UTILITARIANISM AND PRIORITARIANISM II

DAVID MCCARTHY

University of Edinburgh

The priority view has become very popular in moral philosophy, but there is a serious question about how it should be formalized. The most natural formalization leads to ex post prioritarianism, which results from adding expected utility theory to the main ideas of the priority view. But ex post prioritarianism entails a claim which is too implausible for it to be a serious competitor to utilitarianism. In fact, ex post prioritarianism was probably never a genuine alternative to utilitarianism in the first place. By contrast, ex ante prioritarianism is defensible. But its motivation is very different from the usual rationales offered for the priority view. Given the untenability of ex post prioritarianism, it is more natural for most friends of the priority view to revert to utilitarianism.

Although the idea itself is quite old and has been familiar to economists for several decades, Parfit's 1991 *Lindley Lectures* (Parfit, 2000) led to the priority view becoming very popular among moral philosophers. But Parfit's work and much of the subsequent philosophical work on the priority view has been informal, and there is a serious question about how the priority view should be formalized. I will therefore reserve the term 'the priority view' for the cluster of informal and perhaps not fully

1

determinate ideas found in Parfit's work and subsequently adopted by many others. And I will reserve 'prioritarianism' for various attempts to formalize those ideas.

There are at least two theories which deserve to be seen as formalizations of the priority view, what I call ex ante prioritarianism and ex post prioritarianism. I discuss ex ante prioritarianism in a companion to this article (McCarthy, 2006). There I use ex ante prioritarianism to defend the priority view against the conclusion of Broome (1991) that the priority view is meaningless. Here I focus on ex post prioritarianism.

Ex post prioritarianism is the most natural formalization of the priority view. But building on an idea due to Rabinowicz (2001, 2002), I will show that ex post prioritarianism entails a very surprising conclusion. And I will argue that this conclusion is too implausible for ex post prioritarianism to deserve its place as a serious competitor to utilitarianism.

If that is right, it may seem natural for friends of the priority view to shift their allegiance to ex ante prioritarianism. But the ideas which underlie ex ante prioritarianism are different enough from those which underlie ex post prioritarianism that I think it would be more natural for them to revert to utilitarianism. In fact, I doubt that ex post prioritarianism was ever a genuine alternative to utilitarianism in the first place.

To motivate ex post prioritarianism will require a lot of care, so an informal sketch may help to frame the discussion. Ex post prioritarianism rests on three key ideas.

The first idea is that so-called Pigou–Dalton transfers always make histories better. A Pigou–Dalton transfer makes a better off person *A* worse off by some amount, and a worse off person *B* better off by that same amount, while still leaving *A* better off than *B* was to start with. This is where ex post prioritarianism departs from utilitarianism, for utilitarianism is indifferent to Pigou–Dalton transfers.

The second idea is that the contribution each person's life makes to overall goodness is independent of what other people's lives are like. This is where ex post prioritarianism departs from egalitarianism. For according to egalitarianism, at least as usually understood, the contribution someone's life makes to overall goodness can depend on how it affects the patterns of equality, and that can depend on what other people's lives are like.

These two ideas are only about when one entire world history is better than another. To extend them to a complete theory, we need to take risk into account. In other words, we need to be able to say when one lottery over histories is better than another. The third idea is that the account of when one lottery is better than another conforms to the axioms of expected utility theory. This idea is often just taken for granted.

Taken separately, there seems to be a reasonable case for each of these three ideas. But the main claim of this article is that although there is

indeed a reasonable case for each of these ideas, their conjunction is incoherent.

Section 1 presents restricted prioritarianism, a formalization of what the priority view says about when one history is better than another. Section 2 extends restricted prioritarianism to a view about lotteries by adding expected utility theory, resulting in what I call weak ex post prioritarianism. Section 3 rehearses the way the literature has seen the defence of the claims which go into weak ex post prioritarianism. Section 4 argues that friends of the priority view are committed to a further claim, resulting in ex post prioritarianism. Section 5 derives an apparently implausible conclusion from ex post prioritarianism. Section 6 argues that because of this conclusion, we should reject ex post prioritarianism. Section 7 argues that ex post prioritarianism was never a genuine alternative to utilitarianism anyway. Section 8 reassesses the priority view in the light of all this.

1. RESTRICTED PRIORITARIANISM

We are eventually going to be interested in the *betterness relation.* The betterness relation holds between lotteries over histories. In particular, it holds between lotteries L_1 and L_2 just in case L_1 is at least as good as L_2. But most of the discussion of the priority view in the philosophical literature starts by ignoring risk. It focuses on what I will call the *restricted betterness relation*. The restricted betterness relation holds between histories. In particular, it holds between histories h_1 and h_2 just in case h_1 is at least as good as h_2.

This terminology has a natural explanation. Let us first identify any history h with the corresponding degenerate lottery $L = [1, h]$ in which h has probability one. By restricting the betterness relation to degenerate lotteries, we immediately see that the restricted betterness relation is embedded in, or is a special case of, the betterness relation. More generally, I will often use the term 'restricted' to describe a principle about histories which can be regarded as a special case of a principle about lotteries over histories in an obvious way.

The implicit assumption in most of the philosophical literature is that the distinctive ideas of the priority view are all to do with histories. Thus the assumption seems to be that we do not need to discuss lotteries to understand, for example, where and why the priority view disagrees with utilitarianism and egalitarianism. It is very common for discussions of the priority view to follow Parfit (2000) in not discussing lotteries at all. It is also common to regard the extension to lotteries as an independent and fairly straightforward problem. I am eventually going to question whether ignoring lotteries at the outset is a good idea. But my first goal is to articulate what I take to be the most standard version of the priority view.

So in the spirit of the philosophical literature, this section presents what I will call restricted prioritarianism. Restricted prioritarianism is meant to formalize what the priority view says about when one history is at least as good as another.

Some writers try to define the priority view just by stating its main ideas informally, much as Parfit (2000) does. But this leaves it unclear what these ideas jointly entail, to what extent these ideas pick out a unique view, or in what way they need to be supplemented to arrive at such a view. Other writers try to define the priority view with a mathematical formula which expresses a view about when one history is better than another, much as Rabinowicz (2001, 2002) does. The view expressed entails the main ideas behind the priority view, and a formula has the advantages of precision. But it is left unclear what all the ethical ideas which lie behind the view expressed by the formula are. And it is also left unclear whether some of those ideas go beyond the ideas we naturally associate with the priority view.

Representation theorems can help us with such problems, since they can show how various ethical ideas entail a view about when one history is better than another which is expressed by a single formula. This section provides such a theorem, and uses it to motivate the definition of restricted prioritarianism. But first I need to introduce the assumptions which go into the theorem. These fall into five groups: measurement assumptions; an assumption connecting betterness and betterness for individuals; assumptions about the structure of betterness; the restricted Pigou–Dalton principle; and an assumption about what is known as separability across people.

Individual goodness measures. The usual way of expressing interpersonal and intrapersonal comparisons is to say such things as: a history h_1 is better for an individual i than h_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 is easier to group the individual and history into a pair, which we can regard as a life. Thus the *better life relation* holds between lives (i, h_1) and (j, h_2) just in case (i, h_1) is at least as good a life as (j, h_2). Although the better life relation is officially defined between lives, it is sometimes easier to revert to more familiar idioms like "h_1 is at least as good for i as h_2 is for j", and to quantify over the individuals and histories concerned. However, (i, h) denotes a life only if i exists in h. Whenever the more familiar idiom is used, quantification is therefore always understood as tacitly restricted, so that the individual exists in the history in question.

A real valued function v is said to *represent* a binary relation R just in case: for all x and y, Rxy if and only if $v(x) \geq v(y)$. This is a very important concept, and will be used throughout the article. When we use a function

to represent a relation, our real interest is almost always in the binary relation. But it is often more convenient to work with a function which represents the relation rather than work directly with the relation itself.

I will say that a function is a *restricted individual goodness measure* just in case the function takes an individual and a history containing that individual and gives a real number which somehow measures how good the history is for the individual. Restricted individual goodness measures therefore somehow measure how good lives are from the perspective of the individuals leading them. This falls short of a full definition because I have not said what it is to somehow measure how good a history is for an individual. I am not going to get into that topic here because the theory of measurement is complicated. But I will take it for granted that a necessary condition for a function to be a restricted individual goodness measure is that it represents the better life relation. Restricted prioritarianism makes a further assumption which I will call the *measurement assumption*: it assumes that there exists a restricted individual goodness measure, and that any two restricted individual goodness measures are related by what is known as a *positive affine transformation*. That is a transformation of the form $x \rightarrow ax + b$ where a and b are real numbers and $a > 0$. So if $g(i, h)$ and $\tilde{g}(i, h)$ are two restricted individual goodness measures, then $\tilde{g}(i, h) = ag(i, h) + b$ for some $a > 0$ and b. Although there is allowed to be more than one restricted individual goodness measure, it is common to work with just one of them as long as we are careful that whatever we use it to say does not depend on the particular measure chosen.

The measurement assumption entails that the better life relation is complete: for any individuals i and j and any histories h_1 and h_2, either h_1 is at least as good for i as h_2 is for j, or h_2 is at least as good for j as h_1 is for i. Because of doubts about whether it is always possible to make evaluative comparisons of very different sorts of lives, some will find this assumption implausible. But I will take it to be an idealizing assumption which enables us to focus more sharply on other problems.

The final assumption about restricted individual goodness measures can be ignored by anyone not greatly interested in technicalities. Very roughly, it amounts to saying that 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. More precisely, let $g(i, h)$ be a restricted individual goodness measure, and write it in the more familiar form $g_i(h)$. For any population Π let H_Π be the set of all histories whose population is Π. The *rectangular field assumption* is that there exists an interval I of real numbers of positive length such that for any population Π of size $n \geq 1$, $\{[g_i(h), \ldots, g_n(h)] : h \in H_\Pi\} = I''$. Here and throughout I assume that individuals can exist in more than one history. This assumption could be avoided by using the counterpart theory of Lewis (1968), but I stick with the more familiar idea.

