Making the World Safe for Utilitarianism

JONATHAN WOLFF

Introduction

Utilitarianism has a curious history. Its most celebrated founders—Jeremy Bentham and John Stuart Mill—were radical progressives, straddling the worlds of academic philosophy, political science, economic theory and practical affairs. They made innumerable recommendations for legal, social, political and economic reform, often (especially in Bentham's case) described in fine detail. Some of these recommendations were followed, sooner or later, and many of their radical ideas have become close to articles of faith of western liberalism. Furthermore many of these recommendations were made expressly to improve the condition of the deprived, or of oppressed groups. Yet the moral theory which inspired this reforming zeal is, at least officially, utilitarianism, and when we teach this theory to our students we feel it our duty to point out the horrors that could be justified by any theory which assesses the moral quality of actions in terms of the maximization of good consequences over bad. No consequence is so bad that it cannot, in principle, be outweighed by a large aggregation of smaller goods. Hence there are circumstances in which utilitarianism can require slavery, the punishment of the innocent, and redistribution of resources from the poor to the rich, or from the disabled and the sick to the able bodied and healthy. Indeed, in the right circumstances, it can justify pretty much anything you can think of. For all their intelligence and imagination neither Bentham nor Mill seemed to recognise or discuss these catastrophic possibilities.

It is, however, no defence of utilitarianism to say that it was held by people who had fine motives and did not see its consequences. The current orthodoxy is that the central flaw in utilitarianism is that, as Rawls puts it, utilitarianism 'does not take seriously the distinction between persons'.[1] This phrase of Rawls' finds a clear echo in Nozick's remarks that 'There are only individual people,

[1] John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, revised edition 1999), 24.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

with their own individual lives. Using one of these people for the benefit of others, uses him and benefits the others. Nothing more. What happens is that something is done to him for the sake of others. Talk of an overall social good covers this up. (Intentionally?)'[2]

Thus Nozick and Rawls both suggest that utilitarianism proceeds as if morality is some sort of generalised prudence. On such a view just as an individual may accept losses for the sake of greater benefits, so might 'society'. Yet, they counter, in reality all that happens is that one person gains and the other loses. This is no moral compensation. The gains do not justify the losses, even when they are bigger.

Consequently utilitarianism has been out of favour in philosophy for some time. It is remarkable that those with moral and political views as far apart as Rawls and Nozick have united to condemn it on apparently the same grounds (although this is less of a surprise when we realise that they both claim Kantian roots for their theories).[3] In its place we find various views which assign rights to individuals which will block at least some, perhaps even all, applications of maximizing consequentialism. This is true even for some views which are still sympathetic enough to consequentialism to retain the name.[4]

Yet while philosophers have turned away from maximizing consequentialism, public policy decision making has embraced it. Many areas of public policy are dominated by cost-benefit analysis, which at least in its purest from is a particularly crude form of consequentialism: consequentialism of money. Many decisions, large and small, are informed by cost-benefit analysis. The topics range from the building of a new airport to the permissibility of performing a particular animal experiment. This is an important example of where what may well be regarded as an outdated and crude philosophical theory has taken hold—almost as a default or standard theory. This should worry philosophers. And, indeed,

[2] Robert Nozick, *Anarchy, State, and Utopia* (Oxford: Blackwell, 1974), 33.

[3] Of course Nozick also believes that Rawls has not appreciated the full force of this objection, for it undermines Rawls' own theory too, so Nozick argues.

[4] See, for example, the rule-consequentialism of Brad Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000).

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

some have duly reported themselves worried.[5] Yet how we should respond to this situation is by no means clear.

This paper explores the parallels between maximizing consequentialism in philosophy and cost-benefit analysis in public policy decision making. I believe that each area can cast light on the other. My conclusion—to anticipate—is that these maximizing doctrines are very powerful but also very dangerous. Like a powerful but destructive technology, the task is understanding when to use it and when not to. The danger alone is not a sufficient reason to reject the technology if conditions of safe use can be understood and reliably implemented, especially if we are unable to find an alternative which better meets our needs. Hence I want to produce a highly qualified defence of both utilitarianism and cost-benefit analysis. The qualifications I discuss in this paper are these. First, I defend the theories as *decision procedures* rather than as *moral theories* in their own right. Second, they are adequate decision procedures only under certain highly constrained conditions.[6] Indeed, these conditions may rarely, if ever, be met. However if the main argument of this paper is accepted then we have every reason for considering how we can move to a situation where the conditions are in fact satisfied.

Cost-Benefit Analysis And Its Difficulties

What we can call 'pure' cost-benefit analysis takes the following steps:

(i) A qualitative statement of all costs and all benefits of a particular course of action (and its main alternatives, including the alternative of doing nothing) is set out.

[5] Elizabeth Anderson, *Value in Ethics and Economics* (Cambridge, MA: Harvard University Press, 1993), and Henry S. Richardson, 'The Stupidity of the Cost Benefit Standard', in *Cost-Benefit Analysis: Legal, Economic, and Philosophical Perspectives*, M.D. Adler and E.A. Posner (eds.) (Chicago: University of Chicago Press, 2001), 135–167.

