REPLIES

JOHN BROOME

University of Oxford

I am extremely grateful to the five commentators for reading my book and offering such interesting thoughts in reaction. Shortage of space may make my responses seem brusque. But of course they are not meant to be.

1.

Roger Crisp urges me to accept hedonism. Two versions of hedonism are:

Personal hedonism: The goodness of a life depends only on the amounts of pleasure and pain it contains.
General hedonism: The goodness of a distribution of wellbeing depends only on the amounts of pleasure and pain it contains.

Both are consistent with the theory of value contained in *Weighing Lives*, but general hedonism restricts the theory, because it implies a particular value for what I call the 'neutral level for existence'. This neutral level is defined as the level of wellbeing such that adding to the population a person who has that level of wellbeing is equally as good as not adding her. General hedonism implies that this neutral level is the level of a life that contains no pleasure and no pain. That is because adding a person who has no pleasure and no pain leaves unaltered the total amounts of pleasure and pain in the distribution.

It is general hedonism that Crisp urges on me. In *Weighing Lives* I rejected it, but I should not have done. My policy was to avoid taking sides on questions about what is ultimately valuable; the book is not about that but about how ultimate values – whatever they are – aggregate together. So I now withdraw my rejection of general hedonism, but I do not accept it either. I return to a resolutely uncommitted stance.

115

However, I do want to point out that general hedonism leads to a dilemma. If you accept general hedonism, you must either embrace the repugnant conclusion or alternatively deny utilitarianism. I do not mean merely that you must deny some specific version of utilitarianism that comes into play when the population changes. You must deny the core of utilitarianism: the view that, of two distributions that have the *same* population, the one that has the greater total of wellbeing is the better.

Here is the argument. Assume personal and general hedonism. Start with a distribution containing a billion people, each enjoying a great deal of pleasure and experiencing no pain. Call this the high-level distribution. Now take an enormously larger population. Take the same total of pleasure and divide it up equally among all these people. In the resulting distribution, each person will have just a tiny amount of pleasure, and no pain. Call this the low-level distribution. According to general hedonism, the low-level distribution is just as good as the high-level one, since the amount of pleasure is the same in both. This is an instance of the repugnant conclusion.

The only way to avoid it is to deny that pleasure can be divided up in the way I assumed. We could suppose that pleasure itself cannot be divided into minute amounts. The low-level distribution could not then exist. But that would not block the repugnant conclusion if we understood it as the claim that, *if there were* a low-level distribution, it would be just as good as the high-level distribution. This claim is not falsified by assuming that the antecedent of the conditional cannot be true.

The only satisfactory response is to suppose there are higher and lower pleasures, with particular properties. The lower ones can be infinitely divided, but not the higher ones. Let a higher pleasure be reading *Pride and Prejudice*, and let a lower one be drinking lemonade. The lower pleasure can be divided into tiny quantities by drinking tiny quantities of lemonade, but to get the pleasure of reading *Pride and Prejudice* you need to read the whole book. Furthermore, suppose the value of the lower pleasure – however much of it there is – cannot be as great as the value of the higher pleasure gained from one reading of *Pride and Prejudice*. For instance, the function that specifies the goodness of a distribution might be:

$$v = h + l/(1 + l)$$

where h is the amount of higher pleasure and l is the amount of lower pleasure. For positive values of l, this is a continuous function; there is no discontinuity.

Now, in our example, let each of the billion people in the high-level distribution have one unit of higher pleasure. Let each person in the low-level distribution have just a tiny amount of lower pleasure. However many people there are in the low-level distribution, this distribution is inevitably worse than the high-level one. So there is no repugnant conclusion.

Assume now there is a quite enormous but fixed population of people, and compare two distributions. In the first, everyone has no pleasure and no pain, except for a billion people who each have one unit of higher pleasure, and no pain. Call this the new high-level distribution. Remember it is different from the old one because, besides this billion, it contains an enormous number of people who have no pleasure and no pain. In the new low-level distribution, n people each have a tiny amount of lower pleasure, and no pain. Everyone else in the low-level distribution has no pleasure and no pain. Our value function tells us that the new high-level distribution is better than the new low-level one, however large n may be. This is the implication of general hedonism, once we incorporate higher and lower pleasures as I have described them.

