising, consequence is that the attempt to measure utility functions directly threatens the central role of the utility concept in economics. Although utility functions are useful measurable representations of preferences, the more fundamental concept is the perceived distribution function.

From the viewpoint of building an efficient research strategy, it seems obvious to me that demand systems, consumption functions, labour supply functions, *etc.,* are fundamentally misspecified if preference formation is not accounted for. Of course, misspecified models do not contribute greatly to the quality of econometric forecasts.

From a policy viewpoint, it is equally important to take preference formation into account. If, for instance, it is incorrectly assumed that utility functions are constant, a high national income is the remedy for all problems. If we account for the relatively of utility, it becomes clear that, for example, the extent of poverty within a country is not primarily a function of the *level* of national income, but rather of its *distribution.*

For those who have always claimed that incomes have to be redistributed in a more equitable way, the theory provides a warning, however. If the distribution of incomes (or of goods, or of other sources of satisfaction) becomes more equal, perceived distributions become more equal as well. Put differently, although income differences may decrease, the corresponding change in perceived distributions may induce a stronger, rather than a lesser, perception of these differences. The persistence of discussions about income inequality in the Netherlands, during a period in which the income inequality has decreased, appears to confirm this.

The relative nature of my preference formation theory does not imply that economic growth, *i.e.,* an across-the-board rise of incomes, is irrelevant for someone's evaluation of his standard of living. Incomes from the past are part of a perceived income distribution and if the current income exceeds past incomes, the current income ranks highly in the perceived distribution. It is not the level of income that counts, but its growth.

10 CONCLUSION

The central theme of this paper has been that if 'utility' is the central concept of economics, it should be given the attention it deserves. Various examples indicate the importance of doing this. Ignoring preference formation leads to economic models that are misspecified. For policy purposes, the problem of using misspecified models is perhaps only minor compared to the fact that in the formulation of policy goals, preference formation is neglected as well, so that political goals turn out to be systematically unattainable. By means of inadequate tools we try to reach unattainable goals.

The paper is also a plea for interdisciplinary research. Had we watched the activities of our colleagues in the sister social sciences a little better, we might have realized earlier that utility functions can be measured *and* that they change all the time. Economics might have become more useful as a result.

REFERENCES

Bassis, M.S., 'The Campus as a Frog Pond: A Theoretical and Empirical Reassessment,' *American Journal of Sociology,* LXXXII (1977), pp. 1318–1326.

Battalio, R.C., E.B. Fisher, J.H. Kagel, R.L. Bassman, R.C. Winkler and L. Krasner, 'A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases,' *Western Economic Journal,* XI (1973), pp. 411–425.

Battalio, R.C., J.H. Kagel and M.O. Reynolds, 'A Note on the Distribution of Earnings and Output per Hour in an Experimental Economy,' *Economic Journal,* LXXXVIII (1978), pp. 822–829.

Bentham, J., 'An Introduction to the Principles of Morals and Legislation,' Reprinted in: N.A. Page (ed.), *Utility Theory. A Book of Readings,* New York, 1823.

Christensen, L.R., D.W. Jorgenson and L.J. Lau, 'Transcendental Logarithmic Utility Functions,' *American Economic Review,* LXV (1975), pp. 367–383.

Dagenais, D.L., 'Evaluating Public Goods from Individual Welfare Functions,' *European Economic Review,* IX (1977), pp. 123–149.

Darrough, M., R.A. Pollak and T.J. Wales, 'Dynamic and Stochastic Structure: An Analysis of Three Time Series of Household Budget Studies,' *Review of Economics and Statistics,* LXV (1983), pp. 274–281.

Davis, J.A., 'A Formal Interpretation of the Theory of Relative Deprivation,' *Sociometry,* XXII (1959), pp. 280–296.

Davis, J.A., 'The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men,' *American Journal of Sociology,* LXXII (1966), pp. 17–31.

Debreu, G., *Theory of Value,* London, 1959.

Duesenberry, J.S., *Income, Saving and the Theory of Consumer Behavior,* Cambridge, MA, 1949.

Duncan, O.D., 'Does Money Buy Satisfaction?,' *Social Indicators Research,* II (1975), pp. 267–274.

Easterlin, R.A., 'Does Economic Growth Improve the Human Lot? Some Empirical Evidence,' in: P.A. David and M.W. Reder (eds.), *Nations and Households in Economic Growth. Essays in Honor of Moses Abramowitz,* New York, 1974.