[6] There is a third qualification. I believe that in some cases, in order to make adequate decisions, we need to add in factors which would not normally form part of an economic decision making mechanism. This is to say that cost-benefit analysis will often need to be supplemented by other considerations, especially where issues of risk to safety are concerned. This takes us beyond the terrain of this paper. I discuss it in Jonathan Wolff 'Risk, Fear, Blame, Shame and the Regulation of Public Safety', *Economics and Philosophy* forthcoming.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

(ii) These costs and benefits are then rendered in quantitative terms.
(iii) These quantities are then converted into a single currency (usually money) and summed.
(iv) A decision is made on the basis of which alternative provides the greatest net benefit in terms of the designated currency.[7]

Any of these steps can be problematic. First coming to a statement of the full consequences of any course of action is difficult. Even when the possibilities are discerned—and this is difficult in itself—many consequences are uncertain and rendering them in probabilistic terms can be arbitrary and misleading. This is particularly so when outcomes will depend on the possible actions of other human beings, including the decisions they make take, including some in response to the decisions they expect us to take. If we really were to take all possible consequences into account we would very soon be overwhelmed. Therefore simplification must be made, and this always has its dangers. Furthermore, there is a potentially indefinite number of alternatives to any course of action. How do we know which to consider? Onora O'Neill has argued that we will tend not even to consider alternatives we consider to be wrong.[8] Hence even at the first step consequentialism appears parasitic on other moral norms. (However this may be less persuasive in public policy, where the constraints may be simply to stay within the law, than in personal morality where many other considerations may seem relevant.)

Even more contested is the step of converting other values into their monetary equivalent. This is often unsettling. For example when carrying out a risk cost-benefit analysis to see whether it is right to spend money to introduce new safety measures there seems no alternative to putting a financial value on reducing the risk of death. For ease of calculation this is rendered in terms of the saving a 'statistical life'. Currently in the UK government policy requires decision makers to operate with a value of life—or rather a value of preventing a fatality—of a little over a million pounds.[9] To some

[7] For this way of breaking CBA into separate steps see J. Wolff and D. Haubrich, 'Economism and Its Limits', in M Rein, M. Moran and R, Goodin (eds.) *Oxford Handbook to Public Policy* (Oxford:Oxford University Press, forthcoming 2006).

[8] Onora O'Neill, *Faces of Hunger An Essay on Poverty, Development and Justice* (London: Allen and Unwin, 1986), 71–96.

[9] Health and Safety Executive, *Reducing Risk, Protecting People*, 2001, http://www.hse.gov.uk/risk/theory/r2p2.htm

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

this sounds a barbaric way of approaching the issues. A common response is that no price should be put on life. Whether or not one is sympathetic to this claim, and its appropriateness in this extremely difficult context, the fact remains that where values are genuinely unquantifiable or incommensurable cost–benefit analysis will be in grave difficulties. We will return to this below.

Yet even where there is no principled objection to the idea of measurement or conversion, the particular valuations can be highly contested. To give an example a number of studies have tried to estimate the 'cost of a problem gambler' in the context of considering whether the benefits of a liberal gambling regime outweigh the costs. A survey reveals that estimates over the last fifteen years or so, all backed up with detailed costings, and aggressive defence of methodology, vary from an annual cost of £373 to an annual cost of £35,300.[10] It is hard to resist the conclusion that these costings are typically used to support policies advocated on other grounds, rather than being based on some sort of neutral method of valuation.

However even if problems of quantification and commensuration can be solved, CBA may still run into difficulties. Bearing in mind the standard objections to utilitarianism, it is not difficult to see the parallel problems for CBA. A policy could have overall benefits yet be extremely costly to some individuals. Why, it will be asked, should benefits for one party justify costs which fall on others? Consider again the gambling case. Whatever we think about trying to quantify the costs, in qualitative terms they are well understood. Problem gambling can lead to despair, extreme child neglect, family break-up and suicide. Is it really right that we should determine gambling policy purely on the basis of whether the aggregated benefits are greater or smaller than the aggregated costs of the ruined lives of those who suffer? CBA, then, as a form of maximizing consequentialism, shares the defects of other forms.

Now it should at once be admitted that the pure form of CBA is rarely seen. Some theorists and practitioners are often highly

[10] Sir Alan Budd, et al. *Gambling Review Body Report*, DCMS 2001, http://www.culture.gov.uk/global/publications/archive_2001/gamb_rev_report.htm Interestingly this highest figure was mentioned on the front page of the Daily Mail on 16 October 2004 as part of their campaign against the liberalisation of U.K. gambling laws. It was not mentioned that this figure was the highest of a range of estimates, or that it was produced by a group of researchers who were trying to convince their audience that an expensive form of proposed therapy for problem gamblers provided good value for money.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

sensitive to the problems just outlined, and attempt to take appropriate steps. Consequently it is not uncommon to hear the argument that while political decisions should be informed by CBA it should not be the sole input into decision making.[11] In effect this is to decline to take the final step of the pure model. Others, worrying about the third step—the translation of costs and benefits into monetary terms—have argued that because valuations are always contestable different groups should be encouraged to offer their own CBA of the same scenarios.[12]

To avoid some distributional problems, the government has now proposed adding a further step in which financial benefits can be weighted for different groups.[13] So a benefit to the poor is considered to be of greater value than the same benefit to the rich, on the assumption that it will make more difference to their lives. This turns CBA into something closer to classical utilitarianism. Furthermore additional weights can be added to give even greater consideration to those who fare badly, which will move CBA to something closer to Parfit's prioritarianism.[14] In other cases groups considered particularly vulnerable are given special concern. Their vulnerability could be something to do with their relation to the decision—for example people living in a particular location—or more general, the poor or the mentally ill. Hence, in effect, special steps are proposed to attempt to avoid the problem that society's risks and costs are all dumped on a particular group who do not share in the benefits.

