Lessons from economics
John Broome

For *The Oxford Handbook of Ethics and Economics*, edited by Mark D. White, Oxford University Press

1. Introduction

Economics and the methods of economics offer several lessons that moral philosophers could beneficially learn. I shall mention some of them, using two significant examples of topics where they can be useful. Indeed, on these particular topics, ignoring the lessons from economics has been seriously damaging; I shall show that it has led to bad mistakes. So my conclusion is not merely that lessons from economics could be beneficial. They are actually essential for dealing properly with some topics within moral philosophy.

Much of moral philosophy is concerned with the personal relations and interactions of a few people. Economics has little to contribute there. But many moral philosophers these days deal with large-scale subjects that have quantitative aspects. For example, they deal with the ethics of public health, of social inequality and of population. These subjects concern economists too. A branch of the discipline of economics known as 'welfare economics' is concerned with making ethical judgements about them and other subjects.

Public health, social inequality and population each have a quantitative aspect and an ethical aspect. Economists are adept at analysing the quantitative aspects using mathematical methods. To be successful, their methods must be based on good ethical theory. But the academic labour cannot just be divided, with moral philosophers working on the ethical bases of economics, and economists applying philosophers' conclusions to complex, large-scale problems. The ethics and the economics are too closely intertwined to be teased apart.

Each discipline must work with the other. Economists typically have their own moral philosophy. For example, they typically believe that value derives only from people's preferences. Partly as a result, they sometimes have the surprising view that different people's wellbeings cannot be compared together. Philosophy can help them clarify their thinking. The papers in this handbook describe some of the lessons that economists should learn from philosophy.

Conversely, moral philosophers must pay attention to the methods and conclusions of economics. That was the theme of my book *Ethics Out of Economics* (Broome, 1999); I return to it now with new examples. Unfortunately, moral philosophers have been slow to learn from economists. This is partly because some of them are unwilling to employ or even read mathematical formulae, which are the staple of economics. No one need engage in a discipline that does not interest her, but a philosopher who does not engage in mathematics should avoid certain quantitative subjects within moral philosophy. When economists employ mathematics, they usually do so because it is needed.

To illustrate the lessons moral philosophy can learn from economics, I have chosen two particular topics as examples: the ethics of equality and the ethics of population. Sections 3–6 describe how the methods of economics can contribute to our understanding of the value of equality. Section 7 illustrates how they can help with population ethics and some connected topics.

Those are examples of beneficial lessons that philosophers can learn from economics. But this chapter starts in section 2 with a bad lesson from economics that many moral philosophers have already thoughtlessly learnt. It is the use of the world 'utility' to denote a person's good or wellbeing. This usage was invented by economists. They would have done better not to have invented it, and moral philosophers would do better not to copy them. For some reason, many philosophers seem more attracted by this bad habit of economists than by their good ones.

2. Do not use 'utility' to denote a person's good

It is perfectly unnecessary to use the word 'utility' to denote a person's good or wellbeing. We already have the words 'good' and 'wellbeing'. 'Benefit' is another useful word with the meaning of 'add to good'. For explicit comparisons we have 'better for'. We do not need a further artificial term. Moreover, using 'utility' with this meaning leads to mistakes. It has been doing damage in economics since the middle of the twentieth century, and more recently in philosophy too. I first complained about this usage in (Broome, 1991a), at a time when few philosophers had adopted it. But by now many have. It should be eschewed.

In economics, 'utility' is ambiguous. First, it may refer to something that plays a role in economists's ethical judgements or their judgements of value. Much of economic theory is concerned with making valuations of states of affairs and policies, to judge how good or bad they are. Often these valuations are based on a 'social welfare function', which could alternatively be called a 'value function'. Typically a social welfare function is written as a function of people's utilities: $W(u_1, u_2, \ldots u_n)$, where u_i stands for the ith person's utility. If the function is specifically the sum of utilities $W = u_1 + u_2 + \ldots + u_n$, it is referred to as a 'utilitarian' function. Utilitarianism is the view that the value of a state of affairs is the sum of individuals' wellbeings, so this tells us that utility is understood to be wellbeing. Even when the function is not additive, the role of a person's utility within a social welfare function is to denote her wellbeing or what is good for her. That is one meaning of 'utility'.

Mistakes occur because the same word is also used by economists with a quite different meaning. 'Utility' is defined in economic theory as the value of a function that represents a person's preferences. This function *represents* preferences in the sense that one thing is defined to have greater utility than another if and only if the person prefers it. This means that a person's preferences maximize utility, by the definition of utility. Utility defined this way is everywhere in economic theory. The definition appears in the canonical works such as (Debreu, 1959) and (Deaton and Muellbauer, 1980). It plays an important role in consumer theory, general equilibrium theory, game theory and all those other parts of economics where a person's behaviour is supposed to be given by her preferences. In none of these places need there be any implication that utility is a person's good. 'Utility' in these contexts is a totally different, technical term.

However, economists tend to assume that one thing is better for a person than another if and only if she prefers it. This may be because they hold the theory that a person's good consists in the satisfaction of her preferences, or because they assume that a person determines her preferences on the basis of what is better for her. But whatever the reason, their view is no doubt encouraged by their inclination to use 'utility' ambiguously for good and also for a representation of preferences.

That is damaging in so far as economists are wrong to assume that one thing is better for a person than another if and only if she prefers it. Much more serious damage arises when they come to use the more restrictive technical definition of utility that is found in expected utility theory. There, 'utility' is defined to mean, not the value of any function that represents a person's preferences, but specifically the value of a function that represents a person's preferences in an *expectational* way. Let a 'prospect' be the set of possible outcomes that might result from some event, each associated with its probability of occurring. In expected utility theory, the utility of a prospect is defined to be the mathematical expectation of the utility of its possible outcomes. Since, by definition, a person's preferences maximize her utility, this further definition means that her preferences over prospects maximize her expected utility. Again, utility defined this way plays an important role in economists' theory of people's behaviour – this time their behaviour in the face of uncertainty. Again, there need be no implication that a person's utility is her good.

Utility defined this way, which I shall call *expectational utility*, has the property of being *cardinal*. Many different utility functions can represent a person's preferences. But all the functions that represent her preferences expectationally are related to each other in a particular way: they are all affine transforms of each other.

A utility function $u'()$ is defined to be an *affine transform* of another $u()$ if and only if

$u'(x) = au(x) + b$ for all x, where a and b are constants and a is positive.

Utility functions that are affine transforms of each other may differ in what state of affairs they assign zero utility to, and they may differ in the size of the unit of utility, but they do not differ in other ways. In particular, they do not differ in the ratio of utility differences. If the difference in utility between A and B is twice the difference between C and D according to one utility function, that is also so according to every utility function that is an affine transform of it. This means that ratios of differences are significant in an expectational utility function, whereas the zero and unit are not. This is what it means to say that expectational utility is cardinal.

Many economists assume that a person's utility defined this way is her good. More exactly, they assume it measures her good cardinally, by which I mean it is an affine function of her good. That is: $u(w) = aw + b$ where w is the person's good or wellbeing, u her utility, a and b are constants and a is positive. Why do they assume this? On the face of it, it is very implausible. It implies that the goodness of a prospect for a person is just the expectation of the good she may derive from the prospect's various possible results. This is to rule out the possibility that risk to her good is bad for a person.

