The United Nations University

WIDER

World Institute for Development Economics Research

Working Papers No. 101

December 1992

Utilities, Preferences and Substantive Goods

John C. Harsanyi

UTILITIES, PREFERENCES, AND SUBSTANTIVE GOODS

John C. Harsanyi

University of California in Berkeley

UNU/WIDER James S. McDonnell Distinguished Scholar in 1992

SYNOPSIS

Part I. Two approaches to utility theory: happiness theory and preference theory.

1. Introduction. 1
2. Difficulties with happiness theory. 2
3. Autonomous desires. 4
4. Actual preferences and informed preferences. 5
5. Informed preferences about drug addiction. 7
6. Spurious preferences. 9
7. Satisfaction and informed satisfaction. 9

Part II. The nature of our preferences.

8. Our unreasoned preferences and the psychological learning process. 12
9. Our reasoned preferences and the factual-assumptions model. 15
10. Comparisons with Hume's model of human motivation. 17
11. Ends and means. 18
12. Desires and preferences. 20

Part III. Our well-being and the substantive goods of human life.

13. Substantive goods and our basic desires. 24
14. Explanation and normative evaluation of human preferences. 27
15. Interpersonal utility comparisons. 30

Notes. 34

References. 43

Part I. Two approaches to utility theory: happiness theory and preference theory.

1. Introduction.

We owe the concept of utility to the great utilitarian thinkers of the 18th and 19th centuries, to such people as Hume, Adam Smith, Bentham, John Stuart Mill, Sidgwick, and several others. But they could not make proper use of this concept in economic analysis because they lacked the notion of marginal utility. Jevons (1871) was one of the first authors to make a clear distinction between total utility and marginal utility (he called the latter "final utility") and to state the conditions for economic equilibrium under perfect competition in terms of prices and marginal utilities.

Yet, like his utilitarian predecessors, Jevons was a hedonist, who interpreted utility as a measure of a person's happiness, i.e., of the balance of pleasure over pain in his or her[1] life. Accordingly, he tried to explain economic behavior as maximization of one's happiness or utility. This hedonistic approach to utility theory I shall describe as happiness theory.

In contrast, since Pareto and in particular since Hicks (1939), modern economics has explained people's economic behavior in terms of their preferences, considering their utility function merely as a convenient mathematical representation of their preferences. It is still assumed that rational behavior involves maximization of one's utility function. But since the latter is itself defined in terms of one's preferences, it is these preferences rather than utility maximization as such that explain people's behavior. Indeed, strictly speaking, modern economics could be restated, admittedly at the cost of some inconvenience, in terms of people's preferences alone, without any reference to utility to all. But most economists do retain the utility concept as a representation of people's preferences. This

approach to utility theory I shall describe as preference theory.

Economists have been attracted to this approach for two main reasons. One is that it enabled them to dispense with the outdated hedonistic psychology underlying most forms of happiness theory (see below). The other is that, whereas happiness is a psychological concept based on introspection, people's preferences are revealed by their publicly observable choice behavior -- at least if we assume that their actual choices do indicate their real preferences, which, as I shall argue, is not always the case.

2. Difficulties with happiness theory.

Yet, the basic argument against happiness theory is this. What we basically want are not only some desirable subjective experiences, i.e., some desirable states of our own mind, such as happiness, pleasure, and an absence of pain. We also want some desirable objective outcomes in the outside world, such as having money, social status, success, worthwhile accomplishments, as well as other people's love and respect. Moreover, we want not only our own happiness but also that of many others we care about.

In other words, some of the objective outcomes we want amount actually to other people having particular states of mind, such as being happy, or feeling love and respect for us.

Moreover, what we really want is actual achievement of these outcomes, rather than merely the subjective belief of having achieved them when this may not be the case. We do not want to live in a fool's paradise. Rather, we want the real thing: we want to have money, success, love and respect, in actual fact rather than the mere illusion of having

them -- even if this illusion were subjectively indistinguishable from the real thing.

By the same token, we do want happiness. But we want happiness based on the actual facts rather than one based on false illusions. This is inconsistent with happiness theory in its usual subjectivist form. Of course, it is clearly in our interest to take this attitude. It is very dangerous to base one's life on false assumptions because this will lead to making important practical decisions on such assumptions.

Yet, perhaps the most important objection to happiness theory in its usual form is that it cannot admit the possibility of truly altruistic actions. It is committed to the view that whatever we do we always do it for the sake of our own happiness. Yet, this would mean that even our ostensibly altruistic actions meant to benefit other people were done by us primarily to promote our own happiness and our own interests rather than to promote these people's happiness and their interests -- which would mean that they were not truly altruistic actions. But this would be surely an absurd view.

For instance, when parents do various things for their children and perhaps even make great sacrifices for them, they presumably do this mainly for the sake of their children's happiness rather than for the sake of their own. They may feel happy to do these things for their children but their own happiness will be only a secondary consideration. Their primary aim will be to help their children rather than to derive some extra pleasure or some extra happiness from helping them.

Finally, happiness theory cannot explain why many people want not only a happy life but one with some worthwhile accomplishments.

By nature and by choice we are social creatures rather than solitary ones. Perhaps

this is the deeper reason why we have genuine altruistic concern for other people; why we want to make worthwhile contributions to some objectives we share with many other people in intellectual, artistic, social and political life; and why our interests reach out far beyond our own inner experiences and very much extend to states of affairs in the outside world. (As I have already suggested, to pay close attention to the outside world is also clearly in our self-interest.)

3. Autonomous desires.

Another way we can describe the basic difference between hedonistic theories like happiness theory and nonhedonistic theories like preference theory is this. Some of our desires are no doubt based on hedonistic considerations: we may desire something because we expect to derive pleasure from it or because it would be painful for us to be deprived of it. Hedonistic theories claim that *all* our desires are based on such hedonistic considerations. In contrast, nonhedonistic theories maintain that we also have *autonomous* desires *not* based on such considerations.

I have already argued that our *altruistic* desires and our desire for a life with worthwhile *accomplishments* are such autonomous desires.

Moreover, it often happens that our *pleasures* are determined by our *desires*, rather than the other way around as hedonistic theories would suggest. For instance, when we are pleased with a victory of our favorite sports team, we do so because we had a *prior desire* for this team to win, rather than because this team's victory had an intrinsic pleasure-producing quality independent of our prior desire for their victory. Thus, our

desire for them to win was an <u>autonomous</u> desire <u>not</u> based on independent hedonistic considerations.

Finally, let me adduce two examples, involving two of our basic instinctive drives, hunger and curiosity. No doubt, we sometimes eat because we expect that our eating will be a pleasant experience, or because we find our strong hunger unpleasant and want to get rid of it by consuming some food. But on other occasions we eat simply because we are hungry and have a <u>natural desire</u> to eat when we are hungry, even though our hunger is not really unpleasant, and even though we have given little thought to the pleasure we may derive from eating.

Likewise, sometimes we try to find out some information because we think it will be pleasant to have it, or that it will be unpleasant not to satisfy our curiosity. But we may also try to obtain some information simply because we have a <u>natural desire</u> to satisfy our curiosity, regardless of the pleasure we may feel by doing so, and of the displeasure we may feel by not doing so.

Thus, contrary to hedonistic theories, we do have <u>autonomous</u> desires <u>not</u> based on hedonistic considerations.

<u>4. Actual preferences and informed preferences</u>.

In most branches of economics, a person's utility function is usually defined in terms of <u>all</u> his preferences (assuming that the latter satisfy the usual regularity axioms). But in welfare economics and in ethics we want to distinguish between those choices of a person that really express his true preferences and his true interests at a deeper level, from those

choices of his that fail to do so because they are based on incorrect information or on ignorance or neglect of some important information.

Accordingly, I shall distinguish between a person's actual preferences as indicated by his choice behavior and by his verbal statements, and his informed preferences, defined as the hypothetical preferences he would have had he all relevant information and had he made full use of this information.

It is convenient to extend the term informed preferences also to those actual preferences of a given person that agree with his hypothetical informed preferences as just defined, and to use the term mistaken preferences to describe those actual preferences of his that disagree with his informed preferences.

We can define a person's informed desires and his mistaken desires in a similar way.

