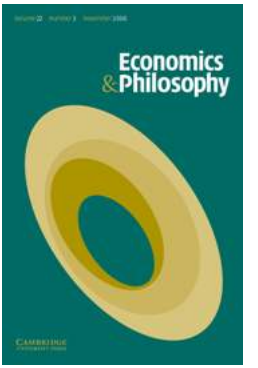

UTILITARIANISM AND PRIORITARIANISM I

DAVID McCARTHY

Economics and Philosophy / Volume 22 / Issue 03 / November 2006, pp 335 - 363
DOI: 10.1017/S0266267106001015, Published online: 11 December 2006

Economics and Philosophy, **22** (2006) 335–363
doi:10.1017/S0266267106001015

UTILITARIANISM AND PRIORITARIANISM I

David McCarthy
University of Edinburgh

Utilitarianism and prioritarianism make a strong assumption about measures of how good lotteries over histories are for individuals, or for short, individual goodness measures. Given some idealizing assumptions about interpersonal and intrapersonal comparisons, they presuppose that any individual goodness measure can be transformed into any other individual goodness measure by a positive affine transformation. But it is far from obvious that the presupposition is correct, so both theories face the threat of presupposition failure. The usual response to this problem starts by assuming that what implicitly determines the set of individual goodness measures is independent of our discourse about utilitarianism and prioritarianism. I suggest reversing this response. What determines the set of individual goodness measures just is the body of platitudes we accept about utilitarianism and prioritarianism. This approach vindicates the utilitarian and prioritarian presupposition. As a corollary, it shows that individual goodness measures are expectational, and provides an answer to an argument due to Broome that for different reasons to do with measurement, prioritarianism is more or less meaningless.

The textbooks say that utilitarianism tells us to maximize the sum of individual goodness. But a problem lurks. For the definition turns out to presuppose that there is a more or less unique arithmetical measure of how

I owe far more than the usual debt of gratitude to John Broome, Wlodek Rabinowicz and Peter Vallentyne for extensive comments and advice about this and related work. Their generosity has been remarkable. I have also benefited greatly from comments by Bertil Tungodden and two anonymous referees. Support for this project was provided by the Arts and Humanities Research Council, and also by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research and the Program for the Investment in the Future of the German Government through a Sofja Kovalevskaja Award in the form of a Visiting Fellowship at the Philosophy, Probability and Modeling Group at the University of Konstanz.

good histories, or lotteries over histories, are for individuals. But it looks very much as if there is no such measure. And if not, utilitarianism suffers from presupposition failure, and turns out to be more or less meaningless. And so do its main competitors, like prioritarianism.

By reformulating a famous theorem of Harsanyi (1955), Broome (1991) claims to have solved the problem for utilitarianism. We do indeed have a more or less unique arithmetical measure of individual goodness. But a surprising byproduct of his argument is his claim that for somewhat different reasons, prioritarianism is still more or less meaningless. Prioritarianism, Broome tells us, distinguishes between individual goodness and how much individual goodness contributes towards overall goodness. But Broome claims that this is a meaningless distinction. For helpful elaboration on Broome's argument, see Jensen (1995).

Broome (2004) withdraws a major component of the solution to the problem for utilitarianism. Since the argument against prioritarianism in Broome (1991) was a corollary to that solution, it is not so clear how much trouble for prioritarianism remains. However, there is no doubt that Broome has identified an important issue. For even if we set aside the threat of presupposition failure, it is hard to believe that we understand utilitarianism or prioritarianism well if we do not understand arithmetical measures of individual goodness.

I am not going to say much more about Broome's approach to the problem for utilitarianism. It is complex, and I have offered my criticisms of it in McCarthy (forthcoming). Here I try to be more constructive and develop a different approach. Not only does this approach solve the problem for utilitarianism, but it simultaneously solves the problem Broome (1991) thinks prioritarianism is faced with. Moreover, these solutions turn out to be mutually reinforcing.

Section 1 outlines a framework for expressing interpersonal and intrapersonal comparisons. Section 2 explains why utilitarianism and prioritarianism appear to suffer from presupposition failure. Section 3 suggests a natural response to the threat of presupposition failure. Sections 4 through 7 develop this response, and argue that in the end utilitarianism and prioritarianism do not suffer from presupposition failure. To this point I will have ignored Broome's (1991) criticisms of prioritarianism. But section 8 shows how the response to the threat of presupposition failure provides a reply, and also addresses Sen's (1976, 1977, 1986) influential views about the relationship between utilitarianism and Harsanyi's theorem.

1. INTERPERSONAL AND INTRAPERSONAL COMPARISONS

The problem utilitarians and prioritarians face will be easier to discuss if we have a framework for expressing interpersonal and intrapersonal

comparisons, and in this section I outline a fairly standard approach. This article is only concerned with what utilitarianism and prioritarianism say about constant populations, so the population of individuals is assumed to be fixed throughout.

We are going to be concerned throughout with lotteries over entire world histories. The usual way of expressing interpersonal and intrapersonal comparisons is to say such things as: a lottery L_1 is better for an individual i than L_2 is for j. When i and j are different individuals, that expresses an interpersonal comparison. When they are the same individual, it expresses an intrapersonal comparison. But it turns out to be much easier to group the individual and lottery into a pair. Thus the *better life lottery relation* holds between (i, L_1) and (j, L_2) just in case L_1 is at least as good for i as L_2 is for j.

It is often thought that there exist values which are very different, and that as a result, there can be two lives such that it is not the case that one life is better than the other, and not the case that the two lives are equally good. But while making room for this possibility does nothing to solve the problems about utilitarianism and prioritarianism this article is concerned with, it does make them harder to discuss. So I am going to idealize this problem away. More precisely, I will assume that the better life lottery relation is *complete*: for any individuals i and j and any lotteries L_1 and L_2, either L_1 is at least as good for i as L_2 is for j, or L_2 is at least as good for j as L_1 is for i.

Completeness is one of the axioms of expected utility theory. It will be easier to say what the rest of those axioms are in a different context. But once we accept the assumption that the better life lottery relation is complete, it is very plausible that the better life lottery relation satisfies all of the axioms of expected utility theory (Hammond 1991; Broome 2004). To say what the formal upshot of this is, we need a concept which will recur throughout the article.

> A real valued function v *represents* a binary relation R just in case: for all x and y: Rxy if and only if $v(x) \geq v(y)$

A function which represents a relation is really nothing more than a mathematical description of that relation. But even though it is the relation we are normally interested in, it is often easier to work with a function which represents it rather than work directly with the relation itself.

The formal significance of the claim that the better life lottery relation satisfies the axioms of expected utility theory is this. There exists a function $u(i, L)$ with the following two properties. First, it represents the better life lottery relation. In other words,

> For all individuals i and j, and all lotteries L_1 and L_2: L_1 is at least as good for i as L_2 is for j if and only if $u(i, L_1) \geq u(i, L_2)$.

Second, for any individual i, and any lottery of the form $L = [p_1, h_1; p_2, h_2; \ldots; p_m, h_m]$ where every possible history h_j under L has a corresponding positive probability p_j and all the probabilities sum to one,

$$u(i, L) = p_1 u(i, h_1) + p_2 u(i, h_2) + \cdots + p_m u(i, h_m)$$

This last condition means that $u(i, L_1)$ is what is known as an *expectational* function, because the value it gives to L from i's point of view is equal to the value it gives to h_1 from i's point of view, the value it gives to h_2 from i's point of view and so on, each multiplied by its corresponding probability then all added up. From now on, $u(i, L)$ will always be this function.

An assumption which I might as well state here will simplify the discussion to enable us to focus more sharply on the philosophical problems. Anyone not interested in technicalities can ignore it. Very roughly, I will assume some lives are better than others, that any individual can lead any possible kind of life, and that all combinations of possible kinds of lives are possible. In technical terms, let I be the image of $u(i, L)$. Then I am assuming that I is an interval of real numbers of positive length, and that for any member $[x_1, x_2, \ldots, x_n]$ of the Cartesian product I^n, there exists a lottery L such that for every member of the population $i = 1 \ldots n$, $u(i, L) = x_i$. This assumption is very similar to what Broome (1991) calls the *rectangular field assumption*, and it is not misleading to use the same name for it. Broome outlines the philosophical questions it raises, but like him, I am simply going to take it for granted. It guarantees that the domain of the betterness relation is simple enough for us to be able to exploit some powerful theorems.

