

Utilitarianism and Its Constraints

High Heaven rejects the lore of nicely calculated less or more.

– Wordsworth

I CONSTRAINTS ON UTILITY MAXIMISATION

Most students of economics will be familiar with the concept of ‘constrained maximisation’. It is the bread and butter of students who take an interest in the basic mathematics of economic theory, where they soon encounter the notion of ‘constrained maximisation’. For this corresponds to fundamental concepts in economics, such as the notion that consumers maximise something that we shall call utility (for old times’ sake and for want of a better word), subject to some constraints including their incomes and the prices of goods. And the difference between Utilitarianism and some competing ethical systems can be seen as being simply all about whether one should just maximise ‘some unique intrinsically valuable objective, utility’, subject to some constraints or whether, instead, there are other independent intrinsically valuable top-level objectives.

Take a very simple example. Suppose a pretty useless person can be killed in order to use some of his organs to save the lives of several very socially valuable people who are waiting for organ transplants such as kidney transplants. (In fact, in some countries it is suspected that some such process actually has had official approval.) ‘But’, you might say, ‘this would conflict with a rule to respect people’s right to life.’ ‘Ah’, says the utilitarian, ‘Not really. You just have to deduct the long-term loss of

utility that society would incur if such a practice became permissible from the shorter-term gain in utility of the rest of society. Easy. What's the problem?'

After all, in the last chapter, the principle of maximisation of utility *per head*, rather than maximisation of *total* utility, would presumably be subject to the constraint that it should not be promoted by cutting off a lot of heads. So why not just accept that utility maximisation – in whichever variety – has to be subject to constraints of one kind or another? The literature is full of examples of situations that may crop up in ordinary daily life of possible conflicts between utility maximisation and some other motivation that seems to have strong appeal either in common sense morality or in a more general moral principle, such as justice, or integrity, or loyalty.

One famous example that has given rise to a whole literature on 'trolleyology' refers to a situation in which, one day, you may be walking over a bridge over a railway line and you notice that a train (or some big trolley) is coming along the line and five people happen to be tied to the line. If you do nothing they will be killed. But, as often happens, a Fat Man is standing near you on the bridge right over the line in question and you could easily push him over on to the line which would stop the train. What do you do? Maximising utility requires that you push him. But other ethical considerations come into play. For example, if you let the train continue on its route, the death of the five people will not have resulted from any action by you. So perhaps you are not morally responsible for them. Whereas if you push the fat man you are responsible for his death.

True, finding oneself next to a fat man (or woman) is becoming quite a common occurrence these days, but the situation in the above example is still unlikely to happen often. So let us consider a scenario that is more common in these days of international travel, such as the story about the man called Jim, who is walking along minding his own business in South America when he comes upon a group of bandits who are about to shoot twenty innocent people. In these days of globalisation and international travel this could happen to anyone. But in this case Jim is told that if he would shoot one of the twenty, the other nineteen would be set free. Should Jim take up this offer? In this case, in addition to the point made in the Fat Man story about who would be actually moral responsible for the deaths in question, there is also the question of Jim's *integrity*. Can this be simply converted into units of utility that can be weighed against the loss of utility that the certain death of twenty innocent people would entail?

Yet another famous example is the one about the sheriff in some small town in the USA. In this story a man is held in jail awaiting trial. An angry mob gathers outside the jail and tells the sheriff that if he does not hand over the prisoner, who they will hang, they will go on a rampage in the town and kill dozens of innocent people. What should the sheriff do? In this story impersonal utility maximisation would have to be balanced against two conflicting common sense moral intuitions. One is the claim of duty – that is, the duty that the sheriff may feel to his office. But, in addition there is also the claim of justice.

The conflict between utility maximisation and justice has come up already in earlier chapters. For example, in [Chapter 6](#), it was shown that a policy to introduce minimum wages in the interests of what might be regarded as a more ‘just’ distribution of incomes could conflict with maximising society’s output. Similarly, [Chapter 7](#) discusses the way that social welfare functions indicate possible trade-offs between total output and a ‘just’ distribution of welfare.

