

# Preference Change and Interpersonal Comparisons of Welfare<sup>1</sup>

Alex Voorhoeve

To appear in *Philosophy* and in Serena Olsaretti (ed.) *Preference Formation and Well-Being*

Cambridge University Press, forthcoming.

## Introduction

Suppose that we agree that for questions of justice in a pluralistic society, we need a public standard of welfare. An appropriate public standard of welfare will have to meet the following two requirements. First, its conception of each person's welfare should, to the greatest reasonable extent, be something that each person can recognise as encompassing the things she wants for herself and as giving these things weights that reflect the relative importance she gives to them. Second, it should be sensitive to the fact that reasonable people hold conflicting conceptions of what constitutes an individual's welfare. It should therefore, to the greatest reasonable extent, respect neutrality of judgement by refraining from endorsing any particular conception of welfare as superior to any other.

In an influential set of essays, Richard Arneson (1990a, 1990b, 1990c) has argued that the following conception of welfare is ideally suited to these requirements: equate each individual's welfare with the degree of satisfaction of her ideally rational, self-regarding preferences. These are the preferences she would have on behalf of herself if she were to engage in ideally extended deliberation with full pertinent information, in a calm mood, while thinking clearly and making no reasoning errors (see Arneson 1990a, pp. 162-163). (For simplicity, in what follows, I will use the term 'preferences', to refer to these ideally rational, self-regarding preferences.)

Arneson argues that this standard of welfare meets the two aforementioned requirements in the best way possible. It meets the first requirement, he argues, because it comes as close as

possible to adhering to a person's own view of her welfare within the constraints set by the need to avoid the intuitively unpalatable move of considering something of value to her that she only considers to be so because of a lack of information or incomplete or erroneous deliberation (1990a, p. 163). It meets the second requirement, he argues, because it does not involve any commitment on the part of the state to a substantive view of what is good for individuals. As Arneson (1990b, p. 450) puts it: 'the good in this conception is an empty basket that is filled for each individual according to her considered evaluations.'<sup>2</sup>

In this paper, I will argue that the fact that people care about which preferences they have, and the fact that people can change their preferences about which preferences it is good for them to have, together undermine this case for accepting a preference satisfaction conception of welfare.

The paper is organised as follows. In section 1, I introduce three concepts of importance to the discussion of a preference satisfaction conception of welfare. First, a person's preference type, which encompasses all the things that determine how she would evaluate, after ideally rational and well-informed deliberation and from the perspective of her self-interest, her situation and her evaluative dispositions. Second, a person's limited preferences, which are her preferences over alternatives in which her situation differs whilst her preference type remains unchanged. Third, a person's extended preferences, which encompass her preferences over alternatives in which either her situation, or her preference type, or both, differ.

In section 2, I argue that the interest in having the preferences one wants to have is of central importance in human life, and that the preference satisfaction approach should therefore attempt to judge a person's welfare by the degree to which her extended preferences are satisfied.

In section 3, I argue that the possibility of a change in a person's extended preferences creates great difficulties for a measure of welfare based on the degree of satisfaction of a person's extended preferences. For if we evaluate a potential change in a person's extended preferences from the perspective of the degree to which her future preferences would be satisfied, then we do not adequately represent each individual's current interest in shaping her future preferences. If, by contrast, we evaluate a potential change in a person's preferences from the perspective of her current extended preferences, then we implausibly disregard the view of her own interests that she would have if the preference change occurred. In sum, it seems that any preference satisfaction measure will have significant drawbacks, since it will involve either neglecting individuals' current interests in shaping their future preferences, or, in their future, judging their welfare by a set of values which might be very alien to them.

In the final section, I argue that this should prompt us to develop alternative measures of welfare. I suggest that one promising candidate is a substantive measure of welfare based on a list of goods and conditions that are recognised as valuable from the perspective of a variety of different conceptions of welfare.

