

Impartiality in Weighing Lives

Richard Bradley
Department of Philosophy, Logic and Scientific Method
London School of Economics
Houghton Street
London WC2A 2AE
r.bradley@lse.ac.uk*

1 Introduction

Weighing Lives examines the problems of choice that we face when the options available to us not only have consequences for different people at different times but also for what people there are at each point in time, for who is alive and when. The principal conclusion of the book is that the goodness of a distribution of wellbeing across a population over some period of time is measured by the sum of the standardized wellbeings of the individuals belonging to it (over that time period), where the standardized wellbeing of an individual is the difference between their wellbeing and a constant that measures the ‘critical level’ of wellbeing above which someone’s existence is a good thing. Broome is not the first to make this claim - it is, for instance, the heart of what Blackorby, Bossert and Donaldson ([1],[2]) call Critical-level Utilitarianism - but his book surely provides the most careful and systematic examination of both its foundations and its consequences.

The idea of a critical or neutral level of wellbeing, of a level at which it is better that someone exists than that they do not, is in apparent contradiction with intuitions that many people have about what they or policy makers should do when considering options that affect the number of living people. Consider the following two examples:

1. The Parents’ Problem: A couple are considering whether to have a third child. Their income is such that the family is reasonably well off and the children do not lack for anything important. Having another child would be expensive and everyone in the family would suffer, but the third child would not be badly off. However even if the new child’s wellbeing was sufficiently great to offset the losses of

*Thanks to Alex Voorhoeve for very useful comments on an earlier draft,

the other members of the family, few of us think that they parents' are obliged to have another child and some of us might even think that they should not.

2. The Policy Problem: The government could increase fertility, by raising taxes and using the revenue to offer inducements to people to have children. In wealthy countries and assuming that the neutral level is not very high, it is likely that the best outcomes would require very substantial transfers of resources, mainly from the old to the young. But few of us think that the government should make these transfers.

The force of these examples could be offset by supposing that the neutral level must be very high. In this case, however, the opposite problem would arise: it would follow that it would be better if many of those alive at the moment had never been born even though from their point of view their lives are perfectly satisfactory. Few of us would accept this.¹

These examples present raw intuitions about what we should do and they could reflect a number of different concerns. One explanation for them is that we simply do not believe that adding people to the world makes things better or, as Broome puts it, "We think intuitively that adding a person to the world is very often ethically neutral" (p. 143).² Broome clearly shares the intuition as he devotes a good deal of space to considering how a suitably formulated version of it, which he calls the principle of equal existence, can be accommodated within his framework. In the end, however, he rejects the intuition of neutrality - "as an intuition about goodness, it cannot be correct" (p. 211). It is difficult to resist Broome's arguments for this conclusion which are characteristically thorough and carefully crafted. But it is equally difficult to shake the intuitions elicited by the two examples. My aim is to explore several routes to reconciling the rejection of the principle of equal existence with them.

2 The Derivation of the Critical Level

Let me start by setting out the basic framework for the derivation of the critical or neutral level. Our concern is with the goodness of prospects, and especially with those properties of it that are determined by the nature of betterness comparisons between prospects. The relevant prospects are sets of possible outcomes for one or more persons, one

¹The difficulty of finding a critical level that did not lead to one kind of 'repugnant' conclusion or another was first noted by Derek Parfit [4].

²All page references are to Broome [3].

for each possible state of nature (I shall ignore the temporal dimension here). An outcome for a person, in a particular state of nature, is a full description of everything that happens to that person that is relevant to their wellbeing, in the event that that state of nature prevails.³ We call this their history in that state of nature. A prospect may be better or worse than another for some individual as well as better or worse overall. Facts of these kinds are represented by personal and general betterness relations on the set of prospects. Broome's argument for the neutral level is driven by the assumptions he makes about these relations.