2. THE THREAT OF PRESUPPOSITION FAILURE

In this section I begin by identifying a presupposition utilitarianism and prioritarianism share. The natural way to do this is to derive it from the way they are standardly formulated. But this forces an immediate choice upon us.

We are going to be concerned with utilitarian and prioritarian accounts of the *betterness relation*. The betterness relation holds between two lotteries L_1 and L_2 just in case L_1 is at least as good as L_2. But there are two sorts of things to take into account in thinking about a lottery over histories: people and uncertainty.

One approach deals with people first. Very roughly, it starts by trying to combine the points of view of all the different people to assess what each possible history is like. That provides some sort of evaluation of what each history is like as a whole. Then it deals with uncertainty. Given the evaluation of what each history is like, it then takes into account how likely each different history is to form an on balance evaluation of the lottery as a whole. Economists often call this an *ex post* approach to aggregation.

The other approach deals with uncertainty first. Very roughly, it starts by trying to combine all the different histories from the point of view of each person, taking into account how likely each different history is. That provides some sort of evaluation of the lottery as a whole from the point of view of each different person. Then it deals with people. Given the evaluations of the lottery from the point of view of each person, it then forms an on balance evaluation of the lottery as a whole. Economists often call this an *ex ante* approach to aggregation.

Ex ante and *ex post* approaches to aggregation are not in general equivalent. Given some sort of distributive idea, such as giving priority to the worse off, we could apply it to people first, then deal with uncertainty, as the *ex post* approach recommends. Or we could deal with uncertainty first, and then apply it to people, as the *ex ante* approach recommends. It turns that for almost all distributive ideas, it matters what the order is. *Ex ante* applications lead to answers which conflict with *ex post* applications (Hammond 1981).

I will say more about this in section 7. I point out the conflict now only to be upfront about how one's initial formulation of a distributive view in the context of uncertainty is almost bound to take sides. With the benefit of hindsight it turns out to be easier to deal with *ex ante* views first. So in their spirit, I am going to start off by assuming that there is a unique measure $g(i, L)$ of how good lotteries are for individuals or, as I will say, an *individual goodness measure*. The function g takes an individual and a lottery and gives a real number which is supposed to be a measure of how good the lottery is for the individual. We are going to be much concerned with what conditions a function has to satisfy to be an individual goodness measure, but for now one thing is obvious. To be an individual goodness measure, a function has to represent the better life lottery relation.

We can now define *ex ante* utilitarianism. Throughout the population of individuals $i = 1 \ldots n$ is assumed to be constant and to contain at least two people, and $\sum_i$ expresses summation over those individuals.

> ***Ex ante* utilitarianism** For all lotteries L_1 and L_2, L_1 is at least as good as L_2 if and only if $\sum_i g(i, L_1) \geq \sum_i g(i, L_2)$

To define *ex ante* prioritarianism we need some terminology. A function f is *increasing* just in case for all x and y in its domain, $x > y \Rightarrow f(x) > f(y)$. And f is *strictly concave* just in case for all distinct x and y in its domain, $f\left(\frac{x+y}{2}\right) > \frac{1}{2}(f(x) + f(y))$. For example, the function $\sqrt{x}$ on the non-negative reals is both increasing (it gets larger as x gets larger) and strictly concave (because it gets flatter as x gets larger). The following definition is standard.

> ***Ex ante* prioritarianism** There is an increasing and strictly concave function w such that for all lotteries L_1 and L_2, L_1 is at least as good as L_2 if and only if $\sum_i w(\mathrm{g}(i, L_1)) \geq \sum_i w(g(i, L_2))$

Parfit (1991) says that the priority view departs from utilitarianism in exactly one way: it claims that benefiting the worse off matters more. To show how the definition of *ex ante* prioritarianism captures this idea, it will help first to formalize the idea. A *Pigou-Dalton transfer of individual goodness* is a transfer which leaves all but two people unaffected, makes the better off of those two people A worse off by some amount of individual goodness and makes the worse off of those two people B better off by the same amount of individual goodness while still leaving A better off than B originally was. So Pigou-Dalton transfers of individual goodness preserve the total sum of individual goodness, but make two individuals closer to being equally well off than they were to start with.

We can now express Parfit's idea as follows: Pigou-Dalton transfers of individual goodness are always improvements. Or, if a lottery L_1 can be obtained from a lottery L_2 by a Pigou-Dalton transfer of individual goodness, then L_1 is better than L_2. It is obvious that *ex ante* utilitarianism is always indifferent to Pigou-Dalton transfers of individual goodness. The only difference between the formulae in the definitions of the *ex ante* versions of utilitarianism and prioritarianism is the presence of w in the definition of prioritarianism. But it easily follows from the definition of an increasing and strictly concave function that *ex ante* prioritarianism always regards Pigou-Dalton transfers of individual goodness as improvements. So the definition of *ex ante* prioritarianism captures Parfit's idea.

It is easy to see that the above formulations of utilitarianism and prioritarianism are *ex ante* views. For according to the formula in each formulation, the first thing that happens in the assessment of a lottery L is that it is converted into a number $g(i, L)$ for each individual i that measures how good the lottery is for i. But to save words, I will drop '*ex ante*' until section 7.

The next question is about how to define utilitarianism and prioritarianism if there exists more than one individual goodness measure. But before addressing this I need to say what it means to say that there is more than one individual goodness measure. In one way, the answer is obvious: there simply happen to be two or more functions which satisfy the predicate 'is an individual goodness measure'. But the real questions are about how that could happen, and if it does happen, how we are to interpret sentences which speak as if there is just one individual goodness measure.

That it is not uncommon for there to be two or more measures of the same thing is clear. We can measure height in inches or in centimeters, for example. It is arbitrary which of these measures of height we choose to use. And that suggests how there can be more than one measure: each of the measures makes what is in some sense an arbitrary choice. In the case of the height measures, the arbitrariness lies in the choice of unit. And that suggests an answer to the question about interpretation. Given a

sentence which speaks as if there is just one measure when in fact there is more than one, to determine its truth value we have to ignore the arbitrary elements in the measures. One standard approach is to *superevaluate* (van Fraassen 1966). To determine the truth value of a sentence like 'The height of this building is twice the height of that building' we first have to work out the truth value of the sentence relative to each measure in the entire class of height measures. According to the method of superevaluation, if the sentence is true relative to each, then it is true. If the sentence is false relative to each, then it is false. And if it is true relative to some and false relative to others, then it is neither true nor false. Other approaches might in this third case say that the truth value of the sentence is indeterminate, but the differences in detail between these approaches are not important here.

Standard formulations of utilitarianism and prioritarianism speak as if there is just one individual goodness measure. Utilitarianism, for example, is said to tell us to maximize the sum of individual goodness. And I will assume that we should cater for the possibility that there is more than one individual goodness measure by superevaluating. But given the idealizing assumption that the better life lottery relation is complete, I know of no suggestion in the literature that utilitarianism and prioritarianism could suffer from truth value gaps. Given superevaluation, the absence of truth value gaps means that utilitarianism and prioritarianism presuppose that if g and $\tilde{g}$ are distinct individual goodness measures, then the deliverances of utilitarianism and prioritarianism do not depend on which one of g or $\tilde{g}$ is used. More precisely, utilitarianism, for example, presupposes

(1) There exists an individual goodness measure $g(i, L)$, and for all individual goodness measures $\tilde{g}(i, L)$ and all lotteries L_1 and L_2, $\sum_i g(i, L_1) \geq \sum_i g(i, L_2)$ if and only if $\sum_i \tilde{g}(i, L_1) \geq \sum_i \tilde{g}(i, L_2)$.

But this condition is quite complicated, and it would be easier to understand what utilitarianism presupposes if we could find a simpler statement about individual goodness measures which is logically equivalent to it.

As it happens, there is one. But the rigorous statement and full proof of this result is unfortunately quite long and technical. But it is important and the theorem it exploits will be used several times in what follows, so I will give a brief sketch.

I begin with a simplifying assumption. I assumed that the better life lottery relation satisfies the axioms of expected utility theory. In part that means I assumed that it is what is known as a continuous relation, and I am now going to assume that the betterness relation is continuous.

To explain, for any two lotteries over histories L and M and any α such that $0 < \alpha < 1$ we can define the *compound lottery* $\alpha L + (1 - \alpha)M$ as the

lottery which scales down L's probabilities by α and M's by $1-\alpha$ and then combines them. For example, if L gave an history h a probability p and M gave h a probability q, then the probability of h under the compound lottery is $\alpha p + (1-\alpha)q$. Then the betterness relation is *continuous* just in case for any lotteries L, M, and N, if L is better than M, and M is better than N, then there exist α and β, each strictly between 0 and 1 such that $\alpha L + (1-\alpha)N$ is better than M, and M is better than $\beta L + (1-\beta)N$.