A person's informed preferences and desires are defined as what his preferences and desires would be like under some hypothetical conditions. This of course means that they are not directly observable empirical variables as his actual preferences and desires are but rather are theoretical constructs.

Suppose that Ron has a bad case of pneumonia, for which A is the best medication. But he erroneously thinks that B is the best medication, even though the latter would be quite ineffective against the type of pneumonia he has. As a result, in terms of his actual preferences, he would prefer to be given B rather than A. Yet, under our definitions, we must say that, in terms of his informed preferences, he would prefer A over B, because this would be his preference if he knew the relevant medical facts.[2]

It seems to me that, at least in welfare economics and in ethics, a person's utility

function should be defined in terms of his hypothetical informed preferences rather than in terms of his actual preferences because some of the latter may be badly mistaken (cf. Harsanyi, 1958, 1977a, and 1977b). Some other authors have made basically similar proposals. In contrast to the actual-preference theory that most economists use in defining people's utility functions, I shall describe this approach as informed-preference theory.[3]

This approach enables us to define a person's utility function and his true interests in terms of his own preferences.[4] Yet, it permits us to bypass his possibly mistaken actual preferences which are contrary to his own interests, in favor of his own, more fundamental, informed preferences, which can be considered as his actual preferences as freed from the distorting effects of his factual errors.[5] [6]

5. Informed preferences about drug addiction

A good test case for our theory is that of drug addiction. A drug addict has strong reasons for trying to break his drug habit because he knows that it greatly reduces the control he has over his own life, and may sooner or later ruin his health, his ability to work, and his family life. Yet, he also has strong reasons for maintaining his drug habit because he may strongly enjoy his drug-induced euphoria, and because he knows that the withdrawal process would be very unpleasant for him.

The problem takes a special form in the case of terminally ill people suffering from a very painful disease (such as cancer), who may get addicted to some pain-killing drugs. For them the case for giving up their pain-killing drugs will be much weaker and the case for their continued use will be much stronger than it would be for an ordinary drug addict.

Yet, in either case, according to informed-preference theory, the only way we can decide what the drug user's true interests are is to find out what his own well-considered rational judgment is in this matter. For instance, a drug addict may decide after careful reflection that he would be much better off if he could break his drug habit, and that this is what he should actually do. Yet, even after making this decision, he may find that, when he feels an urge to take his drug, he seldom has the will power to resist this urge. (Following Aristotle, the philosophic literature does describe this problem as "weakness of the will.") Yet, at least if he seems to have a serious desire to break his drug habit (as shown, e.g., by his willingness to seek medical help in this matter) then we must accept the fact, it seems to me, that this is his real preference. Moreover, on the assumption that he has decided to break his habit after careful consideration of the relevant facts, we must also accept it as his informed preference.

Of course, we must arrive at the opposite conclusion if he decides that he will be better off by not breaking his drug habit -- at least if he appears to have made his decision after careful consideration of the relevant facts.

Likewise, in the case of a patient addicted, or in the danger of getting addicted, to his pain killers, we have to accept his own well-considered rational judgment about what policy would be in his best interest. Presumably, most patients will desire to continue using their pain killers. But as Griffin reports (op. cit., p. 8), Freud refused any strong pain killer during his last and very painful illness, saying, "I prefer to think in torment than not to be able to think clearly."[7]

6. Spurious preferences.

I have suggested that our preferences qualify as informed preferences if they are based on careful consideration of the relevant facts. Yet, it seems to me that we have to add another, very natural, requirement: It is that they should be our genuine preferences rather than spurious ones.

Let me discuss only two classes of spurious preferences. One class consists of the ostensible preferences of people suffering from mental illness or from obsessive-compulsive neurosis, to engage in various forms of compulsive behavior. For example, some neurotics wash their hands astonishingly many times a day for no obvious reason. Their behavior may be to some extent voluntary and to this extent a result of their own preference to act in this way. But at a deeper level it is obviously a result of a more or less irresistible inner compulsion, very much contrary to their true preferences.[8]

Another class of spurious preferences are those based on self-deception. Some people pretend to have, and may eventually even convince themselves to have, some preferences they think to be fashionable or sophisticated to entertain -- even though their real preferences are quite different, or even though they do not themselves know in the end what their real preferences are.

In my view, spurious preferences of either class should not be regarded as real preferences at all, and a fortiori should not be regarded as informed preferences.

7. Satisfaction and informed satisfaction.

Apart from specific pleasant and unpleasant experiences, each person also has a

general feeling of satisfaction or dissatisfaction with his life as a whole. (For our purposes, dissatisfaction can be treated as a rather low level of satisfaction.) It expresses the degree to which his subjective experiences and the objective conditions of his life, as they are now and as he expects them to be in the future, given his own likely behavior in various contingencies, satisfy his own actual preferences at present. Obviously, a person's satisfaction level is an empirical variable, in principle accessible to introspection.

One possible approach to utility theory would be to identify a person's utility level with his satisfaction level. This approach might be described as a satisfaction theory of utility. As is easy to see, it would be exactly equivalent to what we have called the actual-preference theory of utility because both theories would amount to identifying a person's utility level with the degree to which his actual preferences are satisfied.

On the other hand, Brandt (1982, p. 181) has convincingly argued that happiness theory is equivalent to (a particular version of) actual-preference theory. This means that, depending on the specific details of their interpretation, all three theories are equivalent or are nearly equivalent.

Yet, none of these theories is equivalent to our informed-preference theory. But now I propose to show that a modified version of satisfaction theory, based on people's informed preferences rather than on their actual preferences, is equivalent to our theory.

Accordingly, let me define a person's informed-satisfaction level as the degree to which his subjective experiences and the objective conditions of his life, as they are now and as they can be reasonably expected to be at various future times, given his own likely behavior at various contingencies, satisfy his present informed preferences. To put it

differently, a person's informed-satisfaction level is the satisfaction level he would experience if

(1) his actual preferences coincided with his informed preferences and if

(2) he were well-informed about all relevant facts as they are at present and as they can be reasonably expected to be at various future times.

Since a person's informed preferences are not empirically observable variables but rather are theoretical constructs, the same is true for the concept of informed-satisfaction level. That is to say, unlike a person's satisfaction level, his informed-satisfaction level is not accessible to direct introspection. But as I am going to argue, it is a theoretical construct of considerable analytical usefulness.

Yet, before doing so, let me discuss some obvious questions raised by my definition of informed-satisfaction level. First of all, how should we define reasonable expectations about the future? Clearly, nobody can really predict the future. All that anybody can do is to consider a number of alternative possible scenarios for the future and then assign what he considers to be the most appropriate numerical probabilities to all of them. Thus, the notion of reasonable expectations about the future must be interpreted, it seems to me, as a list of carefully selected likely alternative scenarios, together with carefully selected reasonable probabilities for them, both these scenarios and the probabilities assigned to them being selected in the light of the best information available to us.[9]

Another question is, How much weight will a person's informed preferences assign to his pleasures and displeasures based on incorrect information? In my view, this must be decided by his own informed preferences. Presumably, they will assign less weight to

them than to his pleasures and displeasures based on correct information. But the mere fact that he will be mistaken about some important facts will have to be regarded as something undesirable (though it may become something desirable under some special conditions).

A possible approach to utility theory would be to identify a person's utility level with his informed-satisfaction level. We may call it the informed-satisfaction theory of utility. As is easy to verify, this approach is exactly equivalent to our informed-preference theory of utility because both theories amount to defining a person's utility level as the degree to which his informed preferences are satisfied.

As we have seen, satisfaction theory and informed-satisfaction theory are equivalent to actual-preference theory and to informed-preference theory, respectively. But we do gain something by restating these two preference theories as a satisfaction theory or as an informed-satisfaction theory. For when restated in this way, both preference theories will avoid the difficulties pointed out by Brandt (1979, pp. 247-253), which are posed by the changeability of our desires and preferences over time. This is so because both satisfaction theory and informed-satisfaction theory make use only of the relevant person's present actual or informed preferences without making use of his past or future preferences in any way.[10]

Part II. The nature of our preferences.

8. Our unreasoned preferences and the psychological learning process

Some of our choices and preferences are based on more or less careful rational

deliberation. These I shall call reasoned choices and preferences. Others are based on little or no deliberation. These I shall call unreasoned choices and preferences.