In deference to worries about commensurability sometimes the analysis leaps from the second stage to the last, in which costs and benefits are enumerated and quantified but not placed on a common scale. This is the approach taken with respect to animal experimentation.[15] This then leaves decision making to intuitive judgement rather than to the outcome of a quantified formal

[11] Cass Sunstein, *Risk and Reason* (Cambridge: Cambridge University Press, 2002), David Schmidtz, 'A Place for Cost-Benefit Analysis', *Nous* **35,** No. s1 (October 2001), 148–171.

[12] K.S. Shrader-Frachette, *Risk and Rationality* (Berkeley: University of California Press, 1991).

[13] HM Treasury *Green Book, Appraisal and Evaluation in Central Government*, 2003, Annex 5 http://greenbook.treasury.gov.uk/

[14] Derek Parfit, 'Equality and Priority', *Ratio* **10, No. 3** (December 1997), 'Equality and Priority', 202–221.

[15] Michael Banner et. al, *Animal Procedures Committee Recommendations On Cost-Benefit Assessment Under The Animals (Scientific Procedures) Act 1986*, 2003, http://www.apc.gov.uk/

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

method. This type of 'soft' cost-benefit analysis may seem more reasonable, but it loses the advantages that led to the rise of CBA. For it is vulnerable to the criticisms that defenders of pure CBA pose; indeed the same objections which led Bentham to utilitarianism. Bentham, rather scandalously, argues that there are only two alternatives to utilitarianism. The first is the principle of asceticism which is the mirror-image of utilitarianism—maximize suffering—and which Bentham plausibly points out has never been seriously maintained.[16] The second is the principle of sympathy and antipathy, which Bentham also calls the principle of caprice. This, Bentham suggests, approves or disapproves of an action 'merely because a man finds himself disposed to approve or disapprove of them.' The problem, Bentham continues, is that 'What one expects to find in a principle is something that points out some external consideration, as a means of warranting and guiding the internal sentiments of approbation and disapprobation: this expectation is but ill fulfilled by a proposition, which does neither more or less than hold up each of these sentiments as a ground and standard for itself.'[17]

It is hard to disagree with the point that without the rigours of some formal approach to public decision making a great deal must be left to individual judgement—or likings or dislikings—of politicians and even unelected officials. Whether or not this is a problem is arguable. Some will say that in reality we never have an alternative to intuitive judgement, and so CBA, at best, is a cloak for individual judgement, hidden away in the details, and at worst a cloak for abuse of power. On the other hand without some form of formal methodology decisions will seem to lack a firm basis, a series of decisions may be inconsistent, and we may decline beneficial schemes in favour of relatively inefficient ones. Leaving so much to individual judgement opens the way if not always to corruption then to prejudice or lazy thinking. CBA and utilitarianism promise discipline. They provide a means of making decisions rooted in an analysis which can be scrutinised, questioned in public, attacked and defended. Bias and abuse of power can be detected by those scrutinising the calculations. In other words it provides public accountability (something everyone wants for others, even if, less often, for themselves).

[16] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, J.H Burns and H.L.A. Hart (eds.) (London: Athlone, 1970,) 17–21.

[17] Op. cit. note 16, 25.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

What Are the Alternatives?

Economists are well-aware of the accusations against CBA. Their defence often is that they would be more than happy to give it up, if only someone would actually tell them what else to do instead of just complaining. Some critics of CBA have risen to this challenge by proposing 'deliberative democracy' or 'public reason' as a constructive alternative. Yet it is less easy to see what to do with this suggestion. Deliberative democracy is an attempt to substitute an *auditable process* for the *auditable application of a formula*. This has undoubted appeal, yet it is hard to see how this could be applied in the case of each and every decision currently taken by public officials. Discussion will rarely lead to a consensus, and it is hard to see how all legitimate interests will be represented in proportion to their weight. Now there is of course a sophisticated emerging literature on the idea of deliberative democracy, and so this should not be dismissed.[18] Yet although deliberative democracy may be highly suitable for very important issues, such as those concerning national sovereignty, or specialised local decisions affecting a defined population, such as new traffic schemes, it is harder to imagine how it can be used case by case to address the welter of mid-level resource allocation and development decisions faced by central and local government departments. Or at least it is hard to see how it could deliver consistent results. Perhaps this does not matter. However if one hopes to achieve some sort of cumulative effect through decision making it seems unlikely that it can be achieved without some more formal methodology.

What, then, is left? Bentham, it is true, identifies the idea of a formal, accountable, methodology with utilitarianism, and does not consider whether alternative formal methods—quantified or unquantified—are possible. Nevertheless, those who oppose utilitarianism or cost-benefit analysis have a tough time. They need either to propose an alternative accountable methodology, or to be prepared to give up accountability, trusting the intuitions and judgements of officials and politicians, and putting up with the risk of undetectable corruption, prejudice, nepotism and what we might call 'policy drift'. This is a version of a familiar dilemma. For example John Rawls opens *A Theory of Justice* by comparing utilitarianism and intuitionism.[19] Utilitarianism is to be praised for

[18] See, for example, A. Gutmann and D. Thomson, *Why Deliberative Democracy?* (Princeton NJ: Princeton University Press, 2004).