Assuming that the betterness relation is continuous greatly simplifies the mathematics. But it is a mathematical idea which is not easy to explain informally. But, very roughly, to say that the betterness relation is continuous means that no history is infinitely better or merely infinitesimally better than any other. This assumption might not be very plausible if one was dealing with infinitely long lives, for example. An eternity in heaven might be infinitely better than one day in heaven. But if we set such scenarios aside, assuming that the betterness relation is continuous is reasonable and enables us to exploit an important theorem.

Suppose that g and $\tilde{g}$ are individual goodness measures. Then (1) entails that $\sum_i g(i, L)$ and $\sum_i \tilde{g}(i, L)$ both represent the same relation. But notice that each of these has an additive form. For example, $\sum_i g(i, L)$ is just $g(1, L) + g(2, L) + \ldots + g(\mathrm{n}, L)$. For this reason, $\sum_i g(i, L)$ and $\sum_i \tilde{g}(i, L)$ are what are known as additive representations.

But there is a very important theorem which says, roughly, that additive representations are essentially unique. If there are two additive representations of the same relation, then each of the additive representations can be obtained from the other by what is known as a *positive affine transformation*. A positive affine transformation is a transformation of the form $x \mapsto ax + b$ where $a > 0$ and b are real numbers. Broome (1991) provides an informal discussion.

I have left out some technicalities. But it turns out that with all the assumptions that have been made so far, we have all the bits and pieces needed to exploit this theorem rigorously. And the theorem shows that with our assumptions, (1) entails that

(2) There exists an individual goodness measure $g(i, L)$, and all individual goodness measures are positive affine transformations of $g(i, L)$.

In other words, the second clause tells us that if $\tilde{g}$ is an individual goodness measure, then there exist real numbers $a > 0$ and b such that for all individuals i and lotteries L, $\tilde{g}(i, L) = ag(i, L) + b$. Note that if two individual goodness measures are positive affine transformations of $g(i, L)$, then they are also positive affine transformations of each other.

On the other hand, it is easy to show that (2) entails (1). For suppose $\tilde{g}(i, L)$ is an individual goodness measure. By (2), there exist $a > 0$ and b such that $\tilde{g}(i, L) = ag(i, L) + b$. With n the size of the population, we

have: $\sum_i g(i, L_1) \geq \sum_i g(i, L_2) \Leftrightarrow nb + a\sum_i g(i, L_1) \geq nb + a\sum_i g(i, L_2) \Leftrightarrow \sum_i (ag(i, L_1) + b) \geq \sum_i (ag(i, L_2) + b) \Leftrightarrow \sum_i \tilde{g}(i, L_1) \geq \sum_i \tilde{g}(i, L_2)$.

We have therefore shown that utilitarianism presupposes (1), and that with modest assumptions (1) is logically equivalent to (2). Hence utilitarianism presupposes (2).

But how about prioritarianism? Prioritarianism is more complicated because whereas utilitarianism claims that a particular candidate betterness relation is correct, prioritarianism claims that some member of a whole class of candidate betterness relations is correct. Roughly speaking, different prioritarian betterness relations agree that benefiting the worse off matters more, but disagree among themselves about how much more.

But there is no suggestion in the literature that it can be neither true nor false that a particular relation is a prioritarian betterness relation. So I take prioritarianism to presuppose that there exists at least one individual goodness measure, and that for every candidate betterness relation R, either R is a prioritarian betterness relation according to every individual goodness measure, or it is a prioritarian betterness relation according to no individual goodness measure. More precisely, prioritarianism presupposes

(1*) There exists an individual goodness measure, and for every candidate betterness relation R, it is either the case that for every individual goodness measure $g(i, L)$ there exists an increasing and strictly concave function w such that R is represented by $\sum_i w(g(i, L))$, or it is the case that for no individual goodness measure $g(i, L)$ does there exist an increasing and strictly concave function w such that R is represented by $\sum_i w(g(i, L))$.

We have already assumed that the betterness relation is continuous. But with this assumption to hand, one can show that (1*) is logically equivalent to (2). Hence prioritarianism makes the same presupposition about individual goodness measures as utilitarianism. This is a pleasant result, for it vindicates what one would naively expect, namely that utilitarianism and prioritarianism agree about how individual goodness is to be measured, and only disagree about how it should be distributed. Putting everything together, we have

> **The utilitarian and prioritarian presupposition** There exists an individual goodness measure $g(i, L)$, and all individual goodness measures are positive affine transformations of $g(i, L)$.

Notice that this does not say that all positive affine transformations of $g(i, L)$ are individual goodness measures. It only says that the individual goodness measures are a subset of, or lie among, the positive affine transformations of $g(i, L)$, i.e. functions of the form $ag(i, L) + b$ where $a > 0$. The utilitarian and prioritarian presupposition would be true

if the individual goodness measures were precisely the positive affine transformations of $g(i, L)$, in which case $g(i, L)$ is said to be unique up to positive affine transformation. But it would also be true if the individual goodness measures were precisely the positive linear transformations of $g(i, L)$, i.e. functions of the form $ag(i, L)$ where $a > 0$. And it would be true if $g(i, L)$ was the only individual goodness measure.

We can now state the problem utilitarianism and prioritarianism face. We have already noted that a necessary condition for a function to be an individual goodness measure is that it has to represent the better life lottery relation. But it is far from obvious that this is not also a sufficient condition. In other words, it is far from obvious that there are any other constraints a function has to satisfy to be an individual goodness measure beyond representing the better life lottery relation.

In one way, this is good news. Part of the utilitarian and prioritarian presupposition is that there exists an individual goodness measure. And the assumption that the better life lottery relation satisfies the axioms of expected utility theory vindicates this presupposition. For it tells us that there is at least one function which represents the better life lottery relation. We met it in the last section, the function $u(i, L)$.

But in another way, it is bad news. Suppose there exists an individual goodness measure $g(i, L)$. The other part of the utilitarian and prioritarian presupposition says that all of the individual goodness measures have to be positive affine transformations of $g(i, L)$. The trouble comes from the following well-known result, easily proved from the definitions of an increasing function and of representation.

Lemma 1: Suppose a function v represents a binary relation R. Then a function $\tilde{v}$ also represents R if and only if there exists an increasing function f such that $\tilde{v}(x) = f(v(x))$.

In other words, if you apply an increasing transformation to a function which represents a relation, the result is another function which represents that relation. And if any two functions represent the same relation, each of the functions can be turned into the other by applying some increasing transformation. Thus as a representation of a relation, a function can only ever be, in the jargon, unique up to increasing transformation.

But here's the problem. If all a function has to do to be an individual goodness measure is to represent the better life lottery relation, then all of the increasing transformations of $g(i, L)$ are individual goodness measures. But the class of increasing transformations of $g(i, L)$ is vastly larger than the class of positive affine transformations of $g(i, L)$. So if all there is to being an individual goodness measure is representing the better life lottery relation, there are vastly too many such measures for the utilitarian and prioritarian presupposition to be true. Utilitarianism and prioritarianism therefore face the threat of presupposition failure.

3. INTERPRETATIVE STRATEGIES

One response to the threat of presupposition failure concedes immediately that utilitarianism and prioritarianism do indeed rest on a false presupposition. But this a position of last resort because it leads to an error theory about a large body of discourse. A vast amount of ethical theory takes utilitarianism seriously, if not by arguing for it, then by reacting against it. For example, prioritarianism is typically motivated by the claim that utilitarianism is distributively insensitive and ignores the separateness of persons, more permissive views by the claim that utilitarianism is too demanding, more pluralistic views by the claim that utilitarianism is only true in restricted contexts, and so on. But if utilitarianism suffers from presupposition failure, it is more or less meaningless. And the result is an error theory about all this ethical theory.

But standard accounts of charity of interpretation say there is a standing presumption against such error theories (see e.g. Lewis 1974 and Davidson 1984). In other words, there is a standing presumption to interpret utilitarianism in a way that vindicates the pivotal role it plays in ethical theory. So there is a standing presumption to interpret utilitarianism in a way which vindicates what I will call the *platitude about ethical significance*: utilitarianism expresses important, substantive, controversial but reasonably well motivated ethical ideas. And given the role prioritarianism has played in recent ethical theory, I will take the platitude about ethical significance to apply to prioritarianism as well. So because of the strong pressure to vindicate the platitude about ethical significance, there is also strong pressure to vindicate the utilitarian and prioritarian presupposition.

