Prioritarianism and Uncertainty

On the Interpersonal Addition Theorem and the Priority View

Rabinowicz, Wlodek

Published in:
Exploring Practical Philosophy

2001

Document Version:
Peer reviewed version (aka post-print)

Published in
Egonsson, Josefsson, Petersson and Rønnow-Rasmussen (eds.),
Exploring Practical Philosophy: From Action to Values,
Aldershot: Ashgate 2001, 139-165.

Prioritarianism and Uncertainty: On the Interpersonal Addition Theorem and the Priority View*

Wlodek Rabinowicz
Department of Philosophy, Lund University
e-mail: Wlodek.Rabinowicz@fil.lu.se

This paper takes its point of departure from the *Interpersonal Addition Theorem*. The theorem, by John Broome (1991), is a re-formulation of the classical result by Harsanyi (1955). It implies that, given some seemingly mild assumptions, the overall utility of an uncertain prospect can be seen as the sum of its individual utilities. In sections 1 and 2, I discuss the theorem's connection with utilitarianism and in particular (section 1) the extent to which this theorem still leaves room for the *Priority View*. According to the latter, the utilitarian approach needs to be modified: Benefits to the worse off should count for more, overall, than the comparable benefits to the better off (cf. Parfit 1995 [1991]).

Broome (1991) and Jensen (1996) have argued that the Priority View cannot be seen as a plausible competitor to utilitarianism: Given the addition theorem, prioritarianism should be rejected for measurement-theoretical reasons. I suggest, in section 3, that this difficulty is spurious: The proponents of the Priority View would be well advised, on independent grounds, to reject one of the basic assumptions on which the addition theorem is based. I have in mind the so-called *Principle of Personal Good* for uncertain prospects. If the theorem is disarmed in this way, then, as an added bonus, the Priority View disposes of the aforementioned problems with measurement.

According to the Principle of Personal Good, one prospect is better than another if it is better for everyone or at least better for some and worse for none. That the Priority View, as I read it, rejects this welfarist intuition may be surprising to the reader. Isn't welfarism a common ground for prioritarians and utilitarians? Still, as I will argue, the appearances are misleading: The welfarist common ground is better captured by a restricted Principle of Personal Good that is valid for *outcomes,* but not necessarily for uncertain prospects. As will become clear, we obtain this surprising result if we take the priority weights imposed by prioritarians to be relevant only to *moral*, but not to *prudential*, evaluations of prospects. This makes it possible for a prospect to be morally better (i.e. better overall), even though it is worse (prudentially) for everyone concerned. The proposed interpretation of the Priority View thus drives a sharp wedge between prudence and morality. In section 4, I will argue that this

* For comments and discussion I am much indebted to Gustaf Arrhenius, John Broome, Johan Brännmark, Krister Bykvist, Erik Carlson, Roger Crisp, Sven Danielsson, Dan Egonsson, Marc Fleurbaey, Magnus Jiborn, Mats Johansson, Karsten Klint Jensen, Philippe Mongin, Erik Persson, Ingmar Persson, Toni Rönnow-Rasmussen and Derek Parfit. Apart from Broome and Parfit, who figure so prominently in what follows, I have special debts to Jensen and to Ingmar Persson. Pondering on the former's thought-provoking PhD-thesis led me to the idea of this paper, and the latter not only gave me some very good advice, but also made available a copy of Parfit's seminal paper on the Priority View, which by that time was still quite difficult to come by.

divergence between moral and prudential evaluations should be recognized by prioritarians even for Robinson-type cases, in which there is only one person to consider. In sections 3 and 4, I will also contrast the prioritarian morality, on which each person's welfare makes a separable contribution to the overall good, with egalitarianism, which denies such separability.

Finally, in section 5, I will discuss some underlying conceptual commitments of my interpretation of the prioritarian view. Since this interpretation takes very seriously the distinction between uncertain prospects and uncertainty-free outcomes, it goes against the standard decision-theoretical view according to which the distinction in question is more or less provisional and motivated by practical convenience.

1. Interpersonal Addition

Broome's Interpersonal Addition Theorem is inspired by a formally similar aggregation theorem, due to Harsanyi (1955). While Harsanyi was concerned with aggregation of individual *preferences*, Broome (1991) considers aggregation of individual *betterness* orderings.[1] To state his theorem, we need some preparations. Suppose we start from

— a finite set **I** of *individuals*, $\{i_1, \ldots, i_n\}$,
— a finite partition **Σ** of mutually exclusive and jointly exhaustive *states of nature*, $\{S_1, \ldots, S_m\}$, where it may be uncertain which state in fact obtains,
— a set **O** of possible *outcomes,* where an outcome, intuitively, is a specification of what happens to each individual, with respect to the factors that are relevant to his/her welfare.

An (uncertain) *prospect* is any assignment of outcomes to the states of nature. For each possible state in **Σ**, a prospect specifies an outcome in **O** that would be realized if that state were to obtain. A prospect may be seen as a kind of lottery in which outcomes are possible prizes and the actual prize depends on which state happens to obtain. We can represent a prospect x as a vector, $x = (o_1, \ldots, o_m)$, where o_1 is the outcome that results on this prospect if state S_1 obtains, o_2 is the outcome that results if S_2 obtains, etc.

We assume,
— for each individual i in **I**, an ordering B_i of prospects that specifies, for any two prospects, which of them is better for i or whether they are equally good for that individual.

Suppose also, in addition, that prospects are comparable in an impersonal way. That is, there exists
— an ordering B of prospects that specifies, for all prospects, which of them is overall better or whether they are overall equally good.

Thus, apart from the set of *individual* (or *personal*) betterness relations on prospects, one for each individual, there is also an *impersonal*, or - to use another label - *overall* betterness relation on prospects. Note that these betterness relations on prospects indirectly order outcomes as well, since any outcome o may be associated with the "safe" prospect $(o, \ldots, o)$, which assigns this outcome to each state of nature. The ordering of safe prospects induces the corresponding ordering of outcomes.

Suppose we make the following assumptions about the betterness relations on prospects:

---

[1] In addition to this philosophical difference, there is a technical difference as well: While Broome's betterness orderings range over uncertain prospects (= assignments of outcomes to states of nature), Harsanyi's preference orderings range over von Neumann-Morgenstern lotteries, i.e., over probability distributions on outcomes.

(P1) Each *personal* betterness relation B_i satisfies the axioms of expected utility theory. Thus, B_i is representable by a utility function u_i on prospects and a probability distribution p_i on states of nature, where u_i is expectational with respect to p_i and as such represents i's betterness relation uniquely up to positive linear transformations.

Similarly,

(P2) The *overall* betterness relation B satisfies the axioms of expected utility theory. Thus, B is representable by a utility function u on prospects and a probability distribution p on states, where u is expectational with respect to p and as such represents the overall betterness relation uniquely up to positive linear transformations.

That a utility function *represents* an ordering of prospects means that it assigns higher utility values to better prospects. It is *expectational* if the utility value it assigns to a prospect is the weighted sum of the utilities it assigns to its possible outcomes under various states[2], with the weights being the probabilities of the states in question. (Given appropriate axioms on the underlying ordering of prospects, the probabilities of the states are uniquely determinable from that ordering.) Finally, such an expectational utility representation is *unique up to positive linear transformations* if all expectational functions that represent the same betterness ordering are positive linear transformations of each other. As such, they differ at most by the choice of the zero point and of the unit of measurement.

As the last assumption for the theorem, suppose that the overall betterness ordering of prospects is positively dependent on the individual betterness orderings:

(P3) *Principle of Personal Good:*
(a) Prospects that are equally good for each individual are equally good overall;
(b) If a prospect x is at least as good for each individual as a prospect y and if it is better than y for some individual(s), then x is overall better than y.

