Cost-benefit analysis and population

For the *Journal of Legal Studies*, 29 (June 2000)

This paper was presented at the conference on Cost-Benefit Analysis in Chigago, September 1999. I am very grateful to the participants – particularly Matthew Adler and Eric Posner.– for their comments.

John Broome
University of St Andrews
john.broome@st-andrews.ac.uk

1. Cost-benefit analysis as valuation

What is the purpose of cost-benefit analysis? I assume it is to make a valuation. Cost-benefit analysis may be applied to an act, an event, a policy, a project or something else – for uniformity I shall speak of events. It is intended to assess the value of the event. That is to say, it is intended to assess how good or bad it is – I use 'value' as a synonym for 'goodness'. An event may have various good features, which we call its 'benefits', and various bad features, which we call its 'costs', or the 'harms' it does. To assess how good the event is, taken as a whole, we have to put together its costs and benefits. This is a process of aggregation or weighing, and it is the purpose of cost-benefit analysis.

This is the conventional view, but it is not the only one. Cost-benefit analysis may alternatively been seen as a procedure for making a decision. In a democratic country, it may be seen as a democratic procedure: as a procedure the community sometimes adopts in order to put its wishes into effect. Different members of the community will naturally have conflicting preferences about some of the acts the community might choose to do. To arrive in a democratic fashion at a decision whether or not to do some particular act, people's conflicting preferences may have to be aggregated or weighed against each other. That can be seen as the purpose of cost-benefit analysis. It makes cost-benefit analysis parallel to voting or a parliament. The community decides some issues by means of voting, some through their representatives in parliament, and some by means of cost-benefit analysis.

Under the first interpretation of cost-benefit analysis, a particular way of doing the analysis would have to be justified on the basis of the theory of value. We should have to ask just how particular benefits and costs come together to determine the overall value of an event. By contrast, under the second interpretation, a particular way of doing cost-benefit analysis would have to be justified on the basis of democratic theory. We should have to ask what is an appropriate democratic way of putting together people's conflicting preferences in order to arrive at a decision.

The practice of cost-benefit analysis fits the first interpretation better than the second. To be sure, in a democratic country, cost-benefit analysis is a component of the democratic process – what else could it be? But it cannot plausibly be seen as itself a democratic decision-making procedure. In practice, the outcome of a cost-benefit analysis is never a decision; it is always just an input into a broader decision-making process. This is exactly what we should expect if cost-benefit analysis is seen as an evaluation. An assessment of value should be part of a well-functioning democratic process. When they participate in the democratic process, people at all levels – the government, the voters and their representatives – need information about the value of events they are concerned with. That is what cost-benefit analysis can supply them with.

In a democracy, a particular assessment of value should not determine a decision on its own. For one thing, it is not infallible, and the decision process needs to take into account how reliable it is. Secondly, the decision process may need to take into account considerations that are separate from value, such as the rights of property-owners or minorities. Finally, it has to be a genuinely democratic process, and democracy cannot always be expected to make the best decision. For example, the vote may go against the result of a cost-benefit analysis, even if the analysis is a correct assessment of value. Still, in a democracy, the vote should decide.

I dare say we could institute a process of decision-making by means of cost-benefit analysis, and it might be a genuinely democratic process. But that is not the role of cost-benefit analysis at present. As present, it is intended to be an evaluation, and that is how I

shall take it.

If it is to be an accurate evaluation, it needs to estimate costs and benefits at their true value. Consequently, it needs to be founded on a theory of value, aimed at determining what costs and benefits truly are. Cost-benefit analysis is the practical end of valuing. We start from a theory of value, and we put it into practical effect by means of cost-benefit analysis.

In this paper, I shall investigate some aspects of the theory of value that underlies cost-benefit analysis, but I shall take some other aspects for granted. I shall take it for granted that a cost-benefit analysis of an event is a comparison. It compares how good things would be if the event took place with how good they would be if it did not take place. So it compares the goodness of two states of affairs. I shall also take it for granted that the goodness of a state of affairs depends only on which people exist in that state, and how well off those people are. Briefly and roughly: the only good is the good of people. This assumption may be stated formally as something I call 'the principle of personal good', but I do not need to state it formally here.[1] It means that when we assess the costs and benefits of an event, we may concentrate on the wellbeing of people only.

Once we recognize that cost-benefit analysis depends on a theory of value, a question arises. How should we do cost-benefit analysis if no credible theory of value can be found? I ask this as a genuine – not a rhetorical – question. I do not know the answer to it. It is an important question in practice. In an important domain of cost-benefit analysis, no credible theory of value seems available. Yet there are pressing decisions to be made. What should we do?

2. Killing done by global warming

The problem is very widespread, as I shall explain. But it is especially pressing, or at least especially conspicuous, in decision-making about global warming. I shall use that as my example. Global warming will cause various harms and various benefits. Conversely, if we act to control global warming, we shall bring about various benefits, and also suffer various costs. To weigh them, we evidently need cost-benefit analysis.