[19] Op. cit. note 1, 19-36.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

offering a firm, principled, accountable methodology. Unfortunately it delivers some horrendous conclusions. Intuitionism need never deliver any results we don't like, but in the absence of an accountable methodology provides no argumentative purchase against those who disagree with the conclusion. In the case of CBA, the pure theory represents the analogue of utilitarianism, while the various modifications introduced to make its deliverances more appealing begin to turn it into something closer to intuitionism. For one might have little to say in defence of the exact weights and fiddles used to make the sums come out right when they are disputed. Indeed those who, in a spirit of concession, say that decision making should be informed, but not determined, by CBA are often in difficulties in explaining what else one should take into account, if not costs and benefits.

Rawls can be understood as being motivated by the aim of having the best of both worlds: the rigour of utilitarianism, and the intuitive appeal of intuitionism. And this, indeed, may well be decision maker's holy grail; an accountable decision procedure which produces only ethically acceptable results. This is a motivation for both the form and the content of Rawls' lexically ordered principles of justice.

Rawls' general approach is undeniably attractive and on the face of it, there seems little reason why it should not be translatable into a tool for public policy decision making. As already noted pure CBA is applied utilitarianism, and utilitarianism is out of favour in philosophy. The obvious next step is to state a more acceptable ethical or political theory and apply it to public policy.

Let us put aside the problem that not everyone accepts that Rawls' theory of justice is correct. The question which concerns us is whether it really does provide a feasible alternative to cost-benefit analysis for public decision making. So the question of how Rawls' theory might be applied to public policy is at least worth taking seriously. Indeed one of the appealing aspects, for many, of Rawls' theory is that it is designed, at least according to some interpretations, with the idea of feasibility firmly in mind.[20]

Rawls' theory is, of course, that first we should provide everyone with a fully adequate set of equal basic liberties; second that we should arrange for equality of opportunity; and third we should act so as to maximize the wealth and income of the worst off. Now Rawls is clear that the 'we' who act in such case is not you and me.

[20] C. Kukathas and P. Pettit, *John Rawls' 'Theory of Justice' and Its Critics* (Oxford: Polity Press, 1991).

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

Neither is it government officials or policy makers. Rather these are principles for the 'basic structure' of society; principles so fundamental that they inform the content of the constitution. There is no intention that any policy maker should make any direct appeal to the principles of justice. Consequently there is an apparent gap between these principles of justice and any concrete methodology for policy makers.

How can this gap be bridged, and Rawls' principles made to yield a policy makers' handbook? The first thing to note is that a great deal of public policy is addressed to what would fall under the first two of Rawls' principles of justice: the liberty principle and the fair opportunity principle. For example the creation of a secure legal system can be seen as contributing to the implementation of the liberty principle, while education, and at least some aspects of the health system are arguably institutions designed to fulfil the aim of equality of opportunity. In these areas, although there may be questions at the level of the development of detailed policy, the overall guiding aim of the institution should be clear, and so policy makers should be able to follow the principles of their institutions.

The problem I want to bring out centres on the final principle, the Difference Principle, which tells us, to put it crudely, to maximize the wealth and income of the worst off. How can we arrange public policy so decisions will turn out to be of the greatest benefit to the least advantaged in terms of their wealth and income? Now it may seem simple. We should tax the rich and give to the poor. Redistribution will achieve Rawlsian justice. However most public policy decision making, although often redistributive in its effects, will not be redistributive in its intentions. These include, for example, many decisions about management and development, both in economic terms and in terms of the material environment. Brian Barry points out that inflation, in effect, redistributes from those with cash savings to those with cash debts. So a decision to maintain any particular rate of inflation will have redistributive consequences, as may a decision about public transport, a leisure centre or a library.[21] How are decisions such as these to be appraised? Here, then, I am especially concerned with what we might think of as growth and management decisions, rather than those with an explicit redistributive intent and rationale.

[21] Brian Barry, 'Does Democracy Cause Inflation? The Political Ideas of Some Economists', in L. Lindberg and C. Maier *The Politics Of Inflation And Economic Stagnation* (Washington DC: The Brookings Institute, 1985), 280–318.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

As mentioned above Rawls does not argue that public policy decision makers should, in each decision, aim to make the worst off as well off as possible. And reflection on examples shows that to do so can be a crippling strategy. In deciding where to put the next airport, should the decision really be taken purely on the grounds of which location (if any at all) would make the worst off best off (assuming that issues of liberties and opportunities were already dealt with or do not apply)? Even if maximizing the advantage of the worst off is one's long term goal—as it should be for a Rawlsian—it is far from clear that it is best achieved through applying a direct strategy of this sort to every case.[22]

There are, in fact, at least two types of problem. The first is relatively straightforward where a short term loss to the worst off will lead to a longer term gain. For example the expense of having to purchase double glazing may be worth it if the new airport will eventually produce new employment opportunities with a higher wage.[23] This should be reasonably easy to deal with. The second, and much more tricky problem, is where a general policy of allowing the possibility of uncompensated losses for the worst off might be of greater medium term net benefit even to the worst off than a policy which allows them only to gain. What they lose on the swings might be doubled on the roundabouts. For a general policy which allows risk may well generate larger gains for everyone. This requires thinking of decisions as forming part of a series, even when they are taken in an uncoordinated way by different decision makers. What is needed is a general policy to cover such cases, which allows short term loss in the expectation that such a general policy will nevertheless be to the advantage of the worst off. We might call this 'Rule-Rawlsianism' where we sometimes allow the worst off to lose. Yet is may not be at all obvious how we can formulate the algorithm for policy makers which will collectively generate the greatest advantage for the worst off.