But what kinds of facts could constrain the interpretation of 'individual goodness measure' tightly enough to vindicate the presupposition? It is obvious that we do not have anything like a widely held, explicit definition of the term. The only hope is that facts about our use of the term somehow constrain interpretation of the term tightly enough to vindicate the presupposition, and somehow amount to an implicit definition.

But how are we to do that? I think it is fairly obvious that 'individual goodness measure' is an ethical term. That is, its dominant usage is found in our theorizing about ethics. And it mainly features in our theorizing about utilitarianism. For there are number of platitudes about utilitarianism and about how it relates to competitors like prioritarianism. And since 'individual goodness measure' features in the standard definition of utilitarianism, it thereby features in those platitudes, at least implicitly. In fact, as far as I can see, our discourse about utilitarianism contains just about all the platitudes in which 'individual goodness measure' features. So it is those platitudes which are going to implicitly define the term if anything does.

But actually, I have something more specific in mind. I suggest that we first use all the platitudes to make the referent of 'the utilitarian betterness relation' and 'the prioritarian betterness relations' explicit, and only then to make the referent of 'individual goodness measure' explicit. The form of inference this approach deploys will be used several times, so I begin by rehearsing it.

Suppose that knowing only that $v(i, L)$ represents the better life lottery relation without yet knowing that $v(i, L)$ is an individual goodness measure, we nevertheless somehow manage to show that the relation represented by $\sum_i v(i, L)$ is the utilitarian betterness relation. If the utilitarian and prioritarian presupposition is vindicated, there is an individual goodness measure $g(i, L)$ such that $\sum_i g(i, L)$ represents the utilitarian betterness relation. But then $\sum_i v(i, L)$ and $\sum_i g(i, L)$ both represent the utilitarian betterness relation. But by the result about the uniqueness of additive representations already sketched in section 2, it follows that $g(i, L)$ is a positive affine transformation of $v(i, L)$. But if the utilitarian and prioritarian presupposition is vindicated, the class of individual goodness measures is a subset of the set of positive affine transformations of $g(i, L)$. So the class of individual goodness measures implicit in the claim that the relation represented by $\sum_i v(i, L)$ is the utilitarian betterness relation turns out to be a subset of the class of positive affine transformations of $v(i, L)$. I will call this form of inference an *inference to individual goodness*. I will discuss in section 8 whether the set of individual goodness measures just is the set of positive affine transformations of $v(i, L)$, or whether instead it is some proper subset thereof.

It is worth noting two things about this interpretative strategy. In one way, it runs against a widely accepted view. For following the influential work of Sen (1976, 1977, 1986), though it seems to me far from clear that this is Sen's own view, it is often held that we somehow have to figure out what the set of individual goodness measures is without thinking about utilitarianism. He says little about Sen's views, but the best developed approach along these lines is due to Broome (1991). But in McCarthy (forthcoming) I argue that Broome's approach does not succeed, and that the reason for this is just the ignoring of utilitarianism. So here I am suggesting we reverse the usual strategy. In another way, it accords with a widely accepted view. For especially following the influential work of Lewis (1970), it is often thought that the way to figure out the referent of theoretical terms is just to ask what interpretation of those terms best vindicates the body of platitudes in which they feature.

4. THE CANDIDATES

The goal of this section is to show that we can go quite a long way towards understanding and narrowing down the relations which are eligible to

be interpreted as the utilitarian or prioritarian betterness relations just by exploiting the platitude that individual goodness measures represent the better life lottery relation.

If the utilitarian and prioritarian presupposition is correct, there exists an individual goodness measure $g(i, L)$. Utilitarianism then entails that the betterness relation is represented by $\sum_i g(i, L)$. And prioritarianism entails that the betterness relation is represented by $\sum_i w(g(i, L))$ for some increasing and strictly concave w. Since $g(i, L)$ is an individual goodness measure, it represents the better life lottery relation. But so does $u(i, L)$. Therefore, by Lemma 1 it follows that $g(i, L)$ is an increasing transformation of $u(i, L)$. And since w is increasing and $g(i, L)$ represents the better life lottery relation, it follows by Lemma 1 again that $w(g(i, L))$ represents the better life lottery relation. And by another application of Lemma 1, it follows that $w(g(i, L))$ is also an increasing transformation of $u(i, L)$. Therefore, both utilitarianism and prioritarianism entail that the betterness relation can be represented by $\sum_i f(u(i, L))$ for some increasing function f.

But we have already assumed that the betterness relation is continuous. It turns out that this forces the function f in the previous paragraph to be continuous. Therefore, utilitarianism and prioritarianism entail that the betterness relation can be represented by

$$\sum_i f(u(i, L)) \text{ for some increasing and continuous function } f$$

I will call the class of relations represented by a function of that form the *candidates*. Knowing only that individual goodness measures represent the better life lottery relation, we have just shown that if the utilitarian and prioritarian presupposition is correct, then the utilitarian and prioritarian betterness relations have to lie among the candidates.

The rest of the section investigates the candidates more closely. So it will be looking at what the utilitarian and prioritarian betterness relations have in common. We will be using representation theorems to do this, so first I have to introduce the principles and concepts which go into these theorems. Their definitions are tacitly restricted to the case in which the population is constant, which we are assuming throughout. The following is due to Broome (1991).

> **Principle of personal good** If two lotteries are equally good for each person, then they are equally good. And if one lottery is better for one person than another lottery, and at least as good for every person, then the first lottery is better than the second.

The betterness relation is an *ordering* just in case it is *transitive* and *complete*. It is transitive just in case for all lotteries L_1, L_2, and L_3: if L_1 is at least as good as L_2 and L_2 is at least as good as L_3, then L_1 is at least as good as L_3. It is complete just in case for any two lotteries L_1 and L_2: either L_1 is at least as good as L_2, or L_2 is at least as good as L_1.

The betterness relation is *impartial* just in case, roughly, it is indifferent to which individuals are leading particular sorts of lives. More precisely, call a permutation a mapping of the individuals $1 \ldots n$ onto themselves, so that distinct individuals are mapped onto distinct individuals. Suppose that for two lotteries L_1 and L_2 there exists a permutation π of the individuals such that for all individuals i, L_1 is exactly as good for i as L_2 is for $\pi(i)$. Then the betterness relation is impartial just in case in all such cases, L_1 and L_2 are equally good.

The betterness relation is *strongly separable across people* just in case whenever two lotteries L_1 and L_2 are equally good for each member of a group of people, what the lotteries are actually like for those people is irrelevant to the question of whether L_1 is at least as good as L_2.[1]

All of these principles and concepts can of course be stated fully rigorously. But it is very important to note that they only use information supplied by the better life lottery relation.

This is the representation theorem.

Theorem 2: Assume a constant population of individuals $i = 1 \ldots n$ where $n \geq 2$. Assume that the better life lottery relation satisfies the axioms of expected utility theory, so that there exists an expectational function $u(i, L)$ which represents it. And also assume that the rectangular field assumption is true. Then the betterness relation is represented by $\sum_i f(u(i, L))$ for some increasing and continuous function f if and only if the principle of personal good is true and the betterness relation is an impartial continuous ordering which is strongly separable across people (and if $n = 2$ satisfies the hexagon condition).

Theorem 2 can be proved from a well-known theorem about additive representation. It is called the central theorem of additive representation in Wakker (1989), which provides a rigorous introduction to the theorem. A rigorous survey of results which are similar to Theorem 2 and their application to utilitarianism is in Blackorby, Bossert and Donaldson (2002). Theorem 2 is related to their discussion of what they call generalized utilitarianism.

[1] It turns out that for the assumption that the betterness relation is strongly separable across people to have full effect, the population has to contain at least three people. But when there are exactly two people, the full effect can be obtained by assuming that the betterness relation satisfies what is known as the *hexagon condition*. See e.g. Wakker (1989) for a definition and discussion of this condition, which is too complicated to state here. If there is a case for the assumption that the betterness relation is strongly separable across people when there are at least three people, I believe there is an equally good case for the claim that it satisfies the hexagon condition when there are exactly two people. But I cannot discuss this here, and will usually suppress the condition. Anyone who wants to entirely ignore it can assume that the population always contains at least three people.

