John Broome

Extended Preferences[1]

In Christoph Fehige, Georg Meggle and Ulla Wessels (eds),
Preferences, de Gruyter.

Abstract. Ordinalism is generally taken to imply that interpersonal comparisons of good are impossible. But some ordinalists have argued that these comparisons can be made in a way that is consistent with ordinalism, on the basis of extended preferences. This paper shows that this argument is mistaken, and ordinalism is indeed incompatible with interpersonal comparisons of good.

[1] I have greatly benefitted from discussions and correspondence on the subject of this paper with John Harsanyi, Susan Hurley, Serge-Christophe Kolm, Brian Skyrms, and, particularly, Hans-Peter Weikard. The research for this paper was funded by the Economic and Social Research Council under grant R000 23 3334.

1 Introduction

Many economists have adopted a doctrine known as `ordinalism'. Ordinalism insists that we can only know about people's good by means of our knowledge of people's preferences. Many ordinalists believe their doctrine implies we cannot know how one person's good compares with another's. But some have resisted this conclusion; they have argued that we can make interpersonal comparisons of good in a way that is consistent with ordinalism. They think a particular class of preferences can constitute a basis for comparisons between the good of different people. They have in mind people's preferences between very widely defined alternatives, each of which consists of a way of life together with the personal characteristics of a person who lives that life. These are called `extended preferences'. This paper reveals a flaw in the argument that extended preferences can be a basis for interpersonal comparisons of good. It shows that interpersonal comparisons really are inconsistent with ordinalism.

Section 2 of this paper explains ordinalism in more detail. Section 3# describes the notion of extended preferences, and specifies a condition extended preferences must satisfy if they are to serve the purpose they are meant for: everyone must have the same extended preferences. Section 4 quotes an argument of John Harsanyi's that is intended to show everyone will indeed have the same extended preferences. Section 5 explains that Harsanyi's argument contains an error.

There is an alternative approach to interpersonal comparisons implicit in some ordinalist writings, including Harsanyi's. This second approach is not usually clearly distinguished from the approach through extended preferences, but actually it does not depend on extended preferences. Section 6$ describes it. Section 7'', however, shows it is inconsistent with ordinalism. Section 8 argues that this approach is anyway otiose.

My aim in this paper is only to prove that interpersonal comparisons of good are inconsistent with ordinalism. What conclusion should we draw? I draw the conclusion that ordinalism must be incorrect, since interpersonal comparisons of good are clearly possible. But in this paper I shall not try to prove as much as that.

2 Ordinalism

I shall take
ordinalism to be the conjunction of these three claims:

1. *The preference-satisfaction theory of good*. If a person has a high-grade preference for one alternative over another, then it would be better for the person to have the first alternative rather than the second.
2. Given a pair of alternatives, we can (sometimes at least) know whether a person has a high-grade preference for one of them over the other.
3. Our knowledge of people's high-grade preferences is the only way we can come to know anything about how good it would be for a person to have some alternative.

By a `high-grade preference' I mean a preference that passes some test of quality: it is rational and well-informed, or something of that sort. It does not matter for my purposes precisely what the test is, though in Section 7'' I shall mention something it cannot be. Some ordinalists may

not insist on any test at all; for them, any preference is high-grade. From now on in this paper, I shall only speak of high-grade preferences, and the word `preference' is always to be understood as referring to a high-grade preference.

I am treating ordinalism as a theory about what we can know, and not as a theory about meaning. Many authors take ordinalism to include some verificationist view about meaning: perhaps the view that a statement about people's good has no meaning unless we can know whether or not it is true. But this paper is not concerned with meaning, and I shall leave verificationism aside.

3 Extended Preferences and Interpersonal Comparisons of Good

One conclusion that has often been drawn from ordinalism is that we cannot compare one person's good with another's. The argument is this. One person *i*'s preferences will tell us whether one alternative is better for *i* than another. Another person *j*'s preferences will tell us whether one alternative is better for *j* than another. But no one's preferences will tell us whether one alternative is better for *i* than another alternative is for *j*.