The rectangular field assumption is similar to the assumption of the same name in Broome (1991) and McCarthy (2006). Broome outlines the philosophical questions it raises. But like him, I am not going to be overly troubled by it. It is another idealizing assumption. The role of all the idealizing assumptions will be defended at the end of this section. The effect of the rectangular field assumption is to give the restricted betterness relation a relatively simple structure. This will enable us to understand it better by using some powerful theorems.

Individual betterness and betterness. The following principle connects the two, and is a special case of a principle due to Broome (1991).

> **Restricted principle of personal good** For any two histories h_1 and h_2 containing the same population: if h_1 and h_2 are equally good for every member of the population, then they are equally good; and if h_1 is better than h_2 for at least one member of the population, and at least as good for every member of the population, then h_1 is better than h_2.

The structure of betterness. The restricted betterness relation is *transitive* if and only if for any histories h_1, h_2 and h_3: if h_1 is at least as good as h_2, and h_2 is at least as good as h_3, then h_1 is at least as good as h_3. And it is *complete* if and only if for any histories h_1 and h_2: either h_1 is at least as good as h_2, or h_2 is at least as good as h_1. A binary relation which is transitive and complete is known as an *ordering*. Restricted prioritarianism assumes that the restricted betterness relation is an ordering.

Restricted prioritarianism also assumes that the restricted betterness relation is *continuous.* Like the rectangular field assumption, this assumption is adopted to make it easier to use mathematics to understand the restricted betterness relation. Very roughly, it says that small changes in individual goodness only ever amount to small changes in overall goodness. It is too mathematical to state properly here, but it is widely adopted in welfare economics and almost any formal study of orderings will provide a good explanation. See for example Wakker (1989). Since I cannot explain it any further, I will treat it as another idealizing assumption.

Finally, restricted prioritarianism assumes that the restricted betterness relation is *impartial.* That means that it cares only which sorts of lives are being led, not which individuals are leading them. For example, if a good life and a bad life are being led, it is a matter of indifference who is leading the good life and who is leading the bad life. More precisely, suppose histories h_1 and h_2 contain populations Π_1 and Π_2 respectively, and suppose that Π_1 and Π_2 are of the same size. Suppose further that there is some mapping π from Π_1 onto Π_2 such that for all members i of Π_1: h_1 is as good for i as h_2 is for $\pi(i)$. Then the restricted betterness relation is impartial just in case in all such such cases, h_1 is as good as h_2.

The only one of these assumptions about the structure of the restricted betterness relation which most philosophers would not readily grant is the completeness assumption. But like the assumption that the better life relation is complete, it is a commonly made idealizing assumption which enables us to focus on other problems more sharply. I am not going to dwell on these assumptions because the focus of where the priority view is seen as departing from its main competitors, utilitarianism and egalitarianism, lies in the following conditions.

Pigou–Dalton. This is where restricted prioritarianism departs from utilitarianism. Informally, say that a history h_2 can be obtained from a history h_1 by a *Pigou–Dalton transfer of individual goodness* just in case we can obtain h_2 from h_1 by taking some positive amount of individual goodness from some person A and giving exactly that amount of individual goodness to some other person B while still leaving A better off than B was originally, and without affecting anyone else.

> **Restricted Pigou–Dalton principle** For any two histories h_1 and h_2 containing the same population: if h_2 can be obtained from h_1 by a Pigou–Dalton transfer of individual goodness, then h_2 is better than h_1.

The restricted Pigou–Dalton principle is one of the central claims of the priority view. It is a natural way of formalizing what Parfit (2000) sees as the most important component of the priority view, namely that "benefiting people matters more the worse off these people are" (p. 213). Utilitarianism rejects the restricted Pigou–Dalton principle, for utilitarianism regards all Pigou–Dalton transfers of individual goodness as leaving histories exactly as good as they were to start with. But according to the priority view, "we should not give equal weight to equal benefits, whoever receives them. Benefits to the worse off should be given more weight" (Parfit 2000, p. 366). It seems natural to interpret Parfit's remarks as entailing the restricted Pigou–Dalton principle, but I will delay discussing why friends of the priority view accept this principle until section 3. Note that I have not said that Parfit's remarks point to the only way of defending the restricted Pigou–Dalton principle. For example, many egalitarians accept the restricted Pigou–Dalton principle because they believe Pigou–Dalton transfers make inequality less bad, but friends of the priority view reject this argument.

The informal definition of a Pigou–Dalton transfer talks about amounts of individual goodness. More precisely, it seems to presuppose that we can properly talk about when two differences in levels of individual goodness are equal. For example, it seems to presuppose that we can properly say when the difference between A's level of individual goodness in h_1 and A's level in h_2 equals the difference between B's level in h_2 and B's level in h_1.

However, if the measurement assumption is correct, this is unproblematic. For let $g(i, h)$ and $\tilde{g}(i, h)$ be two restricted individual goodness measures. The measurement assumption entails that these are related by a positive affine transformation, so that $\tilde{g}(i, h) = ag(i, h) + b$ for some $a > 0$ and b. But then it is simple algebra to show that

$$g(A, h_1) - g(A, h_2) = g(B, h_2) - g(B, h_1) \quad \text{if and only if}$$
$$\tilde{g}(A, h_1) - \tilde{g}(A, h_2) = \tilde{g}(B, h_2) - \tilde{g}(B, h_1)$$

Therefore, if the difference in individual goodness for A between h_1 and h_2 equals the difference in individual goodness for B between h_2 and h_1 according to one restricted individual goodness measure, it equals it according to all restricted individual goodness measures. In other words, if h_2 can be obtained from h_1 by a Pigou–Dalton transfer of individual goodness according to one restricted individual goodness measure, h_2 can be obtained from h_1 by a Pigou–Dalton transfer of individual goodness according to every restricted individual goodness measure.

In short, the measurement assumption entails that the presuppositions about talking about amounts of individual goodness that seem to be built into the restricted Pigou–Dalton principle are correct. The measurement assumption is far from obvious. But given some assumptions about completeness, it turns out to be entailed by an idea which has received a lot of defence in the literature and which I will discuss in section 5, namely Bernoulli's hypothesis. So I am going to take the measurement assumption on trust.

Strong separability. I need to introduce some terminology. Suppose histories h_1 and h_2 contain the same population, and suppose that there is some subpopulation such that for each member i of that subpopulation, h_1 is exactly as good for i as h_2. The restricted betterness relation is *strongly separable across people* if and only if in any such case, we can ignore what h_1 and h_2 are like for each member of the subpopulation when we ask whether h_1 is at least as good as h_2.

More formally, suppose h_1, h_2, h_3 and h_4 contain the same population Π. Suppose for some subset Σ of Π: (1) for every member i of Σ: h_1 is exactly as good for i as h_2, and h_3 is exactly as good for i as h_4; and (2) for every member j of $\Pi - \Sigma$: h_1 is exactly as good for j as h_3, and h_2 is exactly as good for j as h_4. Then the restricted betterness relation is strongly separable across people if in all such cases: h_1 is at least as good as h_2 if and only if h_3 is at least as good as h_4.

Parfit (2000) said that according to the priority view, the restricted betterness relation is "nonrelational" (p. 372). To illustrate this very loosely, according to the priority view, the urgency of making someone better off is unrelated to how well off other people are. By contrast, Parfit said that according to egalitarianism, the restricted betterness relation

is "relational". For making someone better off can affect the patterns of equality or inequality differently according to how well off other people are.

However, Parfit's discussion is informal, and there is more than one way of trying to formalize what it means for the restricted betterness relation to be nonrelational. The most natural attempt, and the one that seems to fit best with most of Parfit's discussion, is to say that nonrelational just means strongly separable across people. This leads to an understanding of the distinction between the priority view and egalitarianism which is quite widely accepted: according to the priority view, the restricted betterness relation is strongly separable across people; according to egalitarianism, it is not.

In my view, that is not quite right. The claim it makes about egalitarianism is too strong, for one can be an egalitarian while accepting that the restricted betterness relation is strongly separable across people (McCarthy ms). And the claim it makes about the priority view is too weak to be the whole truth. It is obvious from Parfit's discussion that friends of the priority view are committed to the claim that the restricted betterness relation is strongly separable across people. But I think that this does not capture every aspect of nonrelationality. In my view, friends of the priority view are committed to the claim that the restricted betterness relation satisfies a stronger condition than merely being strongly separable across people. I will say more about this in section 5. However, my present goal is to formalize the way the priority view has been standardly understood. So for the time being I am going to stick to ideas that are explicit in the literature. And whether or not they are committed to more than this, friends of the priority view are plainly committed to the claim that the restricted betterness relation is strongly separable across people.

I need to make a qualification which can be ignored by anyone not interested in technicalities. It turns out that the claim that the restricted betterness relation is strongly separable across people is not quite strong enough to do all the work that people ask of it in cases where the population contains exactly two people. The usual way of coping with this problem is to add something known as the hexagon condition to cover the population-size-two case. I believe that anyone who thinks that the restricted betterness relation is strongly separable across people should also accept the hexagon condition in such cases. But I will not try to spell out this complicated condition and I will only mention it in the theorems; for formal discussion, see e.g. Wakker (1989).

The representation theorem uses the following concepts. A real valued function f is *increasing* just in case for all x and y in its domain, if $x > y$ then $f(x) > f(y)$. And f is *strictly concave* just in case for all distinct x and y in its domain, $f(\frac{x+y}{2}) > \frac{1}{2}(f(x) + f(y))$. For example, the function $\sqrt{x}$ on the

positive reals is both increasing (it gets larger as x gets larger) and strictly concave (it gets flatter as x gets larger). Finally, given a population Π, say that the *restricted betterness relation for* Π holds between two histories h_1 and h_2 just in case the population of both h_1 and h_2 is Π and h_1 is at least as good as h_2.