The Principle of Personal Good is based on the intuition that overall good is a function of the personal good of the individuals, and of nothing else (clause (a)). What's more, this function is strictly increasing in each argument (clause (b)): Making a prospect better for some without making it worse for anyone else always makes the prospect better overall.

We are now ready to state the theorem:

Interpersonal Addition Theorem:
P1, P2, P3 $\Rightarrow$
If an expected utility function u represents the overall betterness relation B, then there are expected utility functions $u_1, \ldots, u_n$ that represent the individual betterness relations $B_1, \ldots, B_n$, respectively, such that u is the *sum* of $u_1, \ldots, u_n$:

$$u(x) = u_1(x) + \ldots + u_n(x), \text{ for all prospects } x.$$

This looks very much like *utilitarianism*, according to which the overall goodness of a prospect is the sum of its goodness values (welfare values) for each individual. That we should arrive at utilitarianism in this way is quite astonishing since the assumptions of the theorem seem to be relatively innocuous while utilitarianism is a deeply controversial view. However, as Broome argues, the appearances are misleading. The theorem, as it stands, is not about goodness but about utility. To be sure, the utility function u_i represents the individual betterness ordering B_i, which means that it orders prospects according to how good they are for a given individual. But such a utility function may still not be a proper measure of the

[2] We let the utility of an outcome o be the same as the utility of the corresponding safe prospect $(o, \ldots, o)$.

individual goodness of prospects. It may instead be a strictly increasing but *non-linear* transformation of the underlying goodness function:

For all prospects x, $u_i(x) = w(g_i(x))$.

In this equation, g_i is the goodness function for i and the transformation w of g_i is supposed to be increasing but non-linear: the curve for this transformation slopes upwards but not in a straight line. (For example, w may be a root function, provided that the g_i-values are all non-negative, or a logarithmic function.) That w is increasing, i.e., that

$w(g_i(x)) > w(g_i(y))$ if and only if $g_i(x) > g_i(y)$,

implies that u_i orders the prospects in the same way as g_i, so that both functions represent the personal betterness relation B_i:

$u_i(x) > u_i(y)$ if and only if $g_i(x) > g_i(y)$ if and only if xB_iy.

However, the non-linearity of the transformation w would entail that g_i, unlike u_i, is *not* an expectational function. Remember that u_i was supposed to be unique up to positive linear transformations. Which implies that u_i is a linear transformation of each *expectational* function that orders the prospects in the same way as u_i does.

This means that the derivation of utilitarianism from the Interpersonal Addition Theorem would require an extra assumption. We need to assume

Bernoulli's Hypothesis: Individual goodness is an expectational function.

I.e., the individual goodness of a prospect is the probability-weighted sum of the individual goodness values of its possible outcomes. Given P1, Bernoulli's hypothesis is equivalent to the claim that each individual utility function u_i that appears in the equation $u(x) = u_1(x) + ... + u_n(x)$ is identical with the goodness function for i up to a linear transformation.

In the absence of Bernoulli's hypothesis, Broome argues, we have not yet established utilitarianism. Without that extra assumption, there is room for other theories of the good, such as the Priority View that has been put forward by Derek Parfit (1995 [1991]). The Priority View distinguishes between how good a situation is for an individual and the *contribution* that this individual goodness makes to the overall goodness of the situation. The contribution is positive but not linear according to prioritarians: increased individual benefits have a successively decreasing impact on the overall goodness of a situation. On Broome's interpretation, then, the Priority View accepts the three assumptions of the theorem but denies Bernoulli's hypothesis. On that view, the expectational function u_i measures the contribution made by individual goodness but not the individual goodness itself. The former is supposed to be a non-linear transform of the latter, which means that the latter must be non-expectational, given the addition theorem.

While Broome admits the Priority View as a theoretical option, he is quite skeptical about its viability (cf. Broome 1991, p. 217). Jensen (1996) develops this line of criticism. Roughly, the difficulty with the Priority View is that its defense would require providing an *independent* method of measuring individual goodness. We must be able to measure the latter in some other way than the one we use to measure individual utility. But the two measures would still have to coincide in their ordering of prospects! That an independent order-preserving measure of goodness can be found is doubtful, to say the least.

On the standard measurement-theoretic view, quantitative measures are nothing more than representations of the underlying qualitative orderings and the measures g_i and u_i are supposed to *coincide* in their ordering of prospects. Therefore, the claim that these two measures

essentially differ from each other can be meaningful only if the difference between them can somehow be made good in qualitative terms, when we turn our attention from simple prospect orderings to some more comprehensive or more complex qualitative structures. Suppose two such distinct structures give rise to the same prospect ordering and the prospect measures g_i and u_i are each derived from some numerical representation of its corresponding structure, without being derivable from any numerical representation of the other structure. *Then, and only then,* the two measures may be said to be essentially different. But the difficulty is that it is unclear what the relevant qualitative structures might be.

To find them, we might consider some more complex ordering relations. In particular, we could distinguish between two 'difference orderings' of prospects:

(i) an ordering R_i of the individual welfare differences:

> the change from *x* to *x'* is better for *i* than the change from *y* to *y'*;

(ii) an ordering R'_i of the differences in contributions made by individual welfare:

> the change in *i*'s welfare from *x* to *x'* contributes more to the overall goodness than the change in *i*'s welfare from *y* to *y'*.

In other words, while R_i compares changes in *i*'s welfare, R'_i compares the contributions these changes make to overall goodness. Suppose now that R_i and R'_i are non-equivalent orderings: for some prospects *x*, *x'*, *y* and *y'*, the change from *x* to *x'*, as compared with the change from *y* to *y'*, gives *i* a larger increment in welfare, but this larger increment makes a smaller contribution to overall goodness. Suppose, however, that R_i and R'_i still yield the *same* simple ordering of prospects: the two difference orderings coincide whenever $x' = y$ and $y' = x$.[3] That is, an increment in *i*'s welfare always makes a positive contribution to the overall value of a prospect. Still, if the functions g_i and u_i come, respectively, from the difference measures G_i and U_i that represent these two distinct underlying difference orderings R_i and R'_i[4], and if u_i is not just a linear transformation of g_i, then we could argue that individual goodness and its contribution to overall goodness are non-equivalent concepts. The trouble with this approach, however, is that the comparisons needed to determine both difference orderings are quite demanding. It is by no means clear that we could have access to such sophisticated comparisons to the extent that is needed to distinguish between individual goodness and its contribution to the overall value of a prospect.[5]

---

[3] Defining simple prospect orderings from the corresponding difference orderings is straightforward. Thus, a prospect *x* is better for *i* than a prospect *y* iff the change from *x* to *y* is better for *i* than the change from *y* to *x*. Analogously, *i*'s welfare makes in *x* a larger contribution to overall goodness than in *y* iff, as far as *i*'s welfare is concerned, the change from *x* to *y* contributes more to the overall goodness than the change from *y* to *x*.

[4] Let G_i be any value assignment to pairs of prospects such that $G_i(x,x') > G_i(y, y')$ if and only if the change from *x* to *x'* ranks higher in R_i (i.e., this change is better for *i*) than the change from *y* to *y'*. Set $G_i(x, x)$ to 0, for any *x*. U_i is constructed in the same way, as a representation of R'_i. We define a prospect measure g_i from G_i as follows: Let *x** be an arbitrary prospect. Then, for every *x*, $g_i(x) = G_i(x^*, x)$. The definition of u_i from U_i is analogous.

[5] Note, however, that Broome has recently become much more sympathetic to the possibility of drawing such fine distinctions. He now admits that it makes good sense, at least *conceptually*, to distinguish between a person's good and how much that person's good counts in the overall evaluation (cf Broome, 1999).