Call the *background assumptions* the assumptions that the better life lottery relation satisfies the axioms of expected utility theory, and that the rectangular field assumption is true. With the background assumptions in place, Theorem 2 tells us that the claim that the betterness relation lies among the candidates is equivalent to the claim that the principle of personal good is true and that the betterness relation is an impartial continuous ordering which is strongly separable across people.

But both the principle of personal good and the claim that the betterness relation is an impartial continuous ordering which is strongly separable across people are reasonably well motivated substantive ethical ideas (or in the case of continuity, a mild idealizing assumption). The only real exception is the implicit claim that the betterness relation is complete, which it has to be to be an ordering. If the better life lottery relation is incomplete, the betterness relation will also be incomplete. But this is the only reason I have seen put forward in the literature for doubting that the betterness relation is complete. This needs more discussion, but here I am going to treat the completeness of the betterness relation as an idealizing assumption which is on a par with the idealizing assumption that the better life lottery relation is complete.

This is not to say that all of the ideas are uncontroversial. For example, although the claim that the betterness relation is strongly separable across people is reasonably plausible, it is the hallmark of egalitarianism to deny it. In fact, the claim that the betterness relation is strongly separable across people seems to be the best way of formalizing the informal idea of non-relationality which Parfit (1991) uses to distinguish the priority view from egalitarianism; see Jensen (2003) for further discussion. But the discussion is enough to show that the claim that the betterness relation lies among the candidates expresses substantive, reasonably well motivated ethical ideas. This result will be important when we return to the platitude about ethical significance.

5. LOTTERIES

The previous section looked at what the utilitarian and prioritarian betterness relations agree about. But to sort out what those relations actually are, we have to consider what they disagree about. The next section argues that we ought to interpret their differences in terms of what they say about the distribution of chances. We already know that the utilitarian and prioritarian betterness relations have to lie among the candidates. So this section does some preliminary work by looking at what different candidates say about the distribution of chances.

Consider the following two lotteries. In essence, they were first discussed by Diamond (1967) as a criticism of Harsanyi (1955).

$L_=$	$\frac{1}{2}$	$\frac{1}{2}$
A	1	0
B	0	1

L_A	$\frac{1}{2}$	$\frac{1}{2}$
A	1	1
B	0	0

The numbers are meant to do nothing more than encode information supplied by the better life lottery relation in the obvious way. For example, a history in which A gets 1 is exactly as good for A as a history in which B gets 1 is for B. The lottery $L_=$ gives A and B equal chances of better and worse histories. But L_A gives A the better history for certain, and B the worse history for certain. The lotteries are equally good for everyone else.

The pair of $L_=$ and L_A is really a schema, and could have many different instances. It could involve allocating a lifesaving resource to one of two patients, or a sweet to one of two children. I will say that, for example, $L_=$ is always better than L_A to mean that for any instance of the pair $L_=$ and L_A, the instance of $L_=$ is better than the instance of L_A. Then the betterness relation satisfies the *lottery claim* just in case it says that $L_=$ is always better than L_A. And it satisfies *lottery-neutrality* just in case it says that $L_=$ and L_A are always equally good. These concepts can easily be formalized using only information supplied by the better life lottery relation.

The claim that the betterness relation satisfies the lottery claim has often been defended. The usual argument for it starts with the intuitive appeal of giving each person equal chances. And it is typically backed up by appeals to the separateness of persons. For example, Broome (1990–91) says that giving each person equal chances is better because it is fairer. And fairness is an issue that arises only when the claims of different people are being balanced. But the most extensive defense is in Kamm (1993) who appeals directly to the separateness of persons and the significance of the personal point of view.

The claim that the betterness relation satisfies lottery-neutrality has often been defended as well. The main argument for it is an argument for the claim that the betterness relation satisfies what is known as *strong independence*, the central axiom of expected utility theory. It satisfies strong independence just in case for any lotteries L, M, and N and any α such that $0 < \alpha < 1$, L is at least as good as M if and only if the compound lottery $\alpha L + (1 - \alpha)N$ is at least as good as the compound lottery $\alpha M + (1 - \alpha)N$.

To illustrate, consider the compound lotteries $\alpha L + (1 - \alpha)N$ and $\alpha M + (1 - \alpha)N$ where $0 < \alpha < 1$. And suppose that these compound lotteries are run by first tossing a biased coin, with probability α of heads, and probability $(1 - \alpha)$ of tails. If heads comes up, either the lottery L or the lottery M is run, depending on which of the compound lotteries is chosen; if tails comes up, either the lottery N or the lottery N is run, depending on which of the compound lotteries is chosen. But those two lotteries are the same, so what the compound lotteries are like if tails comes up cannot make

a difference to their relative evaluation. That only leaves heads, which we have to take into account since it has a positive probability. But given that heads comes up, the choice between the two compound lotteries is exactly the same as choice between the two simple lotteries L and M. Therefore, L is at least as good as M if and only if $\alpha L + (1 - \alpha)N$ is at least as good as $\alpha M + (1 - \alpha)N$. Hence the betterness relation satisfies strong independence. Or, at least, that is the good prima facie case for the claim that it does. It is essentially an argument originally set out by Samuelson (1952). For further discussion in the framework of subjective expected utility theory, see the discussion of the sure thing principle in Broome (1991).

But one can show that if the betterness relation is transitive and impartial, then the claim that the betterness relation satisfies strong independence entails that the betterness relation satisfies lottery-neutrality. Taking the claim that the betterness relation is transitive and impartial on trust, the argument for the claim that it satisfies lottery-neutrality is then just the Samuelson-style argument for strong independence.

I am not going to make any attempt to evaluate the relative plausibility of the claim that the betterness relation satisfies the lottery claim and the claim that it satisfies lottery-neutrality. Nor do I have the space to discuss Broome's (1991) attempt to reconcile them by claiming, in his terminology, that unfairness is an individual harm. I believe that this attempted reconciliation does not succeed, but this will have to serve as a promissory note. It is enough for the purposes of this article that both claims are well established in the literature, and reasonably well motivated.

But we are interested in the candidates, and the following theorem teaches us what some of them say about chances. I omit the straightforward proof.

Theorem 3: Assume a constant population of individuals $i = 1 \ldots n$ where $n \geq 2$. And suppose that $u(i, L)$ is an expectational function which represents the better life lottery relation. Suppose the betterness relation can be represented by $\sum_i f(u(i, L))$ for some increasing and continuous function f. Then

(i) The betterness relation is represented by $\sum_i u(i, L)$ if and only if the betterness relation satisfies lottery-neutrality if and only if the betterness relation satisfies strong independence, and
(ii) The betterness relation is represented by $\sum_i w(u(i, L))$ for some increasing and strictly concave function w if and only if the betterness relation satisfies the lottery claim.

Let U denote the relation represented by $\sum_i u(i, L))$, and the Ws denote the relations represented by $\sum_i w(u(i, L))$ for some increasing and strictly concave function w. Theorem 3 then teaches us that U is the only candidate

which satisfies lottery neutrality, and that the *W*s are the only candidates which satisfy the lottery claim.

6. INTERPRETATION

I am now going to argue that *U* is the utilitarian betterness relation, and that the *W*s are the prioritarian betterness relations. Whenever I say that a particular relation says such and such, that will be short for: it says such and such when interpreted as the betterness relation.

We know that both the utilitarian and the prioritarian betterness relations have to lie among the candidates. So to make this argument, we have to argue that interpreting *U* as the utilitarian betterness relation and the *W*s as the prioritarian betterness relations vindicates the various platitudes about utilitarianism and prioritarianism both well enough, and better than any other interpretation from among the candidates.

The first platitude we saw was the platitude about ethical significance. By Theorem 3, we learn that

(i) The betterness relation is *U* if and only if the betterness relation lies among the candidates and satisfies strong independence.
(ii) The betterness relation is one of the *W*s if and only if the betterness relation lies among the candidates and satisfies the lottery claim.

But in section 4 we saw that the claim that the betterness relation lies among the candidates expresses a reasonably well motivated substantive ethical claim. In section 5 we saw that the claim that it satisfies strong independence expresses a reasonably well-motivated substantive ethical claim, as does the claim that it satisfies the lottery claim. Therefore, interpreting *U* as the utilitarian betterness relation and the *W*s as the prioritarian betterness relations vindicates the platitude about ethical significance.