In his earlier book *Weighing Goods* Broome derives a quantitative representation of goodness for a fixed population, P , from the assumption that both personal and general betterness relations satisfy the standard (i.e. Savage's) axioms of expected utility theory and the principle of personal good.⁴ The resulting interpersonal addition theorem says that under these conditions the personal betterness relation for each individual $p \in P$ is represented by an expected utility function, g^p , defined on the set of prospects and unique up to a choice of scale, which may be interpreted as a measure of p 's wellbeing.⁵ Furthermore there exists a function V , also defined on the set of prospects and unique up to a choice of scale, that coheres with the general betterness relation, and is such that:

$$V = \sum_{p \in P} \alpha^p g^p \quad (1)$$

where the α^p are strictly positive weights on the g^p . In effect, general or overall goodness is a weighted average of personal wellbeing.

In *Weighing Lives* things are done slightly differently (presumably for simplicity) with both the general betterness relation and the value function that represents it being defined directly on vectors of quantities of personal wellbeing. This reformulation makes little difference - we simply read the g^p in equation (1) not as functions but as elements of a vector representing distributions of wellbeing - but the fact that the utility measures of wellbeing are not unique does mean that caution is required in interpreting the results. One advantage though is that it makes the formulation of a requirement of impartiality much more straightforward: a betterness relation will be said to be *impartial* just in case if two vectors of quantities of personal wellbeing are permutations

³This should be understood in the widest possible sense. If something's happening to someone else affects my wellbeing, then we count this as happening to me.

⁴The principle says that if one prospect is at least as good as another for every individual then it is at least as good overall as the other and if furthermore it is strictly better for at least one individual, then it is strictly better overall.

⁵This interpretation is what Broome calls Bernoulli's hypothesis.

of one another, then neither is better than the other. The assumption that the general betterness relation is impartial then implies that the α^p in equation (1) are equal. Setting them to 1, we obtain:

$$V = \sum_{p \in P} g^p \tag{2}$$

the ‘Utilitarian’ measure of goodness.

Consider now the more general case of variable populations and in particular the problem of determining the goodness of a distribution x in which some individual $p \in P$ is not alive. Assume that the general betterness relation is complete over the set of all possible prospects, so that this distribution can be compared to ones in which p is alive. Then by varying p ’s wellbeing, but fixing the wellbeings of all other individuals in P , we can find some distribution in the domain of V which is no better or worse than x . Individual p ’s wellbeing in that distribution is her neutral level. But by impartiality it must be the same for everyone else.⁶

If we want to resist this conclusion, namely that there is critical level of wellbeing above which it is better overall that someone exists, then we must reject at least one of:

1. The conditions leading to the representation of general goodness as a weighted sum of personal goodness.
2. The assumption of the completeness of the betterness relation over the entire domain of prospects.
3. Impartiality.

The case for the first assumption. - the derivation of equation (1) from conditions on personal and general betterness relations - belongs to the earlier book and I will not consider objections to it here. The second assumption, the completeness condition, looks a good deal more vulnerable. It is certainly difficult in general to compare prospects which realise quite different values and this is especially true where comparisons are being made across different populations. Sometimes we are inclined to say of two prospects that neither is better than the other, nor are they equally good, because they are good or bad in different ways. Broome in fact canvasses a number of possible explanations for this attitude, rejecting the view that prospects can be incommensurate

⁶A further argument is required to show that the neutral level is not context dependent, but I shall omit discussion of this as it’s the neutrality, not its context independence or otherwise that is in conflict with our intuitions.

in value, and suggesting instead that it is best explained by the vagueness of the betterness relation. If betterness is vague, then so too will be the critical level for existence. This could explain our attachment to the principle of equal existence. It will often be the case that some ways of making betterness precise will make it better to add someone to the world, while others will make it worse. Hence we will want to say that adding that person neither makes things determinately better nor makes things determinately worse.

Unfortunately, however strong the case for admitting vagueness, it will not really solve our problem. For, as Broome shows, the principle of equal existence plus the assumption that betterness is indeterminate has other implausible consequences. Furthermore, our intuition is not of indeterminacy. In the Parents' problem, for instance, it is not the case that the parents cannot compare the prospect of two children with that of three. The problem is that they can, but that they might well feel that it would definitely be wrong if they had another child, simply because the outcome would be worse for the existing members of the family.

