

## Uncertainty and Fairness

John Broome

*The Economic Journal*, Vol. 94, No. 375 (Sep., 1984), 624-632.

Stable URL:

<http://links.jstor.org/sici?sici=0013-0133%28198409%2994%3A375%3C624%3AUAF%3E2.0.CO%3B2-M>

*The Economic Journal* is currently published by Royal Economic Society.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/res.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).



## UNCERTAINTY AND FAIRNESS\*

*John Broome*

It is a popular belief that there is merit in equalising people's utilities: for a given total of utility, the more equally it is distributed to people the better. The same belief is sometimes extended to *expected* utility too: there is merit in equalising people's expected utilities. Peter Diamond (1967) takes this view, so does Roger Myerson (1981), and Peter Hammond (1983) reluctantly endorses it. If it is true it is profoundly important for welfare economics. It means, as I shall explain in a moment, that 'social' preferences do not obey the sure-thing principle,<sup>1</sup> which is generally taken to be an essential requirement of rationality.

If the sure-thing principle fails for social preferences then one consequence will be the breakdown of John Harsanyi's (1955) very powerful argument in favour of utilitarianism. His argument relies on the sure-thing principle as a premise, and Diamond's objection to the principle was explicitly intended to refute Harsanyi. I, on the other hand, believe there is quite a lot to be said for Harsanyi's argument once it is cleared of some unnecessary commitments. I have tried to say it in another paper (Broome 1983), but as a preliminary I need first to shore up the sure-thing principle against Diamond's attack.

A second consequence of the principle's failure would be that, when the government's actions have uncertain consequences, a conflict could arise between a so-called '*ex-ante*' appraisal of the actions and an '*ex-post*' one, and the *ex-ante* would be superior. (This is explained in Broome 1982, p. 211). This, though, is an undesirable conclusion because there are other good reasons, described by Hammond (1982; 1983) and myself (Broome 1982), to prefer an *ex-post* appraisal to an *ex-ante* one if ever they conflict. This explains why Hammond is reluctant about endorsing the claim that equality of expected utilities is beneficial. To support *ex-post* appraisals I need an answer to this claim.

These, then, are the facts that motivate this paper. I hope to show that there is no reason to value equality in expected utilities, nor to give up the sure-thing principle for social preferences.

\* I presented a version of this paper at an Economics Sub-Faculty Seminar in Oxford. I am grateful to everybody who was there for their comments and particularly to Amartya Sen for his advice then and at other times. I have also received valuable written comments from Peter Diamond, James Griffin and two anonymous referees, for which I thank them. The paper was written while I was a Visiting Fellow at All Souls College, Oxford, financed by a grant from the Social Science Research Council. I thank the College for its generous hospitality and the SSRC for its support.

<sup>1</sup> The sure-thing principle is an axiom of the theory of decision under uncertainty. It says that if one action will in both of two exclusive and exhaustive events lead to results that are at least as good as the results of some alternative action, then the first is at least as good an action as the second. See Savage (1954) pp. 21-2.

An important side-effect of my argument is that it offers a new objection to what Amartya Sen (1979) calls 'welfarism'.<sup>1</sup>

To focus the argument I shall adopt an example of Diamond's (1967), which is displayed in Table 1. There are two equally probable states of nature 1 and 2, and two alternative actions *A* and *B* that the government might take. The society consists of Michael and Maggie, two equally deserving people. The table shows what results the actions would have in the different states of nature; the brackets show Michael's utility, then Maggie's.

Table 1  
*Diamond's Example*

Actions	States of nature	
	1	2
<i>A</i>	(1, 0)	(1, 0)
<i>B</i>	(1, 0)	(0, 1)

Either action will lead to an equally unequal distribution of utilities. But action *B* distributes expected utilities equally and action *A* does not. Diamond says that he finds *B* preferable for that reason. If *B* is preferable then we can see at once that the sure-thing principle is violated. If state of nature 1 occurs the results of *A* and *B* will be the same and therefore indifferent, and if state 2 occurs the results will be symmetrical and therefore indifferent. It is a sure thing, then, that the results of *A* and *B* will be indifferent. According to the sure-thing principle the actions themselves should be indifferent. But actually we are supposing *B* is preferable.