But there are other platitudes about utilitarianism and prioritarianism. A good stock is found in the famous discussion of utilitarianism in Rawls (1971). Rawls makes three main claims about utilitarianism: it is distributively insensitive, it treats interpersonal and intrapersonal aggregation in the same way, and it ignores the separateness of persons. These claims have been endorsed by many writers, are very well established in the literature, and I am going to treat them as platitudes about utilitarianism. Prioritarianism is routinely motivated as expressing a form of opposition to the features of utilitarianism Rawls emphasized (see e.g. Parfit 1991 and Rabinowicz 2002). So I will treat the platitudes about utilitarianism as containing contrastive platitudes about prioritarianism. For example, I will take it to be a platitude that in whatever way

utilitarianism is distributively insensitive, prioritarianism is not. I will discuss the three types of platitudes in turn.

Distributive insensitivity. What does it mean to say that utilitarianism is distributively insensitive? The usual answer is this:

> *Standard claim about distributive insensitivity* Utilitarianism is always indifferent to Pigou-Dalton transfers of individual goodness. By contrast, prioritarianism always regards them as improvements.

However, there is a problem. Our goal is to use platitudes about utilitarianism and prioritarianism to argue that a particular member of the candidates should be interpreted as the utilitarian betterness relation, and that a particular subclass should be interpreted as the class of prioritarian betterness relations. But as an answer to the question of how utilitarianism is insensitive to distribution, although this answer is not false, it is useless.

To see this, let V be an arbitrary member of the candidates. It is therefore represented by a function of the form $\sum_i v(i, L)$ where $v(i, L)$ represents the better life lottery relation. Suppose we interpret V as the utilitarian betterness relation. Then by an inference to individual goodness, the individual goodness measures are a subset of the positive affine transformations of $v(i, L)$. But it then follows that V is indifferent to Pigou-Dalton transfers of individual goodness. Furthermore, on this interpretation the prioritarian betterness relations will be the relations represented by a function of the form $\sum_i w(v(i, L))$. And it follows that they all regard Pigou-Dalton transfers of individual goodness as improvements. So although the choice of V was completely arbitrary, interpreting it as the utilitarian betterness relation vindicates the standard claim about distributive sensitivity. The standard claim is therefore useless for discriminating among the candidates. If we are to appeal to the platitude about distributive sensitivity to constrain interpretation, we need an alternate reading of it.

We can get one by thinking about the distribution of chances. Let h_A and h_B be two histories which are equally good for each person apart from A and B, and which from the point of view of A and B are of the form [1 for A, 0 for B] and [0 for A, 1 for B]. Assuming that the betterness relation is impartial, h_A and h_B are equally good. So the choice between two lotteries over h_A and h_B can never make a difference to the goodness of the resulting history. So the only factor which seems relevant to determining how the betterness relation orders these two lotteries has to do the distribution of chances. And we have already seen two views about the significance of this factor in the case of ordering $L_=$ and L_A. One says that it is a matter of indifference how the chances are distributed. The other says that it is better to distribute the chances equally rather than unequally, and to improve the position of the worse off of the only two people affected by the choice.

The first of these views has a utilitarian flavor, and the second has a prioritarian flavor. However, it follows easily from Theorem 3 that of all the candidates, U is the only one which is always indifferent between $L_=$ and L_A, and that the W's are the only candidates which always regard $L_=$ as better than L_A. I think that this is enough to establish that there is at least one way of understanding distributive insensitivity, namely in terms of the distribution of chances, which favors interpreting U as the utilitarian betterness relation and the Ws as the prioritarian betterness relations. But the analogy with the standard claim about distributive sensitivity can be sharpened.

Consider a lottery in which h_A will occur with chance $p > \frac{1}{2}$ and h_B will occur with chance $1 - p$. What I will call a *Pigou-Dalton transfer of chances* transforms the lottery by reducing the chance of h_A occurring to q, so that $p > q$, and increases the chance of h_B occurring to $1 - q$ subject to the constraint that $q > 1 - p$. The constraint guarantees that the person who was the better off of the two people affected by the transfer to start with, A, is left better off than B originally was. One example of a Pigou-Dalton transfer of chances is the transformation of L_A into $L_=$, another is the transformation of L_A into a lottery in which h_A will occur with chance $\frac{3}{4}$ and h_B will occur with chance $\frac{1}{4}$. But now consider

> *Alternate claim about distributive insensitivity* Utilitarianism is always indifferent to Pigou-Dalton transfers of chances. By contrast, prioritarianism always regards them as improvements.

The alternate claim exactly mimics the standard claim, except that it is stated in terms of the distribution of chances rather than individual goodness. But it is very easy to use Theorem 3 to show that of all the candidates, U is the only one which is always indifferent to Pigou-Dalton transfers of chances, and that the Ws are the only candidates which always regard Pigou-Dalton transfers of chances as improvements. Therefore, interpreting U as the utilitarian betterness relation and the Ws as the class of prioritarian betterness relations is the only interpretation which vindicates the alternate claim about distributive insensitivity. I ignore the possibility of interpreting a proper subset of the Ws as the class of prioritarian betterness relations. If the point of the interpretation is to vindicate the alternate claim, this interpretation is gerrymandered and should be ignored (Lewis 1983). But the standard claim about distributive insensitivity does nothing to constrain the interpretation of the candidates. Therefore, interpreting U as the utilitarian betterness relation and the Ws as the class of prioritarian betterness relations vindicates the original and somewhat vague platitude about distributive insensitivity better than any other interpretation, and well enough.

It might be objected that whereas the standard claim about distributive sensitivity is widely regarded as a platitude, the alternate claim is not. But

it is commonly accepted that something can count as the referent of a term without vindicating the platitudes that contain the term perfectly. It is only required that as long as no single interpretation vindicates the platitudes perfectly, the interpretation makes the platitudes approximately true, or makes some sentences that are similar enough to the platitudes perfectly true, and does better than any competing interpretation. This idea is often used in the metaphysics of David Lewis. It is often deployed when there is strong pressure to make sense of a body of discourse and *no single* interpretation results from *no* interpretation. But the problem with using the standard claim about distributive insensitivity was not that no interpretation vindicates it, but that too many interpretations vindicate it. I have already pointed out how much pressure there is to make sense of the utilitarian and prioritarian presupposition, and I am adapting the idea to the case where no single interpretation results from no *single* interpretation.

Interpersonal and intrapersonal aggregation. Theorem 3 shows that the only disagreement between *U* and the rest of the candidates is over whether the betterness relation satisfies strong independence. *U* claims that it does, and the rest of the candidates claim that it does not. Call an individual *i*'s *individual betterness relation* the relation 'at least as good for *i* as' which holds between lotteries over histories. Each individual betterness relation is a special case of the better life lottery relation. But we are taking it to be a background assumption that the better life lottery relation satisfies the axioms of expected utility theory. It then follows that each person's individual betterness relation satisfies strong independence. Therefore, *U* says there is a strong analogy between interpersonal and intrapersonal aggregation: both the betterness relation and individual betterness relations satisfy strong independence. And the rest of the candidates say there is a strong disanalogy: the betterness relation does not satisfy strong independence, but individual betterness relations do. Therefore, interpreting *U* as the utilitarian betterness relation vindicates the platitude that utilitarianism treats interpersonal and intrapersonal aggregation in the same way.

Every candidate apart from *U* treats interpersonal and intrapersonal aggregation differently. So the *W*s are not alone in doing that. But the literature has typically seen the alleged distributive insensitivity of utilitarianism as closely connected with the way it treats interpersonal and intrapersonal aggregation in the same way. So the literature sees the way prioritarianism departs from utilitarianism in its treatment of interpersonal and intrapersonal aggregation as closely connected with the way it is not distributively insensitive.

I have suggested that for ideas about distributive sensitivity to constrain interpretation, distributive sensitivity has to be understood with respect to the distribution of chances. But of all the candidates which reject

the claim that the betterness relation satisfies strong independence, it is precisely the *W*s which do that by satisfying the lottery claim. But it was satisfying the lottery claim that underwrote the claim of the *W*s to be distributively sensitive.

Therefore, interpreting the *W*s as the prioritarian betterness relations does a better job than any other interpretation at vindicating the platitude that prioritarianism treats interpersonal and intrapersonal aggregation differently, and in a way which is closely connected with its distributive sensitivity. So interpreting *U* as the utilitarian betterness relation and the *W*s as the prioritarian betterness relations does a good job at vindicating the various established views about the ways in which utilitarianism and prioritarianism treat interpersonal and intrapersonal aggregation.