Another common feature of examples such as those given above is that certain acts that may maximise utility will constitute acting in a way that conflicts with other basic values, such as certain basic ‘rights’, notably the right to life. For many people would subscribe to a deontological claim that certain ‘rights’ are inviolate – for example, right to life, whether of humans or animals. For example, supporters of ‘animal rights’ would object to the use of animals in medical research even if this impedes medical progress and hence an eventual future reduction in human morbidity or mortality. Other ‘rights’ that may be claimed to conflict with the reliance on utilitarian maximisation could include property rights, or liberty rights. For example, the opening sentence of Robert Nozick’s famous book, *Anarchy, State, and Utopia*, is ‘Individuals have rights, and there are things no person or group may do to them (without violating their rights)’.¹

One of the most famous conflicts between the welfare of society and personal sense of duty is Sophocles’s *Antigone*, who is faced with a conflict between duty to her dead brother and duty to the laws of the state. It is quite likely that a significant proportion of the world’s population believe that ‘duty’ – such as duty to one’s country, to one’s family, to one’s God, to one’s firm (in Japan), to one’s pet dog (Britain), or to one’s political party – over-rides concern with maximising society’s total welfare.

The examples included situations in which certain acts might maximise utility but conflict with the moral obligations and integrity of the person

carrying them out. This is linked to another whole class of possible constraints on impersonal utility maximisation, and one which is far more common in everyday life. For who has not been faced with a conflict between doing what could be expected to maximise society's utility and some personal claim, such as claims of obligations to one's family, or friends, or community, or country, and so on.

2 'SPECIAL OBLIGATIONS' AND 'AGENT RELATIVE ETHICS'

For this reason one of the most important objections to classical *impersonal* Utilitarianism is an appeal to special obligations that one may have to particular people or groups – such as the deontological rule that one should honour one's parents, or fulfill one's duty to one's country. The appeal to 'special obligations' as a constraint on utility maximisation is closely related to what it known as 'agent-relative ethics' or 'agent-centred ethics'. Even David Hume who, as indicated in the previous chapter, was acknowledged by Bentham as the inspiration of his Utilitarianism, gave a prominent place to 'special obligations', or 'agent-relative ethics' in his theory of the evolution of moral conventions (though he did not use these terms).²

For, unlike Bentham, Hume claimed to provide a psychological *explanation* of how moral principles had developed. For he developed at some length his view that morality is firmly based in human nature. Hume emphasised that the utility with which people are primarily concerned originates in agent-relative concerns. For example, he writes, 'A man naturally loves his children better than his nephews, his nephews better than his cousins, his cousins better than strangers, where every thing else is equal. Hence arises our common measures of duty, in preferring the one to the other. Our sense of duty always follows the common and natural course of our passions'.³ Of course, Hume's concept of the 'passions' does not correspond to its current connotation with its association with obsessions or with extreme romantic sentiments. In brief, passion 'is Hume's general term for emotion, attitude and desire'.⁴

However, Hume recognised that giving free reign to the natural instincts of people to pursue their self-interest and that of their family and friends could prevent the development of social conventions that are necessary to enable stable societies to flourish and develop. So something else was needed, even though this would be influenced by expectations of the advantages that people would derive for themselves and their families

and friends. ‘This can be done after no other manner, than by a convention enter’d into by all the members of the society to bestow stability on the possession of those external goods, and leave every one in the peaceable enjoyment of what he may acquire by his fortune and industry’.⁵ Hence, over the ages, the societies that survived and flourished were those which successfully developed such conventions. These conventions reflected – if quite unconsciously – not only the direct interests of individuals and their special relations to others but also their indirect long-run interests in the viability of progressive, stable societies. ‘*Thus self-interest is the original motive to the establishment of justice: but a sympathy with public interest is the source of the moral approbation, which attends that virtue*’⁶ (italics in the original).

Hume presented his explanation of how our sense of morality and justice has evolved in a manner that has since been developed in various modern forms, including socio-biology and the well-known ‘tit-for-tat’ strategy in game theory. (*loc.cit.* sec. II). For example, he writes, ‘Nor is the rule concerning the stability of possession the less deriv’d from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it’.⁷ Indeed, one of the foremost contributors to game theory, Ken Binmore, goes as far as to say that ‘... a game theorist ought to have recognized from the start that Hume is the original inventor of reciprocal altruism – the first person to recognise that the equilibrium ideas now studied in game theory are vital to an understanding of how human societies work’.⁸

But if our moral intuitions and our sense of justice reflect human nature, which must leave room for agent-relative concerns, why should this give them any irresistible moral status? The answer is that it does not. Hume did not claim that it did. He is famous for deploring the tendency of people to jump from ‘is’ propositions (such as propositions about human nature) to ‘ought’ propositions.⁹ He only claimed to explain how certain moral principles have developed in society. In a sense Hume had no *a priori* ‘moral theory’, only a theory of the psychology of how society’s moral conventions have evolved. A purely psychological theory cannot be satisfactory for anyone who seeks ‘true’ moral positions. In Hume’s view, however, truth is not a term that can apply to moral beliefs. How far one accepts that moral principles that are anchored in human nature, and hence tend to give priority to ‘special obligations’ in certain situations, rather than some other impersonal set of principles, such as

Utilitarian or Kantianism, is, of course, a crucial value judgement that plays a major role in certain practical policy issues.