Not all ordinalists have accepted this conclusion. Some have argued that preferences can indeed allow us to compare one person's good with another's. It depends on what sorts of alternatives the preferences are amongst. Preferences amongst cheeses will not do it, but these ordinalists say *extended preferences*, as they call them, will. An extended preference is a preference between *extended alternatives*, and an extended alternative consists of a way of life paired together with particular personal characteristics. For instance, one extended alternative is to live the life of an academic whilst possessing a thirst for knowledge and modest material needs. Another is to live the life of an academic whilst possessing an insatiable desire for wealth. No doubt we have preferences amongst such alternatives; I prefer the first of the two I have just described to the second. These are extended preferences. It has been argued that extended preferences can give us grounds for interpersonal comparisons of good (for instance, in Arrow (1977) and Harsanyi (1977), pp. 57–60).

How would this work? Suppose I have an extended preference for living person *i*'s life with *i*'s personal characteristics over living *j*'s life with *j*'s personal characteristics. According to the preference-satisfaction theory, it would be better for me to lead *i*'s life with *i*'s personal characteristics than to lead *j*'s life with *j*'s personal characteristics. This suggests that *i*, living her life, is better off than *j*, living hers. In this way an extended preference seems to allow a comparison between one person's good and another's. In this case my preference seems to have allowed a comparison between *i*'s good and *j*'s.

This is not a convincing argument without some extra support. The fact that John Broome has an extended preference for *i*'s life and characteristics over *j*'s life and characteristics cannot be sufficient evidence that *i* is better off than *j*. What if someone else had the opposite extended preference to mine? Even if *i* herself has this same extended preference, that would not be sufficient evidence; what if *j* had the opposite one? Evidently, if extended preferences are to be the basis for interpersonal comparisons of good, the extended preferences of different people must coincide to some extent at least.

Those authors who rely on extended preferences as a basis for interpersonal comparisons of good generally insist that everyone will have the same extended preferences, provided

extended alternatives are construed widely enough.[2] We must make sure that our extended alternatives are a full specification of all those features of lives and personal characteristics that anyone can possibly have preferences about. Given that, these authors claim we must all have the same preferences about such very widely extended alternatives.

Why should this be? It certainly seems implausible. I myself prefer to live the life of an academic, with my own academic characteristics, even in the conditions allotted to academics in contemporary Britain, to being a financial analyst with the characteristics of a financial analyst living in the conditions allotted to financial analysts. A financial analyst, with her different values, would no doubt have the opposite preference. So her extended preferences will differ from mine. Both our preferences may well be high-grade: they may be rational, well-informed and so on. The reason I have mine is that an academic has some slight chance of making a worthwhile contribution to knowledge. I recognize that, if I were a financial analyst, with all the characteristics of a financial analyst, I would not then value knowledge as I do now. Nevertheless, I do value knowledge, and that is why I prefer to be an academic. The different values of academics and financial analysts lead us to have different extended preferences. Or so it certainly seems.

Appearances, then, are against the claim that different people must have the same extended preferences. Are there, on the other hand, any arguments in favour of this claim? I know of one. It is based on the causes of preference. It is spelt out in most detail in Harsanyi (1977), pp. 57–60, but other authors have used it too.[3]

4 Harsanyi's Causal Argument

Harsanyi's argument is difficult to interpret, and I do not want to misrepresent it. To reduce the risk of unfairness, I shall quote almost all of it in this section, before going on to explain and criticize it in Section 5. Harsanyi (1977), pp. 58–9, says:

> If all individuals' personal preferences were identical, then we could ascribe the same utility function U to all individuals and could always make interpersonal utility comparisons in terms of this common utility function U. Moreover, all *inter*personal utility comparisons could be reduced to *intra*personal utility comparisons. If we wanted to know whether a given apple would give more utility to Peter (who has just had a heavy meal) than to Paul (who had only a very light meal), we could simply ask whether Peter himself would derive more utility from an apple after a heavy meal or after a light meal.
>
> Of course, in actuality different individuals often have very different personal preferences and very different utility functions. But the possibility of meaningful interpersonal utility comparisons will remain, as long as the different individuals' choice behavior and preferences are at least governed by the *same basic psychological laws*. For in this case each individual's preferences will be determined by the same general causal variables. Thus the differences we can observe between different people's preferences can be predicted, at least in principle, from differences in these causal variables, such as differences in their biological inheritance, in their past life histories, and in their current environmental conditions. This means that if Peter had Paul's biological makeup, had Paul's life history behind him, and were currently subject to Paul's environmental influences, then he would presumably have the *same* personal preferences as Paul has now and would ascribe the *same* utility as Paul does now to each particular situation.
>
> Let P_j again denote individual j's *subjective attitudes* (including his preferences), and let R_j

[2] See the references in note 3.