### **1. Preference-based interpersonal comparisons of welfare**

Let us start with a simple description of what is involved in preference-based interpersonal comparisons of welfare. For simplicity, I will limit the discussion to cases where, from the perspective of his self-interest, a person cares only about his own situation, and is indifferent to other people's situations and their attitudes towards him. Let us begin by introducing the notion of a preference type. A preference type encompasses all the things that determine how a person would evaluate, after ideally rational and well-informed deliberation and from the perspective of his own self-interest, his situation and his own evaluative dispositions. (From now on, I will drop reference to a person's evaluations being those he would have after ideal

and fully informed deliberation and from the perspective of his self-interest, and take them to be so.) A person's preference type therefore tells us how he would rank each combination of his personal situation and evaluative dispositions, and also which evaluative dispositions he has. In a terminology which will shortly be introduced, this is equivalent to saying that two people have the same preference type just in case what I will be calling their 'limited preferences' and their 'extended preferences' are identical.

Let  $\{\underline{A}, \underline{B}\}$  be the set of preference types consisting of artist's preferences ( $\underline{A}$ ) and banker's preferences ( $\underline{B}$ ). Let  $\underline{S}$  be the set of all possible states of the world. A state of the world is a description of all relevant aspects of each person's situation, excluding each person's judgements and other evaluative dispositions. Let  $\underline{S}$  be a state of the world in  $\underline{S}$ . Let  $\underline{I}$  be the set of individuals and  $\underline{i}$  be an individual in  $\underline{I}$ . The set  $\underline{S} \times \underline{I}$  stands for the set of all pairs  $(\underline{S}, \underline{i})$  with  $\underline{S}$  in  $\underline{S}$  and  $\underline{i}$  in  $\underline{I}$ . A pair  $(\underline{S}, \underline{i})$  stands for 'occupying personal position  $\underline{i}$  in state of the world  $\underline{S}$ '. Let  $\underline{u}^t$  be a Von Neumann and Morgenstern utility function defined on the set  $\underline{S} \times \underline{I}$ , representing the preferences of a person with preference type  $\underline{t}$  over combinations of occupying a person's position in a state of the world. It assigns a real number  $\underline{u}^t(\underline{S}, \underline{i})$  to being in person  $\underline{i}$ 's position in state of the world  $\underline{S}$  and is bounded both above and below.<sup>3</sup> Because this function only represents a person's preferences over (state of the world, person's position) pairs in which he has preferences of type  $\underline{t}$ , I will refer to the preferences it represents as a person's 'limited preferences'.

Let  $\underline{v}^t$  be a Von Neumann and Morgenstern utility function defined on the set  $\underline{S} \times \underline{I} \times \{\underline{A}, \underline{B}\}$ . The function  $\underline{v}^t$  represents the preferences of a person with preference type  $\underline{t}$  over (state of the world, person's position, preference type) triples. It assigns a real number  $\underline{v}^t(\underline{S}, \underline{i}, \underline{t})$  to each triple  $(\underline{S}, \underline{i}, \underline{t})$  in the set  $\underline{S} \times \underline{I} \times \{\underline{A}, \underline{B}\}$  and is bounded both above and below. Because the function  $\underline{v}^t$  represents how a person with preference type  $\underline{t}$  would order a set of options that involve occupying various personal positions in various states of the world with

various preference types, I will refer to the preferences it represents as this person's 'extended preferences'.