Thus, while Hume can be claimed to be the father of Utilitarianism he differed from classical Utilitarianism in one major respect. This is that his version of Utilitarianism was not so ruthlessly impersonal. He did not subscribe to the view that all that matters is the maximisation of the total utility of society irrespective of the effect on the utility of particular groups in society, such as those to which people would naturally give priority. Nevertheless, as indicated above, he did not believe that prior concern with special obligations of one kind or another and the role that they had played in the early evolution of society's moral conventions should always over-ride 'sympathy' with the rest of our fellow men or with the longer-run interests of our own favoured groups.

There are other 'contractarian' theories of justice that are based on appeals to intelligent self-interest. David Gauthier, for example, has set out in detail the view that our moral values have to be grounded in human nature and that it makes sense for people to eschew blinkered utility maximisation in favour of constrained maximisation that takes account of other people behaving in a similar manner. This leads to a strategy of co-operating with like-minded people and of 'defecting' in dealings with others. Thus, moral constraints make us all better off in terms of our preferences, whatever these may be.¹⁰ If, therefore, we accept (i) the positive proposition that agent-relative/game-theoretic ethics have helped the development of relatively peaceful and prospering societies, and (ii) the value judgement that this development is a 'good' thing, one has some logical basis for the normative conclusion that some form of agent-relative ethics is desirable.

3 PLURAL VALUES AND INCOMMENSURABILITY

Consequentialism, in general, and Utilitarianism, in particular, are very elastic moral theories that can be stretched in all sorts of directions in order to encompass many supposedly non-utilitarian values into the conception of social utility. For example, the utilitarian economist, Harsanyi, wrote that 'any reasonable utilitarian theory must recognise that people assign a non-negligible positive utility to free personal choice, to freedom from unduly burdensome moral standards trying to regulate even the smallest detail of their behaviour' (Harsanyi, 1982:60). Thus if the maximisation of utility appears to violate personal choice this is no problem; one merely has

to treat the violation of personal choice as a deduction from the overall net utility that that particular course of action may have yielded.

All of the concerns discussed above – such as justice or special obligations – can be, and sometimes have been, dealt with by simply extending the concept of ‘utility’ to encompass them. In this case the means adopted to promote some particular consequences – namely a violation of somebody’s ‘rights’ – would be evaluated in utility terms. It would be argued that the value of respect for ‘rights’ resides only in the contribution that such respect will indirectly make to society’s ‘utility’. The same sort of argument could be used to incorporate respect for ‘special obligations’ in the concept of utility.

But if almost every consequence of any choice – however indirect – could be brought under the heading of utility, the scope for differences of opinion about how much any action contributes to the wider concept of utility will be so great that the theory will add very little to our guidance to ethical action. It would mean diluting the concept of utility to homeopathic levels. Everybody will differ in their conception of ‘the utility’ that is to be maximised. Hence, if the concept of ‘utility’ is to retain any distinctive meaning that is a helpful guide to policy it has to be defined along reasonably restricted lines, such as the classical definition in terms of some fairly homogeneous concept of ‘happiness’. And this has to be distinguished from other values such as freedom, equality, justice, integrity and friendship.

For example, if, say, the government of some country have to weigh up their obligations to promote the welfare of their own citizens against the desire to maximise world utility, it does not help to say ‘Oh, that’s no problem. Just add the utility of your own citizens to the world utility that you ought to maximise, and then you can simply trade off one against the other’. ‘Oh, of course!’ explains the politicians, ‘How silly of us not to have thought of that at once! But, while you are here, could you please tell us how to determine the weights?’ Or suppose the authorities are trying to judge whether, in the interests of freedom, society should permit the dissemination of dangerous propaganda and false information that they believe are extremely likely to reduce society’s total utility as conventionally defined. If they are told ‘What’s the problem? Just redefine “utility” to include the value of freedom and it will all be simple. You could then just trade off one bit of utility against another’. This might not matter if all the values that are brought under the umbrella of ‘utility’ were commensurate with each other. But they are not. Hence, ‘trade-off’ would not be an appropriate term for the resolution of the conflict of values. As Steven

Lukes has put it, a more appropriate concept in some contexts would be ‘sacrifice’ rather than ‘trade-off’.¹¹