[3] The argument appears in rudimentary form in Harsanyi (1955), pp. 17–18 in the reprinted version, and independently in Tinbergen (1957), p. 501. It also appears in Kolm (1972), pp. 79–80.

denote a vector consisting of all *objective causal variables* needed to explain these subjective attitudes denoted by P_i. Our discussion suggests that the extended utility function V_i of each individual i should really be written as $V_i = V_i[A_j, R_j]$ rather than $V_i[A_j, P_j]$. Written in this form, the utility function $V_i = V_i[A_j, R_j]$ indicates the utility that individual i would assign to the objective position A_j if the causal variables determining his preferences were R_j. Because the mathematical form of this function is defined by the basic psychological laws governing people's choice behaviour, this function V_i must be the same for all individuals i, so that, for example,

$$V_h[A_j, R_j] = V_i[A_j, R_j]$$

for each pair of individuals h and i. In the special case in which $h = j$, we can write

$$V_i[A_j, R_j] = V_j[A_j, R_j] = U_j(R_j)$$

That is, individual i (or individual h) would have the *same* preferences and would assign the *same* utility to any objective situation A_j as individual j now does, if the causal variables determining his preferences took the same value R_j as do the causal variables determining j's preferences.

In other words, even though the `ordinary' utility functions U_i and U_j of two individuals i and j may be quite different, their *extended* utility functions V_i and V_j will be identical. This is so because, by the definition of the causal-variables vectors R_i and R_j, all differences between the two functions $U_i(A_i) = V_i[A_i, R_i]$ and $U_j(A_j) = V_j[A_j, R_j]$ must be attributed to differences between the vectors R_i and R_j and not to differences between the mathematical form of the two functions V_i and V_j.

Yet, if the two individuals have the same extended utility function $V_i = V_j = V$, then we are back in a world of identical utility functions. Hence individual i will be able in principle to reduce any *inter*personal utility comparisons that he may wish to make between himself and individual j to an *intra*personal utility comparison between the utilities that he is *in fact* assigning to various situations and the utilities he *would* assign to them *if* the vector of causal variables determining his preferences took the value R_j (which is the value that the vector of these causal variables takes in the case of individual j).

For example, if I want to compare the utility that I would derive from a new car with the utility that a friend would derive from a new sailboat, then I must ask myself what utility I would derive from a sailboat if I had taken up sailing for a regular hobby as my friend has done, and if I could suddenly acquire my friend's expert sailing skill, and so forth. . . .

5 The Causal Determination of Extended Preferences

Harsanyi's argument relies on the notion of a *utility function*, and cannot be expressed without it. So I need first to explain this notion. If a person has preferences over some range of alternatives, and if these preferences conform to a number of conditions, it is possible to *represent* them by a function U, which is called a utility function. U assigns a value called a *utility* to each of the alternatives. To say U represents the preferences means that U assigns higher utilities to alternatives that are preferred. More precisely: the utility assigned by U to one alternative is at least as great as the utility assigned by U to another if and only if the first alternative is preferred or indifferent to the second. This is how Harsanyi (1977), pp. 31–2, explicitly defines `utility'. Some of his expressions in the passage I have quoted suggest he may also have in mind another meaning for `utility', and I shall come to that in Section 6$. But for the moment I shall stick with this one: utility is the value of a function that represents preferences. Harsanyi calls a utility function that represents extended preferences an *extended utility function*, and I shall do the same.

One technical point about utility functions. There is not just one utility function that represents a person's preferences; many do. If a function U represents someone's preferences, then another function U' will also represent her preferences if and only if U' is an *increasing transform* of U. U' is defined to be an increasing transform of U if and only if, for any pair of alternatives A and B, $U'(A) \geq U'(B)$ if and only if $U(A) \geq U(B)$. For instance, if U' is the function U/a, obtained by dividing U by some positive constant a, then U' is an increasing transform of U, and it will represent the same preferences as U.