Among the harms that global warming will do is the killing of very many people. It will kill in three predictable ways. First, infectious tropical diseases will increase their range. Second, there will be more frequent and more devastating floods. Third, because temperatures are higher, there will be an increased number of deaths caused directly by heat waves. Of these causes, the first will almost certainly be the most important, but only the third has been quantified. One estimate is that this third cause will kill around 200,000 people per year by about fifty years from now.[2] Taking all three causes together, it would not be unreasonable to predict a figure of 1,000,000 deaths per year by about fifty years from now, and this rate of killing might continue for centuries.

Taking a very conservative estimate of the number of killings and a very conservative valuation of each death, S. Fankhauser estimates the harm done by killing as 18% of the total harm done by global warming. Taking a less conservative valuation of a death, his estimate rises to 38%.[3] Whatever we may think of these figures, killing is plainly one of the greatest harms that global warming will cause. Any cost-benefit analysis of global warming or of measures to control global warming will have to take account of it.

3. The value of extending or shortening lives: value based on preferences

The harm of death and the benefit of saving life have been incorporated into practical cost-benefit analysis for a long time. As it happens, they have been incorporated badly. This point

is incidental to my argument, but I shall cause confusion if I say no more about it. So this section and the next, which are about the value of extending or shortening lives, are really an aside from my line of argument.

Cost-benefit analysts have been reluctant to acknowledge that a theory of value must underlie their work, and they have tried to keep theory out of their practice as much as they can. Economists are traditionally reticent people, and they do not like to impose their own theories of value on their work; instead, they prefer to leave valuations as much as possible to the individual preferences of the public. In particular, economists have tried not to commit themselves to any substantive theory about the value of a person's life. So when they value people's lives, they like to base their valuation on the people's own preferences about preserving their lives.

A person's preferences can be represented ordinally by a 'utility function'. That is to say, numbers called 'utilities' can be assigned to states of affairs in such a way that one state has a higher utility than another if and only if the person prefers the first to the second. If a cost-benefit analyst is to base values on preferences, she will have to use utility as her measure of value. She will have to take the value of an event to a person to be the increase or decrease in the person's utility that the event causes.

Preferences do not determine utilities uniquely. Given a person's preferences, there is always a wide range of utility functions that will represent them ordinally. Consequently, a cost-benefit analyst who wishes to base values on preferences needs to choose one function out of the range to serve as the measure of value for the person. Since the analyst is dealing with many people, she must choose one function for each person. In doing so, she implicitly does two things. For each person, she treats a particular function as a cardinal utility function for the person, and she takes each person's cardinal function to be interval-comparable with other people's. Less technically, she chooses a particular arithmetic scale to measure each person's preferences and she fixes a particular basis for weighing one person's preferences against another's.

In practice, the particular utility function picked by most cost-benefit analysts for valuing lives is one known as 'willingness to pay'. It is one of a class of functions called 'money-metric utilities'. In more detail, the value of a person's life is based on the money she is willing to pay to reduce a risk to her life. If she is willing to pay $50 to reduce by 1 in 10,000 her risk of being killed in (say) a road accident, then her life is valued at $50 times 10,000: $500,000, that is. A person's willingness to pay to reduce risk is a feature of her preferences about risk and about money. So this method bases the value of a person's life on her preferences, as it is intended to do.

A person's willingness to pay is normally used in cost-benefit analysis to provide both a cardinal scale of value and a basis for interpersonal comparisons of value. The effect of using it for interpersonal comparisons is to treat a dollar to one person as equally as valuable as a dollar to any other person. However, a dollar to one person is manifestly not equally as valuable as a dollar to any other person. A dollar to a peasant in Bangladesh will sustain life for a while, whereas a dollar to an affluent American, who already has all the necessities of life, will buy nothing of significant value. Using willingness to pay as a measure of value can result in absurdity. In a cost-benefit analysis of global warming, willingness to pay would treat an American life as worth ten or twenty Indian lives.[4] Since Americans are much richer than Indians, they are willing to pay ten or twenty times as much for safety.

This absurd conclusion is not inherent in the project of basing values on people's preferences. It results from adopting a money-metric utility function to represent a person's

preferences, rather than some other utility function. It is possible to use the same data of preferences in a better way, by making appropriate adjustments to money-metric utility.[5] Alternatively, a quite different utility function may be used. For example, one known as the 'healthy-years equivalent' has been recommended for health economics.[6]

However, all these utility functions represent preferences, and there are separate reasons why preferences are an unsatisfactory basis for valuing lives. There is plenty of evidence that, in contexts involving risks, people's preferences are generally muddled and incoherent. They cannot be considered rational,[7] which means they cannot be taken as a sound basis for valuation. Moreover, they are particularly dubious as a basis for valuation when it comes to people's lives. The value of a life is a complex thing, involving the aggregation of wellbeing across time, as I shall explain. It is a difficult theoretical problem to know how the value is determined. Few people have thought much about it, and it implausible that most people's preferences reflect a proper valuation of their lives.

The aggregation of wellbeing across time calls for some theoretical analysis; we cannot expect to base it entirely on people's preferences. We need a theory about how wellbeing distributed over time comes together to determine the overall value of a life. So we need a theory of value. Preferences may well play a part within the theory. For example, people's preferences are a plausible basis for determining the relative values of many current goods, such as different sorts of food. But we cannot escape the need for theory by trying to base all our valuations directly on preferences.