Psychologists tell us that our unreasoned choices and preferences are governed by our innate and acquired psychological drives.[11] Our drives are acquired, extinguished, strengthened, and weakened by a largely unconscious learning process (conditioning process) based on the Law of Effect: When our behavior produces satisfying effects then our drive to act in this way in situations like this tends to grow stronger; whereas if it produces frustrating effects then our drive to act in this way tends to grow weaker. In most cases, both of these tendencies are to our benefit because they make us repeat successful behavior and avoid unsuccessful behavior.

We also benefit (though are sometimes also misled) by our natural inclination to follow other people's example and advice.

Even though this learning process does give us beneficial drives most of the time, it may occasionally saddle us with drives of very questionable utility as a result of one particular aspect of this learning process, called stimulus generalization. The latter may make us largely reproduce our original response not only in very similar cases, but also in situations of a less similar type, or even in ones that merely remind us in some way of the original situation.

For example, we have good reasons to be grateful to a good teacher and to be resentful of a bad teacher. But it obviously makes no sense if we extend this positive or negative attitude to another person we meet many years later, simply because his physical appearance or his accent reminds us of our former teacher. Yet, owing to stimulus

generalization, we may inadvertently do such things, perhaps never realizing how we have come to adopt such an unwarranted attitude toward this person.

Yet, even apart from unwarranted stimulus generalization, this learning process will seldom lead us to choices in agreement with what our informed preferences would require. To be sure, it will make us repeat choices we have found satisfying in the past. But typically these will be choices we have found satisfying after a very limited sampling of the available alternatives, and in ignorance of a good deal of other relevant information.

For instance, a habitual coffee drinker will rightly feel that he has good reasons to drink coffee every morning because he has found repeatedly, perhaps over many years, that coffee drinking is a satisfying experience for him. But without comparing the taste of coffee with that of many other possible morning drinks, and without informing himself about many other relevant facts (such as the costs, the availability, and the health effects of various drinks), he obviously cannot conclude that, in terms of his informed preferences, coffee would be in fact the best possible morning drink for him.

On the other hand, in many fields of human choice, including that of foods and drinks, it usually does not matter very much whether our actual choices are really the best possible choices we could make. Should our coffee drinker discover later that he really prefers the taste of another morning drink, it is unlikely that this discovery will significantly increase his utility level.

What is more, in fields where our actual choices usually make little real difference, we actually derive important benefits from the fact that our learning process tends to provide us with clear preferences, even if they may not be the theoretically optimal ones.

For example, by having definite food and drink preferences, we avoid spending a lot of time and effort in deciding what to eat and what to drink. Of course, when we have to make really important decisions then it will be worth our while to invest some time and effort. (Thus, people very fond of good food and of good drinks do find it worth their while to do a careful sampling of the alternatives and to obtain gastronomic information in other ways.)

To sum up, our unreasoned choices and preferences are governed by our drives, native and acquired, shaped by a largely unconscious learning process based on our personal experiences and on various social influences. By and large these work to our benefit but in general do not produce choices and preferences in full agreement with what our informed preferences would be.

9. Our reasoned preferences and the factual-assumptions model.

I now propose to consider the logical relationship between our reasoned preferences and the factual assumptions underlying them. When we form a reasoned preference between two possible outcomes A and B, we always do this on some factual assumptions about the intrinsically desirable and the instrumentally desirable attributes that each of these two outcomes has. Our preference between A and B will be an informed preference only if these factual assumptions agree with the actual facts (as well as cover all important facts).

I call any attribute intrinsically desirable if it makes an object possessing it into a potentially suitable end of our activities; and call an attribute instrumentally desirable if

it makes an object possessing it into an effective _means_ to achieve some of our possible ends.

Sometimes our preference for some outcome A over another outcome B will be based on a _combination_ of intrinsic and instrumental desirability considerations. For example, we may prefer to attend concert A rather than concert B _both_ on the assumption that the former will have superior intrinsically desirable musical qualities _and_ also on the assumption that it will have greater instrumental value in enhancing our musical education, and perhaps also in providing an opportunity to meet our friends.

Of course, _what_ we consider to be desirable attributes and _how desirable_, i.e., _how important_, we consider any given attribute will depend on our own _attitudes_, presumably governed by our innate and acquired psychological drives, which also govern our _unreasoned_ preferences, as I argued in section 8. As a result, our attitudes are determined by our native temperament, by our personal life experiences, and by various social influences.

On the other hand, the extent to which A and B actually possess any one of these attributes is a _factual question_.

Yet, when we have answered all these desirability questions and all these factual questions to our satisfaction, we still have to decide, all things considered, whether to prefer A over B, or B over A, or to be indifferent between the two. Typically this will be a _nontrivial_ decision because we shall have reasons to prefer A but shall also have reasons to prefer B, and our decision will depend on which set of reasons we find _more important_ in the light of all our desirability judgments and of all our factual assumptions.

Of course, in real life we do not form even our reasoned preferences by full examination of all relevant desirable attributes but rather restrict our attention to those attributes we consider particularly important.

I shall call this model the factual-assumptions model for reasoned preferences.

10. Comparisons with Hume's model of human motivation.

Our model may be regarded as a modified version of the Humean model of human motivation. Like the latter, it assumes that, intrinsically, our reason is an "inert" faculty for discovering facts and for inventing concepts to make these facts understandable to us. But, unlike Hume, I assume that our "passions," i.e., our desires and preferences, themselves crucially depend on the factual assumptions suggested by our reason. It is not the "activist" nature of our reason itself that establishes this relationship between our reason on the one hand and our desires and preferences on the other hand but it is rather the dependence of the latter two on such factual assumptions.

Yet, when all factual assumptions are given, it is our psychological drives rather than our reason that determine the direction of our desires and preferences.

Moreover, whereas Hume wants to restrict our reason to suggesting means to our ends, themselves determined by our "passions," my model assumes that our reason not only suggests means to our ends, on the basis of the instrumentally desirable attributes of various alternatives, but also suggests ends worth pursuing, on the basis of the intrinsically desirable attributes of these alternatives.

11. Ends and means.

Clearly, conceptually, our intrinsic values have logical priority over our instrumental values because it is our ends that determine the means we should use. But it is equally true that the ends we can actually attain depend on the means at our disposal. This explains the important empirical fact that our actual preference between two alternatives is often decided by instrumental and, in particular, by economic, considerations.

For instance, even if we feel sure that concert A will greatly surpass concert B in its intrinsically desirable musical qualities, we may very well decide to attend concert B nevertheless, because the price of admission to concert A is more than we can afford, or because getting to it would take us too much time. Many idealistic social experiments and many noble personal projects have come to naught because they failed the satisfy the economic and the other instrumental requirements of viability.

Let me add that, from an empirical point of view, the very distinction between ends and means is much less clear-cut than it might seem to be in the abstract -- mainly because we often develop intrinsic liking for some of our activities that originally had merely an instrumental value for us. I shall call this process transmutation (which was the term used by medieval alchemists to describe the imaginary magical process supposedly converting "base metals" like iron into "noble metals" like gold).

For instance, an engineering student may at first study mathematics merely as an instrumentally useful skill in engineering computations but, as he gets better acquainted with mathematics, might develop a genuine intellectual interest in it. This example involves what may be called unintended transmutation because this engineering student

presumably did not originally intend to develop such an intrinsic intellectual interest in mathematics.

Yet, there are also cases of intended transmutation. For example, suppose a very lonely person without any friends wakes up to the fact that his life might be much richer and much more rewarding if he had a few intimate friends. Therefore, he makes an effort to acquire friends, and let me assume that he succeeds in doing so.

At first his attempt to make friends may be motivated purely by their instrumental value in terms of his own self-interest. But he may realize that his own self-interest will be best served by his developing real friendships with his new friends, and by changing is original narrowly self-centered attitude to one of real appreciation of their positive qualities and of their own friendly behavior toward him, and by developing a genuine concern for their well-being. Thus, he may make an intentional effort to convert his original purely instrumental attitude into one based on the intrinsic value of his new friends and of their friendship.

Such a transmutation of our purely instrumental activities into ones valued for their own sake can make our lives more meaningful and more satisfying. But it can also give some of our instrumental activities some semblance of great intrinsic value they do not really deserve. For instance, it may induce some people to make money making, status seeking, or keeping fit, into the central activities of their lives, pushing aside other activities with much greater intrinsic value -- contrary to their own informed preferences and to their own real interests at a deeper level.