For example, it implies that getting one unit of benefit for sure is equally as good for a person as a gamble that gives her two units or no units with equal probability. If, say, the goodness of a day in a person's life is given by the number of hours she spends fiddling with her phone, or the number of hours she spends thinking through deep problems in philosophy, then an hour for sure engaged in one of these activities is no better than a risky gamble that gives her two hours or no hours with equal probability. This is not an intuitively attractive assumption. Quite plausibly, the risk-free option may be better.

A way of putting it is that risk to good is 'neutral' according to the view that utility measures good cardinally. There is a theoretical defence of this view, which I shall describe at the end of section 5, but this defence was not available to economists when they first adopted this view. The most likely explanation of the economists' assumption is they were confused by their own terminology. By the definition of utility as a cardinal representation of a person's preferences, preferences are neutral about risk to utility. Since economists also use 'utility' for a person's good, they assume risk to good is neutral.

There is no need for this assumption. A person's utility need not be an affine transform of her good. Instead, it could be a strictly concave function of her good (which means that the graph of utility against good curves downwards as it slopes upwards). Then risk to good will be bad, not neutral.

The assumption that risk to good is neutral is an old one. Daniel Bernoulli (1738) seems to have made it in his account of the St Petersburg paradox, which may fairly be considered the beginning of expected utility theory. For this reason, I call this assumption 'Bernoulli's Hypothesis'. Bernoulli was not confused by terms. He wrote in Latin, and what he actually assumed is that risk to *emolumentum* is neutral. 'Emolumentum' is badly translated as 'utility' in the 1954 *Econometrica* version of the paper. 'Good', 'benefit' or 'wellbeing' would be better translations. Bernoulli's achievement was to show that expected utility theory does not have to assume that risk to money is neutral. $100 for sure may be better than a gamble at equal odds on $0 and $200; this is consistent with the theory. It is an easy next step to realize that the theory also does not have to assume that risk to wellbeing or *emolumentum* is neutral. Bernoulli happened not to make that step.

Nor did anyone for the next two centuries, it seems. Bernoulli's Hypothesis persisted at the foundation of expected utility theory for all that time. Exactly because it is intuitively unattractive, expected utility theory fell into disrepute among economists in the first half of the twentieth century (see for instance Tintner, 1942).

It was spectacularly revived by John von Neumann and Oskar Morgenstern's (1944) formulation of axiomatic expected utility theory. (Frank Ramsey's (1931) earlier formulation received little attention.) von Neumann and Morgenstern showed that, provided a person's preferences conform to some axioms, she can be treated as an expected-utility maximizer. That is to say, a utility function can be defined for her that represents her preferences in an expectational way. Nothing in von Neumann and Morgenstern's work suggests that this utility measures the person's wellbeing. Their work gives no support to Bernoulli's Hypothesis, though they themselves did not make this clear.

Some clear-minded economists saw it immediately. For example, Kenneth Arrow (1951, p. 10) says:

> [von Neumann and Morgenstern's] theorem does not, as far as I can see, give any special ethical significance to the particular utility scale found. For instead of using the utility scale found by von Neumann and Morgenstern, we could use the square of that scale; then behavior is described by saying that the individual seeks to maximize the expected value of the square root of his utility. This is not to deny the usefulness of the von Neumann-Morgenstern theorem; what it does say is that among the many different ways of assigning a utility indicator to the preferences among alternative probability distributions, there is one method (more precisely, a whole set of methods which are linear transforms of each other) which has the property of stating the laws of rational behavior in a particularly convenient way. This is a very useful matter from the point of view of developing the descriptive economic theory of behavior in the presence of random events, but it has nothing to do with welfare considerations, particularly if we are interested primarily in making a social choice among alternative policies in which no random elements enter. To say otherwise would be to assert that the distribution of the social income is to be governed by the tastes of individuals for gambling.

Other economists drew the opposite conclusion. They concluded that expectational utility measures wellbeing or good. This means they implicitly adopted Bernoulli's Hypothesis, however implausible it may seem. John Harsanyi is a leading example. This remark of his illustrates his use of 'utility':

> If I want to compare the utility that I would derive from a new car with the utility that a friend would derive from a new sailboat, then I must ask myself what utility I would derive from a sailboat if I had taken up sailing for a regular hobby as my friend has done, and if I could suddenly acquire my friend's expert sailing skill, and so forth. (Harsanyi, 1977, p. 59.)

Harsanyi evidently means to ask himself how much benefit he would derive from a new sailboat if he had taken up sailing. He is using 'utility' for benefit or good. This earlier passage gives a fuller description of his thinking:

> To be sure, the vNM utility function of any given individual is estimated from his choice

> behavior under risk and uncertainty. But this does not mean that his vNM utility function is *merely* an indication of his attitude towards risk taking. Rather, as its name shows, it is a utility function, and more specifically, it is what economists call a cardinal utility function. This means that the primary task of a vNM utility function is *not* to express a given individual's attitude toward risk taking; rather it is to indicate how much utility, i. e. how much subjective *importance*, he assigns to various goals. . . . Consequently, vNM utility functions have a completely legitimate place in ethics because they express the subjective importance people attach to their various needs and interests. (Harsanyi, 1975, p. 600.)

This argument is evidently founded on what Harsanyi takes to be the meaning of the terms 'utility' and 'cardinal utility'. He writes as though he is describing the ordinary meaning of 'utility' in English, but he mistakes economists's jargon for real English. In English 'utility' means usefulness, but that is not the meaning Harsanyi describes. He does not even give us the meaning of 'utility' that is officially defined in economics: the value of a function that represents preferences. Instead, he says that 'utility' *means* subjective importance. Since he is evidently a subjectivist about a person's wellbeing, he takes 'utility' to refer to wellbeing.

He denies that the primary task of a vNM utility function is to express a person's attitude towards risk. But that is indeed its primary function. It is a technical notion that is defined to have exactly this function within von Neumann and Morgenstern's theory. We could understand it as referring to a particular sort of subjective importance for the special purpose of decision making under risk. But as Arrow pointed out 25 years earlier, nothing in the definition implies that 'utility' refers to subjective importance in any way that is relevant to ethics. Harsanyi's thinking seems to have been driven by a misunderstanding about the meaning of 'utility'. In this he is typical of many economists.

I shall explain at the end of section 5 that Harsanyi had available a much better argument against Arrow. Moreover, this argument derives from Harsanyi's own work, as we shall see. However, Harsanyi himself chose to rely on this weak argument based on the meaning of 'utility'.

Now to just one example of modern practice in moral philosophy. Michael Otsuka and Alex Voorhoeve (2009, p. 173) explicitly take 'utility' to refer to how well a person's life is going, and they also assume that utility is expectational (pp. 172–3, note 3). So these authors assume that expectational utility measures wellbeing. That is to say, they commit themselves unhesitatingly to Bernoulli's Hypothesis, despite its implausibility. This seems unwise, given that their aim in this paper is to refute prioritarianism. We shall see at the end of section 5 that the question of how to measure wellbeing – in particular whether to use expectational utility as a measure – is a central issue in judging whether prioritarianism is credible.