Now it might be thought that I have exaggerated the difficulties. Any policy maker should do whatever is best to stimulate the economy; this will then expand the tax base and allow for greater redistribution to the worst off. Problem solved. However this ignores the uncertainty of the outcome of decision making. Consider, for example, the decision whether or not to bid for the

[22] Still less by deliberative democracy, of course.

[23] And indeed where a large surplus is envisaged compensation to those who lose in the short-term may often be possible (although for some qualifications to this see note 32 below).

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

Olympics, restricting attention just to financial consequences.The bid will cost money, and it might not succeed. Worse, we may win the bid but find that the economic consequences are negative. Only under the most favourable outcome will the economy be stimulated and generate economic benefits. Under all others it is likely that the poor will be made a little worse off. Considering this, if our concern is to ensure the position of the worst off, we had better not take this gamble. Yet, it seems, this conservative approach could be damaging over time, if generalised. So we are left with this puzzle: what is the right general approach to adopt?

The significance of this point is that while Rawls' theory seems to provide the best hope for a formal, quantifiable, alternative to utilitarianism, and hence by extension to cost-benefit analysis, it falls down at the vital point. If Rawls' theory—a theory specifically designed with issues of implementation in mind—fails to yield a determinate decision procedure for policy makers, what hope is there for other theories? Hence it is easy to have sympathy for the claim that cost–benefit analysis, in some form or other, seems to provide the only determinate methodology we have, in the sense of providing a formal, accountable decision procedure for public policy.

How important is a determinate methodology, bearing in mind that in the case of maximizing consequentialism its cost can be unacceptable results? What good is a method if it doesn't yield what we want? Here we might be reminded of the old joke about a drunk looking for his lost house key in a well-lit street, even though he dropped it in the adjacent dark street, arguing that 'at least I can see properly here'. Certainly he is right that on the whole it is better to be looking for something in the light than in the dark. But if what he is looking for is somewhere else he will have a frustrating night ahead of him. However two points are worth making. Even though it is true that looking for the key in a well-lit street doesn't help you find it when it is not there, this does not mean that the approach is 'refuted' and should not be used when you have reason to believe that the key is there. Second, and to continue to pursue this rather strained analogy, consider breaking the original search into two parts. In the first the contents of the neighbouring dark streets are trawled and dragged somehow into the well-lit street. In the second the light street is searched. Now this doesn't seem so stupid. This is to attempt to convert cases where the method will not work into cases where it will; an ambitious strategy but not one to be disregarded entirely.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

To put this in context, earlier we briefly explored responses to the fact that forms of maximizing consequentialism can yield counter-intuitive results. We can now see that they broadly fall into two types. The first is to modify the theory so it yields better results, through weights and incorporation of other factors. The second is to try to appeal to a different moral theory altogether. Both responses have grave difficulties. In the first case the more we modify the theory to make it more acceptable the more we lose both the benefits of the maximizing approach and the discipline of a formal method. In the second we lose the method, and unless we find another we are sunk. But now I'm suggesting a third possibility: first try to distinguish the types of circumstances under which the method works from those where it fails. Then we use the method where it works. What about cases where it does not work? One possibility, I have suggested, is that we might attempt to convert cases where it does not work into types of cases where the method works—drag them into the well-lit street. Of course conversion is unlikely always to be possible, and so various secondary strategies will be needed. But nevertheless the sensible strategy must be first, to take another look at the variety of cases, and try to achieve a systematic account of when maximizing consequentialism works and when it does not.

When Does Maximizing Consequentialism Work?

To advance the discussion I want to return to Rawls; this time to his discussion of the choice of principles behind the veil of ignorance. It will be recalled that Rawls argues that the question of what is the correct theory of justice comes to turn on the question of rational choice: the choice of those behind the veil of ignorance, who are ignorant of their own characteristics. The principles people would rationally choose in such circumstances are, Rawls argues, the correct principles of justice, at least for societies in the western liberal democratic tradition. The issue of the correct choice of principles, in turn, reduces to the question of the rational principles of choice to apply in the circumstances of the original position.

For present purposes we need only compare Rawls' favoured option—the maximin criterion, which leads him to propose his theory of justice outlined above—with the principle of maximizing expectations, or maximizing average utility (MAU) which would lead to average utilitarianism.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

Rawls notes that some will be surprised by his preference for maximin over MAU.[24] For, as Rawls explains, MAU is taken almost as a definition of rational behaviour in many contexts, especially within economic theory. He argues, nevertheless, that the special circumstances of the original position make MAU an irrational approach. For present purposes the most important special features are, first, this is a one-off situation, and second, the choice is highly significant in that if it goes wrong it will go very wrong. There are, Rawls says, 'grave risks'.[25] Consequently, it seems implied by Rawls's discussion, if these features were **not** present then MAU would be much more attractive. That is to say, if the situation was regular and repeating, and losses were not catastrophic, then the economists may well be right: MAU is rational under those circumstances. To illustrate suppose that each week you are faced with a lottery in which you can either have the certainty of £100 or a fifty percent chance of £300 (and a fifty percent chance of nothing). If you really were faced with this choice each week, and you were sure that the odds were genuine, then you would clearly end up better off by playing a regular strategy of MAU—gambling—rather than playing safe with maximin.