The separateness of persons. The sole dispute between *U* and the rest of the candidates is over whether the betterness relation satisfies strong independence. Now even among those who deny that the betterness relation satisfies strong independence there is acknowledged to be a burden of proof. For arguments in the style of Samuelson (1952) already rehearsed are widely thought to establish a good prima facie case for the betterness relation satisfying strong independence. Call a *symmetrical contest* a case in which there are two possible histories of the form h_A and h_B. As far as I can see, the only established view which accepts that individual betterness relations satisfy strong independence but denies that the betterness relation satisfies strong independence appeals to the view that it is best to distribute the chances equally in symmetrical contests. And in section 5 we saw that that view is typically backed up by appeals to the separateness of persons, or the significance of the personal point of view.

Therefore, interpreting *U* as the utilitarian betterness relation and the *W*s as the class of prioritarian betterness relations vindicates the platitude about the separateness of persons as follows. Contrary to what is sometimes claimed, utilitarianism need not be seen as resting on a daft view about persons. Rather, it rests on the view that the truth about persons – the fact that they are separate, have their own personal points of view, and so on – ought to be ignored because it is irrelevant. It just does not cut any ice with the Samuelson-style argument for strong independence. Prioritarianism, by contrast, rests on the view that such facts about persons somehow do override the argument for strong independence and therefore ought not to be ignored, as witnessed by the rationales that have been offered for the view that it is best to distribute the chances equally in symmetrical contests.

In summary, interpreting *U* as the utilitarian betterness relation and the *W*s as the prioritarian betterness relations vindicates what are taken to be platitudes about utilitarianism and prioritarianism. It vindicates the

platitude about ethical significance. It vindicates the way the topics of distributive insensitivity, interpersonal and intrapersonal aggregation, and the separateness of persons have been seen as interconnected. And it also vindicates the way in which the rationales for utilitarianism and prioritarianism have typically been seen as contrastive. For the fundamental dispute is over whether the betterness relation satisfies strong independence. Utilitarianism, now understood as the claim that U is the betterness relation, says it does. That is what makes utilitarianism ignore the separateness of persons, treat interpersonal and intrapersonal aggregation in the same way, and be distributively insensitive in the way described. Prioritarianism, now understood as the claim that one of the Ws is the betterness relation, says it does not. In particular, prioritarianism says that the distribution of chances in symmetrical contests matters. That is how prioritarianism expresses the view that the separateness of persons is somehow important, and is what makes it treat interpersonal and intrapersonal aggregation differently and be distributively sensitive in the way described.

There are some more technical arguments for interpreting U as the utilitarian betterness relation and the Ws as the prioritarian betterness relations, but I will stop here as the case for that interpretation is already very strong.

7. *EX POST* VIEWS

However, we need to remember that we have so far been looking at *ex ante* interpretations of utilitarianism and prioritarianism. So what has actually been argued is that U is the *ex ante* utilitarian betterness relation, and that the Ws are the *ex ante* prioritarian betterness relations.

But what about *ex post* views? We could pursue this question by the laborious method of doing for *ex post* views what we just did for *ex ante* views. But a shortcut comes from noticing something special about U. U is represented by the function $\sum_i u(i, L)$. But the second property of $u(i, L)$ described in section 1 (i.e. the fact that $u(i, L)$ is an expectational function) means that we can write $\sum_i u(i, L)$ in an expanded form. For any lottery of the form $L = [p_1, h_1; p_2, h_2; \ldots; p_m, h_m]$, $\sum_i u(i, L)$ equals

$$\sum_i \sum_j p_j u(i, h_j)$$

where the is index the different members of the population and the js index the different possible histories and their corresponding probabilities. This can be seen as expressing an *ex ante* approach to aggregation because the $\sum_j p_j u(j, h_j)$ part of the expression gives us an evaluation of the lottery from the point of view of individual i. The $\sum_i$ part of the expression then combines the evaluations from the points of view of all the different individuals into an overall evaluation of the lottery. So U can be seen as

dealing with uncertainty first, then people, and hence as taking an *ex ante* approach to aggregation.

However, it is just a matter of some elementary algebra to show that $\sum_i u(i, L)$ also equals

$$\sum_j p_j \sum_i u(i, h_j)$$

But now, the $\sum_i u(i, h_j)$ part of the expression combines the evaluations of the history h_j from the points of view of all the different individuals into an overall evaluation of h_j. The $\sum_j p_j$ part of the expression then combines the evaluations of all the different histories by taking into account the probability of each history. So U can be seen as dealing with people first, then uncertainty, and hence as also taking an *ex post* approach to aggregation.

But suppose that we can show that interpreting U as the *ex post* utilitarian betterness relation vindicates the platitudes about utilitarianism better than interpreting any other relation which takes an *ex post* approach to aggregation as the *ex post* utilitarian betterness relation. Then the case for interpreting U as the (now unqualified) utilitarian betterness relation will be beyond doubt.

To argue for that supposition, it will help to sharpen our understanding of U. We have now seen all the ideas that go into expected utility theory. For a binary relation between lotteries satisfies the expected utility axioms just in case it is a continuous ordering which satisfies strong independence. Theorem 4 below is a variation upon the original theorem of Harsanyi (1955). The interpersonal addition theorem of Broome (1991) departs from Harsanyi's theorem by replacing talk of individual preference relations with individual betterness relations. Theorem 4 departs from the interpersonal addition theorem by subsuming talk of individual betterness relations under talk of the better life lottery relation, and by explicitly assuming that the better life lottery relation is complete and that the betterness relation is impartial. The rectangular field assumption turns out not to be needed (Coulhon and Mongin, 1989).

Theorem 4: Assume a constant population of individuals $i = 1 \ldots n$. Suppose that the better life lottery relation satisfies the axioms of expected utility theory; the betterness relation satisfies the axioms of expected utility theory and is impartial; and the principle of personal good is true. Then there exists an expectational function $u(i, L)$ which represents the better life lottery relation such that $\sum_i u(i, L)$ represents the betterness relation.

In other words, if the premises of the theorem are true, then U is the betterness relation.

In order for a candidate betterness relation to be interpreted as the (*ex post*) utilitarian betterness relation, it has to express plausible ethical

ideas when interpreted as the betterness relation. That is because of the platitude about ethical significance. But ignoring the idealizing assumptions about completeness, there are only really two assumptions in the premises of Theorem 4 against which there is a reasonable ethical case. Hence there are really only two ways of generating relations different from U which are candidates to be interpreted as the (*ex post*) utilitarian betterness relation.

The first criticizable assumption, built into the second premise, is the assumption that the betterness relation satisfies strong independence. Because of Diamond's example, denying this assumption is certainly possible. But using Diamond's example in this way is not consistent with *ex post* approaches, for given that the betterness relation is impartial, they will regard the two histories h_A and h_B as equally good and hence will not distinguish between them when they come to factor in uncertainty.

The second criticizable assumption is the principle of personal good. As far as I can see, this is the only place in the premises of Theorem 4 where there is room for a principled *ex post* challenge. This example is due to Broome (1991).

$L_=$	$\frac{1}{2}$	$\frac{1}{2}$	L_E	$\frac{1}{2}$	$\frac{1}{2}$
A	1	0	A	1	0
B	0	1	B	1	0

$L_=$ is exactly as good as L_E for A. And $L_=$ is exactly as good as L_E for B. So the principle of personal good entails that $L_=$ and L_E are equally good. But there is an egalitarian case for the claim that L_E is better than $L_=$. For in L_E there is guaranteed equality of outcome, whereas in $L_=$ there is guaranteed inequality of outcome. Moreover, this challenge takes an *ex post* approach to aggregation. Any *ex ante* approach is going to say that $L_=$ and L_E are equally good because they are equally good for each person.

We can now claim that of all the *ex post* candidate betterness relations, interpreting U as the *ex post* utilitarian betterness relation is the best way of vindicating the platitudes about utilitarianism. This is for three reasons.

First, there is certainly a good prima facie case for taking the principle of personal good to be part of utilitarian doctrine. Offhand it would be astonishing for a utilitarian to deny that, for example, if two lotteries are equally good for each person then they are equally good. However, we have seen that denying the principle of personal good is the only reasonably plausible and distinctively *ex post* challenge to the premises of Theorem 4. But it is easily seen to be a logical truth that, given all of the premises of Theorem 4 apart from the principle of personal good, U is the only candidate betterness relation which is consistent with the principle of personal good. Therefore, all the reasonably plausible and distinctively *ex post* challenges to U involve denying the principle of personal good.