To write that  $\underline{u}^t(\underline{C},i) > \underline{u}^t(\underline{D},i)$  means that, keeping his preference type fixed at type  $\underline{t}$ , a person with preference type  $\underline{t}$  will prefer occupying person  $i$ 's position in state of the world  $\underline{C}$  to occupying person  $i$ 's position in state of the world  $\underline{D}$ . To write that  $\underline{v}^t(\underline{C},i,\underline{A}) > \underline{v}^t(\underline{D},i,\underline{B})$  means that keeping his preference type fixed at type  $\underline{t}$ , a person with preference type  $\underline{t}$  will prefer occupying person  $i$ 's position in state of the world  $\underline{C}$  with preference type  $\underline{A}$  to occupying person  $i$ 's position in state of the world  $\underline{D}$  with preference type  $\underline{B}$ . By way of illustration, suppose Paul has artist's preferences. Suppose that in  $\underline{C}$ , Paul is a struggling artist, and in  $\underline{D}$ , he is a successful banker. Then  $\underline{u}^A(\underline{C},\text{Paul}) > \underline{u}^A(\underline{D},\text{Paul})$  means that, keeping his artist's preferences constant, Paul prefers being a struggling artist to being a successful banker. This preference will be relevant to his choice of career, so long as his choice of career does not change his preference type. By contrast,  $\underline{v}^A(\underline{C},\text{Paul},\underline{A}) > \underline{v}^A(\underline{D},\text{Paul},\underline{B})$  means that, from the perspective of his artist's preferences, Paul prefers being a poor artist with his current artist's preferences to being a rich banker with banker's preferences. This preference would determine, for example, his choice between going to art school (which, let us assume, will maintain his artist's preferences and lead to a career as a struggling artist) and going to business school (which, let us assume, will lead to a preference change to banker's preferences followed by a successful career in banking). We can imagine he holds this preference because he believes that a life devoted to art is superior to one that is not, no matter how successful the latter is. He therefore values having his artist's preferences to such a degree that he would not want to be rid of them and have them substituted by banker's preferences (which, let us suppose, involve a desire to compete and succeed in the world of high finance and no appreciation of artistic values), even at the cost of being poor and

unrecognised in the work he would do as an artist as opposed to wealthy and successful in the career he would choose if he were to acquire banker's preferences.

Both individuals' limited preferences and their extended preferences may differ. Suppose for simplicity that like Paul, Rob is a struggling artist in C and a successful banker in D. Suppose, further that Rob has banker's preferences, and that, keeping his banker's preferences constant, this means he would rather be a successful banker than a struggling artist:

$\underline{u}^B(\underline{C}, \text{Rob}) < \underline{u}^B(\underline{D}, \text{Rob})$ . In addition, suppose Rob is committed to the competitive values that underlie his preference for banking, so that he would not accept an opportunity to acquire artist's preferences, especially not at the cost of then having to live as a struggling artist, so that  $\underline{v}^B(\underline{C}, \text{Rob}, \underline{A}) < \underline{v}^B(\underline{D}, \text{Rob}, \underline{B})$ . In sum, in contrast to Paul, Rob believes it is worse to be a struggling artist with artist's preferences than to be a successful banker with banker's preferences.

It is worth noting that it is not necessarily the case that when individuals' preference types differ, both their limited and extended preferences differ. Two individuals with different preference types might have the same limited preferences, but different extended preferences, or the same extended preferences, but different limited preferences.

As an example of the former, consider the case of two gourmands who both enjoy exactly the same dishes to an equal extent: in environments in which they face only choices about what to eat, they will evaluate all options in exactly the same way, so that (at least in these environments) their limited preferences are the same. However, one of the two would prefer, if given the chance, to give up his taste for fine dining and develop a taste for music instead, so long as he would have an adequate opportunity to enjoy music with his new tastes. The other, by contrast, would not prefer to develop such tastes, so that their extended preferences are different.

As an example of the latter, consider two hedonists, who both rank all (state of the world, person's position, preference type) triples in the same way, viz. according to the pleasure they yield, so that their extended preferences are identical. However, one of them likes music, but takes no pleasure in eating, whereas the other takes no pleasure in music, but enjoys a good meal, so that their limited preferences differ.

Let us now turn to preference-based interpersonal comparisons of welfare. A preference-based standard of welfare involves making judgements about whether occupying Paul's position in state of the world  $\underline{C}$  with preference type  $\underline{A}$  is better, worse, or just as good as occupying Rob's position in state of the world  $\underline{D}$  with preference type  $\underline{B}$ . More precisely, let the function  $\underline{w}$  be a Von Neumann and Morgenstern utility function representing this public standard of welfare. The function  $\underline{w}$  then assigns a real number  $\underline{w}(\underline{S}, \underline{i}, \underline{t})$  to each triple  $(\underline{S}, \underline{i}, \underline{t})$  in the set  $\underline{S} \times \underline{I} \times \{\underline{A}, \underline{B}\}$  and is bounded both above and below. This  $\underline{w}(\underline{S}, \underline{i}, \underline{t})$  stands for the value of occupying person  $\underline{i}$ 's position in state of the world  $\underline{S}$  with preference type  $\underline{t}$ . In attempting to determine these values with reference to a person's degree of preference satisfaction, we face two questions. First, whether we should use the degree of satisfaction of a person's limited or extended preferences in determining his welfare. Second, how we should evaluate options that involve preference change. I address the first of these questions in the next section, and the second one in section 3.