It is the incommensurability that is the crux of the problem. It is this incommensurability of ‘plural values’ that provides perhaps the main challenge to Utilitarianism, as well as to any attempts to present a unique system of moral theory. If the different values are *incommensurate*, then, whether or not one can put them under the ‘utility’ label, it cannot be claimed that Utilitarianism still provides a simple and handy means of reconciling conflicting values and the different ethical intuitions that are attached to them. In other words, an attempt simply to stretch ‘utility’ to encompass incommensurate values would be stretching it beyond breaking point. The concept of ‘plural values’ refers to basic values that are *incommensurate* with each other and with the classical concept of utility that is identified with ‘happiness’. Indeed, one could regard the term ‘plural values’ as just another term for ‘incommensurate values’.

There has been extensive and subtle discussion in the philosophical literature of the concept of *incommensurability* which I shall not attempt to review.¹² But the basic idea is that values (or options of any kind) are ‘incommensurate’ if they cannot be compared in terms of some common and relevant metric. They are regarded as being *qualitatively* different, so although an increase of any one of these values may add to a ‘good life’ it will add to a *different* ‘good life’. Hence, if some values are fundamentally incommensurate, forcing them into some elastic concept of utility or welfare is an illusion. It merely serves to disguise the fact that where there is some conflict between them – as there frequently is – it cannot be resolved by converting them into some common units of utility and then evaluating the relevant trade-offs.

It is, of course, intellectually attractive to try to force the different kinds of reasons one could advance in favour of one course of action rather than another under one umbrella, namely ‘utility’. Sen has recently drawn our attention to the way that ‘Adam Smith complained more than two hundred years ago about the tendency of some theorists to look for a single homogeneous virtue in terms of which all values that we can plausibly defend could be explained’.¹³ He quotes the passage from Smith’s *The Theory of Moral Sentiments* in which Smith refers to ‘... a propensity, which is natural to all men, but which philosophers in particular are apt to cultivate with a peculiar fondness, as the great means of displaying their ingenuity, the propensity to account for all appearances from as few principles as possible’ (ibid.).

4 INCOMMENSURABILITY AND RATIONAL CHOICE

The fact that in personal and public life decisions are made when faced with difficult choices does not mean that, in the end, some precise commensurate numeraire has been found that helps us trade off the conflicting values involved. Sometimes life confronts us with ‘tragic choices’ where whatever choice we make will lead to violation of some important value and there is no way the alternatives can be compared in quantitative terms. And the fact that some options are strictly *incommensurate* in the sense that they cannot be compared in terms of some relevant metric does not invalidate the economist’s concept of rational choice. This is because there is nothing necessarily irrational about ranking options that are fundamentally incommensurate. The axioms of rational choice described in [Chapter 4](#) do not rely on commensurability. Only *comparability* is required. By definition one cannot ‘rank’ options unless one can ‘compare’ them. So ranking choices in order of preference only requires *comparability*, not *commensurability*.

True, incommensurability means that CBA cannot be given a decisive role in any practical problem, but we have already reached that conclusion anyway, on other grounds. Rational decision criteria do not have to be expressed in terms of the maximisation of a smooth, continuous, twice-differentiable utility function. Incommensurability thus limits the role of precise mathematical expression of the conditions for optimal choice. It implies that rational choice has to be seen as maximisation of a utility function subject to some constraints, which may be certain values that cannot be pushed kicking and screaming into units of utility and hence incorporated into the utility function that is to be maximised. And, as a rule, these values cannot be quantified in any terms at all, so that the mathematical tools of maximisation of a function subject to constraints that can be quantified are also inappropriate. There are no ‘Langrangian multipliers’ that can be calculated as part of the maximisation process.

Of course, sometimes it may appear impossible to rank some option as being higher or lower than some other option. In the example given in [Chapter 4](#) some of the characteristics of the career choices facing our student, such as salary, were easily commensurate. But others – such as how interesting would be the work or how satisfying would be the prestige – would not be *commensurate*. But they would be *comparable*. The student would not have to stay undecided for the rest of his life on account of an inability to choose which career he would prefer to follow.

And even if he found that these two features of the options were evenly balanced, some other consideration would eventually occur to him, or turn up, that would push him to choose one or the other. This might be, for example, the discovery of some feature of the location of the two careers, or a desire to placate (or to spite) his parents. In short, several values may be relevant in rational decision-making even if they cannot be incorporated in the precise framework of a numerical CBA. Reason should still play a crucial role in the decisions. It is simply that reason does not imply simple mechanical application of a single basic value.