2. Interpersonal Comparisons and Probability Agreement

Broome's own view is that we should accept Bernoulli's hypothesis. If we do so, we move from the Interpersonal Addition Theorem to a full-fledged utilitarianism.[6]

As a matter of fact, I do not think we can get full utilitarianism that easily, simply by accepting the three assumptions of the theorem together with Bernoulli's hypothesis. The Interpersonal Addition Theorem states that for each expectational representation u of B, there are *some* expectational representations $u_1,\ldots., u_n$ of $B_1, \ldots B_n$ that sum up to u. Now, even if each u_i in the sum $u = u_1 + \ldots. + u_n$ is just a linear transform of the corresponding g_i, it is still possible that we need to use *different* linear transforms for different individual goodness functions in order to obtain such a simple additive formula. For example, suppose that the transformations in question are as follows:

$u_1 = 2g_1$, while for all $i \neq 1$, $u_i = g_i$.

Then we have:

$u = 2g_1 + g_2 + \ldots + g_n$.

In other words, in the calculation of the overall utility, individual *1* counts twice as much as anyone else. This is, of course, alien to the utilitarian way of counting, according to which each individual is to count as one.[7] Still, the assumptions of the theorem together with Bernoulli's hypothesis do not suffice to exclude this anti-utilitarian possibility. Something more is needed. What could it be? What additional assumption would do the job?

Well, we would be home if we could assume *interpersonal* betterness comparisons. Suppose there exists an interpersonal betterness ordering of prospects that specifies, for all prospects x and y and for all individuals i and j, whether x is better or worse for i than y is for j, or whether

---

[6]At least for those cases in which the set of individuals can be assumed to be *fixed* as one moves from one prospect or outcome to another. If different individuals are allowed to exist in different outcomes that are being compared with each other, the situation becomes much more complicated. We shall ignore this complication in what follows. We shall also ignore a number of other complications:
(i) Utilitarianism, as usually understood, contains a normative component: Apart from the welfarist specification of the relationship between individual welfare and overall goodness, it also involves a consequentialist *injunction* to act so as to realise what is best overall.
(ii) Utilitarian theories often involve a list of one or more specific factors that are supposed to determine the orderings of individual betterness, i.e., such factors as happiness, preference satisfaction or, in a more objective vein, standard of living, capabilities, personal relationships, etc.
(iii) Some utilitarians decline to assign overall value to uncertain *prospects,* as opposed to outcomes. (On this view, the normative status of an action depends on the relative value of the outcome that the action would *actually* result in, as compared with the corresponding outcomes of its alternatives, rather than on the relative value of the prospect it is associated with.)

[7]Objection: But if individual *1* is "counted twice" in this way, then the resulting function u will not be right. It will not represent the correct utilitarian overall betterness ordering. We will be able to find two prospects x and y such that x is overall better than y, but the u-function ranks them in the reverse order. The Interpersonal Addition Theorem does not claim that *every* choice of utility representations for individual betterness relations gives us utility functions the sum of which represents overall betterness, but only that some choice like this is possible. If utility representations for different individuals are chosen independently of each other, the sum of utility values will normally *not* represent overall betterness.

Answer: The point of the difficulty raised in the text is different, namely: What is there to guarantee that the overall betterness ordering, which u is supposed to represent, does *not* count some individuals "twice"? The assumptions of the theorem, in conjunction with the Bernoulli hypothesis, do not exclude this anti-utilitarian betterness ordering.

x is as good for *i* as *y* is for *j*. In terms of this underlying *extended* ordering, which really is an ordering of prospect-individual pairs,[8] the different individual orderings can be easily defined. We define i's betterness ordering from the interpersonal ordering as follows:

x is better for *i* than *y* iff *x* is better for *i* than *y* is for *i*.

The definition of "equally good for *i* as" is analogous. We can now impose an impartiality condition on the relationship between overall betterness and interpersonal betterness, from which it follows that any two individuals *i* and *j* count equally from the overall point of view:

> *Impartiality:* For all prospects *x* and *y*, and for any permutation π on the set **I** of individuals, if for all individuals *i*, *x* is as good for *i* as *y* is for $\pi(i)$, then *x* and *y* are equally good overall.

This excludes the possibility that some individuals count for more than others. But to formulate such an impartiality condition, we need to rely on interpersonal betterness comparisons, which Broome (1991) wanted to avoid. When he wrote *Weighing Goods,* he thought that Bernoulli's hypothesis gives us all we need. To fill in the remaining gap between the Interpersonal Additional Theorem and utilitarianism, we should simply deny that there can be any meaningful difference between how good a prospect is for an individual and how much its goodness for that individual *counts* for its overall goodness. Consequently, the interpersonal comparison is achieved as soon as we find the individual utility functions that add up to overall utility. These individual utilities give us interpersonal comparisons. (Cf. ibid., pp. 215-20.)[9]

However, Broome has recently changed his views on this issue:

> …in *Weighing Goods* I offered the wrong account of the meaning of interpersonal comparisons of good. [I suggested that] the size of a benefit [i.e., the size of an increment in individual goodness] is nothing other than the amount the benefit counts in determining the general [= overall] goodness. […] I now think this whole approach to interpersonal comparisons of good must be mistaken. My reason is that it makes good sense to say it is better for some given amount of good to come to one person than to another. […] We can make a distinction, then, between an amount of good and how much that amount counts in general [i.e., overall] good. (Broome, 1999, section 3.3)

The distinction in question is thus possible to make, at least conceptually. In personal communication, Broome has made the same point as follows:

> […] I no longer think that line [from *Weighing Goods*] is successful, because there is a clear difference between a person's good and how much that person's good counts in overall evaluation (between internal and external value, as I now put it). So in my present book [Broome, 1999] I have a different account of interpersonal comparison, which means I need the impartiality assumption [for his formulation of this assumption, cf. Broome, 1999, section 7.1].

In what follows, I shall assume that the problem of interpersonal comparability has been dealt with in a satisfactory way. We may suppose that all personal betterness orderings come from the same underlying extended betterness ordering, which means that each person's goodness is measurable on a common scale. I shall also assume that some form of the impartiality condition

---

[8] Thus, in symbols, $(x, i) \geq (y, i)$ iff *x* is at least as good for *i* as *y* is for *j*.

[9] For a useful discussion of the similar issue of interpersonal comparability in connection with Harsanyi's aggregation theorem, cf. Mongin (1994) and Mongin & d'Aspremont (1998).

is satisfied by overall betterness.[10] In what follows, however, I shall not dwell upon these issues anymore. But we should be aware of the problems that are thereby swept under the rug.

There is another problem I will sweep under the rug: the one that concerns probabilities. Given the Principle of Personal Good, it can be shown that the personal probability distributions p_i on states cannot differ from each other: they must all coincide with the probability distribution p that can be elicited from the impersonal betterness relation on prospects (cf Broome 1991, ch. 7). As Broome points out, this *probability agreement theorem* is a singularly welcome result. The individual *betterness* ordering of prospects, as opposed to an ordering that reflects that individual's *preferences,* should not depend on that person's subjective probability assignment: Instead, it should be a ranking that is based on a probability distribution that does not differ from one individual to another. Unlike individual preferences, individual betterness orderings of prospects do not depend on private and idiosyncratic probability assignments to states.

Thus, the probability agreement theorem gives us just what we need. However, the theorem essentially depends on the Principle of Personal Good and the validity of this principle will be questioned in what follows. Therefore, we need some other way to make sure that the different personal betterness relations are based on a common probability distribution. One such way would be to derive them all from the underlying extended betterness ordering of prospect-individual pairs, as has been suggested above. All of them would then depend on the same probability distribution on states – the one that can be elicited from the extended ordering that underlies them all. Another way would be to use the classical von Neumann-Morgenstern axiomatization of expected utility theory, in which the objects of comparison are not uncertain prospects but objective lotteries, with explicitly specified probabilities of outcomes. But again, in what follows, we shall keep this problem under the rug.