3. The definition of prioritarianism

After the bad lesson, I come to some good lessons that philosophers can learn from economics. I start with all the large number of particular truths that economists have discovered about topics that interest philosophers. These truths mostly lie within the theory of value. The theory of the value of equality is the best example. Economists have studied the value of equality for a long time. Philosophers came more recently to the subject, and can make use of discoveries economists made previously.

In 1991, Derek Parfit gave a Lindley Lecture at the University of Kansas, entitled 'Equality or priority' (Parfit, 1991). He introduced what he called 'the priority view' and distinguished it from what he called 'egalitarianism'. This lecture is described as the '*locus classicus*' of the priority view by Otsuka and Voorhoeve (2018), though these authors do recognize economists' earlier work on the subject. In a footnote, Parfit (1991, p. 41 note 30) referred to some philosophers who had earlier supported the priority view, but not to any economist who had done so.

He appears not to have known that the priority view was sufficiently popular in economics that it could fairly be called the economists' standard account of the value of equality. For example, it is explicitly stated in a standard textbook by Anthony Atkinson and Joseph Stiglitz (1980, p. 340). However, I believe the useful name 'the priority view' was Parfit's invention. It has now given way to the equally useful 'prioritarianism'.

Much earlier, economists had identified a precise way of making the distinction between the priority view and others. It is best expressed as a condition on a betterness ordering among distributions of wellbeing. A distribution can be described by a vector of people's wellbeings $w = (w_1, w_2, \ldots w_n)$. In the literature on equality, it is nearly always assumed that people's wellbeing has an interpersonally comparable measure, and I shall continue with that assumption. So I take each w_i to be a real number. I assume the population of people is constant at n. Betterness is assumed to constitute a *complete ordering* (a transitive and asymmetric relation) on the set of distributions. I am speaking of 'overall' or 'general' betterness: betterness from the point of view of society or the universe. We can specify various condition on this betterness ordering.

For example, we might expect it to be *increasing* in each person's wellbeing. That is:

$(w_1, w_2, \ldots w_n)$ is better than $(w_1', w_2', \ldots w_n')$ if
for some i, $w_i' > w_i$ and for all $j \neq i$, $w_j' = w_j$.

We might also expect it to be *impartial* between people, which means that:

$(w_1, w_2, \ldots w_n)$ is equally as good as $(w_1', w_2', \ldots w_n')$ if
$(w_1, w_2, \ldots w_n)$ is a permutation of $(w_1', w_2', \ldots w_n')$.

The *Pigou-Dalton condition* is a satisfactory way of specifying the condition that equality is valuable. It says that a transfer of wellbeing from a better-off person to a less well-off one makes the distribution better, provided the transfer is not enough to reverse the people's relative positions. That is:

$(w_1, w_2, \ldots w_n)$ is better than $(w_1', w_2', \ldots w_n')$ if,
for some i and some j, $w_i + w_j = w_i' + w_j'$, and $w_i' > w_i \geq w_j > w_j'$, and
for all k such that $k \neq i$ and $k \neq j$, $w_k = w_k'$.

I also need to formulate a condition of *separability*. Let I be a subset of the population made up of the people $i, j, \ldots k$. Let $w_I = (w_i, w_j, \ldots w_k)$ be the distribution of wellbeing over these people, and let w_R be the distribution over the rest of the population. Given impartiality the order of elements does not matter, so we may write the whole distribution as (w_I, w_R). Compare the four distributions (w_I, w_R), (w_I', w_R), (w_I, w_R'), and (w_I', w_R'). The first pair differ only in the wellbeings of the members of I; the second pair also differ only in the wellbeings of those same people. Furthermore, the difference between the first and second member of each pair is exactly the same in each case. The subset of people I is defined to be *separable* in the betterness ordering if and only if, for all distributions, (w_I, w_R) is better than (w_I', w_R) if and only if (w_I, w_R') is better than (w_I', w_R'). Separability means that the influence of the subset I on the ordering is independent of whatever the rest of the distribution may be.

The betterness ordering is defined to be *strongly separable* if and only if every subset of the population is separable in the ordering.

I shall deal only with theories of betterness that imply an increasing and impartial betterness ordering. Of these, the ones that give value to equality also satisfy the Pigou-Dalton condition. These latter theories can be divided into two classes. Those that satisfy strong separability are prioritarian. Those that do not Parfit calls 'egalitarian' but I shall call 'nonprioritarian'.

I have presented this definition of prioritarianism in terms of the structure of the betterness ordering rather than in terms of a value function. This is because a value function does not always exist. A value function is a way of representing an underlying betterness relation in terms of real numbers. Given a betterness relation, a value function $v(w)$ *represents* it if and only if, for any distributions w and w', $v(w) > v(w')$ if and only if w is better than w'. Not all betterness orderings can be represented by a value function. For instance, some orderings with lexical features cannot be.

An ordering that can be represented does not have a unique representation; it can be represented by many different value functions.

When a betterness ordering does have a value function, it is strongly separable if and only if it is *additively separable* (Debreu, 1954). An additively separable ordering is one that can be represented by a value function having the form:

(1) $v(w) = v_1(w_1) + v_2(w_2) + \ldots v_n(w_n)$

(It will also have other representations that do not take this form.) If the ordering is to be increasing and impartial, and satisfy the Pigou-Dalton condition, each of the functions $v_i()$ must be the same, increasing, strictly concave function. $v_i(w_i)$ can be thought of as the contribution of i's wellbeing to general value $v(w)$. Equation (1) clearly shows how each person's wellbeing makes a contribution to general value that is independent of other people's wellbeings. This is the characteristic of prioritarianism that Parfit stresses.

Nevertheless, we cannot define prioritarianism in terms of the additively separable form of the value function, because some betterness ordering that should be counted as prioritarian cannot be represented by a value function. One of these is the so-called *leximin* ordering, which holds an important place in the history of prioritarianism. It is defined by the criterion:

One distribution w is better than another w' if and only if:

the person with the lowest wellbeing in w is better off than the person with the lowest wellbeing in w', or

the person with the lowest wellbeing in w is equally as well off as the person with the lowest wellbeing in w', and the person with the second-lowest wellbeing in w is better off than the person with the second-lowest wellbeing in w', or

the person with the lowest wellbeing in w is equally as well off as the person with the lowest wellbeing in w', and the person with the second-lowest wellbeing in w is equally as well off as the person with the second-lowest wellbeing in w', and the person with the third-lowest wellbeing in w is better off than the person with the third-lowest wellbeing in w', or

and so on.

The leximin ordering satisfies the Pigou-Dalton condition and it is strongly separable. So it is prioritarian by the definition I gave. This is a desirable conclusion. The leximin ordering was introduced by Amartya Sen (1970, p. 138 note) in discussing the 'maximin' ordering, which he ascribed to John Rawls (later published in Rawls 1972). (The maximin ordering has the value function $\min(w_1, w_2, \ldots w_n)$. It does not satisfy the Pigou-Dalton condition.) In an appendix to the Lindley Lecture, Parfit (1991) argues that the leximin, rather than the maximin, ordering represents Rawls's true view about equality. Parfit uses informal arguments to show leximin is prioritarian. Historically, the leximin theory is one of the more prominent prioritarian theories.