THEOREM 1. *Assume:*

(1) *There exists a restricted individual goodness measure, and any two restricted individual goodness measures are related by a positive affine transformation.*
(2) *The rectangular field assumption.*
(3) *The restricted principle of personal good.*
(4) *The restricted betterness relation is an impartial continuous ordering.*
(5) *The restricted betterness relation is strongly separable across people (and satisfies the hexagon condition in population-size-two cases).*
(6) *The restricted Pigou–Dalton principle.*

Then for every restricted individual goodness measure $g(i, h)$ and any $n \geq 1$, there exists an increasing and strictly concave function w_n such that $\sum_{i \in \Pi} w_n(g(i, h))$ represents the restricted betterness relation for any population Π of size n.

The proof is in the appendix. Call premises (1) and (2) the *background assumptions*. For the rest of this article, I am going to assume that they are correct. I will give references to various defences of (1) in section 5. I continue to treat (2) as an idealizing assumption. With the background assumptions in place, we may define restricted prioritarianism as follows.

Restricted prioritarianism Let $g(i, h)$ be a restricted individual goodness measure. Then for any $n \geq 1$ there exists an increasing and strictly concave function w_n such that

$$\Sigma_{i \in \Pi} w_n(g(i, h))$$

represents the restricted betterness relation for any population Π of size n.

Given the background assumptions, restricted prioritarianism is logically equivalent to the conjunction of premises (3) to (6). Theorem 1 establishes one direction of the equivalence, and the other is straightforward. I will therefore take premises (3) to (6) to express the main ethical ideas behind restricted prioritarianism. Utilitarianism has no quarrel with any of them except that it rejects the restricted Pigou–Dalton principle. At least as normally understood, egalitarianism has no quarrel with any of them except that it rejects strong separability across people.

It may seem odd to define restricted prioritarianism in a way which covers cases of population-size-one. But Parfit (2000) and Rabinowicz (2001, 2002) make it clear that they believe that the priority view applies to such cases. However, note that all restricted prioritarianism says about population-size-one cases is that one history is at least as good as another

just in case it is at least as good for the sole member of the population. That follows immediately from the fact that the function w_n in the definition of restricted prioritarianism is increasing. So restricted prioritarianism agrees with utilitarianism about population-size-one cases, and in those cases says nothing more than the restricted principle of personal good. But we will later see that population-size-one cases are more complicated when we introduce risk.

I have mentioned various ways in which both the background assumptions and premises (3) to (5) contain idealizing assumptions. Since I am going to criticize one prominent strand of thinking about the priority view by taking restricted prioritarianism to be its starting point, one might wonder how reasonable it is to load it up with idealizing and perhaps not very plausible assumptions.

My own view is that at least most of the idealizing assumptions are either correct or can be dispensed with. But it would take at least a separate article to argue for that. Instead I will rely on the fact that the idealizing assumptions pick out a wide and important range of cases. If the priority view leads to trouble in those cases, it is in trouble even if the mathematics is too complicated to allow us to rigorously establish the criticism in all possible cases. For example, the mathematics would be much more complicated if we abandoned the assumption that the restricted betterness relation is continuous. But it is very hard to believe that the viability of the priority view depends upon the restricted betterness relation not being continuous. As far as I am aware, nothing remotely like an appeal to the denial of any of the idealizing assumptions appears in any of the arguments that have been put forward for the priority view. So I will carry on regarding the idealizing assumptions as a harmless way of focusing on the core of the priority view.

2. RISK

Anyone sympathetic to restricted prioritarianism will want to take risk into account and to extend restricted prioritarianism to an account of when one lottery over histories is at least as good as another. In other words, they will want to extend restricted prioritarianism from an account of the restricted betterness relation to an account of the betterness relation. This section discusses a natural and seemingly quite plausible way of doing that.

To the extent that the philosophical literature discusses how to extend the priority view to deal with risk, the assumption that the betterness relation satisfies the axioms of expected utility theory is by far the most common approach. I will discuss how the literature sees the defence of this approach in the next section. This section just discusses the formal implications of adding the assumption that the betterness relation satisfies the expected utility axioms to restricted prioritarianism. To keep the mathematics simple, I will assume that all lotteries are so-called *simple*

lotteries, or lotteries with *finite support*. These are lotteries in which only finitely many histories can result, each with positive probability.

If the betterness relation satisfies the axioms of expected utility theory, the main result of expected utility theory tells us that there exists a function $v(h)$ defined on histories such that the function

(1) $\quad W(L) = \Sigma_j p_j v(h_j)$ represents the betterness relation.

Here and wherever I define a function f on lotteries by writing $f(L) = \varphi(p_j, h_j)$ for some expression φ, this is short for: for any lottery L of the form $L = [p_1, h_1; p_2, h_2; \ldots; p_m, h_m]$, where the history h_j has a positive probability of p_j and all the probabilities sum to one, $f(L) = \varphi(p_j, h_j)$. Notice that by treating histories as degenerate lotteries, it follows that $v(h)$ represents the restricted betterness relation.

Let Π be a population of size n. Suppose that restricted prioritarianism is true, and let $g(i, h)$ be a restricted individual goodness measure. Then there exists an increasing and strictly concave function w_n such that

(2) $\quad \Sigma_{i \in \Pi} w_n(g(i, h))$ represents the restricted betterness relation for Π.

But let $v_\Pi(h)$ be the restriction of $v(h)$ to histories whose population is Π. Then $v_\Pi(h)$ also represents the restricted betterness relation for Π. This means that we can exploit a simple but important result: two functions h and k both represent the same binary relation if and only if there exists an increasing function f such that $h(x) = f(k(x))$. This result is well known and its proof is easy. By (2) it follows that there exists an increasing function f_n such that

(3) $\quad v_\Pi(h) = f_n(\Sigma_{i \in \Pi} w_n(g(i, h)))$

Let $W_\Pi(L)$ be the restriction of $W(L)$ to lotteries over histories whose population is Π. By (1) and (3) it follows that $W_\Pi(L) = \Sigma_j p_j f_n(\Sigma_{i \in \Pi} w_n(g(i, h_j)))$ represents the betterness relation for Π. This motivates the following definition.

> **Weak ex post prioritarianism** Let $g(i, h)$ be a restricted individual goodness measure. Then for any $n \geq 1$ there exists an increasing and strictly concave function w_n and an increasing function f_n such that
>
> $$W_\Pi(L) = \Sigma_j p_j f_n(\Sigma_{i \in \Pi} w_n(g(i, h_j)))$$
>
> represents the betterness relation for any population Π of size n.

And we have established the following.

THEOREM 2. *Assume* (1) *restricted prioritarianism is true; and* (2) *the betterness relation satisfies the axioms of expected utility theory. Then weak ex post prioritarianism is true.*

The not very transparent term 'ex post' is well established in economics. It is used to describe approaches to evaluating lotteries over histories which aggregate across people first to reach some kind of on balance evaluation of histories before taking the probabilities of those histories into account by multiplying. In this case, the on balance evaluation of a history h_j containing a population Π of size n is expressed by the $f_n(\Sigma_{i \in \Pi} w_n(g(i, h_j)))$ part of the formula for $W_\Pi(L)$.

In summary, taking the background assumptions on trust, we saw in the previous section the ideas that go into restricted prioritarianism. We get to weak ex post prioritarianism just by adding the claim that the betterness relation satisfies the axioms of expected utility theory.

It is fairly obvious from a quick study of the literature on the priority view that the ideas which weak ex post prioritarianism rests upon are a natural formalization of ideas which are explicit in that literature. But weak ex post prioritarianism is not altogether satisfactory. For example, it is unclear how to interpret the function f_n in the weak ex post prioritarian formula. And weak ex post prioritarianism is consistent with a position which does not seem in the spirit of the priority view. In Parfit's terminology, friends of the priority view believe that it is more urgent to benefit the worse off than the better off, and the choice of a particular increasing and strictly concave w_n in the formula for weak ex post prioritarianism represents a view about how much more urgent when the population is of size n. But nothing in the definition of weak ex post prioritarianism prevents a weak ex post prioritarian from saying, for example, that when the population contains exactly two people, it is very urgent to benefit the worse off, but when the population contains exactly three people, it is only mildly urgent to benefit the worse off.

However, there is a natural idea, and one which I will argue that friends of the priority view are committed to, which solves both of these problems in a single stroke. I will call this idea the extended separability principle. Adding the extended separability principle will get rid of the function f_n, so there is no difficulty about the interpretation of that function. And will force the function w_n, to be independent of population size. But while all of the ideas which go into weak ex post prioritarianism are explicit in the philosophical literature on the priority view, the extended separability principle is not. So before introducing it, I will say some more about why many philosophers have found the priority view plausible.

3. THE DEFENCE OF WEAK EX POST PRIORITARIANISM

As I see it, the standard rationale for the priority view starts off with the view that there is something wrong with utilitarianism. We have often been told that utilitarianism is insensitive to questions about distribution, and following Rawls (1971), it seems to have been taken for granted that this is a problem with utilitarianism. But where exactly does utilitarianism go wrong? The usual answer was that it attaches no importance to equality.

However, following the influential work of Parfit (2000), many philosophers became troubled by some of the features of egalitarianism. For example, Parfit's divided world scenarios involve two communities, each unaware of the other's existence. There is perfect equality within each community, but the members of one community are better off than the members of the other community. Egalitarianism seems committed to the view that such inequality is bad. But this may seem implausible. To give another example, Parfit also claimed that egalitarianism is committed to the view that achieving equality by bringing everyone down to the level of the worst off is in one way good, even though it is better for none and worse for some. This is the so-called levelling down objection to egalitarianism. But Parfit argued that the priority view avoids these alleged problems, while still enabling us to depart from utilitarianism and in many cases say much the same thing as egalitarianism.