3. The Priority View – Two Interpretations

As stated in the introduction, I want to concentrate on the interpretation of the Priority View. To clarify the difference between that view and utilitarianism, it is best, I think, to begin with their respective ways of evaluating *outcomes*. Evaluation of uncertain prospects is a question to which we shall come back later. Forget now about the Interpersonal Addition Theorem for a moment and suppose we try to determine how good an outcome is, overall. Assume that we take the overall goodness of an outcome to be positively dependent on how good this outcome is for each individual. That is,

$g(o) = f(g_1(o), ..., g_n(o))$,

where f is increasing in each of its arguments.

When a utilitarian evaluates an outcome, he considers how good this outcome is for each particular person and then simply adds these individual values:

Utilitarianism for Outcomes: $g(o) = g_1(o) + ... + g_n(o)$

Thus, for a utilitarian, the function f is simple addition. It is different with the Priority View. According to a prioritarian, *in the determination of the overall good, the benefits to the worse*

[10] This remark should be qualified. Note that the impartiality assumption, as formulated above, implies as a special case clause (a) of the Principle of Personal Good. (To derive this clause from Impartiality, just let the permutation π be the identity relation.) Since I will be arguing that the Priority View, if correctly interpreted, accepts the Principle of Personal Good only for outcomes and not for prospects, the impartiality assumption must also be restricted to outcomes, in order to be acceptable for prioritarians.

off count for more than the benefits to those who are better off. As a result, the welfare level of a worse off person is given a higher moral weight in the aggregation:[11]

Priority View for Outcomes: $g(o) = t(w(g_1(o)) + \ldots + w(g_n(o)))$,

where t is some increasing transformation, and the moral weight function w is chosen in such a way that the marginal contribution of increments in individual goodness is always positive but *decreasing*, i.e., w is strictly increasing and strictly concave.[12] On some versions of the Priority View, it may also be assumed that the marginal contributions of such individual increments converge to zero as the individual's goodness level increases to infinity. As for the transformation t, the simplest solution is to let t be the multiplication with 1 (or, what amounts to the same, to remove t altogether). Then the overall goodness of an outcome will just be the sum of its morally weighted individual goodness values. However, as will be seen in section 4, the transformation t might also take a non-linear form. Still, we shall argue that the simplest solution is also the right one.

Unlike utilitarianism, the Priority View *hinders unrestricted interpersonal compensations*: it makes it more difficult, and sometimes outright impossible, to justify sacrificing the worse off for the benefit of the better off. [13] At the very least, the strict concavity of the weighting function w has this implication: If an amount of welfare is transferred from a worse-off person and distributed among the better off as additional increments, the result will always be worse overall, since the marginal contribution of such increments is decreasing.

---

[11] This describes what Parfit (1995 [1991]) calls the *moderate* (teleological) version of the Priority View. He suggests that the extreme form of prioritarianism is Rawls' difference principle, which gives *lexical* priority to the improvements for the worse off, and not just a greater weight. For the reasons to be explained below, at the end of this section, I don't think this is quite right, but, in what follows, we shall concentrate on the moderate version. Also, we will not consider *deontological* versions of the Priority View, according to which giving priority to the worse off is a normative requirement on action, which need not have any direct connection with the overall value of the resulting outcome.

[12] Due to the non-linearity of w, the Priority View for Outcomes might appear to pre-suppose that the individual goodness of an outcome is measured on a common *ratio* scale, rather than just on a mere interval scale. For it is easy to see that the comparisons between the sums of w-weighted individual goodness values are not invariant with respect to the changes in the zero point of the scale, if w is non-linear. Suppose, for example that $w = \sqrt{}$ and consider two outcomes, *o and o'*, with just two individuals involved, *i* and *j*. Let $g_i(o) = 0$, $g_j(o) = 16$, while $g_i(o') = g_j(o') = 4$. With this representation of individual goodness, both outcomes are equally good overall from the prioritarian perspective: $\sqrt{0} + \sqrt{16} = \sqrt{4} + \sqrt{4}$. But if we move down the zero point of the goodness scale by one unit, i.e., if we add one unit to each g_i- and g_j-value, the Priority View for Outcomes implies that *o* is overall better than *o'*. Thus, it appears that a prioritarian cannot allow the choice of the zero point for individual goodness to be arbitrary.

However, this argument assumes, somewhat questionably, that the shape of the weight function w is fixed *independently* of our choice of the numerical representation for individual goodness. If the weight function instead is allowed to undergo appropriate compensatory adjustments as we move from one such representation to another, the need for an absolute zero for individual goodness is obviated. Thus, in our example, moving down the zero point of the individual goodness scale by one unit may be accompanied by an appropriate adjustment in the weight function: instead of $\sqrt{k}$, we might let $w(k) = \sqrt{(k-1)}$, for all values k. This adjustment in the weight function will cancel out the effect of the scale transformation. (I owe this observation to Magnus Jiborn.)

[13] *If* the weighting function is such that the marginal contributions of all increments in personal goodness approach zero (or any other fixed value) when the goodness level increases, then the imposition of weights not only hinders impersonal compensations – it sometimes makes them impossible. If the worse off are much worse off than the better off, then, for a fixed number of individuals, we will not be able to compensate a considerable loss to the former by *any* gains to the latter, *however large.*

What would a proponent of the Priority View say about the individual goodness of *prospects*? How good is a prospect $x = (o_1, \ldots, o_m)$ for an individual i? I would suggest that, for a prioritarian, the goodness of a prospect for i is simply its expected goodness for i:

Prioritarian Individual Goodness of Prospects:

$$g_i(x) = \Sigma_{k=1,\ldots,m} P(S_k) g_i(o_k).$$

Thus, my suggestion is that *for a proponent of the Priority View, individual goodness is expectational* – Bernoulli's hypothesis is satisfied.

The Priority View on *Broome's* interpretation would have a different formula for the evaluation of the individual goodness of prospects. The prioritarian connection between individual utility and individual goodness, for both prospects and outcomes, is according to Broome given by the formula: $u_i = w(g_i)$. Consequently, we get the following derivation:

$u_i(x) = \Sigma_{k=1,\ldots,m} P(S_k) u_i(o_k)$,

which means that

$w(g_i(x)) = \Sigma_{k=1,\ldots,m} P(S_k) w(g_i(o_k))$.

Let m be the *inverse* of w, i.e., m is the function such that, for any real number r, $m(w(r)) = r$. If we now apply the transformation m to both sides of the equation above, we get the following result.

Prioritarian Individual Goodness of Prospects – Broome's version:

$$g_i(x) = m(\Sigma_{k=1,\ldots,m} P(S_k) w(g_i(o_k))).$$

Note that this formula for the *individual* goodness of prospects relies on the function w, which is the same concave moral weighting function that the proponents of the Priority View use in their evaluation of the *overall* goodness of outcomes. Thus, on Broome's reading of prioritarianism, and contrary to my own proposal, moral weights have two roles: They are used not just in the determination of the overall value of an outcome from its individual values, but also in the determination of the individual value of an uncertain prospect from the individual values of its possible outcomes.