12. Desires and preferences.

Our behavior expresses both our desires[12] and our preferences. Both of these play an important role in our motivation system. Of course, desire is a simpler and a more fundamental notion than preference is. For a desire for A expresses a pro-attitude toward one object, A. In contrast, a preference for A over B expresses a ranking of our pro-attitudes toward two different objects A and B (or the fact that we do have a pro-attitude toward A but have none toward B) and that we give our pro-attitude toward A higher priority than we give to our pro-attitude toward B.

Yet, this very fact shows that our preferences provide information about our priorities -- which our desires as such, as I shall argue, fail to provide.

We can define our desires in terms of our preferences. Thus, the statement

(1) "Jack desires A"

can be defined as being equivalent to the statement

(2) "Jack prefers to have A rather than not to have A."

On the face of it, we can also define our preferences in terms of our desires. Thus, we may try to define the statement

(3) "Jack prefers A to B"

as being equivalent to the statement

(4) "Jack's desire for A is stronger than his desire for B."

Yet, this approach will not yield a noncircular definition because a statement that one of our desires is stronger than another is not an unambiguous statement. For instance, statement (4) has at least three different possible meanings:

(a) It can mean that Jack's desire for A has a greater felt intensity than his desire for B has.

(b) Or, it can mean that if he had to make a choice then Jack would prefer that his desire for A rather than his desire for B be satisfied.

(c) Finally, it can mean that his desire for A is more effective than his desire for B is -- in the sense that, when he has to choose, he will actually try to obtain A rather than B.

For instance, suppose that, after a good deal of soul searching, Jack, who is a heavy smoker, decides that the pleasure he derives from smoking is not worth the risks it poses for his health in the long run, and that he should quit smoking. Yet, even after this decision, when he feels the urge to light up, he is seldom able to resist this urge.

Clearly, Jack does have both a desire to quit as shown by his decision to do so, and a desire to go on smoking as shown by his actual behavior. According to my story, his second desire is more effective than his first because he actually persists in his smoking habit for the time being. Perhaps the felt intensity of his second desire is likewise greater than that of his first desire. Yet, he claims that he would very much prefer to quit smoking and let us assume he has shown that this is his genuine preference by seeking medical help for achieving this objective.

Let A denote giving up smoking and B denote continued smoking. Then statement (4) will be true in sense (b) but will be false in senses (a) and (c). Now, taken in sense (b) it will be in fact equivalent to statement (3). Yet, the only way we can conclude that statement (4) is true in sense (b) is by inferring this fact from statement (3) itself. Thus,

it would be circular reasoning to try to use statement (4) to infer the truth of statement (3).

To put it differently, we cannot infer a person's preferences from his desires because the former provide information about his priorities, which his desires in themselves fail to do. (Of course, we could define a person's desires so that they would include full information about his priorities. But this would simply amount to incorporating his preferences into his desires and to obscuring the very useful distinction between the two for no good reason.)

On the other hand, as shown in most textbooks on economic theory, a person's utility function can always be defined in terms of his preferences (as long as the latter satisfy the usual regularity axioms). On the other hand, we cannot define it without knowing his preferences. For we cannot infer that he would derive more utility from some situation A than from another situation B without knowing that he would prefer to be in situation A rather than in situation B.

Yet, we have just seen that we cannot infer a person's preferences from his desires as such. This of course means that his utility function cannot be defined in terms of his desires without knowing what his preferences are.

Apparently, this fact is often overlooked by a number of distinguished philosophers, who write as if it were possible to base utility theory on people's desires as such. This is indicated also by their description of various approaches to utility theory as "desire theory," "desire account," "desire-fulfillment theory," etc., when "preference theory," "preference account," or "preference-satisfaction theory" would have been more appropriate descriptions

(cf. Brandt, 1979; Griffin, 1986; Parfit, 1984, p. 494).

More generally, both in utility theory and in the theory of human motivation, philosophers tend to concentrate much more on our desires than on our preferences whereas economists tend to do the opposite. Yet, both theories would benefit from a more balanced approach.

For, as I have already argued, a theory restricted to our desires cannot yield a well-defined concept of utility at all. On the other hand, a theory restricted to our preferences is bound to overlook important empirical regularities concerning our desires if there are no similar regularities concerning our preferences.

In particular, it is bound to overlook the important empirical fact that human beings seem to have much the same basic desires -- even though with considerable interpersonal variation in intensity -- such as a desire for material comfort and for physical and economic security; for freedom to control their own lives; for having good health; for jobs suitable to their personal abilities and personal interests; for further developing their abilities; for deep personal relations in mutual love, in marriage and in true friendship; for having children and for being a good parent; for knowledge and for understanding the world and their own place in the world; for enjoyment of beauty in nature and in art; for having access to the ordinary pleasures of human life; for worthwhile accomplishments of some kind; and for making their own behavior consistent with their basic moral values. (This is an expanded version of Parfit's list of "good things" for human beings in Parfit, 1984, p. 499.)

Yet, even if people's basic desires are much the same, their preferences, both their

actual and their informed preferences, are often very different. Their preferences may be very different between alternative ways of satisfying the <u>same</u> basic desire. For instance, they may prefer very different jobs. Moreover, they may have very different priorities in satisfying <u>different</u> basic desires. Thus, they may have different priorities in dividing their time between their job and their family.

Part III. Our well-being and the substantive goods of human life

13. Substantive goods and our basic desires.

As we have seen, the 18th and 19th century utilitarians identified a person's utility level with his <u>happiness</u>, i.e., with the balance of pleasure over pain in his life. Most economists now define it in terms of this person's <u>actual preferences</u>. But together with some other economists and some philosophers, I have proposed to define it in terms of this person's <u>informed preferences</u>.

Yet, a distinguished philosopher, Tim Scanlon, has recently argued that all these approaches are unduly <u>subjectivistic</u>, and that a person's well-being (he largely avoids the term "utility") is determined by the <u>substantive goods</u> available to him (Scanlon, 1989, and 1991). Moreover, he defines a "substantive good" as anything that makes people's lives <u>go better</u>. As examples for such substantive goods he refers to Parfit's list of "good things" for human beings, which in section 12 I described as the objects of our <u>basic desires</u>.

Scanlon insists that, in ultimate analysis, what makes these things substantive goods is <u>not</u> the fact that human beings desire them. Rather, the <u>converse</u> is true: human beings desire them because they understand that these things are <u>good</u> for them.

I strongly disagree with this claim. In my own view, if I desire some object A then, in ultimate analysis, this <u>cannot be explained</u> solely by facts concerning A itself without invoking my own <u>pro-attitude</u>, or my own <u>disposition</u> to have such a pro-attitude, toward some of the facts concerning A.

I agree with Scanlon that -- apart from our impulsive desires, which are governed by our inborn and acquired psychological drives rather than by any conscious rational criteria -- our desire for any object A <u>is</u> in fact based on the assumption that it will be <u>good</u> for us to have A. But how do we know that it will be a good thing for us to have A? For instance, how do we know that it is a good thing for us to have <u>friends</u>? The obvious answer is, it seems to me, that we know it from <u>our own experience</u>, or know it from reports about <u>other people's experiences</u> that friends may help us to satisfy our psychological need for good fellowship and for emotional support as well as our desire for having fun and a good time. They also may give us practical help in various ways. Moreover, they may enrich our lives by helping us to pursue some of our old interests and to acquire some worthwhile new interests. In short, they may help us to satisfy some of our important psychological needs, desires, preferences, and interests.

Scanlon objects to what he calls <u>subjectivism</u> in the theory of human values. As I stated in section 2, I share his opposition to any subjectivist or "experientialist" theory of <u>human goals</u>, i.e., to the view that the only goal of intrinsic value to human beings is <u>their own happiness</u> or, more generally, to have <u>subjective experiences</u> of certain kinds. As I argued in section 2, this view is refuted by our altruistic concern for other human beings, and by our desire for worthwhile accomplishments of some objective importance. My own

theory of utility is not an "experimentialist" theory. Rather, it is an informed-preference theory, interpreting a person's utility level as the degree to which both his subjective experiences and his objective conditions satisfy his informed preferences, i.e., his preferences as freed from the distorting effects of factual errors.