Yet it cannot be represented by a value function, whether additively separable or not. So we have to use the broad definition of prioritarianism that I gave initially, in terms of the structure of the betterness relation. This is an example of a useful general lesson moral philosophy can learn from economics. Philosophers often start their arguments from their intuitions about value. But betterness is generally more fundamental than value, and a better place to start. (This is not to deny that particular values often contribute to determining what the betterness ordering is.) Not only are many betterness orderings not represented by value functions, but intuitions about value may not have a clear meaning unless they are anchored in intuitions about betterness. We shall come to an example in section 7.

4. A bad argument for prioritarianism

What arguments are there for prioritarianism? Let us take the value of equality for granted, so that a betterness ordering should satisfy the Pigou-Dalton condition. We are also assuming the ordering is increasing and impartial. Granted these things, the question is what arguments support the claim that the ordering is also strongly separable? This question was raised by Parfit in the Lindley Lecture, but it was raised in economics much earlier (see Sen, 1973, pp. 39–41).

The central argument for prioritarianism in Parfit's lecture is that nonprioritarian theories are subject to the *levelling-down objection*. Compare two distributions where the second is levelled down relative to the first. This means that in the second distribution, some of the better-off people are less well off than they are in the first, but no one is better off in the second than she is in the first. Parfit claims that, though no one is better off in the second distribution, nonprioritarians must think the second is better than the first in one respect – namely, it is more equal. On the other hand, prioritarians think the second is better in no respect. Parfit thinks the former view is implausible.

Why must nonprioritarians think the levelled-down distribution is better in one respect? Take the nonprioritarian theory whose value function is the sum of the products of people's wellbeings, taken in pairs. That is:

(2) $v(w) = w_1w_2 + w_1w_3 \dots + w_1w_n + w_2w_3 \dots + w_3w_n + w_3w_4 + \dots + w_3w_n$

This is not a well-known function. However, it is impartial and increasing, provided everyone's wellbeing is positive, and it satisfies the Pigou-Dalton condition. It is a well-behaved function and not obviously objectionable. It is not strongly separable, so it is nonprioritarian. The hallmark of prioritarianism is that the benefit of increasing the wellbeing of one person is not affected by the wellbeings of other people, but in this theory it clearly is. However, a levelling down raises none of the terms in the formula (2), so on the face of it this theory does not suggest the levelled-down distribution is better in any respect.

It does not do so any more than the familiar prioritarian formula (1) does. An example is the sum of the square roots of people's wellbeings:

(3) $v(w) = \sqrt{w_1} + \sqrt{w_2} + \dots + \sqrt{w_n}$

A levelling down raises none of the terms in this formula, and Parfit tells us that for this reason it is not subject to the levelling-down objection.

Parfit argues that a nonprioritarian thinks inequality is bad, and therefore must think that any change that decreases inequality is better in one respect. But a prioritarian also thinks inequality is bad: a reduction in inequality makes the distribution better if total wellbeing remains constant. If a theory has a value function at all, the badness it ascribes to inequality is easily identified. The theory can be represented by a value function of the form

$$V = (w_1 + w_2 + \dots + w_n) - I\,(w_1, w_2, \dots w_n)$$

The first term in this formula is the total of people's wellbeing. The second is a measure of the badness of the inequality. It can be found simply by subtracting whatever value the theory ascribes to a distribution from the total of people's wellbeing. In the case of the square-root formula (3), the measure is

$I(w_1, w_2, \ldots w_n) = (w_1 + w_2 + \ldots + w_n) - (\surd w_1 + \surd w_2 + \ldots + \surd w_n)$

This can be done for a prioritarian theory as easily as for a nonprioritarian one (Fleurbaey, 2016, section 2). Levelling down decreases the measure of the badness of inequality. So here is a respect in which levelling down makes the distribution better, even for a prioritarian.

Parfit claims that a prioritarian thinks a reduction in inequality is only *instrumentally* better. She thinks inequality is not *intrinsically* bad; it is just that shifting wellbeing from better-off to worse-off people, which she values intrinsically, has the effect of reducing inequality, which she does not value intrinsically. But in what sense is the badness of inequality a mere instrumental consequence of prioritarianism? It is not a causal consequence. It is not even a contingent consequence; it is a mathematically necessary feature of the prioritarian formula. The badness of inequality is intrinsic to it in the way that being the square of three is intrinsic to the number nine. Compare the pairwise-product value function (2) with the square-root function (3). Both give value to equality, because they both satisfy the Pigou-Dalton condition. How does the pairwise-product function make this value intrinsic, whereas the square-root function makes it only instrumental? I see no difference.

The upshot is that the levelling-down objection fails. It gives no support to prioritarianism. My argument for this conclusion depends on formal features of prioritarian and nonprioritarian theories, and indeed on formulae. This is the method of economics.

5. A good argument for prioritarianism

Economics also provides a quite different, powerful argument in favour of prioritarianism. It consists in a theorem first proved by John Harsanyi (1955), and subsequently proved in many different forms by other economists, including Border (1985), Broome (1990), Deschamps and Gevers (1979), Fishburn (1984), Hammond (1981) and Mongin and d'Aspremont (1998). Harsanyi discovered a clever way of getting some extra theoretical leverage on the structure of the betterness relation, by investigating betterness, not just among states of affairs, but among uncertain prospects. A prospect is a portfolio of states of affairs, each with a probability attached to it. Expected utility theory gives us information about the structure of betterness among prospects. This in turn induces some structure on the betterness relation among states of affairs. Indeed, it implies that it is strongly separable. That is Harsanyi's discovery.

Harsanyi presents his theorem in terms of preferences, but I prefer to express it in terms of betterness, as I did in Broome (1991b). The theorem rests on these three premises:

First, each individual's betterness among prospects satisfies the axioms of expected utility theory.

Second, general betterness among prospects satisfies these axioms.

Third, one prospect is better than another if it is better than the other for one person and at least as good as the other for every person, and two prospects are equally good if they are equally good for each person.

I call the third premise 'the principle of personal good'. It is a translation of the Pareto principle into the terminology of betterness.

Harsanyi's theorem is that, given these three assumptions, general betterness can be represented expectationally by a value function $v()$ that is the sum of utility functions $u_1()$, $u_2()$, . . . $u_n()$ that represent betterness for each person expectationally. That is:

$v(x) = u_1(x) + u_2(x) + \ldots + u_n(x)$

The variable x ranges over prospects. Among prospects are those that deliver a particular outcome for sure, and these can be identified with the outcome itself. So x also ranges over outcomes. This equation therefore applies to outcomes as well as prospects.

Let us now interpret the outcomes as distributions of wellbeing such as $w = (w_1, w_2, \ldots w_n)$. Prospects are then portfolios of distributions, each with a probability. Harsanyi's theorem tells us that general betterness among distributions can be represented by a value function that is the sum of utility functions that represent betterness for individuals:

$$v(w) = u_1(w) + u_2(w) + \ldots + u_n(w)$$

Think for a moment of just the first person. One distribution w is better for her than another w' if and only if she has more wellbeing in w than in w'. That is to say: if and only if w_1 is greater than w_1'. This implies that, among distributions, betterness for the first person depends only on her own wellbeing w_1. Utility for the first person represents her betterness, so it too depends only on w_1. The same is true for each person. So we may write the equation:

(4) $$v(w) = u_1(w_1) + u_2(w_2) + \ldots + u_n(w_n)$$

This is an instance of the additively separable form (1) of the value function. Since additive separability implies strong separability, it follows that the betterness ordering is strongly separable. That is the powerful argument for strong separability.