## **2. Extended preferences and welfare**

As mentioned, people typically do not just care about having the world conform to their preferences; they also care about which values, aims, attachments, and therefore preferences they have. This interest in having the preferences one wants to have is, moreover, an important one. Considering people who could be said not to care, or not to care deeply, about their values and aims can illustrate this importance.

Consider first what the life would be like of someone who was completely indifferent about his values and aims. This person's life would be devoid of the kind of commitments and relationships that are a central part of most people's lives. This is evident in cases of commitments to moral ideals: being committed to a cause like eradicating world poverty, for example, involves more than having a particular pattern of desires connected to that cause, such as that it should be realised; it also involves wanting to maintain one's desire for its realisation. But is it also a feature of other commitments that are a central part of people's identity. Being committed to being an artist, for example, does not just involve trying to succeed as an artist, but also to want to maintain and develop one's appreciation of art.

Furthermore, deep friendship involves not just caring about one's friend, enjoying her company, and being ready to help her out when she needs help, but also being prepared to take steps to maintain these attitudes towards her. Similarly, being a loving partner involves not just desiring to share one's life with one's partner, desiring to see him do well, etc. but also to actively maintain and, when necessary, reinforce these desires (see Frankfurt (2004)).<sup>4</sup> In these cases, a person does not just desire to have certain preferences because having them would be instrumental to some other end that she has (eradicating world poverty, being a successful artist, furthering the welfare of one's friend or lover), but also because she regards these as the right preferences for her to have.

More generally, the life of someone who was completely indifferent about his values and aims would be devoid of a particular kind of agency: action directed not merely at shaping his environment to satisfy his desires, but also at shaping himself, in the sense of shaping his values and aims (see Frankfurt (1982, p. 83)). As a consequence, if his life showed any unity of purpose, it would not be the product of any action on his part intended to give his life any particular direction, but rather the product of causes in which he played no active part, or the unintended by-product of his actions.



Similar conclusions apply in the case of a person who, though she has preferences about which preferences she has, ranks her present and potential preferences only on the grounds of the ease with which they can be satisfied (see Dworkin 2000, pp. 292-293). Such a person would also lack the particular attitudes necessary for being substantively committed to any particular cause, relationship, or set of values. As a consequence, she would not purposefully shape her life and herself in accordance with the demands of such commitments.

In sum, the interest in having the preferences one wants to have is essentially the interest in one's ability to shape oneself and one's life in accordance with the demands of the causes, values and relationships to which one is devoted. Given the importance of this interest, we should attempt to base a preference satisfaction measure of welfare on the satisfaction of a person's extended preferences, since these represent both a person's interests in his situation and his interests in his preferences.

### **3. Preference change and the degree of satisfaction of a person's extended preferences**

Let us now turn to a method for determining the degree of satisfaction of a person's extended preferences. Suppose there are four states of the world: one in which Paul is a struggling artist (C), one in which he is a successful banker (D), one in which he is an unsuccessful banker (E), and one in which he is a successful artist (F). For any Von Neumann and Morgenstern utility function, we are free to fix the zeros and units.<sup>5</sup> Once we do so, the utilities of all states of the world are fixed. In order to be able to interpret the number  $v^i(\underline{S}, i, \underline{t})$  as the degree to which occupying person  $i$ 's position in state of the world  $\underline{S}$  with preference type  $\underline{t}$  satisfies the extended preferences of someone with preference type  $\underline{t}$ , we therefore proceed as follows. We set the value of what, from the perspective of type  $\underline{t}$  is the best possible (state of the world, person's position, preference type) triple to one, and the worst triple to zero. For example, suppose that, from the perspective of his current artist's preferences, Paul would consider

being a struggling artist with banker's preferences the worst possible triple, and being a successful artist with artist's preferences the best possible triple. We then take

$$\underline{v}^A(\underline{C}, \text{Paul}, \underline{B})=0, \text{ and } \underline{v}^A(\underline{F}, \text{Paul}, \underline{A})=1.$$