On the other hand, as I have just explained, I do support what Scanlon would regard as a subjectivist theory of the good in that I take the view that what is and what is not good for us is ultimately decided by our own needs, desires, preferences, and interests.

Thus, friendship is an important substantive good for human beings. But it would not be that and would not have any intrinsic value for solitary creatures having no desire for social contacts and unable to enjoy each other's company -- even if they did understand the instrumental value of having access to individuals likely to help them when they needed help.

By the same token, intellectual activities in science, in history, and in philosophy are an important substantive good for many human beings. But it would not be that and would not have any intrinsic value for creatures devoid of any intellectual curiosity and disliking any disciplined intellectual effort -- even if they did understand the instrumental value of some intellectual work, such as that of scientific research when it yields economically valuable discoveries and inventions.

Even for human beings, the value of any particular friendship and of any particular intellectual activity will strongly depend on their own personal interests. Even the friendship of a very fine human being will be of little value to us if we have absolutely no

common interests with him or her; and an opportunity to work on some scientific, historical, and philosophical problems will have little attraction for us if we have no real interest in these problems.

Scanlon (1989, pp. 11-12) admits that his theory does not explain what it is that makes all his "substantive goods" intrinsically valuable to human beings. But the reason is that -- on the basis of some ill-defined anti-subjectivist ideology -- he has arbitrarily excluded any explanation based on human desires and preferences. For in fact there is a rather obvious explanation: What makes all these things intrinsically valuable to us is the fact that they are the objects of our basic desires, which we largely share with other human beings, due to our common human nature and to our common biological and psychological needs.

14. Explanation and normative evaluation of human preferences.

In view of the fact that people's basic desires are remarkably similar, they can be largely explained in terms of our common human nature. In contrast, people's preferences show a good deal of individual variation, which can be explained only by individual differences in their information and in their personal attitudes. Indeed, any variation in their informed preferences must be explained by differences in their personal attitudes alone because, by definition, their informed preferences are the preferences they would have if they had correct information about all the relevant facts, and correct information about anything is of course the same for everyone. This means that differences in people's informed preferences must be due solely to differences in their innate psychological drives

(which are probably rather small) and to differences in their personal life experiences and in the social influences they are exposed to (which might be sometimes quite large).

Apart from trying to understand a person's choices and preferences, we may also try to evaluate them in terms of some normative standards. I propose to argue that there are only two relevant normative standards we can use for this purpose. One is this person's own real interests as defined by his own informed preferences. The other is other people's morally protected interests as defined by the norms of morality.[13]

To be sure, philosophers who oppose defining human values in terms of people's own actual and informed preferences may try to use some normative standards independent of human attitudes, such as standards based on perfectionist or other similar criteria. But I shall try to show that such criteria would not yield viable normative standards.

For example, suppose that Peter is a gifted young man whose abilities would probably enable him either to become a good mathematician or to become a good historian. From a moral point of view, he is free to choose either profession. He is tempted to become a mathematician because he thinks that mathematical research would give him the best opportunity to make use of his outstanding abstract reasoning ability. But he is also tempted to become an historian because he is more attracted to a study of human behavior in past ages than to a study of abstract mathematical concepts.

Obviously he will have to ask himself such questions as, "Do I really have the ability to become a good mathematician or a good historian?" "What are the chances that my present interest in either subject will persist over time?" "What are my chances of obtaining a good job in either field?"

Suppose that in the end he chooses a career in history. On what grounds could an outside observer say that Peter has made the wrong choice? Presumably he has chosen history because he thinks that a career in it will give him a more satisfying life, i.e., a life better satisfying his deepest preferences, than a career in mathematics would. Yet, some observers may feel, and perhaps may rightly feel, that Peter has made the wrong choice, i.e., a choice contrary to his own informed preferences, because he has misjudged his own abilities, or has misjudged the likely persistence of his own interest in historical problems, or has misjudged the chances of his obtaining a good job as an historian, or has misjudged some other relevant facts.

Yet, it would be quite inappropriate for any observer to argue that Peter has made the wrong choice because mathematical work is intrinsically more valuable than historical work is by some perfectionist or other similar criteria -- even though Peter does not have the slightest desire to make his career choice depend on these criteria. Indeed, if Peter has no prior desire to follow these criteria in his choice behavior, it is hard to see how anybody could produce rational arguments to show that he nevertheless should have conformed to these criteria.[14]

To be sure, whereas people's actual preferences can be ascertained by observing their choice behavior and by critically assessing their verbal statements, it is much harder to decide what their informed preferences are -- because it would require us to infer what their preferences would be if they did know some pieces of information they actually do not know.

Yet, we can find out another person's informed preferences in two ways: by

analyzing his own actual preferences, and by considering the preferences of other knowledgeable people. For instance, suppose we know that, in terms of this person's own actual preferences, alternative A has important advantages and no significant disadvantages as compared with another alternative B, yet this person is unaware of this fact. Then it will be a fairly safe bet that, if he did know the relevant facts then he would actually prefer A to B. (Accordingly, we can usually safely assume that, if he had no obvious reasons to the contrary, any patient would prefer the most effective medication if he knew that it was the most effective medication.)

Or, suppose that most knowledgeable people assign a high utility to some benefit A because they know from personal experience or from what they have learned about other people's experiences that A tends to add a lot of extra satisfaction to one's life. Yet, a particular individual seems to have no interest at all in obtaining A. Then, it will be a reasonable assumption that if he were better informed then he would likewise assign a high utility to A, more or less within the same range as other people do -- except if he has some special disability preventing him from taking full advantage of this benefit A. (For example, some people may have special psychological difficulties in making friends and in retaining them over any length of time. Others may be unable to enjoy some of the greatest works of world literature for lack of education, and so on.)

15. Interpersonal utility comparisons.

In earlier publications (Harsanyi, 1955, 1977a (Chapter 4), 1992), I have argued that we are often quite able and quite willing to make interpersonal comparisons of utility.

and that in any case we cannot really avoid making them if we want to make our moral decisions in a responsible manner. I have also suggested that making an interpersonal comparison between the utility levels of two individuals i and j amounts to asking the question of how much satisfaction[15] each of them derives from his own objective position, given his own preferences and, more generally, given his own personal attitudes.

Let A_i and A_j denote the objective positions if i and j, respectively, and let P_i and P_j denote their personal attitudes. Let me call the pairs (A_i, P_i) and (A_j, P_j) the two individuals' extended alternatives. (They are "extended" in the sense that they not only include their objective positions but also include their personal attitudes.) Thus, comparing i's and j's utility levels really amounts to comparing the amounts of satisfaction they derive from their own extended alternatives.

In making such comparisons, in principle we could follow either of two alternative approaches. One is to take a third-person perspective and to try to infer the two individuals' satisfaction levels from the laws of human psychology. Yet, in actual fact, our understanding of these psychological laws is as yet very far from being sufficient for doing so.[16]

This is why I have suggested in earlier publications that we must use an alternative approach by taking a first-person perspective and by trying to achieve an empathetic understanding of what it may be like to be in either individual's objective position with the relevant individual's own personal attitudes. Moreover, in keeping with this first-person perspective, I have also suggested that each of us should ask himself whether we ourselves would prefer to be faced with the extended alternative (A_i, P_i) or with the extended

alternative (A_j, P_j). I have argued that in deciding which way our own preference would go between these two extended alternatives, we should concentrate on these two alternatives themselves, and should try to abstract from our own personal attitudes as much as we can.

Of course, I have realized that we can never get rid completely of our personal biases in making interpersonal comparisons of utility, but have been convinced that we can go a long way in doing so if we really try. The problem is similar to that faced by a liberal or a conservative historian with strong political views of his own who wants to discuss the events of past ages as objectively as he can. Perhaps he cannot completely free himself of his own political prejudices but he can come reasonably close to this objective if he tries hard enough.

More specifically, in trying to understand another individual's personal attitudes and values, we can benefit, it seems to me, from Taylor's (1981) theory of how to achieve a sympathetic understanding of the values of a foreign culture. (It is based on Gadamer's (1975) "fusion of horizons" theory of how to achieve a sympathetic understanding of the values of past ages.) Taylor suggests that we must give a sympathetic hearing to these foreign values, retaining our own old values as much as we feel reasonable to do, perhaps with appropriate modifications, but being also willing to revise our own values in the light of these foreign values whenever this seems to be the proper course to take. By following this approach, in the end we may achieve a broader point of view that does justice both to our own values and to those of the foreign culture we are trying to understand.