We are assuming that equality is valuable, which is to say that the Pigou-Dalton condition holds. We are also assuming that the betterness ordering is increasing and impartial in people's wellbeing. Given these assumptions, adding strong separability gives us prioritarianism. So Harsanyi's theorem provides a powerful argument for prioritarianism.

Each function $u_i(w_i)$ shows the contribution i's wellbeing makes to general value. The Pigou-Dalton condition ensures that each $u_i(w_i)$ is strictly concave in w_i. The degree of concavity determines the degree to which inequality in people's wellbeing is bad: the more concave, the more valuable equality is. This gives us a further consequence of the theorem. The functions $u_i(w_i)$ are defined to be expectational utility functions. This means that their degree of concavity also determines the degree to which risk to people's wellbeing is bad. So the badness of inequality exactly matches the badness of risk.

This consequence of Harsanyi's theorem is remarkable and surprising. It seems intuitively that the badness of inequality and the badness of risk are quite different matters. The badness of inequality is a matter of weighing together the interests of different people. This raises the issue of fairness, whereas fairness has no relevance to a single person's risk. Yet Harsanyi's theorem implies a symmetry between the badness of inequality and the badness of risk. They are both given by the concavity of the same utility functions.

Harsanyi himself drew a different conclusion. He thought his theorem constitutes an argument for utilitarianism. This is because, as I explained in section 2, he assumed that the utility functions in equation (4) measure wellbeing cardinally: that they are affine functions of wellbeing. Given this assumption and also symmetry, (4) could be written:

$$v'(w) = w_1 + w_2 + \ldots + w_n$$

where $v'()$ is a value function that is an affine transform of $v()$. This is a utilitarian formula for value. Utilitarianism gives no value to equality of wellbeing; it does not satisfy the Pigou-Dalton condition.

I explained in section 2 that Harsanyi himself had no good grounds for his assumption that utility measures wellbeing cardinally. It appears to have arisen from a confusion over the meaning of 'utility'. However, some grounds can be provided, and Harsanyi's theorem itself contributes to these grounds. I shall now set them out (for more detail see Broome, 1991b, ch. 10).

Since wellbeing is a quantitative concept, we should have some account of the concept of a quantity of wellbeing. For instance, what does it mean for one quantity of wellbeing to be greater than another? A plausible answer is that it means that the first counts for more than the second in contexts where they are weighed against each other in determining an overall value.

One context where quantities of wellbeing are weighed against each other is in determining the value for a person of uncertain prospects. Here is an example. Suppose one outcome x is better for a person than another y, which is better for her than a third z. And suppose the outcome y for sure is better for the person than a gamble at equal odds between x and z. The gamble is better than y in one respect and worse in another. It is better in that it might give the person x rather than y, and worse in that it might give her z rather than y. Evidently the respect in which the gamble is worse counts for more than the respect in which it is better, since the gamble is worse overall. Since it counts for more in this situation of risk, we might take the actual loss of getting z rather than y to be more than the actual benefit of getting x rather than y. That is: the difference in the person's wellbeing between y and z is greater than the difference in her wellbeing between x and y.

Since y is better than the gamble on x and z, the definition of expectational utility implies that $u(y) > ½u(x) + ½u(z)$. This implies that the difference $u(y) – u(z)$ in utility between y and z is greater than the difference $u(x) – u(y)$ in utility between x and y. The utility differences therefore match the putative differences in wellbeing that I have just described. By generalizing over many gambles, we may conclude that expectational utility measures wellbeing cardinally.

This is an argument for the conclusion that expectational utility measures wellbeing cardinally. But it is weak argument. It simply ignores the possibility in the example that the gamble is worse than the certainty of y, not because of the differences in quantities of wellbeing as I described them, but because risk to wellbeing is itself a bad thing.

However, the case for this analysis of quantities of wellbeing is much strengthened by Harsanyi's theorem. His theorem tells us that people's expectational utilities, added up, determine the relative values of distributions of wellbeing. This means that utility specifies, not only how differences of wellbeing count intrapersonally in determining the values of uncertain prospects, but also how they count interpersonally in determining the values of distributions. This makes it much more plausible that utility measures wellbeing cardinally. It gives strength to Harsanyi's claim that his theorem supports utilitarianism.

Still, we do not have to accept this claim of Harsanyi's. Prioritarians do not accept it. However, it leaves them with the responsibility of providing some alternative analysis of quantities of wellbeing (see Broome, 1991, ch. 11, and McCarthy, 2006). Given Harsanyi's theorem, the issue between prioritarianism, which values equality in wellbeing, and utilitarianism, which does not, comes down to a question about the measurement of wellbeing.

6. The dangers of ignoring economic theory

Each of the three premises of Harsanyi's theorem is *prima facie* very plausible, so the theorem provides a powerful argument for prioritarianism. Given the failure of the levelling-down argument, it is by far the strongest argument for prioritarianism.

The same theorem also sets a challenge for prioritarianism. Prioritarianism has to be held apart from utilitarianism by providing some measure of wellbeing distinct from its contribution to general value. Expectational utility is a natural measure to pick, but Harsanyi's theorem shows that we cannot use that measure without abandoning the value of equality. That means abandoning prioritarianism.

So Harsanyi's theorem sets the agenda for the debate about the value of equality. If you value equality but do not want to be a prioritarian, you must explain which of the theorem's three assumptions you deny and why you deny it. If you want to be a prioritarian, you must explain how you measure wellbeing in a way that allows you to value equality of wellbeing.

The literature of economics contains a considerable body of writing that responds to Harsanyi's theorem, including Coulhon and Mongin (1986), Diamond (1967), Mongin and d'Aspremont (1998), Sen (1976, 1977), Weymark (1991). So does the literature of philosophy, including Broome (1991b), Greaves (2015), McCarthy (2006, 2008), Rabinowicz (2002). Yet within philosophy there is also a large body of literature about prioritarianism that takes little notice of this work or its conclusions. Since it does not build on previous knowledge, this latter body of literature is badly placed to contribute to our understanding of the value of equality.

An example of it is 'Why it matters that some are worse off than others' by Michael Otsuka and Alex Voorhoeve (2009). I pick this paper as an example because it has been highly influential. It was published in the leading journal *Philosophy and Public Affairs*, and subsequently a whole issue of *Utilitas* was devoted to comments on it. Yet this paper made no reference to Harsanyi's theorem, and nor did any of the comments in *Utilitas*.

Otsuka and Voorhoeve argue against prioritarianism. Their main argument is that different considerations are at issue when we weigh together the wellbeings of different people from those that are at issue when we weigh wellbeing for a single person. We weigh wellbeing intrapersonally when we balance a chance of a gain to a person against the chance of a loss to the same person. We weigh wellbeings interpersonally when we balance a gain to someone against a loss to someone else. Interpersonal weighings raise the issue of fairness, whereas intrapersonal weighings do not. The paper supports this point by means of an example in which it seems clear that interpersonal and intrapersonal weighings should be done differently. The authors say that failing to do so is failing to recognize the separateness of persons.