But it is inconsistent with utilitarianism. According to hedonism, each person in the low-level distribution has some wellbeing, which is greater than the wellbeing of those who have no pleasure and no pain. Utilitarianism says that the goodness of the distribution is the total of people's wellbeing. Each person who has some pleasure contributes to this goodness, and each contributes the same amount. So the goodness of the distribution increases in proportion to the number n of people who get some lower pleasure. Provided n is large enough, this goodness will exceed any number at all. In particular, for a large enough n, the low-level distribution will be better than the high-level one. That is the implication of utilitarianism. It contradicts the implication of general hedonism.

True, the required number of people n might be above the fixed population. But that only shows we did not set up the example with a big enough fixed population. We have only to go back and start again with a bigger one.

True too, I am assuming the standard arithmetic of real numbers. If hedonism can be reconciled with utilitarianism only with the help of infinitesimal numbers, that in itself would be remarkable.

How can general hedonism possibly be inconsistent with utilitarianism? Because general hedonism does not recognize the boundaries between persons, whereas in one way utilitarianism does. Utilitarianism can accept the hedonistic values of pleasure. But it can value pleasure only in so far as it increases the wellbeing of the person who has it. Consequently, it can matter for utilitarianism how quantities of pleasure are packaged into different people's lives. On the other hand, general hedonism does not care how pleasure is packaged. It cares only about the quantity of pleasure, however it is distributed over people.[1]

---

[1] I must thank Douglas MacLean for helping me improve my comments on Roger Crisp.

2.

Douglas MacLean does not like consequentialism. But why? Consequentialism as I defined it seems such an innocuous claim. It says simply that the goodness of an act is given by the goodness of its consequences. Since an act's consequences can be defined broadly to include just about everything that could possibly affect its value, this seems innocuous. Anyway, it is a view that is internal to axiology. It is about the internal structure of goodness, not about the relation between rightness and goodness. It makes no claims about how you ought to act. So whatever concerns you have about the relation between rightness and goodness, they should not lead you to doubt consequentialism. For instance, suppose you believe that what you ought to do is unrelated to the goodness of your actions or the goodness of anything else. From your point of view consequentialism would have no practical importance, but you have no reason to deny it.

MacLean says that my version of consequentialism is characterized by two claims:

> The first is that things, prospects, or states of affairs can be good in themselves or the bearers of intrinsic ethical value. The second claim is about the relationship between intrinsic values and ethical reasons. Values give rise to and explain ethical reasons for action.

But if I made those claims it was an accident. On page 32 of *Weighing Lives* I deliberately set aside the question of explanatory priority between values and reasons – or between good and right. As it happens, I am inclined to give priority to the right over the good, as MacLean does. One way of doing that is described in Section 3.1. Start with a theory of the right or a 'normative theory' as I call it. For any choice you might face, this theory will say which of the options you ought to choose. It specifies a 'choice function' as economists say. If this function satisfies some particular conditions (the so-called 'expansion' and 'contraction' conditions), it will determine an ordering of the options such that, when you have a choice among some options, you ought to choose the highest-ranked among them. We may take this to be the theory's betterness ordering. So the good is determined entirely by the right. Goodness has no part in determining what you ought to do.

This method for determining betterness from rightness works only in particular conditions. When a normative theory meets those conditions I call it 'teleological'. When it does not, it may be possible to massage it into teleological form; *Weighing Lives* considers that possibility. But a theory may not be teleologizable in this way. If not, the theory does not determine a betterness ordering in the way I described. It may determine one in some other way. Alternatively, it may determine no betterness ordering at all.

All this is compatible with consequentialism as I defined it. To be sure, if there is no betterness ordering at all, consequentialism will be empty, but that does not make it false.