4. The value of extending or shortening lives: lifetime wellbeing

When a person's life is saved, she lives a longer life than she would have lived. The benefit to her is the difference between the goodness, or value, of her longer life and the goodness, or value, of the shorter life she would have lived. Conversely, if an event kills a person, the harm done her is the difference between the value of the longer life she would have lived and the value of the shorter life she actually lives.

So to do cost-benefit analysis properly, we need a theory about the value of a life. I shall propose one. I have no conclusive arguments for it, and indeed I recognize that it may well be wrong. It is only what I call a 'default theory'. It is a natural, plausible starting point, which I think we should accept unless we find good reason for departing from it. There are some arguments for it, but they are definitely not conclusive, and I shall not set them out in this paper.[8]

First of all, I shall assume that the goodness of a person's life depends only on how long it continues and on how well it goes at each time it is in progress. A person is born at some time and dies at some time, and at each time in between her life goes well (or badly) to some degree. I shall use the term 'temporal wellbeing' for how well it goes at a particular time. I shall use the term 'lifetime wellbeing' for the goodness of the life as a whole. My first assumption is that a person's lifetime wellbeing depends only on the length of her life and on her temporal wellbeing at all times in her life.

To put it another way, lifetime wellbeing is some sort of an aggregate of temporal wellbeing at all times in the life. What sort of aggregate? I assume as my default theory that it is simply the total. To find the value of a life, simply add up how well the life goes at each time. If a person's life is saved, the benefit to her is the increase in her lifetime wellbeing, which is the increase in the total of her temporal wellbeing. This 'total theory' is my default.

I can simplify it a little more if I make the assumption that there is no backwards causation of temporal wellbeing. Can events that occur later in a person's life affect her temporal

wellbeing at earlier times? Imagine you write a book, which later turns out to be influential. The event of its becoming influential may add value to all the time you spent writing it, by making your work during all those times worthwhile. That is arguable at least, but for simplicity let us assume away this type of backwards causation. Then we can say that the benefit of saving someone's life is the total temporal wellbeing that she goes on to enjoy in the rest of her life after she is saved. This is the simplified version of the total theory.

There are many other possibilities besides the default theory. For example, some authors think a life that starts badly but improves is better than one that starts well and deteriorates, even if both have the same total of temporal wellbeing.[9] Some authors discount wellbeing in later years compared to earlier years, which has the opposite effect. Another suggestion is that ups and downs are a bad thing, so that an life of even tenor is better than a variable one.[10] Conflicting with this is the view that a good life must have a high peak; what really matters is the best time in life. One might think that the goodness of a life is its average level of temporal wellbeing, rather than its total. And so on. I find it hard to assess these views. I know no good arguments for any of them; at best some of them seem intuitively attractive. But since some of them conflict with others, we ought to be suspicious of their intuitive attractions. There are some arguments for the total theory, and it is at least a simple theory. I think it is well qualified to be the default.

It only makes proper sense after we have done some preliminary technical work. First, we need to make sure that temporal wellbeing is measured on a cardinal scale. Second, we need to fix a zero of temporal wellbeing, because the zero makes a difference when it comes to comparing lives of different lengths. This means that the scale of wellbeing is actually more than cardinal; it is a ratio scale. Third, we have to make sure that wellbeing is comparable between different times in a person's life; it must actually be *fully* comparable. My own approach to these conditions is not to rely on an intuitive cardinal scale or an intuitive zero, but to define the scale and the zero, and then to defend the default theory specifically on the basis of the definitions.[11] I do not have to go into these technical matters here.

A theory similar to my default is regularly used to value lives in much of health economics. Health economists often measure the benefit of a treatment in terms of qalys, or quality-adjusted life years. If a person's life is extended by some treatment, they take the benefit to be the total number of years she afterwards lives, adjusted by the quality of life in those years. This is not quite the same as my default theory, since what health economists mean by 'quality of life' is not the same as temporal wellbeing. Still, it is similar. At least health economics has the merit that it bases the value of life on a plausible theory of value.

5. The value of adding or subtracting lives

According to the simplified version of my default theory, which assumes away backwards causation, if a person's life is saved, the benefit that results is the wellbeing that the person goes on to enjoy in the rest of her life. Saving her life adds wellbeing to the world in this way, and that is why we value it. But saving a life often adds wellbeing to the world in a different way too. If a young person is saved, she may well later have a child, who would never have existed had this person not been saved. The child will enjoy wellbeing during her life; her life will be good (or bad) to some degree. Why should we not count the child's wellbeing as part of the benefit of saving the existing person's life, if we count the wellbeing of the person herself?

This is just to raise the question. Saving a life adds wellbeing to the world, and so does creating a life. Why should we value one and not the other? Nothing forces us to treat the two

ways of adding wellbeing equivalently, but at least we need to think about the value of bringing a person into existence. If it does indeed have a value, we certainly should not ignore its value in cost-benefit analysis.[12] Cost-benefit analysis must rest on a theory of value, and the theory must account for population changes. Very many events and acts lead to the existence of new people, and many prevent the existence of people who otherwise would have existed. For example, a change in the rate of income tax will influence people's decisions about having children. Global warming will undoubtedly affect the world's population, though it may not yet be clear what its effect will be. So if adding people to the world has value – either positive or negative – its value is bound to be significant in many cost-benefit analyses.