They offer this as an argument against prioritarianism. But it is most directly an argument against the conclusion of Harsanyi's theorem, although the authors do not identify it as such. Harsanyi's theorem asserts a symmetry between interpersonal weighing and interpersonal weighing. They deny this symmetry.

Their objection to symmetry is an old one. It was raised by Peter Diamond (1967) in response to Harsanyi and has been a subject of discussion ever since. Otsuka and Voorhoeve support it with an example but, given the long history, an example does not advance the argument much. On the one hand, the example elicits an intuition that the theorem's conclusion is wrong. On the other hand, there is the attraction of the theorem's three plausible premises. We have to navigate between these conflicting considerations. If we are to accept the intuitive objection to symmetry, we have to identify what is wrong with at least one of the three premises.

The literature has canvassed the options. Different authors have taken difference stances. Diamond himself objected to the second premise that general betterness conforms to expected utility theory. He provided a direct counterexample to this premise. Accepting Diamond's objection, David McCarthy (2006) also gives up the second premise and develops a version of prioritarianism – ex-ante prioritarianism – that he claims to be defensible. On the other hand, Wlodek Rabinowicz (2002) recommends giving up the third premise – the principle of personal good – and he provides a different version of prioritarianism on that basis. (He distinguishes two sorts of intrapersonal weighing of wellbeings in determining betterness among prospects: weighing for determining betterness for the person and weighing for determining general betterness. He denies that the first sort is symmetrical with interpersonal weighing, but accepts that the second sort is.) McCarthy and Rabinowicz have each worked out their own coherent prioritarian theory.

Otsuka and Voorhoeve do not try to navigate the conflict between their intuition and the premises of Harsanyi's theorem. They do not identify which premise they reject. Indeed, they apparently affirm all three at different points in their paper. They apparently affirm the first (expected utility theory for individual betterness) and second (expected utility theory for general betterness) explicitly (Otsuka and Voorhoeve, 2009, p. 172 n. 3 and p. 195 respectively). They do not explicitly affirm the third (the principle of personal good). However, they apparently affirm it implicitly in responding to Rabinowicz's version of prioritarianism. Take two prospects that are equally good for everyone apart from one person. If I understand the authors correctly, they assume that one of these prospects is better than the other if and only if it is better for that one person (Otsuka and Voorhoeve, 2009, p. 178). By easy steps, this assumption implies the principle of personal good.

I say only that the authors make these assumption 'apparently' because they do not use the language of betterness. In the situation where two prospects are equally good for everyone apart from one person (and there is a question of what medical treatment to provide for that person), they say 'it is reasonable to provide [the person] with a treatment that maximizes the expected increase in her utility'. It appears later on the same page that they use 'reasonable' in such a way that doing anything else counts as unreasonable. I therefore take them to mean that the person ought to be provided with a treatment that maximizes the expected increase in her utility. That seems to imply that this treatment gives the best prospect.

So there are *prima facie* grounds for thinking that Otsuka and Voorhoeve have fallen into a trap of inconsistency. They apparently affirm the premises of Harsanyi's theorem but deny its conclusion. This is a trap they set for themselves by choosing to ignore Harsanyi's work. If the inconsistency is real, it vitiates their argument. At least the authors should have explained how they avoid the threat of inconsistency.

The paper appears to contain a further inconsistency. In responding to McCarthy's version of prioritarianism, the authors in effect deny the principle of personal good (Otsuka and Voorhoeve, 2009, p.198–8). Apparently, they affirm this principle and later deny it. They make various separate affirmations through the paper, but seem not to have checked them for consistency. To be sure they are consistent, they would need to set out a theory of their own that conforms to them, as Rabinowicz and McCarthy do but they do not.

7. The uses of mathematics

How can the authors, referees, editors and many readers of this influential paper not have checked its consistency? I am sorry to say that the practice of moral philosophers is not always as analytically tight as it should be. Often we shun formal language, we avoid symbols and we do not use mathematics. This is appropriate for wide swathes of moral philosophy, including those that are concerned with personal relationships among small numbers of people. But it is not appropriate in areas of moral philosophy that involve large numbers of people and have a quantitative dimension. The value of equality is one of those.

In areas of this sort we need mathematical formulae to keep our thinking accurate. Moreover, we can harness the power of mathematical methods. Harsanyi's theorem illustrates their power. Its premises contain no mention of addition, yet its conclusion is that the value function has an additively separable structure. No amount of verbal discussion or probing of examples could arrive at this strong and surprising result. Yet the result can be used to provide powerful support for the philosophical theories of utilitarianism and prioritarianism.

This is another good lesson we can draw from economics: sometimes we need mathematics. Economics became mathematical in the 1940s and 1950s. After that time, it was impossible to become an academic economist without at least being able to understand mathematical formulae. Moral philosophy does not need such a radical revolution in its methods, but we must recognize that some topics within moral philosophy cannot be studied properly without mathematics. Furthermore, many moral philosophers cannot benefit from the discoveries of economists that I have been commending because, without mathematical understanding, they cannot read the papers.

A lack of mathematical understanding also leads to errors at a more mundane level, where no fancy theorems are required. Simple mathematical sensitivity would often be beneficial. Here is an example.

Let q be a variable ranging over qualities and let t be a variable ranging over quantities. The vector (q, t), which I shall call an 'item', denotes a particular quantity of a particular quality. For instance, it may denote a life of quality q that lasts for a time t. Or it may denote a population of t people, each with a lifetime wellbeing of q. Or it may denote an episode of pain of quality q lasting for a period of time t. And so on.

Assume the items within some domain, such as the domain of episodes of pain or the domain of lives, are completely ordered by their goodness. This means in particular that all the items having some standard quantity T are ordered. We can treat the ordering of those particular items as an ordering of the qualities q themselves. That is, we can define q to be better than q' if and only if (q, T) is better than (q', T). Let q_h and q_l be two qualities such that q_h is better than q_l. For example, q_h might be a high quality of life and q_l a mediocre one. Or q_l might be an excruciating degree of pain and q_h the pain of a slight headache.

Suppose it is true that:

> *Premise*. For any item (q, t), there is a quality q' worse than q and a quantity t' such that (q', t') is better than (q, t).

That is, for some diminution in quality, a sufficient increase in quantity more than cancels out the diminution and leads to a result that is better overall.

Even though betterness is a transitive relation, it does not follow from the Premise that:

> *Conclusion*. There is a quantity t_l such that (q_l, t_l) is better than (q_h, T).

That is, it does not follow that a sufficient increase in quantity more than cancels out a diminution in quality all the way from q_h to q_l. Even given the Premis, it may be that an item with quality q_h and quantity T is better than any item with the lower quality q_l, whatever its quantity.

The premise implies that, starting from (q_h, T) there is a sequence of qualities q_i, q_j, . . . that are progressively worse and worse – so they are all worse that q_h – but such that with sufficient increases in quantity the items (q_i, t_i), (q_j, t_j) are progressively better and better, so they are all better than (q_h, T). But the sequence q_i, q_j, . . . may never get as low in the ordering of qualities as q_l.

The same is true if we reverse the direction of change. This is appropriate if the domain consists of bad things such as pains, so that increasing quantity makes an item worse.

> *Reverse Premise*. For any vector (q, t), there is a quality q' better than q and a quantity t' such that (q', t') is worse than (q, t).