In *Weighing Lives* I also left room for hybrid theories in which betterness plays some independent role in explaining what you ought to do; it is not entirely determined by a prior normative theory. I also left room for the opposite extreme, in which the good is entirely prior to the right – where what you ought to do is entirely explained by betterness. The relative priority of the right and the good makes no difference to the arguments of my book, so I did not take sides on it. *Weighing Lives* says nothing about how one ought to act.

I understand very well that MacLean does not want to give priority to the good over the right. But I do not think he should object to consequentialism as I defined it.

3.

Krister Bykvist speaks of a level of life that is neutral for a person. He think it is good for a person to live a life that is above this level, and bad for her to live a life that is below it. Furthermore, he thinks that, according to utilitarianism at least, the question of whether a person's life is good or bad for her is important in determining whether the existence of this person is a good thing or a bad thing.

I did not myself speak of the level of life that is neutral for a person, or good for her, or bad for her. An expression that purports to mean the same as 'a life that is good for a person' is 'a life worth living'. I did not use that expression with this meaning, either. But many people do, and evidently they think they understand it. It occurs a lot in the philosophical discussion of the ethics of population. But you may think you understand an expression without actually doing so. For example, many people think they understand 'life after death', but they may not.

One thing that may lead you to think you understand an expression when you do not is that the expression may have a clear meaning in a different context. This is so with 'a life worth living'. 'A life' may refer to a whole lifetime, or to how a person is living at a particular period in her life. It makes clear sense to say that your life during a particular period is worth living; it means that your life including this period is better than your life would have been had you not lived through this period. So 'a life worth living' has a clear sense when applied to a period of life, and that may encourage people to think it has a meaning when applied to a whole lifetime.

It sounds arrogant to claim that people do not understand the notion of a life worth living when they think they do. But when the meaning of a

term is open to doubt, it is fair to ask for a definition of it, or an analysis, or some sort of explanation of its meaning.

Bykvist asks 'Why does Broome set the neutral level at a positive value? Why shouldn't we instead just . . . set it at the same level as a life that is neutral for people?' At this point in his comments, he has not explained what he means by a life that is neutral for a person, so at that point I cannot make sense of his questions.

Later, Bykvist tentatively adopts a definition of a life that is neutral for a person. It is a life that is equally as good as a 'constantly neutral life'. A constantly neutral life is one that, at every time, is on the borderline of being worth continuing: a life such that, for the person who lives it, dying at any time would be equally as good as continuing to live. This definition endows Bykvist's questions with meaning. The neutral level for existence is the level of lifetime wellbeing such that it is equally good that a person lives a life at that level as that she does not live at all. The question is: 'Why should this neutral level be above the level of a constantly neutral life?'

That is *all* there is to the question. There is no further significance in the question of why I set the neutral level at a positive value. I chose to set my zero of wellbeing at the level of a constantly neutral life, but that was arbitrary – a matter of convenience only.

Still, the question is a good one. I gave my arguments (p. 259), but I do not suggest they are conclusive. However, they are not quite as weak as Bykvist suggests. Bykvist repeats one of my examples; please refer to his figure 3. For that example, I claimed that our intuition suggests the distribution C is better than D. That is surely correct: the person p is worse off in D than in C to the extent of half a lifetime, and that loss to p is intuitively not compensated for by the existence of another short-lived person q in D. If I understand him, Bykvist suggests that our intuition here depends on assuming that p exists already, at the time we make the choice between C and D. But my intuition, at least, does not. To me it seems just as bad for a future person's life to be shortened as it is for a present person's life to be shortened.

Bykvist explains that the intuitively attractive conclusion that C is better than D will emerge from my formula for value, provided we assume that the neutral level for existence is above the level of a constantly neutral life. Otherwise it will not. That was one of my reasons for making this assumption: it accords with intuition in this case.

Bykvist then modifies the example. He changes D to 'modified D' (as I shall call it), in which p does not exist but her place is taken by a different person r, who also lives for the first two times at level 1. Everything else remains the same. Again, my assumption will rank C above the modified D. Bykvist correctly points out that we do not have the same intuition that C is better than modified D, since nobody is worse off in modified D than she is in C.