This leaves the impartiality assumption and it is there that I think we must concentrate our attention. In the following sections I will look at Broome's arguments for impartiality as well as an objection to it deriving from the problem of interpersonal comparisons of wellbeing. But first I want to consider another possibility; namely that there is no contradiction at all between impartiality and the intuitions expressed our examples, because the principle of equal existence does not properly capture them.

3 Impartiality and Teleology

To say that we must act impartially is to require us to treat equally all those affected by our actions, to give equal weight to each of their interests. In particular, if we are required to act impartially in cases where our actions affect who lives or dies, then we must treat the living and those who might live as equals and give equal consideration to their wellbeing in each possible course of events. But intuitively such impartiality is not required of us. There is nothing wrong with privileging the existent by, for instance, regarding the small possible harms to the lives of one's existent children as outweighing the potential well-being of children not yet born.

The point is not confined to the treatment of the non-existent. There are many situations where someone has special obligations to 'insiders' that requires that they be treated differently from 'outsiders': the relationship between parents and their own children, between club commit-

tees and club members and between states and their citizens are all of this kind. We would not, for instance, expect the manager of a cricket club to spend club funds on a football field even if many more non-members would be served by it than would its members by some cricket facility.

It is not, of course, impartiality in action that Broome is arguing for, but the impartiality of betterness and, hence, in our judgements of it. But if we are right to be partial in our actions it must follow that either (i) betterness is itself not impartial, or (ii) what makes an action right or wrong is not simply the goodness of its consequences. As Broome himself argues against what he calls *teleology*; the view that what a person ought to do is determined by the goodness of the acts available to her, it is open to him use (ii) as an explanation for the clash between our intuitions in the Parents' Problem and the Policy Problem. In short, the explanation goes thus: our intuitions in these cases do not concern the goodness of the consequences of the actions or policies available, but what it is right or wrong to do in these circumstances. The principle of equal existence does not therefore properly represent these intuitions (nor could any principle of betterness) and its inconsistency with the theory of goodness proposed in *Weighing Lives* should not cause us any concern.

This proposal for how we might deal with our problem has the advantage of leaving all of *Weighing Lives* intact without requiring us to give up some rather robust intuitions about what we ought to do. In particular we have no need to challenge the impartiality of betterness. Still we may have doubts about this strategy. For one thing the rejection of teleology raises the question of how exactly the rightness of our actions *is* related to the goodness of their outcomes. If rightness and goodness were to diverge to too great an extent this would considerably reduce the significance of the investigation of goodness. Perhaps, as Broome argues, there are conceptual limits on the extent to which they can diverge. But what relationship between the two would explain our intuitions about the decisions parents and policy makers should make concerning number of children? Parents presumably ought not to have another child if the child's life would be a very poor one, but impartial goodness does not appear to figure much beyond this: it is the good of the existent that seems to count in our deliberations. So even a much weaker relationship than teleology between impartial goodness and our duties may be incompatible with our intuitions about these cases. If this were so, Broome would presumably want to conclude that the latter must be rejected. The alternative is to preserve the tight connection between rightness and goodness, but reject impartiality.

4 Is Betterness Relative?

Why should we accept that the general betterness relation is impartial? Broome does not give much of an argument for it.

“General good is not assessed from any particular person’s point of view. Given that, it is hard to think of any doubts that could plausibly be raised against impartiality. Everyone must count equally in nonrelative good” (p. 135).

There are two steps here: the rejection of a relative notion of betterness and the conclusion that betterness must be impartial. We shall consider each in turn.

There are two senses in which betterness could be relative to an individual or group. One prospect may be better than another from one person’s point of view but not from another person’s, or at one time but not another: betterness is betterness *relative to* some point of view. Betterness may also be relative in the more banal sense that one prospect may be better than another for one individual or agent but not for another: betterness is betterness *for* someone. The difference may be brought out by noting that if betterness is relative to a person’s point of view then two people can disagree about what is best overall and both be right, while if betterness is relative only in the second sense then they should agree that on what is best for each. Broome rejects relative betterness of the first kind but, on the face of it, his personal betterness relations are relative in the second sense (however, as we shall see in the next section, this interpretation poses some difficulties).