It does not follow that:

> *Reverse Conclusion*. There is a quantity t_h such that (q_h, t_h) is worse than (q_l, T).

It will be obvious to anyone with a mathematical sensitivity that the Conclusion does not follow from the Premise, or the Reverse Conclusion from the Reverse Premise. But several philosophers have assumed the opposite, and drawn extravagant conclusions. In one of his papers, Larry Temkin (1966) worked with the domain of pains. He accepted the Reverse Premise and assumed it implied the Reverse Conclusion, given that betterness is transitive. He drew the conclusion that betterness is not transitive. Temkin was reproved for his mistake by Ken Binmore and Alex Voorhoeve (2003), and I believe he has avoided it ever since. Nevertheless, Dale Dorsey (2009) later made exactly the same mistake. He concluded that the betterness ordering is discontinuous.

Derek Parfit (1984) dealt with a domain of populations. He chose a very high value for q_h and a value of q_l that he took to make the Conclusion repugnant. He presented persuasive arguments for the Premise. Then he assumed that the Conclusion follows from the Premise. This presented him with a problem (Parfit, 1984, pp. 430 and 435–6). But since the Conclusion does not follow from the Premise, this problem did not really arise. Whatever problem there is with the Repugnant Conclusion, it does not arise from the Premise alone. (The Repugnant Conclusion is indeed implied by a version of utilitarianism.)

In a much later paper, in dealing with the domain of populations, Parfit (2016) did not assume the Conclusion follows from the Premise alone. He added a further premise that I shall mention below. But in the domain of lives of different lengths, he continued implicitly to assume the Conclusion follows from the Premise alone (Parfit, 2016, p. 119).

A counterexample helps to show the error in this assumption more clearly. Suppose q and t each have a numerical measure. Suppose the betterness ordering can be represented by a function $v(q, t)$, and let this function take the specific form

(5) $v(q, t) = qt/(t + 1)$.

Figure 1 shows this function's contours of constant goodness. These could be called indifference curves for goodness. (Another lesson from economics is that indifference curves are a good aid for understanding the structure of value.) With this function, the Premise is true and the Conclusion false.

The specific form of the function does not matter; other formulae would do instead. What does matter is that the contours slope downwards everywhere, and they approach horizontal asymptotes. Their downward slope explains why the Premise is true: a sufficiently small diminution in quality (a downward movement) can always be compensated for by a sufficiently large increase in quantity (a rightward movement). The asymptotes explain why the Conclusion is false. No item with a quality below the asymptote of a particular contour, however great the quantity, can be as good as any item on or above the contour.

The ordering has the property that James Griffin (1986, p. 85) calls 'discontinuity' and defines as 'So long as we have enough of *B*, any amount of *A* outranks any further amount of *B*, or . . . enough of *A* outranks any amount of *B*'. (In the second disjunct read '*A*' as 'q_h' and '*B*' as 'q_l'.) I assume Griffin was aiming at the mathematical notion of discontinuity, but he did not hit it. The ordering in this example satisfies Griffin's definition but it is continuous in the mathematical sense. Discontinuous orderings can be intractable, but this ordering is not.

Indeed, it is very tractable. I commend an ordering with horizontal asymptotes to philosophers who are struggling with population ethics. It offers a way round some of their problems.

Interestingly, the formal structure of Gustaf Arrhenius's book *Population Ethics* (Arrhenius, forthcoming) ensures that the Conclusion actually does follow from the Premise. His 'First impossibility theorem' states as much. In his structure, quality comes in discrete amounts. The consequence is that the contours of the betterness ordering cannot have horizontal asymptotes while always sloping downwards. This is merely an artefact of his idiosyncratic assumption that quality comes in discrete amounts.

Derek Parfit (2016, p. 162) makes an objection to a betterness relation whose contours have horizontal asymptotes. He says it implies that the existence of more and more people has 'diminishing marginal value'. Adding each new person to the world has less and less value, the more people there already are. Parfit thinks this cannot be correct, and he consequently rejects this counterexample to the claim that the Premise implies the Conclusion.

What matters is the form of the betterness relation, not the value function that represents this relation. A value function is arbitrary to a large extent. A value function for the betterness ordering in Figure 1 can be created by giving a value to each contour. Any values at all can be assigned to the contours, so long as a higher contour always gets a higher value than a lower one. Any valuation of contours that sticks to this rule gives a representation. This means the scale of value is arbitrary, so that the marginal value of quantity is also arbitrary. The value function must have the right contours, but it is arbitrary apart from that.

Despite what Parfit supposes, a betterness ordering whose contours have horizontal asymptotes does not imply that quantity has diminishing marginal value. The value function (5) has that implication, but it is only one of the many different value functions that represent the same ordering. Others do not imply diminishing marginal value.

It is true that, when an ordering has contours with horizontal asymptotes, any representation of it has the feature that value is not everywhere linear in quantity. Quantity cannot always have constant marginal value. Indeed, in any representation, as quantity gets bigger and bigger, it eventually has to have diminishing marginal value. This much of Parfit's supposition is true. But it has little significance, since the point at which the marginal value of quantity begins to decline can be extended as far away as anyone might choose.

For example, think of items (q_l, t) denoting populations of size t at the low level of wellbeing q_l. There is a representation that makes the value of (q_l, t) increase in proportion to t until t reaches, say, a trillion trillion. That is to say, the marginal value of adding a person at level q_l is constant until the population reaches a trillion trillion. If that is not enough, we can go further.

To see how this is possible, start by assigning the value of t to the contour that passes through (q_l, t), and do this for every t up to some large number. To do this in Figure 1, take the horizontal line at level q_l, and assign to every contour that cuts this line the value of t at the point where it cuts it. Algebraically, this is done by transforming the value function $v()$ in (5) to $v'()$ where:

$$v'(q, t) = v(q, t)/\{q_l - v(q, t)\}$$

This means that

$$v'(q, t) = qt/\{q_l(t + 1) - qt\}$$

If you were to continue this method of assigning value for every t out to infinity, only contours that cut the line would get a value. There are many contours that do not cut that line. To assign them a value, at some point you will have to change your valuing method. But you can continue with the original method as far as you like. (If you were willing to accept values that are transfinite numbers, you could even take t out to infinity.)

So Parfit's objection to the counterexample is mistaken. To reveal the mistake, I applied a lesson from economics that I mentioned in section 3: I adopted the economists' practice of concentrating on the form of the betterness relation rather than on a value function that represents it.

8. Summary

The main lesson of section 7 is perhaps the most important of the lessons economics can teach moral philosophy, since it is a precondition for learning the others. Moral philosophers should learn that some of the subjects that interest them demand a little mathematics. If you are not interested in mathematics, you should avoid these subjects.

One benefit of knowing some mathematics it that it will make the previous work of economists available to you. Economist have done excellent work on some of the topics that also interest philosophers. There are many particular lessons to be learnt from their results. Philosophy can build on them, but it must absorb them first. I mentioned two examples.

I described how far the moral philosophy of equality can go astray if it ignores the long history of the analysis of equality in economics. In particular, long before philosophers came to the subject, economists had already investigated the foundations of prioritarianism. They made discoveries that depend on remarkable theorems about additivity. These discoveries could not have be been made except by mathematical methods. The philosophy of prioritarianism cannot be successful if it ignores them.