This point is independent of my particular default theory for the value of extending an individual life. However we value extending life, we need at least to consider whether adding a new life to the world has a value. It is perhaps easiest to see this if you think, as I do, that the value of extending a person's life is simply the value of increasing the person's wellbeing. My discussion of the value of extending life was simply intended to lead up to this point. Now I have made it, we can set aside the question of how to value extending life. The default theory is no longer needed. Let us simply take it for granted that each person who lives has some level of lifetime wellbeing, however that may be determined. Then we can press on to consider the value of adding a person to the world. Naturally, this value may well depend on the level of lifetime wellbeing that the added person enjoys.

My terminology may cause confusion. I have been speaking of the value of a person's life, and also called this value the person's lifetime wellbeing. If a person is added to the population, she will have some lifetime wellbeing; her life will have some value. You might think that this value must be the value of the person's living that life; it must be the benefit of her existence, which must be added into our cost-benefit calculation. Once we have a theory of the value of life, why is that not automatically a theory of the value of adding a life to the world, as well as a theory of the value of saving a life? Why is there any question about it?

To clear this up, it may be helpful if I introduce a distinction. When I spoke of the value of saving a life, I meant its value to the person. It is the benefit to her of having her life saved. I shall call this its *personal* value. But when we consider the value of adding a person to the world, we are asking how valuable it is that this person lives. We are not asking for the value to the person herself, but the value 'from the point of view of the universe' as Sidgwick put it. I call this its *general* value. Nothing says that the general value of a person's life must be the same as its personal value to the person. My default theory is a theory of the personal value of a life; it leaves open the question of its general value. In other words, it leaves open the general value of a person's existence.

6. The intuition of neutrality

Now let us face up to this question of general value. What is the value of adding a person to the world? I know of no cost-benefit analysis in practice that has taken account of this value. I think the reason is plain and understandable. I think cost-benefit analysts take it for granted that adding a person has no value. More exactly, they assume it has no value in itself. No doubt adding a person has value for other people. Parents are often benefited by having a child, and on the other hand many people may be harmed by the demands made on the world's resources by an extra person. But these are externalities, and there is no doubt they should be included in a cost-benefit analysis. On the other hand, cost-benefit analysts assume that adding a person has no value in itself, apart from externalities. Apart from externalities,

the value of adding a person is zero.

This is understandable because it is an extremely natural intuition shared by very many people. The intuition is that bringing a person into existence is not in itself either good or bad; it is ethically neutral; it has no ethical value. Many people think that, and there is this argument to support it. Doing something is surely only good if it benefits someone and bad if it harms someone. To benefit a person you must make her better off than she would otherwise have been, and to harm her you must make her worse off than she would otherwise have been. Bringing a person into existence does not make her either better or worse off than she would otherwise have been. Therefore, it neither benefits nor harms her. So it must be neither good nor bad in itself.[13]

Let me spell out a specific implication of this intuition. Take two possible states of affairs, *A* and *B*, which contain the same population, except that *A* contains an extra person who does not exist in *B*. I mean that, besides this one extra person, both states contain exactly the same people, not merely the same number of people. Suppose that each of these other people is equally as well off in *A* as she is in *B*. For example, if the extra person's parents are happy in *A* because they have a child, then they have some compensating happiness in *B*, where they do not have a child. I assumed in section 1 that all value derives from people's wellbeing. So the relative value of *A* and *B* must depend on the wellbeing of the people who exist in both, or else on the extra person's wellbeing. Everyone who exists in both is equally well off in either; so far as these people's wellbeing is concerned, *A* and *B* are equally good. The intuition of neutrality tells us that the extra person's wellbeing has no value in itself. The upshot is that, according to the intuition, *A* and *B* are equally good.

The intuition is not affected by the level of wellbeing enjoyed by the extra person in *A*. Actually, that is not completely true. If the extra person has a life of unrelieved suffering, then most of us would think it a bad thing that she should exist; it would be better if she did not. Then we would think *B* is better than *A*. But provided the extra person's life is not a bad one, we would think *A* and *B* are equally good, whether her life is very good or only moderately good. At least for a wide range of levels of wellbeing, *A* and *B* are equally good. We might say there is a 'neutral range' of wellbeing such that adding a person whose wellbeing is in this range is equally as good as not adding her. The range may be very wide, extending from very mediocre lives up to the best lives imaginable. Intuitions vary about the width of the range, but most people's intuitions agree that there is a neutral range of some extent.

7. A single neutral level of wellbeing

I share this intuition, but I also know it must be false. To see why, think now of a third state *C*. *C* contains all the same people as *A* – all the people in *B* plus one more. Everyone apart from the extra person is equally as well off in *C* as she is in *A* and in *B*. In both *A* and *C* the extra person is within the neutral range. But in *A* she is better off than she is in *C*: her wellbeing is higher up in the neutral range. Undoubtedly, *A* is a better state of affairs than *C*, because it equally as good as *C* for everyone apart from the extra person, and it is better for the extra person. But according to the intuition, *A* is equally as good as *B*, and *B* is equally as good as *C*, because adding a person within the neutral range is neutral. 'Equally as good as' is a transitive relation. It therefore follows from the intuition that *A* is equally as good as *C*. But it is not: *A* is better than *C*. So the intuition is false.