Tungodden (2003) calls this a negative defence of the priority view. It is by far the most commonly expressed rationale for the priority view in the literature, but it is easy to feel that it sneaks the priority view in via the back door. We start by thinking that utilitarianism is problematic because it ignores equality, discover some apparent difficulties with a concern with equality, and go straight on to adopt the priority view. But why shouldn't we instead conclude that utilitarianism is less problematic that we thought? Indeed, Broome (1989) was already complaining that the priority view is ad hoc.

However, it is possible to find a positive defence of the part of the priority view expressed by Parfit's remark that "benefiting people matters more the worse off these people are" and formalized by the restricted Pigou–Dalton principle. This defence goes back to Rawls (1971), and is the only defence of this part of the priority view I have found in the literature. Rabinowicz (2001, 2002) suggests that this part of the priority view is

> . . . a reaction to the well-known Rawlsian objection to utilitarianism: the trouble with the latter, says Rawls, is that it 'does not take seriously the distinction between persons'. A utilitarian assumes the perspective of an impartial spectator who sympathetically identifies with all the persons involved and thereby fuses them into one. Thereby, for a utilitarian, *inter*personal compensations become as unproblematic as the *intra*personal compensations have always been according to rational choice theory: it may

be rational for a person to sacrifice some of her objectives in order to realize her other goals. (Rabinowicz 2001: 10, quoting Rawls 1971: 27.)

So while it is unproblematic for a single person to make tradeoffs within her own life, so that she is indifferent to transferring a unit of individual goodness from one part of her life to another, it is to ignore the separateness of persons to extend this idea to tradeoffs between different people's lives. Thus it is to ignore the separateness of persons to hold, as utilitarianism does, that it is always a matter of indifference to transfer a unit of individual goodness from one person to another. Extending this idea, it is to respect the separateness of persons to hold that a transfer of a unit of individual goodness from a better off person to a worse off person is always an improvement, at least if no one else is affected and the better off person is left better off than the worse off person was originally. Thus if Rawls has correctly diagnosed a problem with utilitarianism, Parfit's remark seems to be a natural response and the restricted Pigou–Dalton principle seems to be well motivated.

I will be briefer about the defence of the other main part of the priority view, summarized in Parfit's (2000) claim that the priority view is not concerned with "how each person's level [of wellbeing] compares with the level of other people" (p. 370), or that the priority view is "nonrelational" (p. 372). Section 1 at least partly formalized this idea as the claim that the restricted betterness relation is strongly separable across people. As far as I can see, the most direct argument Parfit offers for this claim about strong separability is by a criticism of its denial. For suppose that the restricted betterness relation is not strongly separable across people. Given the restricted principle of personal good, we can always find a case in which whether one history is better than another depends upon what the histories are like for each member of some subpopulation even though (i) for each member of the subpopulation, the two histories are equally good; and (ii) the subpopulation is physically isolated from the rest of the population. This is more formal that Parfit's discussion, but his remarks about these divided worlds scenarios suggest that he regards it as implausible that what things are like for the members of such a "quite unrelated" subpopulation can make a difference to the overall evaluation of the two histories (p. 353).

But what about the way weak ex post prioritarianism treats risk? It is striking that one finds little discussion of risk in the philosophical literature about the priority view. The great majority of writings which are friendly towards the priority view follow Parfit (2000) in not mentioning risk at all. And those which do mention it take expected utility theory to provide an obvious and unproblematic way of extending restricted prioritarianism to take risk into account. For example, Scheffler (1982) is one of the earliest expressions of support for the priority view among philosophers, and

provides a rich discussion of various ways in which one could respond to the alleged distributive insensitivity of utilitarianism. But the extension to risk is left to a footnote, and the addition of expected utility theory to cover this is taken for granted. And the only defence of the priority view I am aware of which takes a detailed look at the extension to risk is Rabinowicz (2001, 2002). But Rabinowicz follows the rest of the literature in developing the rationale for restricted prioritarianism independently of the extension to risk, and then adopts expected utility theory to cover that extension without question.

This completes the explanation of why many philosophers have been convinced by each of the three main ideas which go into weak ex post prioritarianism: the restricted Pigou–Dalton principle, strong separability across people and expected utility theory. But these philosophers have also taken a further assumption for granted. The further assumption is that if there is a reasonably good case for each of the three ideas taken separately, there is no difficulty in combining them: together they articulate a coherent and important alternative to utilitarianism. There are hints in the literature that this further assumption might not be so innocent. For example, Hurley (1989) and Broome (1991) suggest that there are deep interconnections between claims about distribution and claims about risk. If that is so, inferences of the form "separately plausible therefore jointly plausible" concerning claims about distribution and risk may be questionable. But this suggestion seems to have been ignored by the philosophical literature which is friendly to the priority view.

4. THE EXTENDED SEPARABILITY PRINCIPLE

This section introduces a principle which I will claim that friends of the priority view are committed to accepting. And accepting this principle has two effects which they might welcome. It gets rid of the features of weak ex post prioritarianism which were said to be unsatisfactory at the end of section 2. And it makes sense of an aspect of Parfit's discussion of the priority view which is not easy to understand. However, being committed to accepting the new principle is a mixed blessing. The following section shows that adding it to weak ex post prioritarianism entails an apparently implausible conclusion.

The rough idea behind the new principle is this. Suppose that for some population Π, L_1 and L_2 are lotteries over histories whose population is Π. And suppose L_1^+ and L_2^+ are just like L_1 and L_2 except that they are lotteries over histories whose population is Π extended to include one extra person who is unaffected by the choice between L_1^+ and L_2^+. Then the rough idea is that if friends of the priority view judge that L_1 is better than L_2, they are committed to judging that L_1^+ is better than L_2^+. Roughly speaking, an

individual is unaffected by the choice between two lotteries if, from her point of view, the lotteries are exactly the same.

More precisely, say that an individual j is *unaffected* by the choice between two lotteries L_1 and L_2 just incase L_1 and L_2 have the form $L_1 = [p_1, h_1; p_2, h_2; \ldots; p_m, h_m]$ and $L_2 = [p_1, k_1; p_2, k_2; \ldots; p_m, k_m]$ and for $s = 1 \ldots m$, the two histories h_s and k_s are equally good for j. Then the following formalizes the rough idea.

> **The extended separability principle.** Suppose that there is some population Π such that L_1 and L_2 are lotteries over histories whose population is Π. Suppose that L_1^+ and L_2^+ are lotteries over histories whose population is Π together with some extra person j not in Π. Suppose (1) for every member i of Π: i is unaffected by the choice between L_1 and L_1^+ and i is also unaffected by the choice between L_2 and L_2^+; and (2) the extra person j is unaffected by the choice between L_1^+ and L_2^+. Then in all such cases: L_1 is at least as good as L_2 if and only if L_1^+ is at least as good as L_2^+.

Not everyone will accept the extended separability principle. Consider the following example.

L_1 and L_2 are lotteries over histories just containing person 1, and L_1^+ and L_2^+ are lotteries over histories containing persons 1 and 2. The entries indicate how good the lotteries are for the various people, and we assume that x and y are not equal. L_1 is exactly as good as L_2, and the extended separability principle then implies that L_1^+ and L_2^+ are equally good. But egalitarians will almost certainly deny that L_1^+ and L_2^+ are equally good (Broome 1991). For although they are equally good for each person, L_1^+ is guaranteed to lead to equality, and L_2^+ is guaranteed to lead to inequality. In fact, in my view this is really the core egalitarian idea (McCarthy ms). Hence egalitarians will almost certainly claim that L_1^+ is better than L_2^+ and reject the extended separability principle.

Nevertheless, the main claim of this section is that friends of the priority view are committed to accepting the extended separability principle. Rabinowicz (2001, 2002) discusses a weaker version of it. Say that an individual is *completely unaffected* by the choice between two lotteries if she is unaffected by the choice, and in addition all the histories that could result from the two lotteries are equally good for her. And call the *weak extended separability principle* the result of replacing 'unaffected' with 'completely unaffected' in clause (2) in the definition of the extended separability principle. In my terminology, Rabinowicz says that denying the weak extended separability principle would build an extreme sensitivity to the presence of completely unaffected people which friends of the priority view are committed to rejecting. For the divided worlds argument offered earlier for the claim that the restricted betterness relation is strongly separable across people will work equally well for the weak extended separability principle: just imagine the extra

	L_1		L_2		L_1^+		L_2^+	
	Heads	Tails	Heads	Tails	Heads	Tails	Heads	Tails
1	x	y	y	x	x	y	y	x
2	–	–	–	–	x	y	x	y

TABLE 1. Egalitarian counterexample to extended separability.

person is physically isolated from the rest of the population, and rerun that argument. And even if some friends of the priority view accept a different argument for strong separability across people, it is hard to see how they could reject the weak extended separability principle.

Although Rabinowicz's claim that friends of the priority view are committed to accepting the weak extended separability principle is convincing, there is a slip in his proof that the weak extended separability principle leads to the apparently implausible conclusion to be discussed in the next section, and in fact it does not lead to such a conclusion. So it is crucial to my argument that friends of the priority view are committed to accepting the extended separability principle.

But the claim that friends of the priority view should accept the extended separability principle emerges naturally from Parfit's informal discussion of nonrelationality. It is very hard to see a case against the extended separability principle apart from the example of Table 1. But the example of Table 1 should only appeal to egalitarians, and not to friends of the priority view. Moreover, the extension of the divided worlds argument for the weak extended separability principle seems to just work just as well as an argument for the extended separability principle. And this is what happens when we add the extended separability principle to weak ex post prioritarianism.