There is also a long history of population ethics in economics, driven by the work of Charles Blackorby, Walter Bossert and David Donaldson. These authors have developed a complete axiology for population (Blackorby, Bossert and Donaldson, 2005). But for the example of population ethics, I concentrated on just one mathematical error that has been damaging. This error also has ramifications outside population ethics, and it has led to some serious mistakes.

I presenting my examples I mentioned some other lessons in passing. An important one is that, in the theory of value, it is best to think first of the structure of the betterness ordering rather than of value itself. Value merely represents betterness. As I put it in (Broome, 1999, pp. 9–11) the lesson is: ‘Think comparatively’. Attend to what is better than what rather than to what is good. Comparative thinking is deeply embedded in economics but not in philosophy.

Finally, I strongly recommend avoiding the economist’s practice of using ‘utility’ to denote a person’s good.

References

Arrhenius, Gustaf, (forthcoming), *Population Ethics*.

Arrow, Kenneth J., (1951), *Social Choice and Individual Values*, Yale University Press.

Atkinson, Anthony B., and Joseph E. Stiglitz (1980), *Lectures on Public Economics*, McGraw-Hill.

Bernoulli, Daniel, (1738), ‘Specimen theoriae novae de mensura sortis’, *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, translated by Louise Sommer as ‘Exposition of a new theory on the measurement of risk’, *Econometrica*, 22 (1954), pp. 23–36.

Blackorby, Charles, Walter Bossert and David Donaldson (2005), *Population Issues in Social Choice Theory, Welfare Economics and Ethics*, Cambridge University Press.

Border, K. C., (1985), ‘More on Harsanyi’s cardinal welfare theorem’, *Social Choice and Welfare*, 2, 279–81.

Broome, John, (1990), ‘Bolker–Jeffrey expected utility theory and axiomatic utilitarianism’, *Review of Economic Studies*, 57, 477–502.

Broome, John, (1991a), ‘“Utility”’, *Economics and Philosophy*, 7, 1–12.

Broome, John, (1991b), *Weighing Goods: Equality, Uncertainty and Time*, Blackwell.

Broome, John, (1999), *Ethics Out of Economics*, Cambridge University Press.

Binmore, Ken, and Alex Voorhoeve (2003), ‘Defending transitivity against Zeno’s paradox’, *Philosophy and Public Affairs*, 31, 272–9.

Coulhon, T, and P. Mongin (1986), ‘Social choice theory in the case of von Neumannn-Morgenstern utilities’, *Social Choice and Welfare*, 6, 175–87.

Deaton, Angus, and John Muellbauer (1980), *Economics and Consumer Behavior*, Cambridge University Press.

Debreu, Gerard, (1954), ‘Representation of a preference ordering by a numerical function’, in *Decision Processes*, edited by R. M. Thrall, C. H. Coombs and R. L. Davis, Wiley, pp. 159–65.
Debreu, Gerard, (1959), *Theory of Value*, Wiley.
Deschamps, Robert, and Louis Gevers (1979), ‘Separability, risk-bearing and social welfare judgements’, in *Aggregation and Revelation of Preferences*, edited by Jean-Jacques Laffont, North-Holland, 145–60.
Diamond, P. A., (1967), ‘Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment’, *Journal of Political Economy*, 75, 765-6.
Dorsey, Dale, (2009), ‘Headaches, lives and value’, *Utilitas*, 21, 36–58.
Fishburn, Peter C., (1984), ‘On Harsanyi’s utilitarian cardinal welfare theorem’, *Theory and Decision*, 17, 21–8.
Fleurbaey, Marc, (2015), ‘Equality versus priority: how relevant is the distinction?’, *Economics and Philosophy*, 31, 203–17.
Greaves, Hilary, (2015), ‘Antiprioritarianism’, *Utilitas*, 27, 1–42.
Griffin, James, (1986), *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford University Press.
Hammond, Peter J., (1981), ‘Ex-ante and ex-post welfare optimality under uncertainty’, *Economica*, 48, 235–50.
Harsanyi, John C., (1955), ‘Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility’, *Journal of Political Economy*, 63, 309–21.
Harsanyi, John C., (1975), ‘Can the maximin principle serve as a basis for morality? A critique of John Rawls’s theory’, *American Political Science Review*, 69, 594–606.
Harsanyi, John C., (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press.
McCarthy, David, (2006), ‘Utilitarianism and prioritarianism I,’ *Economics and Philosophy*, 22, 1–29.
McCarthy, David, (2008), ‘Utilitarianism and prioritarianism II, *Economics and Philosophy*, 24, 1–33.
Mongin, Philippe, and Claude d’Aspremont (1998), ‘Utility theory and ethics’, in *Handbook of Utility Theory* edited by S. Barbera, P. Hammond and C. Seidl Kluwer, vol. 1, 371–480.
Otsuka, Michael, and Alex Voorhoeve (2009), ‘Why it matters that some are worse off than others: an argument against the priority view’, *Philosophy and Public Affairs*, 37, 172–99.
Otsuka, Michael, and Alex Voorhoeve, (2018) ‘Equality Versus Priority’, in *The Oxford Handbook of Distributive Justice*, edited by Serena Olsaretti, Oxford University Press, 65–85.
Parfit, Derek, (1984), *Reasons and Persons*, Oxford University Press.
Parfit, Derek, (1991), *Equality or Priority?*, The Lindley Lecture, University of Kansas.
Parfit, Derek, (2016), ‘Can we avoid the repugnant conclusion?’, *Theoria*, 82, 110–27.
Rabinowicz, Wlodek, (2002), ‘Prioritarianism for prospects’, *Utilitas*, 14, 2–21.
Ramsey, Frank, (1931), ‘Truth and probability’, in his *Foundations of Mathematics and Other Logical Papers*, edited by R. B. Braithwaite, Kegan-Paul, 156–98.
Rawls, John, (1972), *A Theory of Justice*, Harvard University Press.
Sen, Amartya, (1970), *Collective Choice and Social Welfare*, Holden-Day.
Sen, Amartya, (1973), *On Economic Inequality*, Oxford University Press.
Sen, Amartya, (1976), ‘Welfare inequalities and Rawlsian axiomatics’, *Theory and Decision*, 7, 243–62.
Sen, Amartya, (1977), ‘Non-linear social welfare functions: a reply to Professor Harsanyi’ in

Foundational Problems in the Special Sciences edited by R. Butts and J. Hintikka, Reidel, 297–302.

Temkin, Larry S., (1966), 'A continuum argument for intransitivity', *Philosophy and Public Affairs*, 25, 175–210.

Tintner, Gerhard, (1942), 'A contribution to the non-static theory of choice', *Quarterly Journal of Economics*, 56, 274–306.

Von Neumann, John, and Oskar Morgenstern, (1944), *Theory of Games and Economic Behavior*, Princeton University Press.

Weymark, John A., (1991), 'A reconsideration of the Harsanyi-Sen debate on utilitarianism', in *Interpersonal Comparisons of Well-Being*, edited by John Elster and John Roemer, Cambridge University Press, 255–320.

Acknowledgements

I am very grateful to David McCarthy, Karsten Klint Jensen and Wlodek Rabinowicz for detailed advice about this paper. Research for the paper was supported by ARC Discovery Grants DP140102468 and DP180100355.