If betterness were relative in the first sense - to points of view - then it is likely to be partial in a way that would vindicate our judgements in the opening examples. It is also likely to complicate the relationship between betterness judgements and actions. For example, it could be that from the point of view of a family in its current form it is best that no more children are brought into existence, while from the point of view of the possible larger family, it is better that all of its members exist. So relativity of betterness opens up the possibility of actions that are self-defeating or self-justifying because of their effect on who exists. Broome cites this as an argument against relative betterness, but it is not decisive. Such problems of inter-temporal coherence are familiar from individual decision making and I see no reason why similar solutions should not be applied in this context.⁷ Nonetheless, since Broome

⁷For instance, by appeal to a form of ratificationism that rules out choices that produce conditions in which it would be judged that it would have been better had another choice been made.

declares himself not concerned with relative betterness of this kind, I will not try to defend it any further.

The crucial question is whether by rejecting relativity to a point of view, we commit ourselves to impartiality. I do not see why this should be so: just because a judgement is not made from any particular point of view does not mean that it must give equal weight to all. How individual wellbeing figures in overall goodness may depend on characteristics of the individuals; for instance, on whether they deserve to be well-off. In contrast, if we reject the claim that personal betterness is relative in the second sense then general betterness could hardly fail to be impartial. For this would in effect be to say that whether one prospect is better than another does not depend on who is affected by it. Surely however personal betterness *is* relative. A rise in lawyers' fees, for instance, is good for lawyers, but typically not good for everyone else.

Is general betterness relative in the second sense? When we are concerned with a fixed population it is natural to interpret general or overall betterness as betterness for the population as a whole. The interpersonal addition theorem could then be thought of as addressing a question about the dependence of betterness for a group or population on betterness for the individuals making it up. In the variable population case, however, it is less clear how general betterness should be interpreted. One possibility is that general betterness is to be understood as betterness relative to all possible persons, both those that are currently alive and those that might come into existence. Another is that general betterness is not supposed to be relative in any sense. Both possibilities raise problems. Do we have any real grip on the idea of the set of all possible persons and can we really hope to use it as a reference point in our judgements of betterness? On the other hand, the idea of overall betterness that is not betterness for some group or population is hard to grasp.

As before there is no simple relation between endorsing or rejecting the relativity of betterness and endorsing or rejecting impartiality. Betterness that is not relative to any population could be partial in the technical sense, unless impartiality was somehow built into the idea of a non-relative betterness relation. On the other hand, betterness that is betterness for a population can still be impartial (with respect to its members); indeed it is natural to think that it is. I would speculate that for most of us, betterness is always conceived in relation to some individual or group. We find abstract variable population cases very difficult to judge because we are not sure what the reference population for our judgements is supposed to be. In concrete cases there is often a salient reference point: in the Parents' Problem it is that of the existent

family, in the Policy Problem it is that of existing population of the country. This explains our intuitions in these and similar cases: what we judge best is implicitly in reference to some population; often some subset of the existing one. This could be because we find it hard to think in terms of the overall good of possible populations; our partiality reflects our cognitive limitations. Or it could be because betterness is always relative to some population and we employ some other principle to determine which one is relevant or whose interests should count. I think the latter explanation is more plausible.

5 Interpersonal Comparisons

In this final section, we turn to an objection to the impartiality condition based on the observation that no sense can be made of it unless wellbeing is interpersonally comparable. Recall that the utility representations of personal good that we are working with are unique only up to a choice of scale for each of them. So for one set of choices of scales the impartiality condition might be satisfied, for most others it will not. Hence, if the impartiality condition is to be applicable it must be the case that the utility representations of personal good are *co-scaled*. But many would deny that this can be done in any way except by fiat.