Who is right? In defense of my proposal, I would like to point out that, for the proponents of the Priority View, the decreasing weight of individual goodness is essentially an expression of a *moral* concern. The improvements for the worse off are given moral priority as compared with the improvements for the better off. Easy interpersonal compensations are thereby disallowed. I take this view to be a reaction to the well-known Rawlsian objection: The trouble with utilitarianism, says Rawls, is that it "does not take seriously the distinction between persons" (Rawls 1971, p. 27). A utilitarian takes the view of an impartial spectator who sympathetically identifies with all persons and thereby fuses them all into one:

> For it is by the conception of the impartial spectator and the use of sympathetic identification that the principle [of rational choice] for one man is applied to society. It is this spectator who is conceived as carrying on the required organisation of the desires of all persons into one coherent systems of desire; it is by this construction that many persons are fused into one." (ibid., p. 26)

Because "many persons are fused into one", the principle of rational choice for one person can be applied to the society viewed as a unit. 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. But if this diagnosis is right, i.e., if prioritarianism is driven by a concern for the distinctness of persons, then the priority weights should only be

used in an *inter*personal but not in *intra*personal balancing of benefits and losses. In particular, these weights should not be used when we ask whether, from an ex ante perspective, an individual's loss in one possible outcome is compensated, *for that same individual*, by his gain in another possible outcome. Consequently, we should not use priority weights when we calculate the individual goodness of an uncertain prospect.[14] The individual value of a prospect can be identified with the simple expectation of the individual value of the resulting outcome.

Now, given this purely expectational interpretation of individual goodness, it turns out that the Priority View must reject one of the central assumptions of the Interpersonal Addition Theorem – the Principle of Personal Good. Here is an illustration of this point, with two individuals, *i* and *j*, and two equiprobable states of nature, S_1 and S_2. Suppose we compare the following two prospects with each other:

	Prospect *x*		Prospect *y*		
	S_1	S_2	S_1	S_2	$P(S_1) = P(S_2) = ½$
i	10	10	16	5	
j	10	10	5	16	

The values in the matrices specify how good the different outcomes are for each person, *i* and *j*, respectively. In prospect *y*, as compared with prospect *x*, *i*'s welfare is increased by 6 units in S_1 and decreased by 5 units in S_2, while for *j* it is the other way round. In each case, the increase is larger than the decrease, but – if we use priority weights – we may suppose that the 5-unit loss for a person detracts more from the overall goodness of an outcome than the 6-unit gain. Thus, $w(10) - w(5) > w(16) - w(10)$. For each state, then, prospect *x* yields an outcome that is better overall than the corresponding outcome of prospect *y*. If we make a plausible assumption that overall betterness satisfies *dominance* (= "a prospect that results in a better outcome given each state, is better"), it follows that

Prospect *x* is overall better than prospect *y*.

On the other hand, when it comes to individual betterness,

Prospect *y* is better for *i* than prospect *x*,

since its expected goodness for each individual is greater, and we have assumed that the individual goodness of a prospect is just its expected individual goodness. Analogously,

Prospect *y* is better for *j* than prospect *x*.

Thus, we get a counter-example to the Principle of Personal Good:

Prospect *y* is overall worse than prospect *x*, even though *y* is better than *x* for each individual.[15]

This is a counter-example to clause (b) of the Principle of Personal Good: Prospect *y* is overall worse than prospect *x*, even though it is better than *x* for each individual. Can we set up

---

[14] In personal communication, Derek Parfit has confirmed that this is how he himself would interpret the Priority View. The prioritarian weights are moral, not prudential. Therefore, they have no role to play for the individual goodness of prospects.

[15] Roger Crisp has suggested an alternative treatment of this example (in private communication). Like myself, and unlike Broome, Crisp takes prospect *y* to be better for each individual than prospect *x*, from the prioritarian point of view, but he suggests that prioritarians should keep the Principle of Personal Good for prospects intact. Therefore, they must conclude that y is overall better than *x*, even though *y* yields a worse outcome than *x*, overall, under each state of nature. Which means that the prioritarians must reject *dominance*: A prospect is better even though its outcome is worse under each state. Needless to say, this is a very radical suggestion – too radical in my view.

a counter-example to clause (a) as well? Certainly. Just change the example so as to equalise the gains and losses in individual goodness. (For example, replace in the matrix for prospect *y* both occurrences of 16 by 15.) Then prospect *y* is as good as *x* for each individual but *x* still is better than *y* overall: individual losses in each of *y*'s outcomes detract more from the overall goodness of the outcome than the equal-sized individual gains.

Note that this prioritarian counter-example to the Principle of Personal Good only concerns the application of that principle to (uncertain) *prospects*. For *outcomes,* the principle is fully valid. If an outcome *o* is better than an outcome *o'* for some individuals and as good as *o'* for everyone else, the Priority View will imply that, overall, *o* is better than *o'*. Since the overall goodness of an outcome is an increasing function of its priority-weighted individual goodness values, the overall goodness of an outcome will increase when its individual goodness for some person increases. Similarly, if two outcomes are equally good for everyone, they are equally good overall. Thus, the Priority View can still be seen as a fundamentally welfarist position, at least as far as the evaluation of outcomes is concerned. It is only when the Principle of Personal Good is applied to uncertain prospects that the Priority View starts having problems with this principle.

This means that a proponent of the Priority View can escape Broome's and Jensen's criticisms. If individual goodness is expectational, it can be measured in the standard way – in the way we measure utility. There is no need for a measurement of individual goodness that would be independent from the measurement of individual utility. We can directly elicit individual goodness values from the betterness orderings on risky prospects. But the Interpersonal Additional Theorem no longer follows, since one of its principal assumptions, the Principle of Personal Good for prospects, has been rejected.

If we let the individual goodness of a prospect be its expected individual goodness, then it seems natural, if not mandatory, to do the same for overall goodness, i.e., to let the overall goodness of a prospect $x = (o_1, \ldots, o_m)$ be its expected overall goodness:

$$g(x) = \Sigma_{k=1,\ldots,m} P(S_k)g(o_k).$$

Thus, on this reading of prioritarianism, overall goodness will be just as expectational as individual goodness. Measuring overall goodness thus boils down to measuring overall utility. The latter can be identified with the former, up to a positive linear transformation.

In section 1 above, we have seen that prioritarians want to distinguish between a person's good in a prospect and the contribution that person's good makes to the overall goodness. The person's good is represented by function g_i. To represent its contribution, we need some preparations. As we have suggested above, the overall goodness of a prospect is the probability-weighted sum of the overall goodness values of its possible outcomes under various states. The overall goodness of an outcome is an increasing transform of the sum of its morally weighted individual goodness values. If we simplify for a moment and assume that the increasing transform *t* just consists in the multiplication with 1 (in section 4 it will be argued that such a simplification is well-motivated), we obtain the following formula for the overall goodness of a prospect $x = (o_1, \ldots, o_m)$:

$$g(x) = \Sigma_k P(S_k)g(o_k) = \Sigma_k P(S_k)t(\Sigma_i w(g_i(o_k))) = \Sigma_k P(S_k)\Sigma_i w(g_i(o_k)) = \Sigma_i \Sigma_k P(S_k)w(g_i(o_k)).$$

This allows us to separate the contributions made by each individual's welfare to the overall value of a prospect:

$$c_i(x) = \Sigma_k P(S_k)w(g_i(o_k)).$$

This individual contribution to the overall value of x is the probability-weighted sum of the contributions made by i's welfare to the overall values of the possible outcomes of x, under various states. The overall value of x is the sum of such individual contributions:

$g(x) = \Sigma_i\, c_i(x)$.

Note that the function c_i, which measures the individual contribution to overall value, is expectational, just like function $g_i(x) = \Sigma_k P(S_k) g_i(o_k)$, which measures individual goodness. The contribution of i's welfare to the overall value of a prospect equals its expected contribution to the overall value of the resulting outcome:

$c_i(x) = \Sigma_k P(S_k) c_i(o_k)$.[16]

However, the two functions g_i and c_i are *not* just positive linear transformations of each other, since the two prospects orderings that are represented by these functions do not coincide. For a prioritarian who rejects the Principle of Personal Good, a larger increment in individual goodness of a prospect can make a smaller contribution to overall goodness. In our example above, a prospect y is better for i than a prospect x,

$g_i(y) > g_i(x)$,

but the contribution made by i's welfare to the overall value of y is smaller than the corresponding contribution to the overall value of y,

$c_i(y) < c_i(x)$.[17].