The value of all other (state of the world, person's position, preference type) triples will then be determined as follows. We then take the number assigned to any other triple to be equal to the probability  $p$  that would render Paul indifferent between accepting that triple and a lottery with probability  $(1-p)$  of ending up in his position in  $\underline{C}$  with preference type  $\underline{B}$  and probability  $p$  of ending up in his position in  $\underline{F}$  with preference type  $\underline{A}$ . In this way, each (state of the world, person's position, preference type) triple is assigned a number between zero and one, which we can call the degree to which this triple satisfies Paul's current type  $\underline{A}$  extended preferences over (state of the world, person's position, preference type) triples. For example, if with artist's preferences Paul would be indifferent between being a struggling artist with artist's preferences and a lottery with a probability of 0.2 of ending up in his position in  $\underline{C}$  with preference type  $\underline{B}$  and a probability of 0.8 of ending up in his position in  $\underline{F}$  with preference type  $\underline{A}$ , then  $\underline{v}^A(\underline{C}, \text{Paul}, \underline{A})=0.8$ .

Now, we face a difficulty in deciding how to move from the degree to which each triple would satisfy Paul's current extended preferences to an assessment of how well off he would be if each of these triples were realised. The difficulty is that if we assess each triple by Paul's current, type  $\underline{A}$  extended preferences, this assessment may differ from his own assessment of these triples once he is in the situation characterised by this triple. For when these triples involve a preference change to preference type  $\underline{B}$ , then though Paul will now regard this change as making him worse off, once he has preferences of type  $\underline{B}$ , he may regard this change in preferences as making him better off. For example, from the perspective of his current, artist's preferences, he might assign the situation in which he is a successful banker with banker's preferences a value of 0.3:  $\underline{v}^A(\underline{D}, \text{Paul}, \underline{B})=0.3$ . But if he ended up in this

situation through a process of preference change that, from the perspective of his new preferences, he does not regard as in any way a bad one to have undergone, then we may suppose that from the perspective of his new preferences, he would assess this situation as the best possible one:  $v^B(\underline{D}, \text{Paul}, \underline{B})=1$ . The question is, then, whether we should take Paul's pre-preference change, or post-preference change evaluation as determining his welfare in such cases.

Before we attempt to deal with this question, we should note that in order to assess the impact of a change in a person's extended preferences on his welfare, it is important to assess the conditions under which it takes place. If the preference change was a result of the subversion of Paul's cognitive capacities, or of coercion or oppression, or was a response to an unduly limiting environment, then this might discredit Paul's post-preference change view of his own welfare. I will assume, however, that all preference changes under consideration are not the result of the subversion of a person's cognitive capacities, of coercion, oppression, or unduly limiting circumstances. Preference changes of this kind can occur throughout people's lives; one might, for example, have artist's preferences and through contact with one's friends or one's social environment, or simply through the passage of time, find one's preferences changed to banker's preferences. I will also assume that from the perspective of preference types A and B, having had one's preferences change in this way is not viewed as a bad or a good thing in itself.

Now, the possibility of this change in an individual's evaluation of a particular (state of the world, person's position, preference type) triple means we have two possible ways of judging an individual's welfare by the degree of satisfaction of his extended preferences. The first is to equate the welfare level of each triple with the degree of satisfaction of the extended preferences that he has in that triple. The second is to equate the welfare level of each triple

with the degree to which this triple satisfies his current extended preferences. I will discuss each method in turn.