The standard argument for the impossibility of interpersonal utility comparisons is formulated for a somewhat different context - where utility measures strength of preference rather than goodness - but much of it can be carried over. The essential point is that a utility measure, whether of personal goodness or of preference, represents only comparisons between prospects and can thus only express how well off some outcome makes that person, or how strongly it is preferred by them, relative to the other possible outcomes and given some choice of reference point (which determines the zero of the scale for measuring wellbeing or preference strength). What it does not express is how well off an outcome makes one person relative to how well off it makes another person, or how strong their preference is relative to someone else's. For example, suppose we measure the goodness of a set of prospects for two people by assigning in each case a zero measure to their worst options. Does it follow that their wellbeing is equal at their worst outcomes? Clearly not. And nothing about the betterness ranking of the two people tells us which prospect of makes the second person as well off as the first is made by her worst outcome, or vice versa. So the betterness rankings, by themselves, provide no basis for comparing wellbeing.

As ranking information does not determine utility comparisons, some other basis must be found for them. In the context of utility measures of preference strength it is generally agreed that no empirical basis can

be provided, since the preference rankings or the choices which reveal them represent all the available evidence there is for agents' preferences.⁸ Hence interpersonal comparisons are either indeterminate or partially determined by value judgements. Similarly, since interpersonal comparisons of betterness are not determined by the properties of betterness rankings, we must either accept that such comparisons are indeterminate or find some other basis for them. In *Weighing Goods* Broome seems to accept this and argues that what determines interpersonal comparisons of wellbeing is none other than the impartiality of general betterness. But in *Weighing Lives* this argument is rejected in favour of one based on what he proffers as a conceptual truth:

“A person’s wellbeing is how good things are for that person. It depends on - supervenes on - how things are for her. If things for one person are exactly the same as they are for another, those two people have the same wellbeing” (p. 94)

This gloss on wellbeing as how good things are for a person does not on the face of it rule out the view that wellbeing depends not only on access to certain kinds of goods, but also how they are evaluated subjectively. On this latter view, in order to compare my wellbeing to someone else’s, the value I attach to things must be compared. But possibility of doing so is precisely what is in question. For what evidence would determine a judgement as to whether I value something to a greater, lesser or identical extent to you? Rather than provide the basis for interpersonal comparisons, Broome’s argument seems to presuppose that they are possible.⁹

To block this objection we must construe prospects in a particular way. A history must specify not just what happens to a person in a possible state of nature but also any features of that person that determine how the evaluations of these events are relevant to her wellbeing. Suppose that only a person’s preferences are relevant. Then an outcome for a person must take the form of something like “At time t , with preferences p , the following happens to you”. If two such outcomes are identical for two agents then their wellbeing must be the same. If they

⁸See, for instance, Rossi [5].

⁹An useful analogy is with the argument that it is impossible to compare one person’s perception of a colour to another person’s because no empirical evidence can determine whether the quality of their experiences are the same or different. An appeal to the ‘conceptual truth’ that if one person perceives colour in exactly the same way as another, then they have the same colour experience, does not seem very helpful here. As in the utility case, we simply can’t say when the antecedent ‘sameness’ condition is satisfied.

were not, it would be the case that a person's preferences are not the only subjective feature relevant to their wellbeing and further specification of the outcome would be required.

This construal of prospects assures interpersonal comparability of wellbeing by considerably restricting the extent to which the personal betterness relations defined over them can differ. If it is better for you to have one history along with certain personal characteristics rather than another history with these same characteristics, then it must be so for me too. For the characteristics that distinguish us and are responsible for our different attitudes to these histories are 'screened out' by the embedding of characteristics in the outcomes. Hence our betterness relations, and the utility functions representing them, differ only in a rather trivial way: your relation to a prospect will be different to mine only because the prospect specifies a different history-characteristic combination for me than for you. While it does not follow that personal betterness relations cannot be construed as representing betterness for that person, it does follow that the goodness of an outcome, qua history-characteristic combination, is fundamentally person-independent.

Two remarks are in order. First it may be questioned whether every coupling of a person and a prospect is coherent. What does it mean for *me* to be better or worse off with fundamentally different characteristics to my own? Are there really no characteristics of me that are partially constitutive of my identity and relevant to the relative goodness for me of possible outcomes? I do not find it very plausible that there are none. Nonetheless, given the person-independent construal of prospects under consideration, this is a consequence of the axioms of expected utility used to derive the interpersonal addition theorem.¹⁰

Secondly if the goodness of an outcome is person-independent, then what utility representations of a person's betterness rankings measure is not their wellbeing in possible histories, but the goodness of being a person with certain characteristics in that history. This in turn implies that our presentation of the interpersonal addition theorem itself was somewhat misleading. For what it must now be read as establishing is not so much a relation between overall and personal goodness, but a relation between a utility measure of the overall goodness of a prospect and sub-utility measures of the goodness of the different characteristic-history components of it. Assuming impartiality, this relation is better expressed as follows.