For the other individual, j, the case is analogous.

That's how my interpretation of the Priority View is supposed to work. But is such an interpretation plausible? I have been arguing that a prioritarian will not use moral weights in the *intra*personal balancing of benefits and losses in various possible outcomes. Moral weights have no role to play in the determination of the individual goodness of a prospect. But consider the following objection to my claim. On the Priority View, being worse off is morally bad. In this context, however, 'being worse off' is not meant to refer to a relation between one individual and another. Rather, it refers to a relation between how an individual is and how he might have been. In this respect, the view in question differs from *egalitarianism*.[18] As Parfit puts it:

> […] on the Priority View, we do not believe in equality. We do not think it in itself bad, or unjust, that some people are worse off than others. [...] We do of course think it bad that some people are worse off. But what is bad is not that these people are worse off than *others*. It is rather that they are worse off than *they* might have been.[19] [...] on the Priority View, benefits to the worse off matter more, but that is only because these people are at a lower *absolute* level. It is irrelevant that these people are worse off than others. Benefits to them would matter just as much even if there *were* no others who were better off." (Parfit, 1991 [1995], p.23)

---

[16] Proof: For any outcome o, $c_i(o)$ equals i's contribution to the safe outcome, $c_i(o, \ldots, o)$. Since the latter equals $\Sigma_k P(S_k) w(g_i(o_k)) = w(g_i(o))$, it follows that, for any x, the expected contribution of i, $\Sigma_k P(S_k) c_i(o_k)$, equals $\Sigma_k P(S_k) w(g_i(o_k))$, which equals $c_i(x)$.

[17] As we have assumed, $w(10) - w(5) > w(16) - w(10)$. This implies that contribution $c_i(y)$, which equals $\frac{1}{2}w(16) + \frac{1}{2}w(5)$, must be smaller than $c_i(x)$, which equals $\frac{1}{2}w(10) + \frac{1}{2}w(10)$.

[18] This difference between the Priority View and egalitarianism is emphasised in Persson (1996).

[19] This shows, by the way, that it is be incorrect to interpret Rawls' difference principle as the extreme, lexical form of the Priority View. Rawl's principle gives absolute priority to those people who are worse off than all the *others*.

Now, this prioritarian emphasis on the comparisons of how an individual is with how he "might have been", as opposed to the egalitarian comparisons of one individual with others, might suggest that Parfit would want us to use priority weights even in the determination of the individual goodness of a prospect. It might seem that the moral weights should apply not just to interpersonal balancing but also to those intrapersonal balancings that determine an individual's welfare in a prospect.

I disagree. What Parfit says certainly suggests that moral weights should be used in the determination of the individual contribution to the *overall* goodness (or badness) in all cases, even when there are no others who are better off than the individual under consideration. This need not imply that the moral weights should be used to determine how good a prospect is *for a given individual*.

Still, the above quote makes it clear that the function of the moral weights for Parfit cannot simply be to hinder unrestricted interpersonal compensations (gains for the better off at the expense of the worse off). If this were their only role, moral weights could be made dependent on the *relative* levels of individual goodness. Benefits and losses to a given person could be allowed to have varying moral impact depending on how that person fares in comparison with other persons: Ceteris paribus, they would have larger impact if the others were better off.[20] Insofar as this relativization to others is disallowed, however, the function of moral weights must be more far-reaching. They must be taken to express a moral concern for the welfare of each individual taken separately – a concern that decreases with increases in that individual's welfare, without taking into account the welfare of others.

4. Prioritarianism and the overall goodness: one-person case

Probably, then, Parfit would maintain that the moral weighting function should be used to determine the overall goodness of an outcome not only when no others are better off (cf. the quote above) but also when there *are* no others, i.e., when the outcome only involves one person. On such a view, we can still distinguish between the *individual* goodness of a one-person outcome and that outcome's *overall* goodness. The same distinction should therefore be possible to make for prospects that involve just one person. It should be possible to distinguish between how good a prospect is for this person and how good it is morally (i.e., overall). It is only for this latter issue that moral weights are allowed to play a role. [21]

To elaborate on this point, let us again consider the prioritarian formula for the overall goodness of outcomes:

Priority View for Outcomes: $g(o) = t(w(g_1(o)) + \ldots + w(g_n(o)))$

If we let t be idle (i.e. if we let t = the multiplication with 1), the overall goodness of an outcome will simply be the sum of its weighted individual goodness values. The moral weights will then play a role for overall goodness even in a one-person case (i.e., a case when $\mathbf{I} = \{i\}$, for some individual i). We shall have, for that case, the following formula:

Robinson Outcomes: $g(o) = t(w(g_i(o)) = w(g_i(o))$.

The overall goodness of a one-person outcome will thus differ from its individual goodness. This difference, however, will not show in the ordering of Robinson outcomes that involve one

[20] I am indebted to Ingmar Persson for this observation.

[21] In personal communication, Derek Parfit has confirmed that this is how he himself would interpret the Priority View.

and the same person *i*. For any two such outcomes *o* and *o'*, *o* is overall better than *o'* iff *o* is better for *i* than *o'*.

But what about *prospects*? If *t* is idle, then prioritarianism, on my interpretation, will lead to a quite striking divergence between prudence and the prioritarian morality. The two will diverge even in some of the cases in which the agent *i* is the only individual involved. Thus, suppose that there are just two possible states, S_1 and S_2, each of them equally probable, and let *i* have a choice between a risky prospect $y = (o', o'')$ and a safe prospect $x = (o, o)$ that yields the same outcome whatever happens. Assume that *o'* is better for *i* than *o*, which in turn is better for *i* than *o"*. In particular, suppose that *i*'s goodness values for the different outcomes are related to each other as follows:

$g_i(o') - g_i(o) > g_i(o) - g_i(o'') > 0$.

Then, prudence dictates that i should choose the risky *y*: $g_i(x) > g_i(y)$ (given that individual goodness is expectational, as we have assumed). As compared with *o*, which is the guaranteed outcome of *x*, his gain in *o'* is larger than his loss in *o"* and these two possible outcomes of *y* are equiprobable.

Suppose, however, that this larger gain weighs less than the smaller loss, in terms of *w*:

$w(g_i(o')) - w(g_i(o)) < w(g_i(o)) - w(g_i(o''))$.

Or, what amounts to the same,

$w(g_i(o)) > \frac{1}{2}w(g_i(o')) + \frac{1}{2}w(g_i(o''))$.

Then the prioritarian morality prescribes *x*, even though prudence dictates *y*. The safe prospect *x* is better overall:

$g(x) = g(o,o) = P(S_1)g(o) + P(S_2)g(o)$	[since overall goodness is expectational]
$= \frac{1}{2}t(w(g_i(o))) + \frac{1}{2}t(w(g_i(o)))$	[by the Priority View for Outcomes]
$= \frac{1}{2}w(g_i(o)) + \frac{1}{2}w(g_i(o))$	[given that t is idle]
$= w(g_i(o))$	
$> \frac{1}{2}w(g_i(o')) + \frac{1}{2}w(g_i(o''))$	[by assumption]
$= \frac{1}{2}t(w(g_i(o'))) + \frac{1}{2}t(w(g_i(o'')))$	[given that *t* is idle]
$= P(S_1)g(o') + P(S_2)g(o'')$	[by the Priority View for Outcomes]
$= g(o', o'') = g(y)$	[since overall goodness is expectational]

That prudence and morality can diverge in Robinson-type cases may seem counter-intuitive for some prioritarians. If, contrary to Parfit's suggestion, the only function of moral weights were to hinder unacceptable interpersonal compensations, then it would be natural to expect the coincidence of prudence and morality in Robinson cases.