THEOREM 3. *Assume:*

(1) *There exists a restricted individual goodness measure, and any two restricted individual goodness measures are related by a positive affine transformation.*
(2) *The rectangular field assumption.*
(3) *The restricted principle of personal good.*
(4) *The restricted betterness relation is an impartial continuous ordering.*
(5) *The restricted betterness relation is strongly separable across people (and satisfies the hexagon condition in population-size-two cases).*
(6) *The restricted Pigou–Dalton principle.*
(7) *The betterness relation satisfies the expected utility axioms.*
(8) *The extended separability principle.*

Then for any restricted individual goodness measure g(i, h) there exists an increasing and strictly concave function w such that $W(L) = \Sigma_j \Sigma_{i \in \Pi} p_j w(g(i, h_j))$ represents the betterness relation for any population Π.

The proof is in the appendix. Taking the background conditions on trust, this motivates the following definition.

> **Ex post prioritarianism.** Let $g(i, h)$ be a restricted individual goodness measure. Then there exists an increasing and strictly concave function w such that
>
> $$W(L) = \Sigma_j \Sigma_{i \in \Pi} p_j w(g(i, h_j))$$
>
> represents the betterness relation for any population Π.

Given the background assumptions, premises (3) through (6) are equivalent to restricted prioritarianism. The addition of premise (7) makes them equivalent to weak ex post prioritarianism. And the further addition of premise (8) makes them equivalent to ex post prioritarianism. There is some redundancy in the premises of Theorem 3 since the addition of (8) makes it possible to do without (5), but I have not eliminated it as it makes the path from restricted prioritarianism to ex post prioritarianism easier to follow. For what has just been said establishes

THEOREM 4. *Assume the background assumptions. Suppose* (1) *restricted prioritarianism is true;* (2) *the betterness relation satisfies the axioms of expected utility theory; and* (3) *the extended separability principle is true. Then ex post prioritarianism is true.*

Here is the first benefit of accepting the extended separability principle. The function f_n which appears in the definition of weak ex post prioritarianism has disappeared in the definition of ex post prioritarianism. So there is no longer a problem about its interpretation. And the function w_n which appears in the definition of weak ex post prioritarianism has been replaced by a function w which is independent of population size. In Parfit's terminology, this means that the urgency of improving the lots of the worse off is independent of the number of other people around. Thus ex post prioritarianism is not only simpler than weak ex post prioritarianism, but is closer to the spirit of the priority view.

5. THE COINCIDENCE PRINCIPLE

This section presents the apparently implausible conclusion which ex post prioritarianism entails. To begin, here is what seems to be a very plausible account of when one lottery is better than another when the population contains just one person.

> **The coincidence principle.** For any population containing exactly one person and any lotteries L_1 and L_2 over histories whose population contains just that one person: L_1 is at least as good as L_2 if and only if L_1 is at least as good for the sole person as L_2.

Say that a person's *individual betterness relation* holds between two lotteries just in case the first lottery is at least as good for the individual as the second lottery (Broome 1991). Then the coincidence principle says that in population-size-one cases, betterness and individual betterness coincide.

The next part of the argument adopts an account of when one lottery is better for a person than another which has been argued for by many people.

> **Bernoulli's hypothesis.** For any lotteries L_1 and L_2, L_1 is at least as good for a person as L_2 if and only if L_1 gives that person at least as great an expectation of restricted individual goodness as L_2.

Given a restricted individual goodness measure $g(i, h)$, the expectation of restricted individual goodness a lottery $L = [p_1, h_1; p_2, h_2; \ldots; p_m, h_m]$ gives an individual i is given by the function

$$G(i, L) = \Sigma_j \, p_j g(i, h_j)$$

In other words, it is just the restricted individual goodness i gets in each history that can result from the lottery, multiplied by the probability of that history, then all added up. The background assumptions guarantee that Bernoulli's hypothesis is consistent. For the measurement assumption entails that if $g(i, h)$ and $\tilde{g}(i, h)$ are any two restricted individual goodness measures, then L_1 gives i at least as great an expectation of restricted individual goodness as L_2 according to $g(i, h)$ if and only if it does so according to $\tilde{g}(i, h)$. Another way of looking at Bernoulli's hypothesis is to note that it tells us that the function $G(i, L)$ represents i's individual betterness relation.

I adopted the measurement assumption with little comment in section 1. But the measurement assumption makes a very strong claim which is both crucial to the meaningfulness of the priority view and far from obviously true (Broome 1991, McCarthy 2006, 2007). However, Bernoulli's hypothesis turns out to entail the measurement assumption, and my own view is that it is the case for Bernoulli's hypothesis which makes the case for the measurement assumption. Bernoulli's hypothesis is not at all obvious. Nevertheless, it has been argued for (Harsanyi 1975, 1977; Broome 1991; Hammond 1991; Jensen 1995; Risse 2002; McCarthy 2006, 2007). And it is accepted by Rabinowicz (2001, 2002) in his discussion of ex post prioritarianism. Here I am simply going to take it for granted. But it creates some trouble for ex post prioritarianism, as the following theorem indicates.

THEOREM 5. *Assume the background assumptions. Then Bernoulli's hypothesis and ex post prioritarianism imply that the coincidence principle is false.*

PROOF. Assume a population just containing person 1. Let $g(i, h)$ be a restricted individual goodness measure. By the background assumptions there exist histories h_0, h_1 and h_2 such that h_0 is better for person 1 than h_1, and h_1 is better for person 1 than h_2. Simple algebra shows that there exists p such that $0 < p < 1$ such that $g(1, h_1) = pg(1, h_0) + (1 - p)g(1, h_2)$. By Bernoulli's hypothesis the degenerate lottery $L_1 = [1, h_1]$ is exactly as good for person 1 as the lottery $L_2 = [p, h_0; (1 - p), h_2]$. By ex post prioritarianism, there exists some increasing and strictly concave function w such that L_1 is better than L_2 iff $w(g(1, h_1)) > pw(g(1, h_0)) + (1 - p)w(g(1, h_2))$. But this inequality holds by a well known fact about concave functions (Hardy *et al.* 1934). Hence L_1 is better than L_2, contrary to the coincidence principle.

Note that the proof does not require the background assumptions; it only requires the very weak assumption that there exist the three histories mentioned. So given Bernoulli's hypothesis, ex post prioritarians are forced to reject the coincidence principle. In fact, it is easy to modify the proof (by making p a tiny bit smaller) to show that given the existence of the three histories mentioned, the betterness relation and the better-for-the-sole-individual relation sometimes have to say strictly opposite things: one lottery can be better than another, even though it is worse for the sole individual.

Here is the second benefit of accepting the extended separability principle. As noted in section 1, both Parfit (2000) and Rabinowicz (2001, 2002) suggest that the priority view has distinctive conclusions in population-size-one cases. But this suggestion is far from straightforward. Suppose, for example, that we characterize the priority view via the slogan that benefiting the worse off matters more. There is an immediate difficulty when we ask: what exactly does this amount to in one-person cases? We can make at least partial sense of the slogan in two-or-more person cases if we take the slogan to entail (even if is not equivalent to) the restricted Pigou–Dalton principle. But by itself, the restricted Pigou–Dalton principle has no implications in one-person cases.

In fact, neither Parfit nor Rabinowicz succeed in articulating a version of the priority view which says something distinctive in one-person cases in the following sense: neither articulates a version of the priority view which is inconsistent with what utilitarianism says about one-person cases. Parfit does not discuss risk at all, and any version of the priority view accepts what I will call *restricted coincidence principle*: in population-size-one cases, one history is at least as good as another if and only if it is at least as good for the sole individual. This is just the restriction of the restricted principle of personal good to populations of size one. But utilitarianism

also accepts the restricted coincidence principle, so it is hard to know what Parfit has in mind for one-person cases.

Rabinowicz's position is more subtle. He makes the very important point that one is going to have to discuss risk to make sense of a disagreement between utilitarianism and the priority view in one-person cases. The version of prioritarianism he ends up with is, in my terminology, the conjunction of weak ex post prioritarianism and the weak extended separability principle. He claims that this version of prioritarianism is inconsistent with the coincidence principle. But it isn't. To see why not, suppose person 1 is the sole member of the population, and that the image of the restricted individual goodness measure $g(1, h)$ on histories containing person 1 is the positive reals. For all n, take the functions f_n and w_n in the definition of weak ex post prioritarianism to be the exponential and logarithmic functions respectively. Given Bernoulli's hypothesis, it is easy to show that this form of weak ex post prioritarianism is consistent with (in fact, entails) both the weak extended separability principle and the coincidence principle. But utilitarianism accepts the coincidence principle. So the version of the priority view Rabinowicz articulates is not strong enough to force a disagreement with utilitarianism in one-person cases.

The second benefit of accepting the extended separability principle is therefore this. Given Parfit's remarks about nonrelationality, it seems that friends of the priority view are committed to the extended separability principle anyway. But Theorem 5 shows adding it to weak ex post prioritarianism results in a version of the priority view which does say something distinctive about one-person cases. So the addition of the extended separability principle and the consequent rejection of the coincidence principle has the benefit of interpreting the priority view in a way which makes sense of Parfit's and Rabinowicz's suggestions.

However, the coincidence principle seems very plausible. So plausible, in fact, that it is tempting to regard the forced rejection of the coincidence principle as a reductio of ex post prioritarianism.

6. ASSESSING EX POST PRIORITARIANISM

Ex post prioritarians have an obvious response. They can claim that the premises of ex post prioritarianism are plausible enough to make a good case for rejecting the coincidence principle. So the most important question now seems to be: what can we say about the relative plausibility of the premises of ex post prioritarianism and the coincidence principle?

But first I need to bypass a complication pointed out by Broome (2004). Whether one life is a better life than another life surely depends only on what each life is like from a perspective which is somehow internal to the life in question. But whether it is better that one life exists rather than another life exists may depend on features of the world that are external

to this perspective, such as the time of existence. To illustrate, those who believe in temporal discounting will have to say that in population size-one-cases, it is worse for an individual's existence to be delayed when the delay will make no difference to the individual's life from the internal perspective and will therefore not be worse for the individual. Hence an established position is inconsistent with the coincidence principle.