This argument shows there cannot be a neutral range. There can only be a single level of wellbeing such that adding a person at that level is equally as good as not adding her. There is only a single 'neutral level', we may say. Adding a person at a level of wellbeing above the

neutral level is better than not adding her; it has a positive value. Adding a person below the neutral level is worse than not adding her; it has a negative value.

For all I have said so far, this neutral level may depend on the starting point. My example was concerned with adding an extra person to a state of affairs *B*; *B* constitutes the starting point. In it, a particular number of people exist, and each person has a particular level of wellbeing. Given this number of people and their levels of wellbeing, there is a single neutral level such that adding a person at that level is equally as good as not adding her. If we set out from a different starting point, there will still be only a single neutral level, but it may be a different one.

However, there are arguments to show, or at least strongly suggest, that the neutral level is in fact constant; it is independent of the starting point. This is not essential to my conclusions in this paper, so I shall not rehearse these arguments here.[14] But it is a convenience and I shall take it for granted; I shall assume there is a constant neutral level. Adding a person above this level is a good thing; adding a person below this level is a bad thing.

I shall make a further assumption that is not essential to the argument, but is a great convenience. I shall assume that, when we compare two states of affairs that have the same population, the one that has the greater total of people's lifetime wellbeing is better than the other. This is a utilitarian principle, applied only to a constant population. I do not assume it gratuitously. It can be supported by arguments that I find reasonably persuasive.[15]

From this constant-population utilitarian principle, together with a constant neutral level, it is easy to derive a specific formula for the value of a state of affairs.[16] I do not regard this formula as merely a default, because its two premises are supported by reasonably persuasive arguments. It is:

(*) $$(g_1 - \nu) + (g_2 - \nu) + \ldots + (g_n - \nu).$$

In this formula, $g_1, g_2, \ldots g_n$ are the lifetime wellbeings of all the people who exist, and ν is the neutral level. The theory of value represented in this formula was originally called 'critical-level utilitarianism' by Charles Blackorby and David Donaldson.[17] I prefer to call it the 'normalized total principle'.[18] To value a state of affairs, it tells us first to calculate, for each person, the difference between her wellbeing and the neutral level. Then we take the total of all these differences. This total is the value of the state of affairs.

We may treat the zero of wellbeing as arbitrary. If we were to set it at the neutral level, then (*) would be simplified. It would be simply the total of people's wellbeing – hence my name 'the normalized total principle'. However, for a technical reason that will appear (I wish to treat the neutral level as vague) I cannot actually make this normalization. So I shall assume the zero of wellbeing is assigned arbitrarily. It has no significance.

8. A vague neutral level

The conclusion that there is only a single neutral level, whether constant or not, is inescapably counterintuitive. What level of wellbeing could it plausibly be? Suppose it is at the level of a fairly good life. Adding someone whose life would be a little less good than this would be a bad thing. It would be worth some small sacrifice on the part of existing people to prevent the existence of a person at this level. Yet her life, if she lived it, would be a fairly good one. How could it be a bad thing that a person lives a fairly good life? Or suppose alternatively that the neutral level is at the level of a mediocre life. Adding a person whose life would be a little better than this would be a good thing. It would be worth some small sacrifice on the

part of existing people – a small reduction in their wellbeing – to bring this person into existence. Yet her life would be mediocre. It is implausible that existing people's wellbeing should be sacrificed for the sake of creating a person whose life would be mediocre. Either high or low, it is hard to believe there is just a single neutral level. Besides all this, there remains the argument I mentioned in section 6, that adding a person must be ethically neutral because it neither harms nor benefits anyone.

The argument is not conclusive. It rests on the assumption that an event is ethically neutral if it neither harms nor benefits anyone, and we could reject that assumption. Furthermore, we could conclude that our intuition about the value of adding people is unreliable. Not all intuitions are correct, and if they cannot be fitted into a coherent theory of value, they must be rejected. Still, intuitions are an important source for moral philosophy, and it would be unwise to give this one up without a struggle. It is worth seeing how far it can be accommodated in our theory of value.

Because the intuition is so powerful, a great deal of literature tries to make this accommodation. Larry Temkin argues that the relation 'equally as good as' may not be transitive. This would refute the argument I set out in section 7 for the conclusion that there is only a single neutral level.[19] Partha Dasgupta rejects the whole idea of goodness 'from the point of view of the universe', which the argument relies on. Dasgupta thinks that goodness must be understood in a relative way: goodness from the point of view of a particular population.[20] A third potential accommodation can be found in a notion of conditional goodness that may be drawn from Bernard Williams's notion of conditional desires.[21] I have investigated all these attempted accommodations and others, and concluded that all but one of them are unsuccessful.[22]

One is moderately successful. It stems from an idea that can be found in Derek Parfit's *Reasons and Persons*.[23] It has been developed in one direction by Charles Blackorby, Walter Bossert and David Donaldson.[24] I have developed it in a different direction in my *Weighing Lives*.[25] Here is an outline of my development.

I suggest there is indeed only one neutral level, but this level is vague. I favour the supervaluationist account of vagueness,[26] and I shall explain my suggestion in terms of that account. So first I need to outline supervaluationism.