To guarantee such coincidence, we could make the transform *t* depend on the weight function *w*, in such a way that the two functions cancel out in the one person-case. (I owe this suggestion to John Broome, but, as we shall see below, Broome no longer thinks it to be reasonable.) If we let *t* be the *inverse* of *w*, i.e., if we let $t = m$, the overall goodness of an outcome will reduce to its individual goodness in the Robinson-type cases:

$g(o) = t(w(g_i(o)) = m(w(g_i(o)) = g_i(o)$.

In the same way, on this suggestion, the overall goodness of a *prospect* $x = (o_1, \ldots, o_m)$ that involves only one individual reduces to the individual goodness of that prospect:

$g(x) = \Sigma_k P(S_k) g(o_k)$ [since overall goodness is expectational]

$= \Sigma_k P(S_k) t(w(g_i(o_k)))$ [by the Priority View for Outcomes]

$= \Sigma_k P(S_k) m(w(g_i(o_k))) = \Sigma_k P(S_k) g_i(o_k))$ [if t is the inverse of w]

$= g_i(x)$ [since individual goodness is expectational]

Which view, then, should be taken by prioritarians? Depending on their choice of the transformation t, they can reach conflicting conclusions as to the relationship between prioritarianism and prudence. If t is idle, prudence and the prioritarian morality will diverge even when the agent is the only person involved. But letting t instead be the inverse of w removes this divergence.

In my view, the former alternative is the more reasonable one. According to the latter option, as long as he is alone on his island, the prioritarian morality gives Robinson purely prudential recommendations: He should choose a risky prospect rather than a riskless one, if the former has for him a higher expected goodness value. But as soon as the man Friday comes into the picture, Robinson is no longer be allowed to go for the risky prospect, even though – as we might suppose - his choice *would not affect Friday in any way.* If Robinson's larger gain in one state weighs less, morally, than his smaller loss in the other (equiprobable) state, and Friday's welfare given each state is the same whatever prospect Robinson chooses, then, with Friday present, the risky prospect is overall worse than the riskless one.[22] This extreme sensitivity to "other persons being present" is counter-intuitive. Surely, if Robinson's choice cannot affect what happens to Friday, bringing Friday into the picture should not matter, morally, according to prioritarianism. The riskless prospect should be morally (i.e. overall) better in the absence of Friday if and only if it is morally better in his presence. We get this desirable result if we let t be idle. Therefore, prioritarians would be well advised to view the overall value of an outcome as the simple sum of its morally weighted individual values.

---

[22] To see this, let us compare a prospect $y = (o', o'')$ with a prospect $x = (o_1, o_2)$, where both states of nature are supposed to be equiprobable. Prospect x is safe for Robinson: $g_r(o_1) = g_r(o_2)$. Prospect y, on the other hand, is risky for him:

(i) $g_r(o') - g_r(o_1) > g_r(o_1) - g_r(o'') > 0$ (and similarly for o_2).

(i) implies that the risky prospect y is better for Robinson than the safe alternative x. If t is set to m (the inverse of w), this entails that, *in the absence of Friday, y is overall better than x*.

Let us now bring in Friday into the prospect and assume that Friday's goodness function g_f is such that, from his point of view, the choice between x and y does not matter, whatever the state happens to be the case. I.e., $g_f(o') = g_f(o_1) = k_1$ *and* $g_f(o'') = g_f(o_2) = k_2$. If $t = m$, the overall goodness of y now equals

(Gy) $\frac{1}{2}m(w(g_r(o') + w(k_1)) + \frac{1}{2}m(w(g_r(o'') + w(k_2))$,

which can be compared with the overall goodness of the safe prospect x:

(Gx) $\frac{1}{2}m(w(g_r(o) + w(k_1)) + \frac{1}{2}m(w(g_r(o) + w(k_2))$.

Now, (i) is fully compatible with:

(ii) $w(g_r(o')) - w(g_r(o_1)) < w(g_r(o_1)) - w(g_r(o''))$.

In view of (ii), Robinson's larger gain in o' (as compared with o_1) weighs less than his loss in o''(as compared with o_2). Since function m is strictly increasing, it follows that the increase in the first term of (Gy), as compared with the first term of (Gx), is smaller than the corresponding decrease in the second term in (Gy), as compared with the second term of (Gx). Which means that, *in the presence of Friday, x is overall better than y*. But this change cannot be explained by reference to Friday's welfare, which is assumed to be the same under each prospect, whatever state happens to obtain.

There is also another reason for this simplification. In section 9.3 of *Weighing Goods*, Broome considers a choice between the following two prospects, which involve two individuals and two equiprobable states:

	Prospect *x*		Prospect *y*		
	S_1	S_2	S_1	S_2	$P(S_1) = P(S_2) = ½$
i	2	1	2	1	
j	2	1	1	2	
	o_1	o_2	o_3	o_4	

For an *egalitarian*, as Broome points out, prospect *x* should be preferable to prospect *y*: the former guarantees equality, in each outcome, while the latter guarantees inequality. For a *utilitarian*, on the other hand, the two prospects are equally good. Each of them gives each individual the same expectation of individual goodness: 2 with probability ½ and 1 with probability ½.

A natural question to ask is what a *prioritarian* should say about this case. (I am indebted to Marc Fleurbaey for raising this issue.) *If* we let *t* be idle, then the Priority View, just as utilitarianism, will imply that prospects *x* and *y* are equally good overall. The reason is simple: As we have seen above (section 3), if *t* is idle, then the overall goodness of a prospect is the sum of the individual contributions to the overall goodness of this prospect:

$g(x) = \Sigma_i\, c_i(x)$,

where $c_i(x) = \Sigma_k P(S_k) w(g_i(o_k))$. The contribution of each individual is counted *separately* and independently of the contribution of others, just as in utilitarianism (with the only difference being that, in utilitarianism, $c_i(x)$ is set to $\Sigma_k P(S_k) g_i(o_k)$).

Now, since in our example the contribution of each individual to each of the prospects is the same,,

$c_i(x) = ½w(2) + ½w(1) = c_i(y)$ and $c_j(x) = ½w(2) + ½w(1) = c_j(y)$,

it follows that $g(x)$ must be equal to $g(y)$.

However, if *t* instead is some non-linear transformation, it becomes impossible to separate the individual contributions to the overall value of a prospect. In particular, if *t* is strictly convex, which it must be if it is the inverse of the strictly concave function *w*, *x* will be overall better than *y*:

$$
\begin{aligned}
g(x) = ½g(o_1) + ½g(o_2) &= ½t(w(2) + w(2)) + ½t(w(1) + w(1)) \\
&= ½t(2w(2)) + ½t(2w(1)) \\
&> t(w(2) + w(1)) \qquad \text{[if } t \text{ is strictly convex}^{23}\text{]} \\
&= ½t(w(2) + w(1)) + ½t(w(1) + w(2)) = ½g(o_3) + ½g(o_4) = g(y)
\end{aligned}
$$

In other words, with a convex *t*, prioritarianism will be considerably closer to egalitarianism. Surely, this is *not* a desirable consequence. As we have seen, Parfit has been at pains to point out that the benefits to a person should count as much independently of how other people fare. Consequently, on the Priority View, and contrary to egalitarianism, the welfare of each individual in various outcomes should make a *separable* contribution to the overall goodness of a prospect, independent of the welfare of others, which could not be the case if *t* were a

[23] Explanation: For any strictly convex *t* and any real numbers *k* and *l*, it holds that $½t(2k) + ½t(2l) > t(k + l)$. Therefore, $½t(2w(2)) + ½t(2w(1)) > t(w(2) + w(1))$.

non-linear transform. This point has recently been forcefully made by Broome (2001).[24] Therefore, Broome now agrees with me that my interpretation of the Priority View will have to lead to the divorce of morality from prudence. The reconciliation between the two in one-person cases cannot be achieved by letting the transformations *t* and *w* cancel out. However, Broome still prefers his own interpretation of prioritarianism, which rejects Bernoulli's hypothesis and accepts the Principle of Personal Good. On that interpretation, of course, morality and prudence automatically coincide in all Robinson-type cases.