Still, the conclusion that *C* is better than modified *D* follows easily from our previous intuition that *C* is better than unmodified *D*, if we add just one further assumption that is itself intuitively attractive. We have only to assume that value is impartial between people: that the value of a distribution is not affected by the identify of the people in it. It follows that modified *D* is equally as good as unmodified *D*. Since our intuition has told us that *C* is better than unmodified *D*, *C* is better than modified *D*.

This illustrates the point of theory: it goes beyond our direct intuitions.

4.

In *Weighing Lives* I described a common intuition that I called 'the intuition of neutrality'. It is the intuition that adding a person to the population of the world is, as a general rule, neither better nor worse than not adding her. I mentioned three doubts I had about it. Qizilbash interprets neutrality as 'parity', and uses this interpretation to respond (here) to two of my doubts. He suggests that, as a general rule, adding a person to the population is on a par with not adding her.

The first of my doubts is what Qizilbash calls my 'alternative intuition'. He presents this intuition, quoting me, as 'Neutrality is most naturally understood as equality of value.' That remark of mine was meant to apply only in a particular context. There are many examples of comparisons where different values are at stake. The famous dilemma of Sartre's student is one. In those cases, neutrality is naturally understood as incommensurateness of value, since the different values may be incommensurable. But when only one value is at stake, that is not a natural understanding. If you have a choice between two drinks of lemonade, and neither is better for you than the other, the natural thing to suppose is that the alternatives are equally good. That was the context of my remark.

I said that, when we are thinking about adding a person to the world, the only value at stake, if it is a value, is the number of people. So if neutrality in this case is to be understood as anything other than equality of value, we need an explanation of why. There may indeed be an explanation. Qizilbash gives the beginning of one when he says the comparison has some complexity. But I would like to know more about this complexity, since there appears to be only one value at stake.

My second doubt about the neutrality intuition is that neutrality is implausibly 'greedy' if it is understood as something other than equality of value. Qizilbash describes accurately what I mean. Please refer to the example of mine that he sets out in his comment. Compare the options *a* and *c*. There are two differences between them. One is that *c* contains an extra person; that difference is supposed to be of neutral value. The other is that one person is worse off in *c* than in *a*. That is a bad thing. A neutral thing and a bad thing should together make a bad thing. So, if neutrality

behaved as it should, *c* would be worse than *a*. But actually *c* is not worse than *a*. Neutrality does not behave as it should; it swallows up the bad thing. That is my doubt.

Qizilbash suggests in response that neutrality might plausibly swallow some badness, so long as it is not significant badness. Maybe that is so. But in this example, neutrality swallows significant badness; the person who is worse off in *c* is much worse off, and that is significant. Nothing in Qizilbash's account of parity prevents this from happening. So I do not think Qizilbash has a real response to this doubt.

<div align="center">

5.

</div>

It is not the job of philosophy to give direct practical advice either to people or to governments. Nevertheless, moral philosophy is immensely significant in practical matters. It influences the way we think and act, but only slowly as it filters through the process of public debate. I hope *Weighing Lives* will have a practical influence, but it is not meant to be a directly practical guide.

However, decisions about life and death are made daily. Individuals decide what measures to take to extend their lives, and what risks to take with their lives. Governments make similar decisions on behalf of their people. If a government's decisions are made badly, they will be very wasteful of people's lives. For example, the UK government seems willing to impose very high standards of safety on the railways, so that a lot of money is spent on the railways for each life saved, whereas in the UK health service, life is much cheaper. Many more lives would be saved if resources were moved from the railways to the health service.

So we urgently need these decisions to be made well. Fortunately, there are good practical thinkers who contribute to making them. They look for practical ways to assign value to the saving of lives. Michael Jones-Lee is one of the most influential. He has done us a great service by improving public decision-making in matters of life and death.