According to supervaluationism, the meaning of a vague term such as 'bald' consists of a range of 'sharpenings', each of which is a potential interpretation of the term. The sharpenings of 'bald' include: having fewer than 1000 hairs on the head, having fewer than 1001 hairs on the head, having fewer than 990 hears on the head, and so on. According to supervaluationism, we may assert a statement if and only if it is true under every sharpening of its terms. For instance, we may say that Serge is bald if and only if Serge has fewer than 1000 hairs on the head, and fewer than 1001, and fewer than 990, and so on. Correspondingly, we may deny a statement if and only it is false under every sharpening. If a statement is true under some sharpenings and false under others, we cannot assert or deny it. If Serge has 999 hairs on his head, we cannot say he is bald, and nor can we deny it.

Remember that the neutral level of lifetime wellbeing is the level such that, if a person lives at that level, her existing is equally as good as her not existing. If a person lives at a higher level, it is better that she lives than that she does not. If a person lives at a lower level, it is worse that she lives than that she does not. I suggest that the term 'the neutral level' is vague. It has many sharpenings, each of which is a particular level of wellbeing. These sharpenings fall within some range. The range has a lower limit at some level of wellbeing, and an upper limit at some higher level. I do not rule out the possibility that the upper limit

might be infinite. Also, the limits themselves are likely to be vague, but I shall ignore that complication.

We may assert a statement if and only if it is true whatever level of wellbeing within the range we interpret as the neutral level. For example, we can assert that it is better that a person exists rather than not, if and only if the person's level of lifetime wellbeing is above the entire range. We can deny it if the level is below the entire range. If the level lies somewhere within the range, we can neither assert nor deny this statement. Conversely, we can say it is worse that a person exists rather than not, if and only if the person's level of lifetime wellbeing is below the entire range. We can deny it if the level is above the entire range. If it lies somewhere within the range, we can neither assert nor deny this statement

I said earlier that intuitively there is a neutral range of levels of wellbeing, rather than a single neutral level. As I originally interpreted the neutral range, for each level of wellbeing within the range, if a person will live at that level, her living is equally as good as her not living. Neutrality was interpreted as equality of goodness. Now, with the idea that the neutral level is vague, I have constructed a different sort of a neutral range. There is only one neutral level, but the level is vague. Its vagueness is spread over a range, which we may call a neutral range. For each level of wellbeing within this range, if a person will live at that level, we cannot either assert or deny that it is better she should live rather than not, and we cannot either assert or deny that it is worse. This is a different sort of neutrality, but it goes some way towards accommodating the intuition. I believe it is the best that can be done to accommodate it.

9. An incredible conclusion

It does not accommodate it fully, however. This sort of neutrality is not really neutral enough for intuition. To make this point I shall use a little example.

Suppose the neutral range of wellbeing is between 0 and 10. If a person's wellbeing is 7, adding that person to the population is neutral in my reconstructed sense: we cannot say it is better that she is added, and we cannot say it is worse. Nor can we deny either of these things. Now suppose a person is added at level 7, and at the same time some harm is done to an existing person. Suppose the existing person's wellbeing is reduced by 5 units as a result. How do we evaluate this combined change?

Let us adopt the normalized total principle, expressed in (*), as our theory of value. It is easy to see that the overall increase in the normalized total is $(7-\nu) - 5$. This is positive for values of ν less than 2 and negative for values of ν above 2. Now we are taking the neutral level *v* to be vague, we can assert a statement only if it is true under all sharpenings of 'neutral level'. That is to say, it must be true for all values of ν from 0 to 10. But it is not true for all values of ν from 0 to 10 that the combined change is beneficial overall, and nor is it true that it is harmful overall. So we cannot conclude that this change is beneficial, nor that it is harmful. It is neutral in our newly defined sense.

But this conclusion is not as it intuitively should be. The change is to add one person at 7 and reduce one person's wellbeing by 5 units. Adding a person at 7 is supposed to be neutral. Reducing a person's wellbeing by 5 units is a bad thing. So we are doing one neutral thing and one bad thing. Intuitively, the result should be a bad thing. But according to our theory, it is not; it is neutral. The neutral act of adding one person turns out to neutralize the definitely harmful act of harming another. This is not how neutrality should behave intuitively. It is not neutral enough; a neutral event should not cancel out other definitely good or definitely bad events.

This is a major failing in practice. It means that whenever any event affects the world's population, cost-benefit analysts cannot justifiably ignore this effect. Traditionally, they do ignore it, and I suggested the reason they do so is that they assume changes in population are ethically neutral. But even if changes in population are indeed ethically neutral, it turns out that they may neutralize other good or bad effects. This is because the only theoretically coherent account of neutrality is the one I gave in terms of vagueness, and that account allows neutrality to swallow up other good or bad effects. So changes in population cannot be ignored.

Take global warming again. As I said, global warming will kill many people; it will shorten many people's lives. I discussed how this harm might be evaluated, and offered nothing more than a default theory for that purpose. But however we do the evaluation, the conclusion must be that the killing is unambiguously a bad thing. There are no doubt exceptions; some people suffer so much that a shorter life is better for them than a longer one. But for nearly everyone, it is a bad thing for their lives to be shortened. So there will be this one definite bad effect of global warming.