Let $\langle x^1, x^2, \dots, x^n \rangle$ be a vector of characteristic-history pairs, one for

¹⁰This is very similar to the criticism that is often made of the assumption, intrinsic to the version of decision theory due to Savage and which Broome draws on, that the value of an outcome is independent of the state of nature in which it is realised.

each person in an n -person population. For each person p , let g^p be a utility function on the set of such prospects that represents p 's personal betterness relation. Now $g^p(\langle x^1, \dots, x^n \rangle)$ depends only on x^p and hence, in light of the argument for person-independence given above, $g^p(\langle \dots, x^p, \dots \rangle) = g^q(\langle \dots, x^q, \dots \rangle)$ if $x^p = x^q$. So we can define a function g on the set of characteristic-history pairs by setting $g(x_p) = g^p(x_p)$. We now restate the interpersonal addition theorem as follows: under the assumed conditions, including impartiality, there exists a utility function G and sub-utility g such that G coheres with the general betterness relation and:

$$G(\langle x_1, x_2, \dots, x_n \rangle) = \sum_{p \in P} g(x_p)$$

In effect, general or overall goodness is the sum of the goodness of each characteristic-history combination. Personal goodness simply drops out of the picture.

I submit that, so construed, the interpersonal addition theorem does not address the decision problems that we set out at the beginning of the paper and which Broome's own examples suggest that he is addressing. For these concern what to do when our actions have different consequences for people with different characteristics, not what to do when our actions produce different history-characteristic combinations. It is the difference between the question of what we should do to make things go as well as possible for people as they are and what we should do to produce the optimal combination of person types and things that happen to them. Only in the context of the former problem is teleology plausible. We do feel that we should make people as well-off as possible. But we do not feel we should make the sorts (or numbers) of persons that are most susceptible to wellbeing.

6 Conclusion

I have argued that there are two weaknesses in Broome's defence of impartiality. The first is that the relation between relative betterness and impartiality is more nuanced than he acknowledges. Only if personal betterness were not relative in the sense of representing betterness for that person would the impartiality of overall betterness follow. But is implausible that it is not relative in this sense and Broome does not claim that it is. On the other hand, his solution to the problem of interpersonal comparisons of wellbeing does seem to require that the goodness of outcomes be person-independent because it requires that prospects be construed in a particular way. This route to interpersonal comparability leads to an interpretation of the interpersonal addition theorem that does not address the decision problem that *Weighing Lives*

appear to set for itself. This is the second weakness.

These objections are not conclusive. Another basis for interpersonal comparability could be offered and more conclusive arguments given for the impartiality of betterness. Even then we would not necessarily have to give up on our intuitions about what to do in Parents' and Policy problems. For we could reject teleology and argue that what we ought to do is not determined by the impartial goodness of the consequences of our actions. Indeed if we think that overall goodness is always goodness for some reference population, some (presumably non-teleological) principle would have to be offered to determine whose goodness should be object of our concerns. In this regard we have many intuitions, but little theory.

References

- [1] Blackorby, C., W. Bossert and D. Donaldson (1997) "Critical Level Utilitarianism and the Population Ethics-dilemma", *Economics and Philosophy* 13: 197-230
- [2] Blackorby, C., W. Bossert and D. Donaldson (2005) "*Population Issues in Social Choice Theory, Welfare Economics, and Ethics*", Cambridge University Press
- [3] Broome, J. (1995) *Weighing Goods*, Basil Blackwell, Oxford
- [4] Parfit, D. (1984) *Reasons and Persons*, Oxford University Press, Oxford
- [5] Rossi, M. (2006) Interpersonal Comparisons of Utility, *Choice Group working papers* 2 (2).