The first method involves using the extended preferences of type A to evaluate a situation that involves Paul occupying his position in a state of the world with preference type A, and the extended preferences of preference type B to evaluate a situation that involves Paul occupying his position in a state of the world with preference type B. This would mean taking  $\underline{w}(\underline{C}, \text{Paul}, \underline{A})$  to be equivalent to  $\underline{v}^{\underline{A}}(\underline{C}, \text{Paul}, \underline{A})$  and  $\underline{w}(\underline{C}, \text{Paul}, \underline{B})$  to be equivalent to  $\underline{v}^{\underline{B}}(\underline{C}, \text{Paul}, \underline{B})$ , and so on.

Doing so means that at every point in time, our standard of welfare will agree with each individual's own assessment of his welfare at that time. Moreover, this standard of welfare will always respect each individual's preferences over options that do not involve changes in his preferences. However, this standard will not always agree with an individual's preference change assessment of the value of options that involve preference change. For this measure will count a change from a situation in which Paul is a struggling artist with artist's preferences to a situation in which he is a successful banker with banker's preferences as an improvement in Paul's welfare, since the degree of satisfaction of the extended preferences he has after the change is larger than the degree of satisfaction of his extended preferences before the change:  $\underline{w}(\underline{C}, \text{Paul}, \underline{A}) = \underline{v}^{\underline{A}}(\underline{C}, \text{Paul}, \underline{A}) = 0.8 < \underline{w}(\underline{D}, \text{Paul}, \underline{B}) = \underline{v}^{\underline{B}}(\underline{D}, \text{Paul}, \underline{B}) = 1$ . But from the perspective of his current, artist's preferences, Paul will disagree with this judgement.

It follows that this measure does not adequately represent Paul's interest in having the preferences he wants to have: it will not consider the goods and conditions that enable him to sustain his preferences, or develop them in the direction he wants, as contributing to his welfare, unless his sustaining or developing these preferences will contribute to a higher degree of satisfaction of whatever preferences he ends up having. For example, so long as Paul has artist's preferences, this measure will regard the resources and conditions that help

him sustain or reinforce his artist's preferences as of less value to him than the resources and conditions that would lead him to develop banker's preferences, when the latter could be more easily satisfied. Given the importance of the interest in shaping one's tastes, values and aims in the direction one wants, this represents a severe drawback of this version of the satisfaction measure of welfare.

The second method assesses each (state of the world, person's position, preference type) triple from the perspective of his current preferences. To illustrate this method, suppose again that Paul's current extended preferences are those that go with type A. We regard these extended preferences as determining the welfare of all future (occupying Paul's position in a state of the world, preference type) triples. We would then take  $\underline{w}(C, \text{Paul}, A)$  to be equivalent to  $\underline{v}^A(C, \text{Paul}, A)$ ,  $\underline{w}(C, \text{Paul}, B)$  to be equivalent to  $\underline{v}^A(C, \text{Paul}, B)$ , and so on.

This method obviously represents Paul's current interests in his future preferences. However, it does so at the cost of not always respecting Paul's post-preference change extended preferences. Suppose Paul's preferences at time 0 are artist's preferences. Suppose further that we take the degree of satisfaction of his extended preferences at time 0 as the measure of his current and future welfare. Then we will evaluate a change from a situation at time 0 in which he is a struggling artist to a situation in which at time 1 he is a successful banker as making him worse off. As noted, after this change, Paul will disagree: at time 1 he will regard his new situation as the best possible one. Now, suppose this change does take place, and at time 1 we can present Paul with an opportunity to change his preferences back to artist's preferences at time 2. From the perspective of Paul's extended preferences at time 0, this would be an opportunity to increase his welfare. But from the perspective of his extended preferences at time 1, taking this opportunity would make him worse off. By making his preferences as time 0 normative throughout these periods, we would not be respecting his judgements at time 1. We thus can represent Paul's interests in his future preferences at time 0

only at the cost of disregarding his extended preferences at time 1 (including, of course, his interests in his future preferences at that time).