However this alleged counterexample to the coincidence principle works by being inconsistent with the restricted coincidence principle. So as things stand, the alleged counterexample is of no help to ex post prioritarians for it is inconsistent with a principle they themselves accept. Ex post prioritarians can respond in one of two ways. They can either reject temporal discounting, or they can modify their own position to allow for it. The modification will claim that the restricted coincidence principle only applies to lives which take place over the same interval of time etc. But the rejection of the coincidence principle had nothing to do with time, and this modified version of ex post prioritarianism will have to reject a similarly modified version of the coincidence principle. Hence appealing to temporal discounting or any similar view about time is a useless diversion in trying to defend ex post prioritarianism against the accusation that it is forced to reject an apparently very plausible principle. So I will henceforth ignore it.

Returning now to the main argument, let T be the conjunction of Bernoulli's hypothesis and all of the premises in Theorem 3 apart from the restricted Pigou–Dalton condition. T is therefore composed of the background assumptions we are taking on trust, Bernoulli's hypothesis, and claims which ex post prioritarianism entails. And we already have

(1) Given T: ex post prioritarianism $\Leftrightarrow$ the restricted Pigou–Dalton principle.

As already discussed, there is little prospect for defending ex post prioritarianism by rejecting T. Therefore, we will have an argument against ex post prioritarianism if we can show that given T, the restricted Pigou–Dalton principle is less plausible than its negation.

T is a powerful theory. A routine adjustment to the proofs of Theorems 1 and 3 yields

(2) $T \Rightarrow$ there exists an increasing function k such that for any population Π, the betterness relation for Π is represented by $K(L) = \Sigma_j \Sigma_{i \in \Pi} p_j k(g(i, h_j))$.

Given the background assumptions and Bernoulli's hypothesis, the following definition of utilitarianism is uncontroversial.

Utilitarianism. Let $g(i, h)$ be a restricted individual goodness measure. Then

$$U(L) = \Sigma_j \Sigma_{i \in \Pi} p_j g(i, h_j)$$

represents the betterness relation for any population Π.

Note that T is also composed of the background assumptions we are taking on trust, Bernoulli's hypothesis, and claims which utilitarianism entails. We can therefore treat T as a background theory which is common ground to both utilitarianism and ex post prioritarianism. And we have

(3)   Given T: utilitarianism $\Leftrightarrow$ the coincidence principle.

PROOF. Left to right is immediate by Bernoulli's hypothesis. For right to left, let Π be a population of size 1. Then (2), the coincidence principle, and Bernoulli's hypothesis force the function k in (2) to be linear, hence the result.

Therefore, we will have an argument against ex post prioritarianism if we can show that given T, the coincidence principle is more plausible than the restricted Pigou–Dalton condition. This is an interesting issue to address because we have seen that the rationales for the priority view are typically based on a contrast with utilitarianism. Utilitarianism has suffered greatly in moral philosophy from the assumption that the best argument for it is the impartial spectator argument, an argument which faces serious difficulties. But (3) shows that given a background theory on which both utilitarianism and ex post prioritarianism can agree, utilitarians only need to argue for the apparently plausible coincidence principle.

Let us begin with the possible rationales for the restricted Pigou–Dalton condition given T. I can think of three. The restricted Pigou–Dalton condition could be said to be (i) motivated by the alleged distributive insensitivity of utilitarianism; (ii) motivated by an appeal to the separateness of persons; or (iii) just self-evidently plausible. Of course, these rationales have to work against the background of T.

Distributive insensitivity. Philosophers often say that utilitarianism is distributively insensitive, but this can be interpreted in different ways. Every complaint I am aware of falls into one of two classes. The first class is exemplified by the complaint that utilitarianism implies that a small benefit to each member of a sufficiently large group of people outweighs one person bearing a large burden. But every theory which accepts T has this implication because of the additive form of the function in (2). More generally, complaints in the first class are distinguished by applying to all theories which accept T, so are useless for motivating ex post prioritarianism as an alternative to utilitarianism.

The second class is exemplified by the complaint that utilitarianism cares only about the sum of individual goodness, and not about how that

sum is distributed. But (2) implies that every theory which accepts T cares only about the sum of some function of individual goodness, and not about how that sum is distributed. In the case of utilitarianism, the function is the identity function; in the case of ex post prioritarianism, the function is a strictly concave function. So for the complaint to apply to utilitarianism and not to ex post prioritarianism, it has to say: mistaken to care only about the sum of individual goodness; not mistaken to care only about the sum of a concave transformation of individual goodness. But given T, this is just to assert the correctness of the restricted Pigou–Dalton condition. More generally, complaints in the second class are distinguished by taking the restricted Pigou–Dalton principle for granted. They therefore do not motivate the restricted Pigou–Dalton principle.

Separateness of persons. Section 3 sketched a motivation for the restricted Pigou–Dalton principle which rested upon Rawls's separateness of persons criticism of the impartial spectator argument for utilitarianism. But there are now two serious objections to this. The first is that given T, the derivation of utilitarianism uses nothing like the impartial spectator argument. It simply adds the coincidence principle, and it is hard to believe that to assert the coincidence principle is to ignore the separateness of persons. In response, a friend of ex post prioritarianism might claim that there is a separateness of persons objection to the coincidence principle given T, for given T the addition of the coincidence principle yields a theory, namely utilitarianism, that is said to treat interpersonal and intrapersonal aggregation in the same way. And the objection continues: to treat these in the same way is to ignore the separateness of persons.

But what exactly does it mean to say that utilitarianism treats interpersonal and intrapersonal aggregation in the same way? The objection is plainly attempting to point to some aspect of the way utilitarianism aggregates which is independent of whether the aggregation involves just one person or whether it involves two or more people. Let L be some lottery, and $(i_1, h_{j_1}), \ldots, (i_k, h_{j_k})$ be an arbitrary list of some of the lives that could result from L. We can think of these lives as locations of goodness which somehow need to be factored into the evaluation of the overall goodness of L. According to utilitarianism, the contribution of these lives to the overall goodness of L is always the same independently of whether $i_1 \ldots i_k$ are all the same individuals, all different, or some mixture thereof. This is immediate from the definition of utilitarianism. And as far as I can see, this is the only feature of utilitarianism which really invites the claim that utilitarianism treats interpersonal and intrapersonal aggregation in the same way. But it is immediate from the definition of ex post prioritarianism that ex post prioritarianism shares this feature with utilitarianism. In fact, (2) entails that any consistent theory containing T already has this feature. So given T, any motivation for the restricted Pigou–Dalton principle which

is based on the assertion that interpersonal and intrapersonal aggregation should be treated differently is self-defeating.

Self evidence. I suspect that many philosophers simply find the plausibility of the restricted Pigou–Dalton principle self-evident. But the issue here is whether it is plausible given T, and I strongly suspect that such philosophers would find the coincidence principle even more self-evidently plausible. But we have seen that given T, the restricted Pigou–Dalton principle and the coincidence principle are jointly inconsistent. The proof of that involves some fairly complicated mathematics, so it is hard to take seriously the view that the plausibility of the restricted Pigou–Dalton principle given T is somehow self-evident.

My own view is stronger. Someone who finds the restricted Pigou–Dalton principle self-evidently plausible is committed to individual goodness having a lot of quantitative structure. Along with many other people, I accept this claim. But such a person is also committed to a view I call *distributive intuitionism*: the plausibility of distributive claims about simple cases which exploit the quantitative structure of individual goodness is transparent to intuition. I reject distributive intuitionism, so I do not believe that there are any grounds for claiming that the restricted Pigou–Dalton principle is self-evidently plausible. It would take a detailed discussion of the general theory of measurement to explain why I reject distributive intuitionism, and I cannot do that here. But it turns out that the best defence of utilitarianism does not rely upon any distributive claims which exploit the quantitative structure of individual goodness, and is consistent with the rejection of distributive intuitionism.

However, it is not easy to explain why that is true. The defence of utilitarianism which proceeds by arguing for T and then arguing for the coincidence principle is not very efficient. A more efficient defence is provided by a direct appeal to the famous aggregation theorem of Harsanyi (1955) and the status of Bernoulli's hypothesis as a claim about meaning, not distribution. And this defense is the best one to focus on to argue that the case for utilitarianism does not rely on distributive intuitionism. But I lack the space to discuss this here. For more details, see Broome (1991) and McCarthy (2006). However, I can at least give a hint about what is going on. Given T, we need the restricted Pigou–Dalton condition to get to ex post prioritarianism. The restricted Pigou–Dalton principle plainly exploits the quantitative structure of individual goodness. But to get to utilitarianism, we only need the coincidence principle. But there is a good case for this principle even if there is no quantitative structure to individual goodness, for this principle only involves the comparatives 'at least as good for a person as' and 'at least as good as'. So unlike the imagined defence of the restricted Pigou–Dalton principle, the case for the coincidence principle does not rely upon distributive intuitionism.

Since I cannot argue against distributive intuitionism here, I cannot claim to have provided a decisive objection to the claim that the restricted Pigou–Dalton principle is self-evident. But I can claim that the imagined defence of ex post prioritarianism is bearing a burden not borne by the best defence of utilitarianism. For as far as I can see, no one has tried to make the case for distributive intuitionism. Such a case would surely have to take a close look at the meaning of quantitative talk about individual goodness, but this topic has been rather neglected. Until someone tries to make this case, there is a large gap in the defence of ex post prioritarianism which appeals to the self-evidence of the restricted Pigou–Dalton principle.

I can now draw a conclusion. Given T, appeals to neither the alleged distributive insensitivity of utilitarianism nor the separateness of persons provide any motivation for the restricted Pigou–Dalton principle. And given T, the plausibility of the restricted Pigou–Dalton principle is not self-evident. Thus given T, the restricted Pigou–Dalton principle is not well motivated. And given that adding the restricted Pigou–Dalton principle to T results in the rejection of the apparently very plausible coincidence principle, I conclude that ex post prioritarianism is not a viable alternative to utilitarianism.