We should note that choosing MAU does not entail that you believe that the correct theory of prudence is to maximize the average. On the contrary, you might be convinced that prudence requires one to maximize the worst reasonable possibility (i.e., maximize the worst outcome but ignoring possibilities with freak odds). Interestingly, though, in this case both a reasonably cautious and an adventurous theory converge on the same decision procedure. This is an important, and potentially powerful, result. We can detach the *theory* of maximizing consequentialism from the *decision procedure*, for the same decision procedure may, under certain conditions, be compatible with a variety of theories.

There are, of course, cases where a cautious attitude would dictate a different choice. If, at the time of making a decision, you had no other money, and no way of surviving without money, then

[24] Op. cit. note 1, 134.

[25] Op. cit. note 1, 147–8. Indeed after this paper was substantially drafted I noticed that in these pages Rawls attributes a version of the view defended here to Edgeworth, and then points out and challenges the underlying assumptions. Some of the arguments of this paper can, therefore, be read as responses to Rawls' criticisms of Edgeworth.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

gambling would be very bold.[26] But the main point is that over a long series, with true odds and no grave risks, the policy of maximizing average expectations—be it money or utility—will almost certainly lead to a policy of maximizing actual money and utility.[27] Under such conditions the policy of making the worst thing that can (reasonably) happen as good as possible tells you to follow a maximizing strategy. Consequently, to put it starkly, there are times when a Rawlsian should adopt the methods of a utilitarian.

Can we be more precise about these circumstances? Four conditions seem to stand out:

1. There need to be regular opportunities of a similar nature. (Call this the assumption of 'many chances'.)
2. No single loss (or likely repeated series of losses) creates a type of level of harm for any individual from which recovery is very difficult or impossible. (The assumption of 'recoverable loss'.)
3. There is no reason to doubt that the probabilities run true. (The assumption of 'true odds'.)
4. All relevant gains and losses can be quantified and compared to each other. (The assumption of 'weak commensurability'.)

Each of these conditions requires elaboration. First, 'many chances'. Clearly it will be rare that anyone will be presented with exactly the same type of decision time and again. Rather what I have in mind, in relation to public policy, is some sort of 'routine decision for government and its agencies' so that the decision whether or not to bid for the Olympics, whether to build a high-speed rail link, whether to build a shopping centre, and what

[26] However although a different choice is called for, it is less obvious that this case shows that the policy of maximizing average utility has been abandoned. Those with no money at all have steeply diminishing marginal utilities for money, and so 50% chance of £300 will, under these circumstances, have less utility than £100.

[27] Can it be that the policies of maximizing expected money and maximizing expected utility come apart? In the short term this is possible, and it is even possible to think up examples in which one maximizes actual utility by following the policy of maximizing not expected utility but expected money (where marginal utilities decline one will refuse risks that would be profitable in utility terms in a longer series). However in a longer series the calculations are more awkward as the utility of money depends not only on how much you receive in total but whether you receive it when you need it or can otherwise make good use of it.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

interest rate policy to pursue all fall within the scope of the class to be taken collectively. One-off decisions, in the intended sense, may be much rarer, and may include such things as a decision to go to war, although even this, arguably, could be included. The excluded class also includes those decisions taken explicitly to achieve aims of social justice.

The second assumption of 'recoverable loss' needs further thought, as no path through life is risk free, and maiming or death is not recoverable for the individual involved. Everyday we are exposed to myriad tiny risks; risks tiny enough that we can discount them for most practical purposes. Yet many public decisions will lead to deaths. If we win the Olympic bid and engage in major construction, it is close to certain that at least one construction worker will die. Nevertheless we are unlikely to think this a sufficient reason for not making the bid.[28] Clearly we need to formulate the assumption in such a way that very small risks are permitted, yet it remains a puzzle how to deal with this in detail.[29]

The assumption of true odds also needs explanation. In the case of tossing a coin, where there are true odds one can expect to be a winner roughly as often as being a loser, at least in the longer run. True odds would be violated if it turned out that men, or white people, or the educated, or the rich, won more often than women, or black people, or the uneducated, or the poor. Another way of putting the assumption is that it holds only if in the longer run one's actual payout will come close to matching the statistically expected payout.

Now it may be that this assumption may be quite unrealistic. Indeed I have argued elsewhere disadvantages tend to cluster together, and those who do badly in some respects will come to find that they do badly in others in others.[30] Where we have a society where some groups tend to win and others tend to lose, then the application of cost-benefit analysis can continue to reinforce this. Such clustering of disadvantages needs to be addressed directly as part of a social justice agenda. Hence the arguments of this paper in defence of cost-benefit analysis are most pertinent when we have

[28] Does this reveal some sort of class bias? Martin O'Neill asks whether we would accept the loss so easily if we could predict the death of a teacher or doctor?

[29] For further discussion see Wolff, op. cit. note 6.

[30] Jonathan Wolff and Avner de-Shalit, *Disadvantage* (Oxford: Oxford University Press, forthcoming 2006). For similar arguments, see Brian Barry, *Why Social Justice Matters* (Oxford: Polity Press, 2005).

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

reason to believe that clustering effects will not skew the odds, or for a hypothetical future where we have overcome such effects. We will return to this.

Finally, I have referred to the fourth condition as 'weak commensurability', allowing for sufficient comparability between different goods so that cost–benefit analysis can be applied. Philosophically this is often regarded as highly dubious,[31] yet it is important to understand what exactly is at stake.