A person j has preferences about ways of life, which may be represented by a utility function U_j. Different people have different preferences, and i will have preferences represented by a different function U_i. But there is a causal explanation of why each person has the preferences she has. Let R stand for the causal variables that determine the form of a person's preferences (her upbringing, friends, sporting ability and so on). A person who is subject to causal influences R will have preferences that can be represented by a utility function U_R. Any person subject to the same influences R will have the same preferences, so there is no need to index the function U_R by the name of the person whose utility function it is. This function assigns a utility $U_R(A)$ to ways of life A. An alternative notation is $U(A; R)$. I separate A and R by a semicolon rather than a comma for a reason that will appear.

The function U is a universal function, the same for everybody. It represents the preferences that a person subject to causal influences R will have about ways of life A. It certainly does not represent preferences of any sort, belonging to anybody, about the combination (A, R) of a way of life and causal influences.

I hope this is obvious, but an example may help to make it more so. Suppose people have preferences about opportunities to contribute to knowledge, l, and money, m. Suppose a person's preferences can be represented by a utility function that has this particular form:

$$U_\alpha(l, m) = \alpha\log l + (1-\alpha)\log m$$

Not everyone has exactly the same preferences, because the parameter α (which lies between 0 and 1) differs from person to person. There is a causal explanation of why a person has the particular value of α she has. Let us suppose her α is determined by the age when she was weaned. In fact, let α be equal to w, the age at which the person is weaned (everyone is weaned before one). A person's utility function $U_w(l, m)$, or in an alternative notation $U(l, m; w)$, is:

$$U(l, m; w) = w\log l + (1-w)\log m.$$

This function represents the preferences that a person weaned at w has about l and m. It does not represent anyone's preferences about w. It may be that no one even has any preferences about when she is weaned. The semicolon separates the objects of preference l and m from the cause of preference w. Objects of preference and causes of preference have quite different roles in the utility function.

There is a formal way of demonstrating this point. Let

$$U_\alpha'(l, m) = \log l + (1/\alpha - 1)\log m.$$

U_α' is obtained by dividing U_α by α, which, as a parameter of the function, is a constant. U_α' is therefore an increasing transform of U_α. So U_α' represents the preferences of a person with parameter α, just as U_α does. But, as a causal matter, α is equal to the person's age of weaning w. So the preferences about l and m of someone weaned at w can be represented by the function

$$U'(l, m; w) = \log l + (1/w - 1)\log m.$$

$U'(l, m; w)$ is an increasing transform of $U(l, m; w)$ if the arguments of the function are taken as l and m. But if l, m and w are all taken as arguments, then $U'(l, m; w)$ is not an increasing transform of $U(l, m; w)$. Consequently, these two functions cannot represent the same preferences over the three variables l, m and w together. This shows they are no representing preferences over these three variables at all.

Return to the general case. $U(A; R)$ as I defined it is a universal function, the same for

everyone. But it is not a universal utility function representing preferences about A and R together. In the argument I quoted, Harsanyi calls $V_i[A_j, R_j]$, which is my $U(A; R)$ expressed in his notation, an extended utility function. He implies that it represents extended preferences over ways of life A and causal variables R together. But it does not. Harsanyi hoped to exhibit universal extended preferences, the same for everyone. But he fails to do so.

There is a complication. The things that have a causal influence over people's preferences may also be things that people have preferences about. For instance, people have preferences about the friends they have, and their friends influence their preferences. Harsanyi's separation between `objective position' A and causal variables R is too sharp. I must reformulate my point to take this complication into account.

Let A be a variable that stands for ways of life. Let P stand for personal characteristics. Then (A, P) is an extended alternative as I originally defined it: a way of life lived with particular personal characteristics. People have extended preferences about these extended alternatives. Let us make sure that A and P are defined broadly enough to include anything that anyone has a preference about. (It does not matter what is included in A and what in P, so long as everything comes into one or the other.) A person j has extended preferences about (A, P) that can be represented by an extended utility function $V_j(A, P)$. I have given V_j the index j because as yet we have no reason to think j's extended preferences are necessarily the same as anyone else's.

Now, there are causal variables R that determine people's extended preferences. A person who is subject to causal variables R will have extended preferences $V_R(A, P)$. An alternative notation is $V(A, P; R)$. Since anyone subject to the same causal variables will have the same extended preferences, there is no need to index this function by j. V is a universal function.