In trying to understand another person's attitudes and values, we must follow a

similar approach. Even though presumably the scholarly values of Socrates are intrinsically much superior to the narrowly hedonistic values of the fool, Socrates cannot meaningfully compare the fool's satisfaction level with his own unless he has a sympathetic understanding of the fool's attitudes and values -- yet without losing sight of the latter's obvious limitations.

When different people make interpersonal comparisons of utility between the same two individuals or between the same two social groups, they may perhaps arrive at somewhat different conclusions, but most of the time their conclusions will be close enough to show that making such comparisons is not an altogether hopeless undertaking.

NOTES

1. In what follows, for stylistic reasons from similar phrases I shall often omit the female pronoun.

2. Suppose that Ron has access to both medications and, being unaware of the relevant medical facts, does choose B. Obviously, under our assumptions, by doing so he will act <u>against</u> his true interests and, in this sense, will make a <u>mistake</u>.

Of course, by discussing the <u>theoretical</u> distinction between people's <u>informed</u> preferences and their <u>actual</u> preferences, we have not considered the <u>practical</u> -- i.e., the moral and political -- implications of this distinction. To what extent should society follow a <u>liberal</u> policy, permitting people to "make their own mistakes," and to what extent should it follow a <u>paternalistic</u> policy, trying to prevent people from self-damaging behavior?

In my own view, in a democratic society, <u>positive</u> paternalism, which would try coercively to prevent self-damaging behavior, can be justified only if such behavior would inflict utterly intolerable damage on the agent or would seriously damage innocent bystanders. In contrast, in general we are morally free to practice <u>negative</u> paternalism by refusing to help and to subsidize activities we consider to be against the agent's own interests. I am under no obligation to buy another drink for an alcoholic, even though I have no right to prevent him by force from drinking.

3. In some earlier publications, I used the term "true preferences" to describe what I now call "informed preferences," to indicate that these better express his true preferences than his actual preferences do. But now I am describing them as "informed preferences" because they are actually defined as the preferences he would have if he were better informed about the actual facts. (The adjective "informed" I have borrowed from Griffin (1986), even though he speaks of "informed desires" rather than of "informed preferences.")

4. In my own view, any acceptable definition of a person's utility function and of his true interests must be in terms of <u>his own</u> informed preferences rather than in terms of any other standard. It certainly should not be in terms of <u>our</u> preferences and <u>our</u> value judgments. In my own view, this is both a moral and an epistemological requirement. It is a <u>moral</u> requirement because it would be a morally unacceptable imposition to ask another person, grown-up and mentally competent, to conduct his own life in accordance with our preferences and our value judgments rather than his own. But more fundamentally, it is also an <u>epistemological</u> requirement because we have no rational grounds for defining another person's real interests in terms of <u>our</u> preferences and <u>our</u> value judgments rather than his own -- unless we have convincing arguments to show that the former are a <u>superior standard</u> for defining his real interests. Yet, it is hard to see how anybody could produce such arguments. Indeed, it is hard to imagine what

the logical structure of such arguments could possibly be.

5. Suppose that I want to define the <u>informed</u> preferences of another individual <u>i</u>. According to the definition of his informed preferences, these are the preferences he <u>would</u> have if he had correct information about all relevant facts. But I have no direct access to "correct information" about these facts. All I can do is to define <u>i</u>'s informed preferences as the preferences he would have if his beliefs about the relevant facts were correct <u>as judged by the best information I have</u> about these facts, which of course might contain some errors. This in turn means that different individuals will define <u>i</u>'s informed preferences differently if the best information they have about the relevant facts is not the same.

On the other hand, individual <u>i</u> himself will define <u>his own</u> informed preferences in terms of <u>his own</u> best information about the relevant facts, i.e., in terms of the same information he will use in forming his own <u>actual</u> preferences. Therefore, his definition of his <u>informed</u> preferences will coincide with his <u>actual</u> preferences.

6. Griffin, whose own views are quite close to what I am calling informed-preference theory, has nevertheless argued that this theory would run into serious difficulties in some cases (Griffin, 1986, p. 11). But it seems to me that on closer inspection these difficulties entirely disappear.

He adduces the example of a person who right now has no liking for caviar though he would have a strong liking for it once he had trained his own palate to savor its very special refined taste. Griffin suggests that, in terms of his _actual_ desires (or, as I would say, in terms of his actual preferences), this person has no liking for caviar, yet that, in terms of his _informed_ desires (his informed preferences), he would have a strong liking for it.

Now, according to informed-preference theory, this person's utilities must be defined in terms of his _informed_ preferences. Yet, when we do this then we obtain the conclusion that this person will derive a lot of utility from eating caviar. But this is inconsistent with the original story, according to which he will be quite unable to enjoy caviar _before_ he has trained his palate in a suitable manner.

No doubt, this conclusion _is_ inconsistent with the original story. Yet, the trouble does not lie in our theory but rather in the way Griffin proposed to define this person's informed preferences. His suggestion that, in terms of his _informed_ preferences, this person would have a strong liking for caviar needs an important qualification. The truth is that, in terms of his _informed_ preferences, he will have a strong liking for eating caviar with a properly _trained_ palate but will have no liking for eating it with an _untrained_ palate.

Once we define this individual's informed preferences in this alternative way, our theory will yield exactly the right conclusion: It will tell us that the utility this individual will

derive from eating caviar will crucially depend on whether he does so with a properly _trained_ palate or not, and it will depend on this precisely in the way described by the original story.

Thus, Griffin's interesting example poses no real difficulty for our theory. But it does show that our theory works only if our definition of informed preferences is in terms of _all_ relevant information.

7. Brandt's (1979, pp. 110-129) distinction between _rational_ and _irrational_ desires is presumably meant to serve the same purposes as my distinction between _informed_ and _mistaken_ preferences. Yet, he defines a rational desire as one that will survive a hypothetical psychological procedure he calls _cognitive psychotherapy_. Under this procedure, each person would confront any desire of his in his own mind with the likely consequences of satisfying this desire and with those of leaving it unsatisfied, representing these consequences to himself as vividly as possible, and doing this repeatedly over some reasonable period of time.

Thus, a drug addict's desire to go on consuming his drug would count as a _rational_ desire if it could not be extinguished by cognitive psychotherapy -- even if he himself came to the conclusion that he would be much _better off_ if he could break his drug habit. Presumably the implication is that if his drug habit cannot be broken by using cognitive psychotherapy then he does not have to worry about it any longer.

I find this view unacceptable. Why should a drug addict

permit his decision on what to do about his drug habit to be determined, not by his own _rational judgment_ on what decision would best serve his interests, but rather by an arbitrary psychological procedure whose outcome is not under his own rational control? If he finds that cognitive psychotherapy does not make his drug addiction go away yet thinks that he would be much better off if it did go away, then why should he not try alternative methods for achieving this objective, such as use of suitable medication, hypnosis, behavior therapy, or whatever?

8. It sometimes may be hard to decide whether some incomprehensible and pointless-looking behavior is a result of a genuine, even if very unusual, preference, or is a result of a psychological disturbance. But the more incomprehensible it is and the more pointless it seems to be the greater the plausibility of the latter kind of explanation. Rawls's (1971, pp. 432-433) famous example of a man spending all his time in counting the blades of grass certainly comes under this heading.

9. Suppose I am considering alternative possible scenarios for the future of another individual _i_. Then, I must choose probabilities for these scenarios in the light of the best information _I have_ about _i_'s future. Yet, this means that I must use _my own_ subjective probabilities for this purpose, rather than _i_'s or anybody else's subjective probabilities.

10. Parfit (1984, p. 494), Griffin (1986, pp. 16-17), and Scanlon (1989, p. 4) suggest that the desires and preferences we use in defining a person's utility level should be restricted to those for conditions significantly affecting _his own life_. Thus, desires and preferences for the well-being of members of his family and for that of his close friends should be included whereas desires and preferences concerning conditions in the distant future or conditions in remote parts of the universe should be excluded. Let me call the latter _extraneous_ desires and preferences.