7. FROM EX POST PRIORITARIANISM TO UTILITARIANISM

But actually, I doubt that ex post prioritarianism is any kind of alternative to utilitarianism at all. I continue to take Bernoulli's hypothesis on trust.

Both utilitarianism and ex post prioritarianism imply that there exists a function v which represents the better life relation such that for any population Π, the betterness relation for Π is represented by $V(L) = \Sigma_j \Sigma_{i \in \Pi} p_j v(i, h_j)$. Suppose that there is such a function.

Utilitarianism will turn out to be correct if and only if v is a restricted individual goodness measure. Ex post prioritarianism will turn out to be correct if and only if v is an increasing and strictly concave transformation of a restricted individual goodness measure. Setting aside other possibilities (e.g. that v is a convex transformation of a restricted individual goodness measure), there are three possibilities: (i) utilitarianism is correct; (ii) ex post prioritarianism is correct; and (iii) both are determinately neither correct nor incorrect, or it is indeterminate which is correct, or somesuch. What kind of fact would settle which of these obtains?

If v is a restricted individual goodness measure, we have:

(1)   For any restricted individual goodness measure $g(i, h)$ and any population Π, the betterness relation for Π is represented by $V(L) = \Sigma_j \Sigma_{i \in \Pi} p_j g(i, h_j)$.

If v is an increasing and strictly concave transformation of a restricted individual goodness measure, we have

(2)   For any restricted individual goodness measure $g(i, h)$ there exists an increasing and strictly concave function w such that for any population Π, the betterness relation for Π is represented by $V(L) = \Sigma_j \Sigma_{i \in \Pi} p_j w(g(i, h_j))$.

It seems clear that (1) gives us a simpler characterization of the betterness relation than (2). Moreover, those who accept (2) have to live with the added complication that in population size one cases, betterness and betterness for the sole individual do not coincide. And I suggest that this greater simplicity decides the case in favour of utilitarianism, not on any ethical grounds, but on grounds of meaning. The ex post prioritarian interpretation takes restricted individual goodness measures to provide us with an important way of characterizing ethics, but does it in an apparently gratuitously complex way for it is everywhere uniformly discounted. Until ex post prioritarians give us an explanation of the theoretical point of this extra complexity, I do not believe they have managed to articulate a genuine alternative to utilitarianism. However, it would take a separate article to give these questions about meaning a proper treatment, so for now I leave it as an invitation for ex post prioritarians to respond. For related arguments, see Harsanyi (1977), and the arguments for Bernoulli's hypothesis in Broome (1991, 2004) and McCarthy (2006, 2007), although note that this section takes Bernoulli's hypothesis for granted.

8. THE PRIORITY VIEW RECONSIDERED

I have been taking some assumptions on trust, either as relatively uncontroversial or as harmless idealizing assumptions. The focus has been on the effect of adding three further assumptions: the restricted Pigou–Dalton condition, the extended separability principle (which actually entails the separability assumptions which go into restricted prioritarianism), and the claim that the betterness relation satisfies the expected utility axioms. I have claimed that the resulting theory, ex post prioritarianism, is indefensible. But I have not tried to locate the fault in any one of the three further assumptions. In fact, it is only the addition of all three which leads to serious trouble. The addition of any two is consistent with a defensible theory.

   Utilitarianism is consistent with (in fact, entails) the extended separability principle and the claim that the betterness relation satisfies the expected utility axioms, but it rejects the restricted Pigou–Dalton condition. This is immediate from the definition of utilitarianism given earlier. As already noted, the best defence of utilitarianism is based on the

aggregation theorem of Harsanyi (1955), essentially involves claims about risk, and does not rely on distributive intuitionism.

Egalitarianism is consistent with (though I believe it does not entail either of) the restricted Pigou–Dalton condition and the claim that the betterness relation satisfies the expected utility axioms, but it rejects the extended separability principle. I believe that there is a reasonable defence of the form of egalitarianism which accepts the restricted Pigou–Dalton condition and the claim that the betterness relation satisfies the expected utility axioms. This defence likewise essentially involves claims about risk, and does not rely on distributive intuitionism (McCarthy ms).

A third defensible view is consistent with (in fact, entails) the restricted Pigou–Dalton condition and the extended separability principle, but rejects the claim that the betterness relation satisfies the expected utility axioms. I discuss this view at length in McCarthy (2006), so here I will give a brief summary. The key idea is due to Diamond (1967), who suggested that when two contestants for a good are perfectly symmetrically placed, it is better to bring about a fair lottery than to give the good directly to one of the two people. Call this the *lottery claim*. But *pace* Broome (1991), the lottery claim leads to the rejection of the claim that the betterness relation satisfies the central expected utility axiom, the so-called strong independence axiom or sure thing principle. The lottery claim is therefore inconsistent with ex post prioritarianism. And with the extended separability principle, the lottery claim leads to

Ex ante prioritarianism For any restricted individual goodness measure $g(i, h)$ there exists an increasing and strictly concave function w such that

$$P(L) = \Sigma_j \Sigma_{i \in \Pi} w(p_j g(i, h_j))$$

represents the betterness relation for any population Π.

Notice that the probability p_j lies inside the scope of the function w. The definition of ex post prioritarianism is the same, except that there the probability p_j lies outside the scope of w. This difference in position reflects a large ethical difference, manifested in Diamond's example. Ex ante prioritarianism implies that it is better to toss the fair coin, but ex post prioritarianism is indifferent. These forms of prioritarianism are equivalent when restricted to histories, entailing restricted prioritarianism and hence the restricted Pigou–Dalton principle. But this defence of the restricted Pigou–Dalton principle is not available to ex post prioritarianism, for it essentially involves the lottery claim. This defence of ex ante prioritarianism does not rely on distributive intuitionism, for the plausibility of the lottery claim is independent of whether there is any quantitative structure to individual goodness, and it essentially involves claims about risk.

However, although ex ante prioritarianism is a defensible view and was popular with economists decades before philosophers became interested in the priority view, this should be of little consolation to those who have been tempted by the ideas which go into ex post prioritarianism. The rationales that have been offered for ex post prioritarianism are very different from the rationale for ex ante prioritarianism. The appeal of ex post prioritarianism is illusory, and the closest defensible position is utilitarianism.

If my brief remarks about utilitarianism, egalitarianism and ex ante prioritarianism are correct, it follows that claims about risk lie at the heart of a wide range of reasonably plausible theories about distribution. This article has been a case study in the converse. Ignoring the complex logical relations between claims about risk and claims about distribution can lead to extremely implausible theories even when the premises of those theories are by themselves reasonably plausible. I earlier noted two tendencies in the literature about the priority view: making claims about distribution while postponing thinking about risk, and assuming that separately plausible claims about risk and distribution are jointly plausible. I suggest that while these tendencies look quite innocent, they turn out to contain serious mistakes.

APPENDIX

Proof of Theorem 1. Let $\Pi = \{1, \ldots, n\}$ be a population of size $n \geq 2$, for the case $n = 1$ is immediate from (3). Let $g(i, h)$ be a restricted individual goodness measure, and write it as $g_i(h)$. Let H_Π be the set of histories containing Π. For any $i \in \Pi$, (2) implies that $I = g_i(H_\Pi)$ is an interval of real numbers of positive length. By (2) through (5) there exists a continuous ordering $\succeq$ of I^n which is strongly separable (and satisfies the hexagon condition if $n = 2$) such that: (i) for all $h_1, h_2 \in H_\Pi$: h_1 is at least as good as h_2 if and only if $[g_1(h_1), \ldots, g_n(h_1)] \succeq [g_1(h_2), \ldots, g_n(h_2)]$. By the central theorem of additive representation (Wakker, 1989) there exist continuous functions $w^i : I \to R, i = 1 \ldots n$ such that: (ii) for all $x, y \in l^n$: $x \succeq y$ iff $\Sigma_i w^i(x_i) \geq \Sigma_i w^i(y_i)$. By (4) the w^i's are all identical; write $w_n = w^1 = \cdots = w^n$. By (3) w_n is increasing. By (6) w_n is strictly concave. By (i) and (ii) $\Sigma_{i \in \Pi} w_n(g_i(h))$ represents the restricted betterness relation for Π. Let $\tilde{g}_i(h)$ be some other individual goodness relation. It is a positive affine transformation of $\tilde{g}_i(h)$, hence $\tilde{g}_i(h) = ag_i(h) + b$. But then $\Sigma_{i \in \Pi} \tilde{w}_n(\tilde{g}_i(h))$ represents the restricted betterness relation for Π where $\tilde{w}_n(x) = w_n(\frac{x-b}{a})$. Since w_n is increasing and strictly concave, so is $\tilde{w}_n$. If Π_1 is some other population of size n, (4) entails that $\Sigma_{i \in \Pi}, w_n(g_i(h))$ represents the restricted betterness relation for Π_1. This completes the proof.

Proof of Theorem 3. Let $g(i, h)$ and I be as in the proof of Theorem 1. Let $\Pi_k = \{1, \ldots, k\}$ be a population of size $k \geq 1$, and let $\Pi_{k+1} = \{1, \ldots, k+1\}$

be the population obtained by adding person $k+1$ to Π_k. If H is some nonempty set of histories, write $\Delta(H)$ for the set of lotteries with finite support over H.