Jones-Lee is particularly associated with the 'willingness to pay' approach to valuing lives. To speak very roughly, this approach values a person's life on the basis of what she is willing to pay to extend it. An alternative is popular among those who work in public health. They value lives in terms of qalys – quality-adjusted life years – or similar quantities. One limitation of the qaly approach is widely recognized. Its unit of value, the qaly, is not comparable with the value of economic goods. In particular, it is not comparable with the costs of extending people's lives in the health service, or on the railways or elsewhere. This means the qaly approach cannot tell us how much it is worth spending to save lives. It can tell us how life-saving resources can best be used, but not what quantity of

resources we should devote to life-saving in the first place. On the other hand, willingness to pay gives the value of lives directly in terms of money.

In recent years there has been some convergence between the two approaches. Jones-Lee mentions a project to set a willingness-to-pay money value on a qaly. This will be an important advance. For one thing, the willingness-to-pay approach will be improved by recognizing that saving a person's life is more valuable when she has many years to live than when she has few. For another, it will assign qalys a value that is comparable with the value of other things such as the costs of life-saving.

But I must reiterate a cautionary remark I made in *Weighing Lives*. Willingness to pay makes the value of qalys comparable with other things. However, without adjustments, it is no good for assessing the relative values of things to different people – for 'interpersonal comparisons', as economists say. Suppose one person is willing to pay £100,000 to extend her life by one year, and another £10,000. It does not follow that a year of the first person's life is worth ten times a year of the second person's. The difference may be caused entirely by the different values that money has for the two people. Money is generally more valuable to a poor person than to a rich one, because a poor person has more urgent needs to satisfy. Money is generally more valuable to someone who is more remote from death (younger, perhaps) than to someone who is closer to death, because she has more opportunities to spend her money. Willingness to pay cannot properly be used to value lives or anything else, unless the differing values of money are corrected for.

There are ways of correcting for them. One crude correction is mentioned by Jones-Lee: the UK Department for Transport assigns the same value to each person's life, whether she is rich or poor, near to death or far from it. This is too crude, since not all lives have the same value. Still, to treat each person's life as having the same value is certainly closer to the truth than to treat money as having the same value to everyone.

However, although corrections can be made, practitioners of willingness to pay seem unable to purge themselves entirely of the assumption that money has the same value to everyone. Jones-Lee makes this assumption himself at one point in his comment.'For simplicity', he says (section 2), 'we set the marginal social welfare of wealth ... equal to unity across all individuals'. That is the assumption, expressed in economists' terms.[2] It remains pervasive and potentially dangerous. For example, one report of the Intergovernmental Panel on Climate Change suggested that American lives might be assigned 15 times the value of Bangladeshi lives, on the grounds that Americans are willing to pay 15 times as much as Bangladeshis to extend their lives.[3] Here is a more

---

[2] Jones-Lee and his colleagues correct it in a later paper: Baker *et al.*, Valuing lives equally.
[3] Pearce *et al.* 1996: 195–8.

recent example. The US Office of Management and Budget instructs federal agencies to assign more value to a year of an older person's life than to a year of a younger person's, on the grounds that 'senior citizens may have accumulated savings to spend on their health and safety'.[4] What sort of grounds are those? The OMB must be saying that an older person, having accumulated savings, may be willing to spend more on extending her life. But this does not show a year of her life is worth more. It shows money is worth less to her, because she has more of it.

If we are to trust the willingness-to-pay approach, it must cast off this prejudice about the value of money.

REFERENCES

Baker, Rachel, Susan Chilton, Michael Jones-Lee and Hugh Metcalf. Valuing lives equally – a theoretical introduction

Office of Management and Budget. 2003. Circular A–4: regulatory analysis. 17 September

Pearce, D. W., W. R. Cline, A. N. Achanta, S. Fankhauser, P. K. Pachauri, R. S. J. Tol and P. Vellinga. 1996. The social costs of climate change: greenhouse damage and the benefits of control. In *Climate change 1995, Volume III: Economic and social dimensions of climate change*, 179–224. Intergovernmental Panel on Climate Change, Cambridge University Press

---

[4] Office of Management and Budget. Circular A–4, section E.