Many of the variables in R will be things that people have preferences about. So they will also appear in A or P. Indeed, let us now go back and enlarge A and P to make sure that they include, not only anything that anyone has a preference about, but also anything that has any causal influence on people's preferences. (A, P) is now a very comprehensive specification of a life and a person living that life. In $V(A, P; R)$ all the variables contained in R will now also appear in A or P.

Still, R is not redundant as an argument of the function. Although all the variables in R will also appear in A or P, it will be different values of the variables in each case. R contains the values of the causal variables that a person is actually subject to. (A, P), on the other hand, contains the values of the variables that the person contemplates as objects of her preference.

$V(A, P; R)$ is a universal function. But it does not represent universal extended preferences. It does not represent any preferences at all over A, P and R taken together. It represents preferences over A and P, but many different preferences, one set of preferences for each value of R. So, even after taking account of the complication, we have not found universal extended preferences.

Serge-Christophe Kolm (1972), pp. 79–80, says:

> At bottom, all individuals have the same tastes, the same desires. Without doubt, this assertion requires explanation.
>
> If two persons have preferences which appear to differ, there is a reason for this, there is something which makes them different from each other. Let us place this `something' within *the object of the preferences* which we are considering, thereby removing it from the parameters which determine the structure of these preferences. The preferences of these two persons defined in this way are necessarily identical.
>
> We may carry out this operation in the case of any society: namely, the operation of placing in the object of preferences everything which would cause differences between the preferences of different members of society. An identical preference of all members of this society obtained

> in this way is called `a fundamental preference' of the members of this society. It is a property which describes the tastes and needs of the `representative individual' of this society.
>
> If this society includes all human beings, then that which discerns this common preference is at bottom `human nature'.[4]

Kolm evidently has an idea like Harsanyi's. The causes of people's preferences may indeed also be objects of their preferences. But by including causes amongst objects of preference, we do not stop them from being causes. We do not remove them from the parameters which determine the structure of people's preferences. They retain their role as causes, and they may cause one person's preferences to be different from another's. My preferences are influenced causally by the life I lead. Since I am an academic, I have preferences that differ from the preferences of a financial analyst. The life I lead is also amongst the objects of my preferences: I prefer the life of an academic to the life of a financial analyst. But simply recognizing that the cause of my preferences is also an object of my preference is not going to make my preferences identical to a financial analyst's. No doubt the financial analyst has the opposite preference to mine. It is just not true that, at bottom, all individuals have the same tastes and desires.

Kolm seems to think that, by treating the causes that act upon me as objects of my preference, I can somehow withdraw myself from their influence. But I cannot escape from my own causal situation.

The causal argument was offered as a demonstration that everyone will have the same extended preferences. It fails.

6 The Causal Determination of Good

What then, is the attraction of the causal argument? Why have so many authors put forward an argument that is so plainly mistaken? I suspect they have confused it with another argument that is not mistaken. They have been looking for grounds for interpersonal comparisons of good. And there is a truth about causation that is germane to interpersonal comparisons of good. It is true, but not germane, that if two people were in the same causal situation they would have the same preferences. It is also true, and germane, that if two people were in the same causal situation, they would be equally well off.

This latter truth might possibly give us a route to interpersonal comparisons of good. It would allow us to move from comparisons of good for a single individual to comparisons of good between individuals. Suppose we are able to determine, for a single person *i* and for any pair of alternatives *A* and *B*, whether *A* or *B* is better for *i*, or whether the two are equally good for her. Suppose, that is to say, we can put all the alternatives in order according to their goodness for *i*. Now suppose the alternatives we are dealing with are very broadly defined. They are extended alternatives, including both ways of life and personal characteristics, and also including all the causal variables that could have an influence on how well off a person is. One of these alternatives could not possibly be any better or worse for *i* than it would be for anyone else. Take two of these extended alternatives, *E* and *F*. Would *E* be better for one person, say *j*, than *F* would be for a different person, say *k*? This would be so if and only if *E* would be better for *i* than *F* would be for *i*. It would be so, that is to say, if and only if *E* comes higher than *F* when the alternatives are ordered by their goodness for the single person *i*.

[4] I take this translation from Rawls (1982), p. 174.

When we order the alternatives by their goodness for *i*, therefore, we are also producing an interpersonal ordering. Provided our alternatives are defined sufficiently widely, an ordering by goodness for a single person is also an interpersonal ordering. Simply, it is an ordering of the alternatives by their goodness.