5. Uncertainty All the Way Down?

On my interpretation of prioritarianism, the distinction between prospects and outcomes is taken seriously. For an outcome, its overall value is the sum of its morally weighted individual values. But for a prospect, the relationship between its overall value and its value for various individuals is much less straightforward and indirect. In fact, there is no functional dependence in this case: two prospects may be equally good for each individual and still differ as far as their overall value is concerned.

This means that, on my interpretation, a prioritarian must take seriously the distinction between prospects and outcomes. He cannot treat outcomes as "small worlds" in Savage's sense (cf. Savage 1972 (1954), section 5.5), i.e., as situations that are assumed to be free from uncertainty only provisionally, for the problem at hand. He cannot adhere to the view, so popular among many decision theorists, that uncertainty really is present "all the way down" and that it can always be discerned, in any outcome, if we only use a sufficiently strong magnifying glass. To illustrate this popular view, consider the famous example of *Savage's omelet* (ibid., section 2.5): I have broken five eggs into the bowl, to make an omelet. Should I do the same with the remaining sixth egg, or should I break it into a separate saucer? If I do the former and the last egg turns out to be rotten, I will have to throw out everything. For the purposes of my decision problem I can treat this as one possible outcome. From a more discerning perspective, however, the 'no omelet' outcome can be seen as shot with uncertainties: What will happen if I don't get my omelet? Will I miss my lunch altogether, and, if so, how will it affect my mood and behavior? Will I overeat in the evening, will I be irritable for the rest of the day, etc? And what might *this* lead to, in turn? Thus, if we wish, we could re-describe the outcome in question as an uncertain prospect which, depending on various factors, can result in different more specific outcomes. These can in their turn be re-described as uncertain prospects, and so on.

It is in this sense that Savage's outcomes are "small" possible worlds, as he puts it, rather than "grand worlds" without any residuum of potential uncertainty. The reason he adduces for the small-worlds approach is that of practicality: the decisions we make in real life are never made with a view to all the uncountable uncertainties that may arise in connection with our actions (cf. ibid., section 5.5). Behind this practical point, one might add (even though Savage himself might not be prepared to go as far as that), there is a more fundamental fact: Our preferences as regards various occurrences are preferences for these occurrences *under descriptions* (cf. Schick 1982). Pace behaviorism, preferring is an intentional attitude, which means that preference at least to some extent is *representation-dependent.* Since our representational capacities are limited, "grand worlds" cannot be fully represented, in all their details. Therefore, if a decision theorist wanted to start from an agent's preferences over grand outcomes, he would have to allow for the possibility that one and the same grand outcome

[24] See, however, Fleurbaey (2001) for an interesting criticism of the idea that the Priority View is fundamentally opposed to egalitarianism.

might be valued differently by the agent depending on how this outcome is represented. He would also have to accept that the agent's preference ranking would contain huge gaps, as the agent's ability to discern between different grand outcomes is severely limited: As ordinary agents, with finite powers of conceptual discrimination, we can only discern between *classes* of grand possible worlds, but not among individual worlds of this kind. [25] Restriction to small outcomes allows the decision-theorist to avoid these difficulties. As long as he keeps to small outcomes, the outcomes and their propositional representations need not be distinguished from each other.

Prioritarians who take seriously the distinction between prospects and outcomes must instead opt for the "grand" interpretation of outcomes. The outcomes must be comprehensive possible worlds, which, in principle, contain a determinate answer to every question of fact. Otherwise, if an outcome might just as well be seen as a prospect, with larger magnification, it would be difficult to defend a theory that treats prospects and outcomes differently. Thus, the question arises: Can a prioritarian assume the existence of univocal (i.e. representation-independent) betterness orderings on *grand* outcomes?

The answer, it seems, must in part depend on the connection between betterness and preference. It may be that this connection is quite close and that, in particular, an ordering of betterness is grounded in pro-attitudes of some kind, which are just as intentional as individual preferences (cf Schick 1982). If these attitudes are supposed to be directly aimed at the relata of the betterness ordering, rather than at various general good-making features of the relata, the prospects for a univocal betterness ordering of grand outcomes look very bleak indeed. But the possibility remains that the relation between the ordering of betterness and our pro-attitudes is not as straightforward. It may well be that betterness orderings of comprehensive possible worlds should be seen as theoretical constructs. If at all, such constructs would only be indirectly based in our pro- and contra-attitudes of certain kinds. Rather than aiming directly at comprehensive possible worlds, the pro- and contra-attitudes that underlie the construction are immediately directed at various general features (in particular, various aspects of individual well-being) that such grand worlds may exhibit. Such indirectly constructed betterness orderings need not be fundamentally representation-sensitive. As long as this possibility remains, the interpretation of prioritarianism suggested in this paper does not make this view doomed from the start.

References

Broome, John, 1991, *Weighing Goods,* Basil Blackwell, Oxford.

Broome, John, 1999, *Valuing Lives,* book manuscript.

Broome, John, 2001, "Equality versus Priority: A Useful Distinction", http://aran.univ-pau.fr/ee/page3.html

Harsanyi, John, 1955, "Cardinal Welfare, Individualistic Ethics, an Interpersonal Comparisons of Utility", *Journal of Political Economy* 63: 309–21; reprinted in John Harsanyi, *Essays on Ethics, Social Behavior and Scientific Explanation,* Reidel, Dordrecht 1976, pp. 6 – 23.

Fleurbaey, Marc, 2001, "Equality versus Priority: How Relevant is this Distinction?", http://aran.univ-pau.fr/ee/page3.html

[25] Another worry in connection with 'grand' outcomes is whether the idea of such outcomes is conceptually coherent, to begin with. Here, I assume that the answer is yes, i.e., that it is meaningful to postulate such comprehensive possible ways for the world to be. But I am fully aware that this assumption itself is controversial.

Jensen, Karsten Klint, 1996, "Measuring the Size of the Benefit and Its Moral Weight", in Wlodek Rabinowicz (ed.), *Preference and Value – Preferentialism in Ethics,* Studies in Philosophy 1996:1, Dept. of Philosophy, Lund University, pp. 114–45.

Mongin, Philippe, 1994, "Harsanyi's Aggregation Theorem: multi-profile version and Unsettled Questions, *Social Choice and Welfare* 11: 331-54.

Mongin, Philippe, and Claude d'Aspremont, 1998, "Utility Theory and Ethics", in Barbera, S., Hammond, P. and C. Seidl, *Handbook of Utility Theory*, vol.1, Kluver, Dordrecht, pp. 371-481.

Parfit, Derek, 1995 [1991], "Equality or Priority?", The Lindley Lecture 1991, Department of Philosophy, The University of Kansas.

Persson, Ingmar, 1996, "Telic Egalitarianism vs. the Priority View", in Wlodek Rabinowicz (ed.), *Preference and Value – Preferentialism in Ethics,* Studies in Philosophy 1996:1, Dept. of Philosophy, Lund University, pp. 146–61.

Savage, Leonard J., 1972 (1954), *The Foundations of Statistics,* Dover Publications, New York, 2nd ed.

Schick, Frederic, 1982: "Under Which Descriptions", in Amartya Sen and Bernard Williams (eds.), *Utilitarianism and Beyond,* Cambridge University Press, pp. 251-60.