Now, in some cases of preference change, we might have reason to regard a person's initial preferences in sequences of this kind as normative; an example might be a case in which the preference change between time 0 to time 1 was a result of the subversion of Paul's cognitive capacities, or of coercion or oppression, or was a response to an unduly limiting environment. But we have assumed that the process of preference change was not of this sort. In such cases, it does not appear appropriate to judge Paul's welfare at time 1 from the perspective of his very different extended preferences at time 0: this would be judging his welfare by a set of values that he no longer holds.

Though I cannot discuss all possible methods of dealing with the case of preference change that fall within the family of possible preference satisfaction measures, it seems that all possibilities that involve a compromise between these two approaches will suffer from some combination of the drawbacks of these two methods. For example, consider determining a person's welfare in a given period by the degree of satisfaction of a weighted average of the extended preferences he has over that period, with the weights determined by the relative amount of time he holds certain preferences.<sup>6</sup> This would involve both limiting the degree to which a standard of welfare represents a person's interest in his future preferences, and assessing his welfare at some points in time by a set of values which he no longer holds. It would, for example, imply that if Paul had artist's preferences for 30 years, and then developed banker's preferences later in life, then (re)developing his taste for art by enrolling in evening classes of art appreciation would improve his welfare even if he had his banker's preferences for 10 years, and saw no value at all in taking these classes.

In sum, it seems that any preference satisfaction measure will have significant drawbacks, since it will involve either neglecting individuals' current interests in shaping their future

preferences, or, in their future, judging their welfare by a set of values which might be very alien to them.<sup>7</sup>

#### **4. A substantive conception of welfare?**

This conclusion should, I believe, prompt us to consider alternative measures of welfare. In closing, I would like to outline one approach that strikes me as worth pursuing. This is to construct a public conception of welfare from a list of goods and conditions that can be recognised by people with divergent values as generally important constituents of a good life (see also Scanlon 1991). Some of its constituent elements will be broad categories that can be realised in different ways by people with different values, such as developing one's capacities, leading the life one wants with family and friends, job satisfaction, and achieving success in one's main endeavours. It will also include more specific goods like health, leisure, and wealth, which are generally judged to be important elements of a good life. Our discussion also teaches us that this list should include goods and conditions that generally enable people to maintain or develop the tastes, values, and aims they want to have.

By including categories, goods and conditions that are valuable from the perspective of different views of the good life, this list accommodates both the diverse views of the good life that are held within a population, and the different views of the good life that a person may have during his lifetime. The standard objection to a conception of welfare of this kind is that by using the same list of goods and conditions to assess everyone's welfare, and by using the same weights for these goods and conditions in each person's case, it fails to fully respect each person's view of their own welfare. The force of this objection depends on the assumption that there is an alternative measure of welfare—the preference satisfaction measure—which does fully respect each person's view of their own good. But our discussion makes clear that given the possibility of change in a person's extended preferences, no form

of the preference satisfaction measure can fully respect each individual's judgements of her own welfare. It may be, therefore, that a substantive conception of welfare of this kind meets our first requirement (to respect, to the greatest reasonable extent, each person's view of their own welfare) because this just is the furthest we can go towards respecting each person's view of their own good.

A standard of welfare of this kind also respects our second requirement of neutrality of judgement (see Scanlon 1991, pp. 39-40 and Otsuka 2003, pp. 110-112). For, in attempting to accommodate to the greatest extent possible different views of the good life, it is constructed without the assumption that any particular conception of the good or set of conceptions of the good is the right one to the exclusion of others. In sum, the difficulties that the possibility of change in a person's extended preferences causes for a preference satisfaction conception of welfare render a substantive conception of welfare more attractive.

## References

- Arneson, R. (1990a) 'Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare', in Philosophy and Public Affairs 19, pp. 158-194.
- Arneson, R. (1990b) 'Primary Goods Reconsidered' in Noûs 24, pp. 429-454.
- Arneson, R. (1990c) 'Neutrality and Utility' in Canadian Journal of Philosophy 20, pp. 215-240.
- Binmore, K. (1991) Fun and Games. Lexington, Mass: D.C. Heath.
- Brandt, R. (1979) A Theory of the Good and the Right. Oxford: Oxford University Press.
- Dworkin, R. (2000) Sovereign Virtue. Cambridge: Harvard University Press.
- Frankfurt, H. (1982) 'The Importance of What We Care About', reprinted in his The Importance of What We Care About, Cambridge: Cambridge University Press, pp. 80-94.
- Frankfurt, H. (2004) The Reasons of Love. Princeton: Princeton University Press.