Edwards, P. (ed.), *The Encylopaedia of Philosophy,* Vol. 8, London, 1967.
Grether, D.M., 'Recent Psychological Studies of Behavior under Uncertainty,' *American Economic Review,* Papers and Proceedings, LXVIII (1978), pp. 70–74.
Helson, H., *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior,* New York, 1964.
Helson, H., 'Adaptation-Level Theory: 1970 – And After,' in: M.H. Appley (ed.), *Adaptation-Level Theory: A Symposium,* New York, 1971.
Helson, H. and A. Kozaki, 'Anchor Effects using Numerical Estimates of Simple Dot Patterns,' *Perception and Psychophysics,* IV (1968), pp. 163–164.
Houthakker, H.S., 'Revealed Preference and the Utility Function,' *Economica,* XVII (1950), pp. 159–174.
Houthakker, H.S. and L.D. Taylor, *Consumer Demand in the United States: Analysis and Projections,* 2nd. ed., Cambridge, MA, 1970.
Jorgenson, D.W. and L.J. Lau, 'The Integrability of Consumer Demand Functions,' *European Economic Review,* IIX (1979), pp. 115–147.
Kapteyn, A., 'A Theory of Preference Formation,' Ph.D. thesis, Leyden University, 1977.
Kapteyn, A., T.J. Wansbeek and J. Buyze, 'Maximizing or Satisficing?,' *Review of Economics and Statistics,* LXI (1979), pp. 549–563.
Kapteyn, A., T.J. Wansbeek and J. Buyze, 'The Dynamics of Preference Formation,' *Journal of Economic Behavior and Organization,* I (1980), pp. 123–157.
Kapteyn, A. and T.J. Wansbeek, 'Empirical Evidence on Preference Formation,' *Journal of Economic Psychology,* II (1982), pp. 137–149.
Koopmans, T.C., 'Representation of Preference Orderings over Time,' in: C.B. McGuire and R. Radner (eds.), *Decision and Organization, A Volume in Honor of Jacob Marschak,* Amsterdam, 1973.
Lea, S.E.G., 'Animal Experiments in Economic Psychology,' *Journal of Economic Psychology,* I (1981), pp. 245–271.
Manser, M.E., 'Elasticities of Demand for Food: An Analysis Using Non-additive Utility Functions Allowing for Habit Formation,' *Southern Economic Journal,* XLVII (1976), pp. 879–891.
Merton, R.K. and A.S. Kitt, 'Contributions to the Theory of Reference Group Behavior,' in: R.K. Merton and P.F. Lazarsfeld (eds.), *Continuities in Social Research, Studies in the Scope and Method of 'The American Soldier,'* Glencoe-II, 1950.
Muellbauer, J., 'Testing the Barten Model of Household Composition Effects and the Cost of Children,' *Economic Journal,* XXCVII (1977), pp. 460–487.
Muellbauer, J., 'The Estimation of the Prais-Houthakker Model of Equivalence Scales,' *Econometrica,* IIL (1980), pp. 153–176.
Phlips, L., 'A Dynamic Version of the Linear Expenditure Model,' *Review of Economics and Statistics,* LIV (1972), pp. 450–458.
Phlips, L., *Applied Consumption Analysis,* Amsterdam, 1974.
Pollak, R.A., 'Endogenous Tastes in Demand and Welfare Analysis,' *American Economic Review,* Papers and Proceedings, LXVIII (1978), pp. 374–379.
Ross, M. and J. Thibaut, 'Determinants of Standards of Judgment,' *Journal of Personality,* XLII (1974), pp. 383–398.
Runciman, W.G., *Relative Deprivation and Social Justice,* London, 1966.
Samuelson, P.A., 'A Note on the Pure Theory of Consumer's Behavior,' *Economica,* V (1938), pp. 61–71.
Schoemaker, P., 'The Expected Utility Model: Its Variants, Purposes, Evidence and Linitations,' *Journal of Economic Literature,* XX (1982), pp. 529–563.

Stigler, G.J., 'The Development of Utility Theory,' *Journal of Political Economy,* LVIII (1950), pp. 307–327, 373–396. Reprinted in: A.N. Page (ed.), *Utility Theory. A Book of Readings,* New York.

Stigler, G.J. and G.S. Becker, 'De Gustibus non est Disputandum,' *American Economic Review,* LXVII (1977), pp. 76–90.

Stouffer, S.A., E.A. Suchman and L.C. DeVinney, S.A. Star, R.M. Williams Jr., *The American Soldier: Adjustment during Army Life,* Princeton NJ, 1949.

Van Herwaarden, F.G. and A. Kapteyn, 'Empirical Comparison of the Shape of Welfare Functions,' *European Economic Review,* XV (1981), pp. 261–286.

Van der Leeuw, J.L., 'De toekomst in retrospectief. Over de voorspelkwaliteit van de macro-economische verkenning,' *Economisch Statistische Berichten,* LXIX (1984), pp. 160–163.

Van Praag, B.M.S., *Individual Welfare Functions and Consumer Behavior,* Amsterdam, 1968.

Van Praag, B.M.S., 'The Welfare Function of Income in Belgium: An Empirical Investigation,' *European Economic Review,* II (1971), pp. 337–369.

Van de Stadt, H.A. Kapteyn and S.A. van de Geer, 'The Relativity of Utility: Evidence from Panel Data,' *Review of Economics and Statistics,* forthcoming (1985).

Von Neumann, J. and O. Morgenstern, *Theory of Games and Economic Behavior,* 3rd ed., New York, 1944.

Summary

UTILITY AND ECONOMICS

Although 'utility' has been the central concept in economics, economists have paid relatively little attention to its measurement. Generally, utility is measured indirectly via the revealed preference approach. We discuss problems with this approach and next introduce alternative 'direct' measurement methods. The direct measurement methods are seen to spawn a so-called theory of preference formation, which explains differences in utility functions of different individuals. The similarities of this theory with related theories in sociology and psychology, and various sorts of empirical evidence, are reviewed. The paper concludes with a discussion of the implications of these findings for economic theories.