This is no doubt a sensible requirement. But I am not sure that it is actually needed as a special requirement. For a sensible person will presumably assign to his extraneous desires and preferences (if he has any) _very low priority_ anyhow so that they will have virtually no effect on his utility level. This will be the case even more if we define his utility level in terms of his _informed_ desires and preferences.

On the other hand, even if we do not formally exclude extraneous desires and preferences form people's _individual_ utility functions, we should certainly exclude them from our _social_ utility function, together with all other _external_ preferences (cf. Harsanyi, 1986a, pp. 8-10).

11. I shall describe only a highly simplified version of current psychological theories, omitting details not needed for my present purposes.

12. I am using the term _desire_ in a technical sense, one covering _all_ our pro-attitudes.

13. I am a supporter of rule utilitarian theory and would define the norms of morality in terms of the rule utilitarian moral code (see Harsanyi, 1977a (Chapter 4), 1985, 1986b, 1992).

14. Another distinguished philosopher, Jim Griffin, whose views are somewhat similar to Scanlon's, seems to think that the values we assign to various things should depend on their _nonnatural qualities_ (Griffin, 1991, p. 63). But he does not tell us how we can recognize these nonnatural qualities and how we know even that there are such things, and why we should be guided by them even if there were such things as nonnatural qualities.

15. In this section, by "satisfaction" I shall always mean "informed satisfaction" as deifned in section 7. But for convenience, I shall omit the adjective "informed."

16. Scanlon's theory of substantive goods does not help us in making interpersonal comparisons of utility (or of welfare) for the following reasons. Let B_i and B_j denote the assortments of "substantive goods" available to the two individuals.

(i) Scanlon's theory does not tell us how to decide whether B_i or B_j is intrinsically "more valuable" (or whether the two are "equally valuable").

(ii) By the same token, it does not tell us whether a rational individual _will_ prefer, or _should_ prefer, B_i or B_j (or whether he will or should be indifferent between the two).

(iii) Even less does Scanlon's theory tell us what legitimate role if any a rational individual's own personal attitudes _will_ play, or _should_ play, in deciding his preference between B_i and B_j.

REFERENCES

Brandt, R. B., 1979. A Theory of the Good and the Right. Clarendon Press: Oxford, England.

__________, 1982. "Two concepts of utility." In H. B. Miller and W. H. Williams (eds.), The Limits of Utilitarianism. University of Minnesota Press: Minneapolis, MN; pp. 169-185.

Gadamer, H.G. 1975. Truth and Method. Seabury Press: New York.

Griffin, J., 1986. Well-being. Clarendon Press: Oxford, England.

__________, 1991. "Against the taste model." in J. Elster and J. Roemer (eds.), Interpersonal Comparisons of Well-Being. Cambridge University Press: Cambridge, England; pp. 45-69.

Harsanyi, J.C., 1955. "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility." Journal of Political Economy, 63, pp. 309-321.

__________, 1958. "Ethics in terms of hypothetical imperatives." Mind, 67, pp. 305-316.

__________, 1977a. Rational Behavior and Bargaining Equilibrium in Games and Social Situations. Cambridge University Press: Cambridge, England.

__________, 1977b. "Rule utilitarianism and decision theory." Erkenntnis, 11, pp. 25-53.

__________, 1985. "Rule utilitarianism, equality, and justice." Social Philosophy and Policy, 2, pp. 115-127.

______________, 1986a. "Individual utilities and utilitarian ethics." In A. Diekmann et al. (eds.), Paradoxical Effects of Social Behavior: Essays in Honor of A. Rapoport. Physica-Verlag: Vienna and Heidelberg; pp. 1-12.

______________, 1986b. "Utilitarian morality in a world of very half-hearted altruists." In W. P. Heller et al. (eds.), Social Choice and Individual Decision Making: Essays in Honor of K. J. Arrow. Vol. I. Cambridge University Press: Cambridge, England; pp. 57-73.

______________, 1992. "Game and decision theoretic models in ethics." In R. J. Aumann and S. Hart (eds.), Handbook of Game Theroy. Vol. 1. Elsevier (North-Holland): Amsterdam; Chapter 19.

Hicks, J. R., 1939. Value and Capital. Clarendon Press: Oxford, England.

Jevons, W. S., 1871. Theory of Political Economy.

Parfit, D., 1984. Reasons and Persons. Clarendon Press: Oxford, England.

Rawls, J., 1971. A Theory of Justice. Harvard University Press; Cambridge, MA.

Scanlon, T. M., 1989. "Value, desire and quality of life." MS.

______________, 1991. "The moral basis of interpersonal comparisons." In J. Elster and J. Roemer (eds.), Interpersonal Comparison of Well-being. Cambridge University Press: Cambridge, England; pp. 17-44.

Taylor, C., 1981. "Understanding and explanation in the 'Geisteswissenschaften'." in S. H. Holtzman and C. M. Leich (eds.), Wittgenstein: To Follow a Rule. Routledge and Kegan Paul: London.

WIDER Working Papers

WP 1. Amartya Sen: Food, Economics and Entitlements, February 1986
WP 2. Nanak Kakwani: Decomposition of Normalization Axiom in the Measurement of Poverty: a Comment, March 1986
WP 3. Pertti Haaparanta: The Intertemporal Effects of International Transfers, April 1986
WP 4. Nanak Kakwani: Income Inequality, Welfare and Poverty in a Developing Economy with Applications to Sri Lanka, April 1986
WP 5. Pertti Haaparanta: and Juha Kähkönen: Liberalization of Capital Movements and Trade: Real Appreciation, Employment and Welfare, August 1986
WP 6. Pertti Haaparanta: Dual Exchange Markets and Intervention, August 1986
WP 7. Pertti Haaparanta: Real and Relative Wage Rigidities – Wage Indexation in the Open Economy Staggered Contracts Model, August 1986
WP 8. Nanak Kakwani: On Measuring Undernutrition, December 1986
WP 9. Nanak Kakwani: Is Sex Bias Significant? December 1986
WP 10. Partha Dasgupta and Debraj Ray: Adapting to Undernourishment: The Clinical Evidence and Its Implications, April 1987
WP 11. Bernard Wood: Middle Powers in the International System: A Preliminary Assessment of Potential, June 1987
WP 12. Stephany Griffith-Jones: The International Debt Problem – Prospects and Solutions, June 1987
WP 13. Don Patinkin: Walras' Law, June 1987
WP 14. Kaushik Basu: Technological Stagnation, Tenurial Laws and Adverse Selection, June 1987
WP 15. Peter Svedberg: Undernutrition in Sub-Saharan Africa: A Critical Assessment of the Evidence, June 1987
WP 16. S. R. Osmani: Controversies in Nutrition and their Implications for the Economics of Food, July 1987
WP 17. Frédérique Apffel Marglin: Smallpox in Two Systems of Knowledge, Revised, July 1987
WP 18. Amartya Sen: Gender and Cooperative Conlicts, July 1987
WP 19. Amartya Sen: Africa and India: What Do We Have to Learn from Each Other? August 1987
WP 20. Kaushik Basu: A Theory of Association: Social Status, Prices and Markets, August 1987
WP 21. Kaushik Basu: A Theory of Surplus Labour, August 1987
WP 22. Albert Fishlow: Some Reflections on Comparative Latin American Economic Performance and Policy, August 1987
WP 23. Sukhamoy Chakravarty: Post-Keynesian Theorists and the Theory of Economic Development, August 1987
WP 24. Georgy Skorov: Economic Reform in the USSR, August 1987
WP 25. Amartya Sen: Freedom of Choice: Concept and Content, August 1987
WP 26. Gopalakrishna Kumar: Ethiopian Famines 1973–1985: A Case-Study, November 1987
WP 27. Carl Riskin: Feeding China: The Experience since 1949, November 1987
WP 28. Martin Ravallion: Market Responses to Anti-Hunger Policies: Effects on Wages, Prices and Employment, November 1987
WP 29. S. R. Osmani: The Food Problems of Bangladesh, November 1987
WP 30. Martha Nussbaum and Amartya Sen: Internal Criticism and Indian Rationalist Traditions, December 1987
WP 31. Martha Nussbaum: Nature, Function and Cabability: Aristotle on Political Distribution, December 1987
WP 32. Martha Nussbaum: Non-Relative Virtues: An Aristotelian Approach, December 1987
WP 33. Tariq Banuri: Modernization and its Discontents, a Perspective from the Sociology of Knowledge, December 1987
WP 34. Alfred Maizels: Commodity Instability and Developing Countries: The Debate, January 1988
WP 35. Jukka Pekkarinen: Keynesianism and the Scandinavian Models of Economic Policy, February 1988
WP 36. Masahiko Aoki: A New Paradigm of Work Organization: The Japanese Experience, February 1988
WP 37. Dragoslav Avramovic: Conditionality: Facts, Theory and Policy – Contribution to the Reconstruction of the International Financial System, February 1988
WP 38. Gerald Esptein and Juliet Schor: Macropolicy in the Rise and Fall of the Golden Age, February 1988
WP 39. Stephen Marglin and Amit Bhaduri: Profit Squeeze and Keynesian Theory, April 1988
WP 40. Bob Rowthorn and Andrew Glyn: The Diversity of Unemployment Experience since 1973, April 1988
WP 41. Lance Taylor: Economic Openness – Problems to the Century's End , April 1988
WP 42. Alan Hughes and Ajit Singh: The World Economic Slowdown and the Asian and Latin American Economies: A Comparative Analysis of Economic Structure, Policy and Performance, April 1988
WP 43. Andrew Glyn, Alan Hughes, Alan Lipietz and Ajit Singh: The Rise and Fall of the Golden Age, April 1988
WP 44. Jean-Philippe Platteau: The Food Crisis in Africa: A Comparative Structural Analysis, April 1988
WP 45. Jean Drèze: Famine Prevention in India, May 1988
WP 46. Peter Svedberg: A Model of Nutrition, Health and Economic Productivity, September 1988
WP 47. Peter Svedberg: Undernutrition in Sub-Saharan Africa: Is There a Sex-Bias?, September 1988
WP 48. S. R. Osmani: Wage Determination in Rural Labour Markets: The Theory of Implicit Co-operation, December 1988
WP 49. S. R. Osmani: Social Security in South Asia, December 1988
WP 50. S. R. Osmani: Food and the History of India – An 'Entitlement' Approach, December 1988