At the level of public policy, for example, it is unlikely that a strict CBA by a rich country of giving aid to some poor country devastated by an earthquake or storms would be positive. In such cases other, *non-Paretian*, criteria would be invoked, such as humanitarian considerations, or respect for international obligations. Many countries carry out projects in the interests of national prestige, as when they spend money on projects such as building millennium domes or hosting the Olympic Games. In such cases even though costs may exceed the benefits the projects may still deserve to be carried out. The contrasting situation is where some non-economic value could also outweigh a favourable CBA. This would be the case, for example, in connection with the probable favourable CBA of the practice in Ancient Rome of throwing Christians and other troublemakers to the lions.

There are many instances, therefore, where a decision must take account of non-economic and probably incommensurate criteria. The use of CBA in many situations does not mean that it is the only input into the decision-making process. As has often been said, 'CBA is a decision tool, not a decision rule'. It provides some indication of the net cost to society of subscribing to the values in question. For example, if it were found that half of our national income would have to be used in order to avoid the extinction of some particular species of bird in a remote part of the country, one might sacrifice the birds. But at the other extreme, if a small amount of money would prevent the wholesale disappearance of the entire British bird population, most people would be willing to preserve the birds. It should not, however, be thought that concern to preserve a particular species of bird (or beetle or whatever) illustrates a compassionate concern for the birds in question. It is highly unlikely that a bird that is being pursued by a huntsman is likely to suffer more if it thinks it is almost the last surviving member of his species than if it thinks 'Oh, what the hell. There are millions of other birds of my species still around'.

In this bird extinction example one would avoid the extremes and concede that somewhere in between there might be a point at which the decision would be very difficult. This is why economists tend to dislike lexical orderings. We are concerned with identifying marginal costs and benefits as far as is feasible. And the balance to be struck could be the economic costs of saving the birds against some other value that may be difficult or impossible to quantify. But it is an illusion to think that a utilitarian calculus enables the balance can be struck in a precise manner.

5 CONCLUSIONS

Given that Utilitarianism has been the most influential moral theory to have been developed in the Western world over the last two centuries or more, and is still – in one form or another – one of the most widely accepted moral theories, it is inevitable that a vast literature about it has accumulated. Some of this has taken the form of refinements, additions, expositions and defences of one kind or another. And some of it has consisted of criticism.

I shall not attempt to summarise all the main criticisms since many admirable expert critical surveys are available. But for the purposes of its role in economic analysis the most important criticism concerns the utilitarian assumption that all values can be made commensurate with each other and with a unique basic value, ‘utility’. If, instead, it is accepted that there are independent plural values – such as freedom, personal integrity, loyalty, truth, personal relationships, duty, and so on – that are incommensurate, Utilitarianism can no longer command a monopolistic position in moral theory. This limits the role in economic theory of the maximisation of utility functions. Instead a place has to be given for other independent intrinsic values such as those mentioned above at various points. As I have emphasised in [Chapter 3](#), such values cannot be confirmed or refuted by any scientific and objective observation.

NOTES

1. Nozick, 1974: *ix*.
2. See, for example, the various contributions to Scheffler, S. [(ed.) 1988], such as those by Nagel, Nozick, Scanlon, Sen, Williams and others, to make a rather random selection. Reference to ‘agent-relative’ concepts of ‘the

good' is also made by Broome, in a carefully articulated discussion of inter-generational justice and discounting, in Broome, 1992:42–43 and *passim*.

3. Hume. 1739:3.2.2.10 and 3.2.11.18. He also gives a detailed account of why we tend to attach less value to distant benefits than to present benefits in *ibid.*3.3.7.2, and why this is a regrettable weakness.
4. Blackburn, 2009:55.
5. Hume, 1739:3.2.2.9.
6. *Ibid.*: 3.2.2.24.
7. For example, *ibid.*:1739:3.2.1.27.
8. Binmore, 2005:ix.
9. However, as some philosophers have argued, it would be wrong to interpret this as meaning that Hume did not attach normative significance to his description of the development of moral beliefs or that he failed to spell out the normative basis for a moral system anchored in human nature. See, for example, Hunter, G. 'Hume on *is* and *ought*', and others in a distinguished collection of contributions on this subject to Hudson (ed.), 1969.
10. See Gauthier, 1986.
11. Lukes, 1997.
12. For a compendium of different views on incommensurability and an extremely helpful and illuminating survey one could hardly do better than begin with Chang, R., (ed.), 1997.
13. Sen, 2009:394.