If I were a financial analyst living the life of a financial analyst, subject to all the causal influences that determine how well off a financial analyst is, then I should be exactly as well off as anyone else would be if she occupied that position. The same is true for the alternative of being an academic in the causal situation of an academic. Therefore, if it is better for me to live the life of an academic, it would be better for anyone. The life of an academic would, simply, be better.

So, if we can discover an ordering of widely extended alternatives by their goodness for any single person, we can make interpersonal comparisons of good: we can know whether one person is better off than another. This conclusion is only conditional. *If* we can discover an ordering for a single person, we can make interpersonal comparisons. How we can discover an ordering for a single person is another matter, which I shall come back to in Sections 7" and 8.

I suspect this possible route to interpersonal comparisons has been in the minds of several of the economists who have more explicitly followed the route of extended preferences. It is difficult to be sure, because of the ambiguous use of the word `utility' that is common in economics. (See Broome (1991)). I defined `utility' in Section 5 as the value of a function that represents preferences. This is the definition most often made explicit. But economists also often speak of a person's `utility' when they mean the person's good. In his discussion of `extended sympathy' as a basis for interpersonal comparisons of good, Kenneth Arrow (1977) says:

> We may suppose that everything which determines an individual's satisfaction is included in the list of goods. Thus, not only the wine but the ability to enjoy and discriminate are included among goods. . . If we use this complete list, then everyone should have the same utility function for what he gets out of the social state.

What does Arrow mean when he says everyone should have the same utility function? If he is thinking of a utility function as representing a person's preferences, he must mean that everyone has the same preferences – in this case the same extended preferences. As I say, that is false. If, on the other hand, he is thinking of a utility function as measuring a person's good, he may mean that each extended alternative is as good for one person as it is for another. As I say, that is true, and leads to a different route to interpersonal comparisons of good. I am inclined to think Arrow means to follow this second route, but I am not perfectly sure.

The ambiguity of `utility' is acute in Harsanyi's writings. It is the main reason why the argument I quoted in Section 4 is so hard to interpret. When Harsanyi says `if Peter had Paul's biological makeup, had Paul's life history behind him, and were currently subject to Paul's environmental influences, then he would presumably have the *same* personal preferences as Paul has now and would ascribe the *same* utility as Paul does now to each particular situation', I think he is probably making two quite different points. The first is that if Peter had the same biological makeup and so on as Paul, he would have the same preferences as Paul. The second is that if Peter had the same biological makeup and so on as Paul, then any particular situation would be exactly as good for Peter as it would be for Paul.

The first of these points led Harsanyi to the argument that is most explicit in the passage I quoted: the argument that everyone must have the same extended preferences. This argument is mistaken, as I explained in Section 5. The second point is submerged through most of the

passage but surfaces at the end, when Harsanyi wants to compare the benefit a friend would get from a new sailboat with the benefit he himself would get from a new car. He says, `I must ask myself what utility I would derive from a sailboat if I had taken up sailing for a regular hobby as my friend has done . . .' Harsanyi plans to ask himself, not what preferences he would have if he were in his friend's causal situation, but how well off he would be if he were in his friend's causal situation and had bought a new boat. He knows that if he, Harsanyi, were in his friend's causal situation and had bought a new boat, he would be exactly as well off as his friend would be if he had bought a new boat. So, on this occasion, the route he is taking to interpersonal comparisons is not via extended preferences. He is taking the alternative route I have described in this section.

7 Ordinalism and Goodness for a Person

It is not surprising that this second route to interpersonal comparisons is not made very explicit in the writings of ordinalists. It is inconsistent with ordinalism. The route goes like this. First, we must have extended alternatives ordered by their goodness for one person. We know that an extended alternative is equally as good for one person as it is for any other person. So the ordering for the single person will also be an interpersonal ordering. But how do we find an ordering for the single person to start with? The answer cannot be consistent with ordinalism.