WP 51. Grzegorz W. Kolodko: Reform, Stabilization Policies, and Economic Adjustment in Poland, January 1989
WP 52. Dariusz Rosati and Kalman Mizsei: Adjustment through Opening of Socialist Economies, January 1989
WP 53. Andrei Vernikov: Reforming Process and Consolidation in the Soviet Economy, January 1989
WP 54. Adam Török: Stabilisation and Reform in the Hungarian Economy of the late 1980's, March 1989
WP 55. Zhang Yuyan: Economic System Reform in China, March 1989
WP 56. Amitava Krishna Dutt: Sectoral Balance: A Survey, March 1989
WP 57. Robert Pringle: Financial Markets and Governments, June 1989
WP 58. Marja-Liisa Swantz: Grassroots Strategies and Directed Development in Tanzania: The Case of the Fishing Sector, August 1989
WP 59. Aili Mari Tripp: Defending the Right to Subsist: The State vs. the Urban Informal Economy in Tanzania, August 1989
WP 60. Jacques H. Drèze, Albert Kervyn de Lettenhove, Jean-Philippe Platteau and Paul Reding: A Proposal for "Co-operative Relief of Debt in Africa" (CORDA), August 1989
WP 61. Kaushik Basu: Limited Liability and the Existence of Share Tenancy, August 1989
WP 62. Tariq Banuri: Black Markets, Openness, and Central Bank Autonomy, August 1989
WP 63. Amit Bhaduri: The Soft Option of the Reserve Currency Status, August 1989
WP 64. Andrew Glyn: Exchange Controls and Policy Autonomy – The Case of Australia 1983–88, August 1989
WP 65. Jaime Ros: Capital Mobility and Policy Effectiveness in a Solvency Crisis. The Mexican Economy in the 1980s, August 1989
WP 66. Dan W. Brock: Quality of Life Measures in Health Care and Medical Ethics, August 1989
WP 67. Robert Erikson: Descriptions of Inequality. The Swedish Approach to Welfare Research, August 1989
WP 68. Onora O'Neill: Justice, Gender and International Boundaries, August 1989
WP 69. Bernard M. S. van Praag: The Relativity of the Welfare Concept, August 1989
WP 70. Hilary Putnam: Objectivity and the Science/Ethics Distinction, August 1989
WP 71. John E. Roemer: Distributing Health: The Allocation of Resources by an International Agency, August 1989
WP 72. Charles Taylor: Explanation and Practical Reason, August 1989
WP 73. Gerald Epstein and Herbert Gintis: International Capital Markets and the Limits of National Economic Policy, October 1989
WP 74. A. D. Cosh, A. Hughes and A. Singh: Openness, Innovation and Share Ownership: The Changing Structure of Financial Markets, October 1989
WP 75. Robert B. Zevin: Are World Financial Markets more Open? If so Why and with What Effects?, October 1989
WP 76. Lance Taylor: Gap Disequilibria: Inflation, Investment, Saving and Foreign Exchange, October 1989
WP 77. Andrei Vernikov: Soviet Economy: Opening up and Stabilization, October 1989
WP 78. Kaushik Basu: The International Debt Problem: Could Someone Please Explain It to Me?, October 1989
WP 79. C. K. Omari: Rural Women, Informal Sector and Household Economy in Tanzania, October 1989
WP 80. Partha Dasgupta: Well-Being: Foundations, and the Extent of Its Realization in Poor Countries, October 1989
WP 81. Grzegorz W. Kolodko, Marian Ostrowski, Dariusz Rosati: Stabilization Policy in Poland. Challenges and Constraints, February 1990
WP 82. S. R. Osmani: Food Deprivation and Undernutrition in Rural Bangladesh, February 1990
WP 83. Kálman Mizsei, Adám Török: Modified Planned Economies at the Crossroads: The Case of Hungary, March 1990
WP 84. Partha Dasgupta: The Environment as a Commodity, March 1990
WP 85. V. M. Moghadam: Determinants of Female Labor Force Participation in the Middle East and North Africa, May 1990
WP 86. Lauri Siitonen: Political Theories of Development Cooperation – A Study of Theories of International Cooperation, July 1990
WP 87. Valentine M. Moghadam: Gender and Restructuring: Perestroika, the 1989 Revolutions, and Women, November 1990
WP 88. Walter C. Labys and Alfred Maizels: Commodity Price Fluctuations and Macro-economic Adjustments in the Developed Countries, November 1990
WP 89. Siddig A. Salih, William H. Branson and Yusuf H. Al Ebraheem: Oil Driven Macroeconometric Model of Kuwait, March 1991
WP 90. Simon Johnson and Bakhtior Islamov: Property Rights and Economic Reform in Uzbekistan, September 1991
WP 91. Simon Johnson: Spontaneous Privatization in the Soviet Union. How, Why and for Whom?, September 1991
WP 92. Jeffrey D. Sachs: Accelerating Privatization in Eastern Europe: The Case of Poland, September 1991
WP 93. Ardo H. Hansson: The Emergence and Stabilization of Extreme Inflationary Pressures in the Soviet Union, September 1991
WP 94. Ardo H. Hansson: The Importance of Being Earnest: Early Stages of the West German Wirtschaftswunder, September 1991
WP 95. Madhura Swaminathan: Understanding the "Informal Sector": A Survey, December 1991
WP 96. Stanislav V. Zhukov and Alexander Yu. Vorobyov: Reforming the Soviet Union: Lessons from Structural Experience, January 1992
WP 97. Siddig A. Salih: Managing Renewable Natural Capital in Africa, February 1992
WP 98. Haris Gazdar: Rural Poverty, Public Policy and Social Change: Some Findings from Surveys of Six Villages, May 1992
WP 99. Valentine M. Moghadam: Development and Patriarchy: The Middle East and North Africa in Economic and Demographic Transition, July 1992
WP 100. Bela Bhatia: Lush Fields and Parched Throats: The Political Economy of Groundwater in Gujarat, August 1992

WP 101. John C. Harsanyi: Utilities, Preferences and Substantive Goods, December 1992