Another effect of global warming is that it will undoubtedly change the world's population. I mean 'population' in a timeless way, to include all the people who ever live, at any time. When global warming kills someone, in one sense it directly reduces the population by one person, but not in this timeless sense. However, by killing people it does reduce the timeless population less directly, because some of the people it kills would later have had children, had they survived. Global warming will affect the population in other ways, too. It will alter many people's conditions of life in ways that will undoubtedly have demographic effects. Take one example. If global warming is not checked, large areas of Bangladesh will be drowned. Bangladeshis will have to move elsewhere in the world, and such a vast migration cannot conceivably happen without changes in the size of the population.

I shall not even try to predict whether global warming will increase or decrease the population; I predict only that it will alter the population. How should we evaluate this alteration? The intuition I have been pursuing is that we should value it at zero: adding people to the population of the world and subtracting people from it are both ethically neutral. I have explained that this intuition is hard to fit into a coherent theory, because theoretical arguments show there can only be one neutral level. But we can go some way towards accommodating the intuition if we assume the neutral level is vague. So let us make that assumption.

Let us also assume that most people's level of lifetime wellbeing is within the neutral range. That is to say, it lies with the range of vagueness of the neutral level. We assume, then, that the neutral range is wide enough to encompass most lives that are lived. It cannot plausibly encompass every life. Some people's lives are so full of suffering that it would have been better if they had not lived. Lives at this level are below the neutral range. But we might plausibly assume that every other life is within this range. For none of these lives can we say it is better that they are lived rather than not, and nor can we say it is worse that they are lived rather than not. They are neutral in this sense.

That is as far as we can go in accommodating the intuition that adding people is neutral. It allows to assume that the changes global warming will bring to the world's population are neutral.

The killing that global warming will do is unambiguously bad, and the effect it has on population is neutral. That makes one bad effect and one neutral effect. We should expect these two effects taken together to be on balance bad. Indeed, since the effect on population is neutral, we should expect to be able to ignore it and concentrate on the killing only, when we

do a cost-benefit analysis. But we cannot. Take a crude example again. Suppose the neutral range is between 0 and 10 units of wellbeing. Suppose global warming kills 100 million people, and on average the harm done these people by their death is 3 units each. Suppose global warming also causes the world's population to be reduced. Suppose it reduces the number of people who live at some time or other by 100 million. Suppose on average each of these people would have lived at level 5. The net harm done by these changes is 100 million times (5–ν) + 3, according to formula (*). This is positive for some values of ν within the neutral range (between 0 and 8) and negative for others (between 8 and 10). So we cannot say the change is harmful. Nor can we say it is beneficial.

Yet intuitively it is harmful, because it kills 100 million people. It seems obvious to me that we are entitled to say it is harmful, and a theory that does not yield this conclusion is incredible. Yet the only coherent theory of value that gives some respect to our intuitions does not permit us to draw this conclusion. So what are we to do?

Notice that the problem is very widespread. The example of global warming perhaps raises it more obviously than other problems for cost-benefit analysis. But it will infect the cost-benefit analysis of any event that alters the world's population. Bear in mind that any effect on population is likely to be very large in total. The world's population is not limited by some stable process that causes small changes to die out over time. Instead, any small change will persist for generations and perhaps for ever. A few tens of extra people now will grow to a large number of extra people over the generations. It is true that the problem can be mitigated by discounting future harms and benefits, since many of the people added will not exist until far in the future. But discounting lives is an arbitrary device, with little to justify it.

I conclude there is a serious problem in the foundations of cost-benefit analysis. We can hope that more work in the theory of value will come up with a solution. In the meantime, I do not know what cost-benefit analysts should do.

Notes

1. See my *Weighing Lives*, chapter 6.
2. See the survey in Pearce et al, 'The social costs of climate change'.
3. See Fankhauser, *Valuing Climate Change*, reported in Pearce et al, p. 197.
4. See Pearce et al, 'The social costs of climate change', pp. 195–8. This consequence of the willingness-to-pay utility function caused justified consternation in meetings of the Intergovernmental Panel on Climate Change. See the IPCC's 'Policymakers' summary', p. 11.
5. See Drèze and Stern, 'The theory of cost-benefit analysis'. In 'The social costs of climate change', p. 206, Pearce et al point out that weights can be used to adjust the willingness-to-pay valuations.
6. See Mehrez and Gafni, 'Quality-adjusted life years, utility theory and healthy-years equivalents'.
7. For instance, see Tversky and Kahneman, 'Rational choice and the framing of decisions'. An example that creates problems specifically for valuation appears in Jones-Lee, Loomes and Philips, 'Valuing the prevention of non-fatal road injuries'.
8. See my *Weighing Lives*, chapters 14–16.
9. For instance, Velleman in 'Well-being and time'.
10. See Sen, 'Utilitarianism and welfarism'.
11. See my *Weighing Lives*.