According to ordinalism, we should have to find the single person's ordering through her preferences. The ordering we are interested in is the ordering of extended alternatives by their goodness for the person. We could find it through the person's extended preferences if the preference-satisfaction theory of good (see section 2) were true. This theory implies that the ordering of extended alternatives by their goodness for a person coincides with the person's extended preference ordering. But by now I am able to say with confidence that the preference-satisfaction theory is false, at least when applied to extended preferences. If it were true, one person's extended preference ordering would coincide with the ordering of the alternatives by their goodness for the person. That is to say, it would coincide with their ordering by, simply, their goodness. Consequently, it would coincide with *everyone*'s extended preference ordering. So the preference-satisfaction theory implies that everyone has the same extended preferences. But people do not all have the same extended preferences. Therefore, the preference-satisfaction theory is false. If follows that we cannot use a person's extended preferences to find an ordering of extended alternatives by their goodness for that person. We cannot find this ordering in a way consistent with ordinalism.

The argument in the previous paragraph depends on my assertion that people do not all have the same extended preferences. Up to Section 5, I took seriously the claim that everyone necessarily has the same extended preferences, despite the counterexample of the academic and the financial analyst. But now, after the collapse of the causal argument, I need no longer take this claim seriously. I am happy to use the clear fact that extended preferences differ – for instance, that different people have different preferences between the life of an academic and the life of a financial analyst – as evidence against the preference-satisfaction theory of good. The preferences of all of us are determined by our causal situation, and in the world as it is, our preferences are caused to differ. I come out preferring to be an academic, and an analyst comes out preferring to be an analyst.

To be sure, the preference-satisfaction theory is concerned only with high-grade preferences (see section 2). If either my preferences or the analyst's were not high-grade, our differing

preferences would not confute the preference-satisfaction theory. But both our preferences might be high-grade by any reasonable test. We might have thought long and hard about the relative merits of the two alternatives, scratching our ears, applying our best mental powers and using all the information available. The difference between us is our values: our different causal situations cause us to have different values. But that is no reason to think one of us irrational or ill-informed.

It might be that one of the alternatives that face us is better than the other as a matter of objective fact. I doubt it, but suppose it is for a moment. Then either I or the analyst has got it wrong: one of us prefers an alternative that is objectively worse. Let it be the analyst. Could we say the analyst's preference fails to be high-grade on that account? Could our test of quality for preferences be that they must be in line with objective goodness? That would be inconsistent with the epistemology of ordinalism. It would imply that we could not know whether a preference between two alternatives was high-grade unless we first knew which of the alternatives was better. According to ordinalism, however, we could know nothing about the goodness of any alternative unless we already knew some preferences and knew that those preferences were high-grade. There is no way to break into this circle. So, even if one of our preferences is objectively wrong, it cannot follow that it is not a high-grade preference. A high-grade preference cannot be defined in such a demanding way.

We have to conclude that the preference-satisfaction theory of good is false when applied to extended preferences. This is no serious problem for ordinalism itself; ordinalists can perfectly well decline to apply ordinalism to the arcane domain of extended preferences. But it does mean that there is no route to interpersonal comparisons of good that is consistent with ordinalism. Granted their own assumptions, those ordinalist who deny the possibility of interpersonal comparisons are right.

8 From Individual Orderings to Interpersonal Orderings?

In Section 6$, I suggested we might be able to move from a ordering of extended preferences by their goodness for a single person to an interpersonal ordering. In Section 7'', I explained that this idea is inconsistent with ordinalism because we could not derive a single person's ordering from preferences. Is there any other way we could find out a single person's ordering, in order to discover an interpersonal one?

What must we do to order extended alternatives by their goodness for a person? The task has two parts. First, we must do some theoretical work in ethics to discover what a person's good consists in. Then we must do some empirical work to see how much of this good is delivered by each particular alternative. The nature of the empirical research we shall need to do depends on what conclusion we have reached at the first stage, in deciding what a person's good consists in.

For instance, suppose we conclude that ethical hedonism is correct, and a person's good consists in having good feelings. Then we shall have to do some psychological research into how good are the feelings that different ways of life produce. One way of proceeding would be to take a subject, put her into various causal situations, and ask her which ones make her feel better or worse. This would give us an ordering of the alternative situations by their goodness for the subject. However, the types of alternative we are interested in are whole lives together with the personal characteristics of the people living them. No experiment would allow us to put one subject into a range of alternative situations of this sort. Instead, our

empirical research will require us to compare the feelings of a person living one life with the feelings of a different person living another life, to see which feelings are better. But many people – most ordinalists amongst them – are sceptical about the possibility of comparing the feelings of different people. If they are hedonists, this will lead them to doubt the possibility of interpersonal comparisons of good. However, they might think they can overcome the difficulty by undertaking an imaginary experiment rather than a real one. Instead of causing an actual subject to lead various lives, they would try to imagine themselves leading various lives, and see how they feel when they do. That way, they could find out which lives would be better or worse for themselves: they could order lives by their goodness for themselves. From this personal ordering, they could move to an interpersonal one.