Otsuka, M. (2003) Libertarianism Without Inequality. Oxford: Oxford University Press.

Scanlon, T. (1991) 'The Moral Basis of Interpersonal Comparisons', in Interpersonal Comparisons of Well-Being, eds. Jon Elster and John Roemer. Cambridge: Cambridge University Press, pp. 17-44.

Voorhoeve, A. (2003) 'Harry Frankfurt on the Necessity of Love', in Philosophical Writings 23, pp. 55-70.

## Footnotes

<sup>1</sup> Earlier versions of this paper were presented at the Royal Institute of Philosophy conference on Preference Formation and Well-Being at St. John's College Cambridge in July 2004, in the LSE Choice Group Seminar in January 2005, and the ECAP 5 Workshop on Philosophy, Economics, and Public Policy in Lisbon in August 2005. I am grateful to those present at these meetings, and especially to Richard Arneson, Luc Bovens, Marc Fleurbaey, Christian List, Andrew Williams, and Jo Wolff for their comments. I also thank Richard Bradley, Roger Crisp, Michael Otsuka, and Serena Olsaretti for detailed comments on earlier versions of this paper. The Analysis Trust and the AHRB supported my work on this paper.

<sup>2</sup> See also Otsuka (2003, pp. 110-112).

<sup>3</sup> The fact that the utility function is bounded both above and below means that there exists some numbers  $\underline{a}$  and  $\underline{b}$  such that  $\underline{a} \leq u^i(S, i) \leq \underline{b}$  for each  $S$  in  $\underline{S}$ . This means that in no case is being in person  $i$ 's position ascribed a utility of negative or positive infinity. This assumption avoids certain decision-theoretic paradoxes that arise when utilities of negative or positive infinity are permitted. See Binmore (1991).

<sup>4</sup> See also Voorhoeve (2003) for a discussion with Harry Frankfurt of his views on love.

<sup>5</sup> For an introductory discussion of Von Neumann and Morgenstern utility functions and their properties, see Binmore (1991, chapter 3).

<sup>6</sup> Something akin to this possibility, though without the use of the distinction between limited and extended preferences, is discussed by Richard Brandt (1979, pp. 247-253) and Arneson (1990a, pp. 162-167). See also the following footnote.

<sup>7</sup> It may be of interest to note how my discussion of the difficulties which preference change creates for a preference satisfaction measure differs from Brandt's. Brandt (1979) also argues that the possibility of preference change undermines the case for a preference satisfaction measure of welfare. His discussion differs from mine, however, in not making use of the

distinction between a person's limited and extended preferences. This distinction is, I believe, crucial to the problem. For not just any change in a person's preferences generates a problem for a preference satisfaction metric. A change in a person's limited preferences is not sufficient to undermine the preference satisfaction measure; what is required is a change in a person's extended preferences. To see this, consider the case of a hedonist, who ranks all (state of the world, person's position, preference type) triples on the basis of the pleasure they yield for her. Suppose that, at time 0, she wants to celebrate her birthday at time 2 with a dinner in her favourite fish restaurant rather than in a steak house, since her current limited preferences are for fish over meat, and she does not expect to undergo a preference change between now and time 2. However, suppose she does undergo a change in her limited preferences, so that at time 2 she prefers to dine in a steak house. So long as this is a change in her limited preferences only, we have no difficulty assessing the welfare of these two options. For we can assess the welfare associated with the four options (eating fish at time 2 with a taste for fish), (eating meat at time 2 with a taste for fish), (eating fish at time 2 with a taste for meat), and (eating meat at time 2 with a taste for meat) from the perspective of her unchanged extended preferences, that is to say, by equating the welfare of each option with the pleasure it yields.