12. In *The Economics of Safety and Physical Risk*, Jones-Lee suggests it may be permissible to ignore the value of adding people to the population when the effects on population are unpredictable. But in many cases, the effects are quite well predictable. For example, most people have children, so it is predictable that saving the life of a young person will commonly bring it about that she later has children.
13. An argument like this appears in Narveson's 'Utilitarianism and new generations'.
14. See Blackorby, Bossert and Donaldson, 'Intertemporal population ethics', and a similar argument in my *Weighing Lives*, chapter 12.
15. See my *Weighing Goods* and *Weighing Lives*, chapter 7.
16. The derivation is in *Weighing Lives*, chapter 13.
17. Blackorby and Donaldson, 'Social criteria for evaluating population change'.
18. *Weighing Lives*, chapter 13.
19. Temkin, 'Intransitivity and the mere addition paradox'.
20. Dasgupta, 'Savings and fertility'.
21. Williams describes conditional desires in his 'The Makropulos case'. I developed the idea of conditional goodness and applied it to the problem of population in my 'The value of a person'.
22. See my 'The value of a person' and *Weighing Lives*, chapters 8–11.
23. Parfit, *Reasons and Persons*, pp. 43–2.
24. 'Quasi-orderings and population ethics'.
25. *Weighing Lives*, chapter 10.
26. See Fine, 'Vagueness, truth and logic'. Strong objections to supervaluationism appear in Williamson's *Vagueness*. My own version of supervaluationism differs from Fine's, in order to overcome Williamson's objections. Details appear in my 'Supervaluation reconstructed'.

References

Blackorby, Charles and David Donaldson, 'Social criteria for evaluating population change', *Journal of Public Economics*, 25 (1984), pp. 13–33.

Blackorby, Charles, Walter Bossert and David Donaldson, 'Intertemporal population ethics: critical-level utilitarian principles', *Econometrica*, 65 (1995), pp. 1303–20.

Blackorby, Charles, Walter Bossert and David Donaldson, 'Quasi-orderings and population ethics', *Social Choice and Welfare*, 13 (1996), pp. 129–50.

Broome, John, 'Supervaluation reconstructed', unpublished.

Broome, John, 'The value of a person', *Proceedings of the Aristotelian Society*, Supplementary Volume 68 (1994), pp. 167–85. Reprinted in my *Ethics Out of Economics*, Cambridge University Press, 1999, pp, 228–42.

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

Broome, John, *Weighing Lives*, forthcoming.

Dasgupta, Partha, 'Savings and fertility: ethical issues', *Philosophy and Public Affairs*, 23 (1994), pp. 99–127.

Drèze, Jean, and Stern, Nicholas, 'The theory of cost-benefit analysis', in *Handbook of Public Economics, Volume II*, edited by Alan J. Auerback and Martin Feldstein, North–Holland, 1987, pp. 909–89.

Fankhauser, S., *Valuing Climate Change: The Economics of the Greenhouse*, Earthscan, 1995.

Fine, Kit, 'Vagueness, truth and logic', *Synthese*, 30 (1975), pp. 265–300.

Intergovernmental Panel on Climate Change, 'Policymakers' summary', in *Climate Change 1995: Economic and Social Dimensions of Climate Change*, edited by James P. Bruce, Hoesung Lee and Erik F. Haites, Cambridge University Press for the Intergovernmental Panel on Climate Change, 1996.

Jones–Lee, M. W., *The Economics of Safety and Physical Risk*, Blackwell, 1989.

Jones–Lee, M. W., G. Loomes and P. R. Philips, 'Valuing the prevention of non-fatal road injuries: contingent valuation vs standard gambles', *Oxford Economic Papers*, 47 (1995), pp. 676–95.

Mehrez, A., and A. Gafni, 'Quality-adjusted life years, utility theory and healthy-years equivalents', *Medical Decision Making*, 9 (1989), pp. 142–9.

Narveson, Jan, 'Utilitarianism and new generations', *Mind*, 76 (1967), pp. 62–72.

Parfit, Derek, *Reasons and Persons*, Oxford University Press, 1984.

Pearce, D. W., W. R. Cline, A. N. Achanta, S. Fankhauser, R. K. Pachauri, R. S. J. Tol and P. Vellinga, 'The social costs of climate change: greenhouse damage and the benefits of control', in *Climate Change 1995: Economic and Social Dimensions of Climate Change*, edited by James P. Bruce, Hoesung Lee and Erik F. Haites, Cambridge University Press for the Intergovernmental Panel on Climate Change, 1996.

Sen, Amartya, 'Utilitarianism and welfarism', *Journal of Philosophy*, 76 (1979), pp. 463–89.

Temkin, Larry S., 'Intransitivity and the mere addition paradox', *Philosophy and Public Affairs*, 16 (1987), pp. 138–87.

Tversky, Amos, and Daniel Kahneman, 'Rational choice and the framing of decisions', *Journal of Business*, 59 (1986), pp. 250–78, reprinted in *The Limits of Rationality*, edited by Karen Cook and Margaret Levi, Chicago University Press, 1990, pp. 60–89.

Velleman, David, 'Well-being and time', *Pacific Philosophical Quarterly*, 72 (1991), pp. 48–77.

Williams, Bernard, 'The Makropulos case: reflections on the tedium of immortality', in his *Problems of the Self*, Cambridge University Press, 1973, pp. 82–100.

Williamson, Timothy, *Vagueness*, Routledge, 1994.