Hence of all the candidate betterness relations which take an *ex post* approach, it is only U which when interpreted as the *ex post* utilitarian betterness relation is consistent with what looks like a central part of utilitarian doctrine. This creates a strong presumption for interpreting U as the *ex post* utilitarian betterness relation.

Second, I take it to be a platitude that the main way utilitarianism departs from egalitarianism is in having no non-instrumental concern with equality. But any non-instrumental concern with equality is almost bound to see L_E as better than $L_=$ and reject the principle of personal good. So in accepting the principle of personal good U can be seen as expressing no non-instrumental concern with equality. Therefore, interpreting U as the *ex post* utilitarian betterness relation vindicates the main platitude about how utilitarianism is opposed to egalitarianism.

Third, we have already seen the platitude that utilitarianism is distributively insensitive. But interpreting U as the *ex post* utilitarian betterness relation provides a new way of vindicating this platitude. For as its indifference between $L_=$ and L_E illustrates, when two histories are equally likely, U is indifferent to the way the goods of each individual are distributed across the two histories.

For the purposes of the present argument, I take the second and third points to be added bonuses of interpreting U as the *ex post* utilitarian betterness relation. The first point is the strongest reason for that interpretation. But together, these make a very strong case for interpreting U as the *ex post* utilitarian betterness relation. And that means the evidence for the claim that U is simply the utilitarian betterness relation is overwhelming.

8. CONCLUSION

But now we can discharge the two main tasks of this article. First, if the relation represented by $\sum_i u(i, L)$ is the utilitarian betterness relation, it follows by an inference to individual goodness that the individual goodness measures are a subset of the positive affine transformations of $u(i, L)$. The utilitarian and prioritarian presupposition is vindicated. In addition, $u(i, L)$ is an expectational function. A basic result from expected utility theory tells us that the positive affine transformations of $u(i, L)$ are precisely the expectational functions which represent the better life lottery relation. So we have a defense of a claim which is often made without a shred of justification, namely that individual goodness is expectational.

Second, I have already argued that the Ws are the *ex ante* prioritarian betterness relations. But now the evidence for the claim that U is the utilitarian betterness relation is stronger, so is the claim that the Ws are the *ex ante* prioritarian betterness relations. For given that the individual goodness measures are a subset of the positive affine transformations of

$u(i, L)$, it follows directly from the definition of *ex ante* prioritarianism we began with that the *W*s are the *ex ante* prioritarian betterness relations.

And we now have an answer to Broome's (1991) claim that prioritarianism is more or less meaningless. Broome claims that prioritarianism distinguishes between individual goodness and how much individual goodness contributes towards overall goodness. But he argues that is a meaningless distinction. My account of *ex ante* prioritarianism nowhere takes such a distinction for granted, so does not make the claim Broome criticizes. But if a further answer is needed, it is this. Utilitarianism tells us to maximize the sum of individual goodness. *Ex ante* prioritarianism tells us to maximize the sum of some strictly concave transformation of individual goodness. And if an intuitive meaning means to be supplied to the transformation, it is straightforward.

Very roughly, the addition of a concave transformation to the utilitarian formula reflects the view that the betterness relation satisfies the lottery claim, so that the distribution of chances matters, at least in symmetrical contests. But if the distribution of chances matters in symmetrical contests, it is very plausible that it matters in at least some contests which are not entirely symmetrical. And different concave transformations express different opinions about the distribution of chances in such cases. For example, suppose that two contestants *A* and *B* to a single indivisible good are equally well off to start with, but that *A* would benefit more from getting the good than *B*. Without going into details, one can show that one version of *ex ante* prioritarianism (where the concave transformation is the square root function) recommends giving *A* and *B* chances in proportion to how much they would benefit from the good. Another version (where the concave transformation is the logarithmic function) recommends giving *A* and *B* equal chances. *Ex post* prioritarianism raises somewhat different issues which I hope to discuss elsewhere.

The use of representation theorems in general, and Harsanyi's theorem in particular, to figure out what utilitarianism is has a bad reputation, mainly because of influential remarks due to Sen (1976, 1977, 1986) and elaboration by Weymark (1991). But I think Sen should have no complaints about the approach taken here. In my terminology, not his, Sen claims that prior to substantive argument, it is an open question whether any particular function which represents the better life lottery relation, such as $u(i, L)$, is an individual goodness measure. And he claims that given an account of what the individual goodness measures are, it is a further question whether utilitarianism is correct.

Agreed. But this does not show that representation theorems, Harsanyi's in particular, do not play a crucial role in figuring out what the individual goodness measures are. We can use representation theorems, including Harsanyi's, to figure out which relations are the utilitarian and prioritarian betterness relations, and hence which functions are

the individual goodness measures. I have claimed that this approach shows that the individual goodness measures lie among the positive affine transformations of $u(i, L)$. But this is the conclusion of substantive argument; it is not the kind of arbitrary definition Sen was complaining about. And it does not follow that utilitarianism is correct. For one can accept everything said here while still regarding the truth of the premises of Harsanyi's theorem, and hence the truth of utilitarianism, as a further question.

Finally, I have not said whether the individual goodness measures are all of the positive affine transformations of $u(i, L)$, or whether they form some strict subset of those functions. So for example, I have not said whether individual goodness measures are unique up to positive affine transformation, or whether they satisfy some tighter condition, such as being unique up to positive linear transformation. The most obvious way of pursuing this would be to extend the approach taken here to figure out what the utilitarian betterness relation is when the population size is allowed to vary. It is arguable that this forces the set of individual goodness measures to be unique up to positive linear transformation; see McCarthy (forthcoming) for a sketch of this argument.

REFERENCES

Aczél, J. and J. Dhombres. 1989. *Functional equations in several variables*. Cambridge University Press

Broome, J. 1990–91. Fairness. *Proceedings of the Aristotelian Society* 91:87–102

Broome, J. 1991. *Weighing goods*. Blackwell

Broome, J. 2004. *Weighing lives*. Oxford University Press

Coulhon, T. and P. Mongin. 1989. Social choice theory in the case of von Neumann-Morgenstern utilities. *Social Choice and Welfare* 6:175–87

Davidson, D. 1984. Radical interpretation. In *Inquiries into truth and interpretation*. Oxford University Press

Diamond, P. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment. *Journal of Political Economy* 75:765–6

Hammond, P. 1981. *Ex-ante* and *ex-post* welfare optimality under uncertainty. *Economica* 48:235–50

Hammond, P. 1991. Interpersonal comparisons of utility: why and how they are and should be made. In *Interpersonal comparisons of well being*, ed. J. Elster and J. Roemer. Cambridge University Press:200–54

Harsanyi, J. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63:309–21

Jensen, K. K. 1995. Measuring the size of a benefit and its moral weight. *Theoria* 61:25–60

Jensen, K. K. 2003. What is the difference between (moderate) egalitarianism and prioritarianism? *Economics and Philosophy* 19:89–109

Kamm, F. 1993. *Morality, mortality*. Oxford University Press

Lewis, D. 1970. How to define theoretical terms. *Journal of Philosophy* 67:427–46

Lewis, D. 1974. Radical interpretation. *Synthese* 23:331–44

Lewis, D. 1983. New work for a theory of universals. *Australasian Journal of Philosophy* 61:343–77

McCarthy, D. Forthcoming. Measuring life's goodness. *Philosophical Books*

Parfit, D. 1991. *Equality or priority?* Lindley Lectures, University of Kansas
Rabinowicz, W. 2002. Prioritarianism for prospects *Utilitas* 14:2–21
Rawls, J. 1971. *A theory of justice*. Harvard University Press
Samuelson, P. 1952. Probability, utility, and the independence axiom. *Econometrica* 20:670–78
Sen, A. 1976. Welfare inequalities and Rawlsian axiomatics. *Theory and Decision* 7:243–62
Sen, A. 1977. Non-linear social welfare functions: a reply to Professor Harsanyi. In *Foundational problems in the special sciences*, ed. R. Butts and J. Hintikka. Reidel:297–302
Sen, A. 1986. Social choice theory. In *Handbook of mathematical economics*, ed. K. Arrow and M. Intriligator. Amsterdam, North Holland. 3:1073–81
van Fraassen, B. 1966. Singular terms, truth-value gaps, and free logic. *Journal of Philosophy* 63:481–95
Wakker, P. 1989. *Additive Representations of Preferences*. Kluwer
Weymark, J. 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In *Interpersonal comparisons of well being*, ed. J. Elster and J. Roemer. Cambridge University Press: 255–320