So this is one putative way of finding an order for a person and deriving an interpersonal one from it. But this procedure is dubious for at least two reasons. First, the imaginary experiment itself is dubious. You are supposed to imagine yourself leading a different life from your actual life, with different personal characteristics, and find out how good you feel by observing yourself. However, if the characteristics are much different from your actual characteristics, it would actually be impossible (not just causally impossible, but even metaphysically impossible) for you to have those characteristics, because anyone with those characteristics would not be you. So the situation you are supposed to imagine in this experiment is impossible. Just how you are supposed to conduct this imagined impossible experiment is obscure. Probably the best you could do is imagine a person, not particularly yourself, leading the life and possessing the characteristics you are interested in. You could estimate how good that person would feel. But whatever ordering by goodness you derive from this estimation would not particularly be a personal ordering by goodness for yourself. Yet that is what you are supposed to be finding.

The second reason why the procedure is dubious is that it assumes the dubious doctrine of ethical hedonism. I doubt that hedonism would be the right conclusion to draw from our ethical investigation into what a person's good consists in. I think it is more likely that a person's good consists in a number of more traditional good things: health, having friends, material comfort, varied and interesting experiences, and so on. The most difficult part of our work is likely to be the theoretical job of deciding what these goods are, and how they weigh against each other. After that, we shall come to the empirical investigation into how much of the goods each life delivers to the person who leads that life. But if the goods turn out to be overt things like the ones I mentioned, this may not be so difficult. Hedonists may find the empirical task difficult because they believe the good things in life are feelings, which are often supposed to be covert. But overt goods are more easily detected.

I should like to offer two particular speculations about the conclusions we are likely to come to. I think one conclusion is likely to be that there is no objective fact as to whether the life of an academic is better or worse than the life of a financial analyst. Different goods, such as wealth and opportunities to contribute to knowledge, are surely incommensurable to some extent, from an objective point of view. Consequently, there is room for the analyst and me to have different extended preferences, without either of our preferences being open to objective criticism.

A second conclusion is likely to be that nothing can be gained in our research by starting from the good of an individual and then moving to interpersonal comparisons. The question is simply which lives are better than others, and nothing can be gained by asking first which would be better than others for a particular individual. Indeed, this is a peculiar question in the

first place.[5] Since it would not have been possible for me to live a life far different from my actual life, it is peculiar to ask how good such a life would be specifically for *me*. All we can sensibly ask is, simply, how good the life would be. Interpersonal comparisons are unlikely to be a real problem, because our investigation of good is likely to be interpersonal from the start. Section 6$ mentioned a possible route to interpersonal comparisons through the goodness of lives for individuals; I am now suggesting this route is redundant.

References

Arrow (1977). Arrow, Kenneth J.: `Extended sympathy and the possibility of social choice', *American Economic Review: Papers and Proceedings* 67 (1977), pp. 219–25.

Broome (1991). Broome, John: `"Utility"', *Economics and Philosophy* 7 (1991), pp. 1–12.

Harsanyi (1955). Harsanyi, John C.: `Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility', *Journal of Political Economy* 63 (1955), pp. 309–21; reprinted in Harsanyi (1976), pp. 6–23.

Harsanyi (1976). Harsanyi, John C.: *Essays on Ethics, Social Behavior, and Scientific Explanation*, Dordrecht 1976.

Harsanyi (1977). Harsanyi, John C.: *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge 1977.

Kolm (1972). Kolm, Serge–Christophe: *Justice et Équité*, Paris 1972.

Rawls (1982). Rawls, John: `Social unity and primary goods', in Sen/Williams (1982), pp. 159–86.

Sen/Williams (1982). Sen, Amartya, & Williams, Bernard, (eds): *Utilitarianism and Beyond*, Cambridge 1982.

Tinbergen (1957). Tinbergen, Jan: `Welfare economics and income distribution', *American Economic Review: Papers and Proceedings* 47 (1957), pp. 490–503.

[5] Susan Hurley impressed this point on me.