For the purposes of making decisions, values of different sorts have to be compared in order to decide where our priorities lie in cases of conflict over use of resources. As we noted, for example, in safety decision making, a monetary value of around one million pounds is put on saving a life. Yet we should not assume that means that as a society we put a monetary value on each life, and that value is one million pounds, as if one million pounds somehow equates to a life and compensates for a death. It seems important, then, that we regard this as an equation used for making safety decisions rather than an all-purpose identification. Indeed this form of what I call 'weak commensurability' is consistent with what we could call 'weak incommensurability', which is the idea that the loss of one good—health say—cannot properly be made up for by any amount of some other good, such as money. Many people may be prepared to assent to this form of incommensurability. Nevertheless, the acceptance of such a doctrine does not stop us from arguing that some health interventions provide good value for money while others do not. Yet that judgement, of course, seems to presuppose weak commensurability. Hence it is possible to accept both weak commensurabilty (comparisons for some purposes) and weak incommensurability (non-comparisons for other purposes).[32] We could define the strong versions of each as the denial of the weak version of the other.

[31] See, for example, A. Sen and B. Williams, Introduction to *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982). And, indeed, I have argued for incommensurability myself, see Jonathan Wolff, Addressing Disadvantage and the Human Good, *Journal of Applied Philosophy 2002,* **19**, No. 3, (2002), 207–218.

[32] If this is right then it explains why the apparently appealing idea of 'cost-benefit analysis with real compensation'; (i.e., those who would lose are compensated with money so that no one actually loses) is more problematic than it may seem. Of course compensation with money can be better than no compensation (although not in every case, see Bruno Frey,

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

Jonathan Wolff

Consequentialism Reassessed

It appears that the notorious counter-examples to maximizing consequentialism occur under circumstances when at least one of the first three assumptions—many chances, recoverable loss, and true odds—are violated. The classic 'scapegoating' objection to utilitarianism—hanging the innocent to calm the mob—probably violates all three in its classic version: it is one-off situation, the loss of being hanged is irrecoverable, and a person is picked not at random but because he is a member of a suspected group. Let us call circumstances in which these three assumptions, plus the further assumption of weak commensurability, hold 'fortunate circumstances', and where at least one is violated 'unfortunate circumstances'. Of course unfortunate circumstances come in degrees. When several of the assumptions are violated—as in the case of scapegoating—the circumstances are deeply unfortunate.

In fortunate circumstances—where the assumptions hold—gambling and the preparedness to lose from time to time pays. To put this another way, the strategy of taking the rough with the smooth should ensure more 'smooth' over the medium and long term, even for the worst off, than following any other strategy, such as the direct maximin policy of always trying to make the rough as little rough as possible.

With this in mind it is worth revisiting the 'separateness of persons' objection to utilitarianism. Nozick and Rawls, we saw, argued that utilitarianism falsely assumes that one person's gain can morally override another's loss. However we now see that in fortunate circumstances no such assumption is required. One person's loss will be, if not compensated, than in some sense outweighted by that same person's gain sometime earlier or later. When this is true, utilitarianism—as a decision strategy—really is prudence. There are gains which outweigh the losses within each life. However, in unfortunate circumstances—when the assumptions do not hold—the criticism is apt. Some may suffer uncompensated losses, of a very significant nature. How should we respond to these difficulties?

At least within philosophy, the most common response is to treat the lack of generality of utilitarianism as a reason for rejecting it.

Not Just for the Money (Cheltenham: Edward Elgar, 1997)) but unless strong commensurability holds it will not 'return' people to their baseline situation.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

For if it can lead to unacceptable consequences it cannot be true. However, rejecting a theory is not, in itself to provide an alternative. And although there are plenty of more appealing approaches to personal morality, we do not seem to have many candidate alternatives for public policy decision making, as we have already seen.

A second alternative is to adjust the theory so that the difficult cases cannot, or are least less likely, to occur. As we saw the, the idea of weighting benefits to the poor may avoid some unfairness, and this approach has been explored in Treasury advice.[33] Unfortunately, though, there are two problems. First, and most obviously, unless some weights are absolute, horrors can still occur. Second, and more subtly, while weighting in one-off cases is to be recommended, to adopt weighting in what I have termed fortunate circumstances will do more harm than good, over time, even for the worst off. For it will impede risk-taking that would reasonably be expected to benefit everyone over time, including the poor (remember that in fortunate circumstances probabilities run true). Now if we are never in fortunate circumstances then this is less of a concern. But if we are to adopt weighting as a universal rule we begin to drain away some of the benefits of maximizing consequentialism.

These two problems push in opposite directions. To deal with the problem of catastrophic consequences we may wish to assign very strong individual rights. However once we trim the scope of consequentialism in this way we will lose even more of its benefits. Indeed giving people rights not to lose even in what I have termed fortunate circumstances is highly problematic. We are used to hearing about the tragedy of the commons—that land owned in common will be over-grazed, that oceans owned in common will be over-fished etc—and that private property rights of some sort are the necessary remedy. What we need to hear more about is the opposite—the tragedy of privatisation. Indeed there may be several tragedies of privatisation, but here I mention only one, concerning risk. Suppose, as is increasingly common, people are not prepared to accept the risk of (non-disastrous) loss or harm. In the earlier image, suppose people are simply not prepared to take the rough with the smooth, and insist on rights against even modest damage, rather than 'pooling' exposure. Well, the smooth will not be so smooth for anyone. It will no longer be the case that 'what you lose on the swings, you gain on the roundabout'. Making sure you never

[33] See note 13.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

lose out means that you will lose out, compared to what might have been. Policy making will become more risk-averse, and in what I have called fortunate circumstances we can all expect to lose over time.