Step 1: By Theorem 1, (1) through (6) entail that for any $n \geq 1$ there exists an increasing and strictly concave function w_n such that $\Sigma_{i\in\Pi} w_n(g_i(h))$ represents the restricted betterness relation for any population Π of size n. Case (a): suppose $k \geq 2$. Let $\succeq_k$ and $\succeq_{k+1}$ be orderings on I^k and I^{k+1} respectively such that (i) for all $h_1, h_2 \in H_{\Pi_k} : h_1$ is at least as good as h_2 if and only if $[g_1(h_1),\ldots, g_k(h_1)] \succeq_k [g_1(h_2),\ldots,g_k(h_2)]$; and (ii) for all $h_1, h_2 \in H_{\Pi_{k+1}} : h_1$ is at least as good as h_2 if and only if $[g_1(h_1),\ldots,g_{k+1}\ (h_1)] \succeq_{k+1} [g_1(h_2),\ldots,g_{k+1}\ (h_2)]$. Then $\succeq_k$ and $\succeq_{k+1}$ are represented by $\sum_{i=1}^{k} w_k(x_i)$ and $\sum_{i=1}^{k+1} w_{k+1}(x_i)$. Let $c \in I$ and let $\succeq_{k+1}^c$ be a relation on I^k defined by: for all $[x_1,\ldots x_k], [y_1,\ldots,y_k] \in I^k: [x_1,\ldots x_k] \succeq_{k+1}^c [y_1,\ldots,y_k]$ iff $[x_1,\ldots x_k, c] \succeq_{k+1} [y_1, \ldots, y_k, c]$. By (8), $\succeq_{k+1}^c = \succeq_k$. Hence $\sum_{i=1}^{k} w_k(x_i)$ and $\sum_{i=1}^{k} w_{k+1}(x_i)$ represent the same relation on I^k. By the uniqueness of additive representations (see e.g. Wakker 1989) there exist $a > 0$ and b such that such that $w_k(x) = a w_{k+1}(x) + b$ for all $x \in I$. Hence for all $k \geq 2$, w_{k+1} can be obtained from w_k by a positive affine transformation. Let w be a positive affine transformation of w_2 such that $w(k) = 0$ for some $k \in I$. We have shown that $\Sigma_{i\in\Pi} w(g_i(h))$ represents the restricted betterness relation for any population Π of size $n \geq 2$. Case (b): (3) entails that $\Sigma_{i\in\Pi} w(g_i(h))$ represents the restricted betterness relation for any population Π of size one.

Step 2: By (6) and Theorem 2, for any $n \geq 1$ there exists an increasing function f_n such that $W_n(L)$ represents the betterness relation for any population Π of size n where $W_n(L) = \Sigma_j p_j f_n(\Sigma_{i\in\Pi} w(g_i(h_j)))$ for any lottery $L = [p_1, h_1: \ldots; p_m, h_m] \in \Delta(H_\Pi)$.

Step 3: Let $J = w(I)$, and note that since I is an interval of positive length and w is continuous, J is also an interval of positive length, and note also by definition of w that $0 \in J$. Now let $\succeq_k$ be an ordering on $\Delta(J^k)$ such that for any $L_1, L_2 \in \Delta(H_{\Pi_k})$: L_1 is at least as good as L_2 if and only if $\rho_k(L_1) \succeq_k \rho_k(L_2)$ where $\rho_k([p_1, h_1;\ldots;p_m, h_m]) = [p_1, [w(g_1(h_1)), \ldots, w(g_k(h_1))]; \ldots; p_m, [w(g_1(h_m)), \ldots, w(g_k(h_m))]]$. Define $\succeq_{k+1}$ similarly, and note that for any $[p_1, x_1;\ldots; p_m, x_m], [q_1, y_1;\ldots; q_{m'}, y_{m'}] \in \Delta(J^{k+1}): [p_1, x_1; \ldots; p_m, x_m] \succeq_{k+1} [q_1, y_1; \ldots; q_{m'}, y_{m'}]$ iff $\Sigma_{j=1}^{m} p_j f_{k+1}(\Sigma_{i=1}^{k+1} x_{ji}) \geq \Sigma_{j=1}^{m'} q_j f_{k+1}(\Sigma_{i=1}^{k+1} y_{ji})$. Let $M_1^+, M_2^+ \in \Delta(J^{k+1})$ be such that $M_1^+ = [\frac{1}{2}, [y, 0, \ldots, 0, x]; \frac{1}{2}, [x, 0, \ldots, 0, y]]$, $M_2^+ = [\frac{1}{2}, [x, 0, \ldots, 0, x]; \frac{1}{2}, [y, 0, \ldots, 0, y]]$ for some $x, y \in J$. Let $M_1\ M_2 \in \Delta(J^k)$ be such that $M_1 = [\frac{1}{2}, [y, 0, \ldots, 0]; \frac{1}{2}, [x, 0, \ldots, 0]]$, $M_2 = [\frac{1}{2}, [x, 0, \ldots, 0]; \frac{1}{2}, [y, 0, \ldots, 0]]$. There exist $L_1, L_2 \in \Delta(H_{\Pi_k})$, $L_1^+, L_2^+ \in \Delta(H_{\Pi_{k+1}})$ such that $\rho_k(L_1) = M_1$, $\rho_k(L_2) = M_2$, $\rho_{k+1}(L_1^+) = M_1^+$, $\rho_{k+1}(L_2^+) = M_2^+$. By (7), L_1 is at least as good as L_2 if and only if L_1^+ is at least as good as L_2^+. But L_1 and L_2 are equally good by (3) and (7). Hence L_1^+ and L_2^+ are equally good. So $\frac{1}{2} f_{k+1}(y + x) + \frac{1}{2}$

$f_{k+1}(x + y) = \frac{1}{2} f_{k+1}(x + x) + \frac{1}{2} f_{k+1}(y + y)$. So we have shown that for all $x, y \in J$, $2 f_{k+1}(x + y) = f_{k+1}(2x) + f_{k+1}(2y)$. Letting $u = 2x$, $v = 2y$ transforms this to Jensen's equation on an interval. Since f_{k+1} is increasing it is linear by Aczél (1966). Hence for any population Π of size ≥ 2, the betterness relation for Π is represented by $W(L) = \Sigma_j \Sigma_{i \in \Pi} p_j w(g_i(h_j))$. Now let $M_1^+, M_2^+ \in \Delta(J^2)$ be such that $M_1^+ = [p_1, [x_1, 0]; \ldots; p_m, [x_m, 0]]$. $M_2^+ = [q_1, [y_1, 0]; \ldots; q_{m'}, [y_{m'}, 0]]$ for some $x_1, \ldots, x_m, y_1 \ldots y_{m'} \in J$. Let M_1, $M_2 \in \Delta(J)$ be such that $M_1 = [p_1, x_1; \ldots; p_m, x_m]$, $M_2 = [q_1, y_1; \ldots; q_{m'}, y_{m'}]$. By (7), M_1 is at least as good as M_2 if and only if M_1^+ is at least as good as M_2^+. Hence $\Sigma_{j=1}^m p_j f_1(x_j) \geq \Sigma_{j=1}^{m'} q_j f_1(y_j)$ if and only if $\Sigma_{j=1}^m p_j x_j \geq \Sigma_{j=1}^{m'} q_j y_j$. Hence f_1 is linear. So for any population of Π of size 1, the betterness relation for Π is represented by $W(L) = \Sigma_j \Sigma_{i \in \Pi} p_j w(g_i(h_j))$.

Step 4: Let $\tilde{g}_i(h)$ be some other individual goodness relation. By (1), with the same notation and argument as the last part of the proof of Theorem 1, for any population of Π, the betterness relation for Π is represented by $W(L) = \Sigma_j \Sigma_{i \in \Pi} p_j \tilde{w}(\tilde{g}_i(h_j))$. This completes the proof.

REFERENCES

Aczél, J. 1966. *Lectures on functional equations and their applications.* New York: Academic Press.

Broome, J. Basingstoke: 1989. What's the good of equality? In *Current Issues in Microeconomics*, ed. J. D. Hey. Basingstoke: Macmillan. 236–262.

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

Broome, J. 2004. *Weighing lives.* New York: Oxford University Press.

Diamond, P. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment. *Journal of Political Economy* 75: 765–66.

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

Hardy, G. H., J. Littlewood and G. Polyà. 1934. *Inequalities.* London: Cambridge University Press.

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

Harsanyi, J. 1975. Nonlinear social welfare functions: do welfare economists have a special exemption from Bayesian rationality? *Theory and Decision* 6: 311–32.

Harsanyi, J. 1977. Nonlinear social welfare functions: a rejoinder to Professor Sen. In *Foundational problems in the special sciences*, eds. R. Butts and J. Hintikka. Dordrecht: Reidel, 293–296.

Hurley, S. 1989. *Natural reasons.* New York: Oxford University Press.

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

Lewis, D. 1968. Counterpart theory and quantified modal logic. *Journal of Philosophy* 65: 113–26.

McCarthy, D. 2006. Utilitarianism and prioritarianism I. *Economics and Philosophy* 22: 335–63.

McCarthy, D. 2007. Measuring life's goodness. *Philosophical Books* 48: 303–19.

McCarthy, D. (ms). Distributive equality.

Parfit, D. 2000. Equality or priority? In *The ideal of equality*, eds. M. Clayton and A. Williams. Basingstoke: Macmillan, 347–86.

Rabinowicz,   W. 2001. Prioritarianism and uncertainty: on the interpersonal addition theorem and the priority view. In *Exploring practical philosophy: from action to values*. eds. D. Egonsson, J. Josefsson, B. Petersson and T. Rϕnnow-Rasmussen. Aldershot: Ashgate.

Rabinowicz, W. 2002. Prioritarianism for prospects. *Utilitas* 14: 2–21.

Rawls, J. 1971. *A theory of justice*. Cambridge: Harvard University Press.

Risse, M. 2002. Harsanyi's 'utilitarian theorem' and utilitarianism. *Noûs* 36: 550–77.

Scheffler, S. 1982. *The rejection of consequentialism*. Oxford: Clarendon Press.

Tungodden, B. 2003. The value of equality. *Economics and Philosophy* 19: 1–44.

Wakker, P. 1989. *Additive representations of preferences*. Dordrecht: Kluwer.