Consequently assigning rights seems not to get us where we need to be. How about the strategy hinted at above, of conversion—dragging everything into the well-lit street? The idea, then, would be not to adjust the decision procedure but to adjust the world so that it fits the procedure. How can this be done? The main problem we face may well be the fact that disadvantages cluster which makes the assumption of true odds unlikely to be true. Now it seems to me that we have independent reasons, based on distributive justice, to wish to challenge and break up such clustering and thereby help bring about true odds. Although it may be utopian to think we can ever complete such a task, the closer we get to it, the more appropriate it will be to use cost-benefit analysis, for the benefit of everyone.

Hence if we cannot bring about favourable circumstances—and in fact even if we could—it is worth considering the possibility of approaching decision-making through two waves of policy. The first wave uses cost-benefit analysis; the second mops up afterwards, or even better acts concurrently to ensure that no harm is done to those already relatively badly off. Some time ago I proposed a version of this, calling it a 'weak equity axiom'; that is 'If a change generates a surplus then those already towards the bottom of the distribution should not suffer as a result of the change.'[34] I would now want to interpret this as allowing loss in the short term, but not the medium. Having thought of this as a modification to utilitarianism, I was rather astonished to discover that Bentham had made a similar point. When advocating the introduction of the printing press in Tripoli and Greece, he warned 'care should be taken that the employment given to it should not be

[34] Jonathan Wolff, 'Rational, Fair and Reasonable in P.J. Kelly (ed.) *Impartiality, Neutrality and Justice* (Edinburgh: Edinburgh University Press, 1998), 35–43. This is, of course, very similar to the idea of cost-benefit analysis with real compensation, mentioned in footnote 32.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

such as to throw out of employment any of the existing scribes, except in so far as other employment not less advantageous is found for them'.[35]

From a classical utilitarian point of view, why worry about the scribes? Bentham, no doubt, had some idea of indirect maximizing strategies in mind: noting the value of security, and also appreciating that mitigating the loss of the scribes—which they will feel very heavily—will cost relatively little. But this type of secondary policy can be supported on Rawlsian as well as utilitarian grounds, and is appropriate in large, one-off, cases and circumstances where probabilities do not run true. Indeed it may be no accident that Rawls sees as possibly the most plausible alternative to his theory a hybrid view in which a liberty principle is supplemented by utilitarianism subject to a social minimum.[36] Perhaps several theoretical approaches, including Rawls's own concern for the worst off, all converge on the same decision procedure for public policy.

The idea of two waves of policy may bring to mind Musgrave's classic distinction between the allocative branch and the distributive branch of economic governance. Clearly there is a great deal in common between my approach and Musgrave's.[37] However, there are also differences worth emphasising too. My account does not appeal to the value of efficiency in its own right at any point. Hence, unlike Musgrave I make no observations about personal incentives, labour supply, marginal tax rates, and so on. Rather I point out that in order to make the worst off as well off as possible society needs to take risks which may in fact, over the short term, make the worst off worse off. Hence my concern even at the first stage is distributive. For that reason it is important that we have already taken some steps to bring us closer to fortunate circumstances, especially in relation to clusters of disadvantage. Efficiency, then, turns out to be a very welcome by-product. However given that we are unlikely to meet all the conditions of favourable circumstances, and, even if we were improbable runs of

[35] Jeremy Bentham, *Securities Against Misrule and Other Constitutional Writings for Tripoli and Greece*, (ed.) Philip Schofield (Oxford: Clarendon Press, 1990), 38. I thank Philip Schofield for drawing my attention to this text.

[36] Op. cit note 1, xiv.

[37] I am very grateful Shepley Orr for bringing this to my attention. Richard Musgrave, *The Theory of Public Finance* (New York: McGraw Hill, 1959).

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press

odds may leave some people badly off, a second wave of policy, as for Musgrave, will be needed to mop up ill effects of the first. So there are differences and parallels between Musgrave's ideas and those presented in this paper.

Conclusion

By way of conclusion it is worth reflecting on Will Kymlicka's interesting diagnosis of the problem with utilitarianism.[38] According to Kymlicka the fundamental impulse of utilitarianism is Bentham's dictum that everyone is to count for one and no one for more than one. Bentham, argues Kymlicka, was too quick to interpret this egalitarian premise in additive terms, summing advantages. The problem, we have seen, is this leads to unacceptable outcomes in certain types of case. Kymlicka's response is that Bentham should have been an egalitarian in the tradition started by Rawls and completed by Dworkin. My reply is that the counter-examples are good reasons for limiting utilitarianism's scope as a decision procedure, but need not force its complete rejection. There are two reasons why this is welcome. First, no one else has told us how to approach public policy decision making (except decisions of an explicitly redistributive kind, typically involving transfers of money). Second, we can all be better off Bentham's way. Under fortunate circumstances Utilitarianism—and with it cost-benefit analysis—can make us all better off.[39]

[38] Will Kymlicka, *Contemporary Political Philosophy,* 2nd edition (Oxford: Oxford University Press, 2001).

[39] I would like to thank audiences at the R.I.P. lecture in London, and Cambridge for their comments and criticisms. I would especially like to thank Veronique Munoz Darde, Martin O'Neill, Shepley Orr, and Michael Otsuka, for extremely helpful written comments.

https://doi.org/10.1017/S1358246106058012 Published online by Cambridge University Press