But is *B* really preferable to *A*? Diamond says that *B* gives Maggie 'a fair shake', and it is clear that if *B* is preferable that must be because it is in some way fairer. So it is fairer? There seem to be two reasons why one might think so, both mentioned, or hinted at, by Diamond. One is simply that expected utilities are equal in *B*, and that may be fairer for the same reasons as equality in utility is often thought to be fair. The other is that *B* embodies a fairer 'process of choice' as Diamond puts it. Diamond evidently thinks of *B* as embodying some random process for choosing whether Michael or Maggie should get the unit of utility, and he assumes random choice to be fair in this case. If this is right, then valuing equality in expected utilities may be the right way of capturing in welfare economics the value of random selection.

My argument will answer these two points in turn. Section I shows that, even granted that equalising people's utilities is a way to be fair, that is no reason to think equalising expected utilities is. Section II shows that, even granted the fairness of random selection on appropriate occasions, this fairness is not captured by valuing equality in expected utilities. Section III presents an example to support my case (to add to those already given in Section II), and also mentions, by way of illustration, a controversial application of it.

<sup>1</sup> Harsanyi's (1955) argument is, of course, committed to welfarism. But that is one of the unnecessary commitments it needs to be purged of. See Broome (1983).

## 1. THE NATURE OF EXPECTED UTILITY

Why should we value equality in expected utilities? I think Diamond and other authors have simply taken it for granted that doing so is a natural extension of valuing equality in utilities: since equalising utilities improves fairness, equalising expected utilities must do so too. Roger Myerson (1981, p. 884), for instance, does not even distinguish one from the other.

But this is a mistake. It is perfectly consistent to attach value to equality in utilities but none to equality in expected utilities. Whatever grounds there may be for favouring equality in utilities there is no reason to think they are also grounds for favouring equality in expected utilities. An expected utility is not at all the same sort of thing as a utility, so what applies to one does not necessarily apply to the other. To make this clear I shall need to remind you of just what sorts of things utilities and expected utilities are.

'Utility' is a technical term that is used in different senses by different authors, and I shall have to deal with its various senses. Actually, I only need to distinguish two. On the one hand there is utility as a mental state of some sort, such as enjoyment, pleasure, happiness, satisfaction in the sense of a feeling (as one feels satisfied after a meal), or something else. On the other hand there is utility as desire-satisfaction: an object's utility to a person is the degree to which he wants it and consequently the degree to which his wants will be satisfied if he gets it. No doubt 'utility' is used in other senses too. For instance, it might stand for 'well-being', which might be some index of health, wealth, number of friends, and so on. But I do not need to give these other senses a separate discussion because what I shall say about utility as a mental state can easily be adapted to cover them too. It is only utility as desire-satisfaction that needs to be treated slightly differently.

Suppose first, then, that utility stands for a mental state. Then expected utility is not a mental state. It is the mathematical expectation of what a person's mental state will be. Of course, if a person knows his expected utility, that can *affect* his mental state. He may have feelings of hope, despair, worry, expectation in the sense of a feeling (as one feels expectant before a meal), and so on. These may raise or lower his utility. Feelings like this must be taken account of in specifying people's utilities. Look, for instance, at Diamond's example in Table 1. Let us suppose that if the government takes action *B* that will make Maggie more cheerful than *A* would, until the state of nature becomes apparent. Then this cheerfulness of hers must have been incorporated into the utilities Maggie receives from action *B* in both states of nature. (The table shows, actually, since Maggie's utility in state 1 is the same under both actions, that she is not more cheerful under *B*, or else that *B* has some offsetting disadvantage for her.) Once the influence of expectations on utility has been taken into account like this, it is plain that expected utility does not then separately measure any sort of mental state.

It is instructive to look at the treatment of Diamond's example by Deschamps and Gevers (1979, pp. 149-50). They claim that there can be no reason to prefer action *B* to *A* unless Michael and Maggie know, before the final outcome

is known, which action has been chosen. If they do, these authors say, there will be a period of time, before the people get the utilities displayed in the table, during which they will be deriving utility from their expectations. They implicitly assume that the utility a person derives from his expectations depends positively on the expected value of his eventual utility. Now, they say, there is a good reason to prefer action *B* to *A*, because during this period of expectations action *B* leads to more equal utility. But this treatment robs Diamond's example of its point. The interest of the example is in whether action *B* is preferable to *A* even though both distribute actual utilities equally unequally. Deschamps and Gevers simply assume that the answer to this question is no. Then they distort the example by making action *B* distribute utilities more equally than *A*, taking account of this period of expectations. Then of course (if we are egalitarian about utility) we shall prefer *B*. But that was not what interested us, and it poses no threat to the sure-thing principle.

Now suppose that utility represents desire-satisfaction. This interpretation seems to put utility and expected utility much closer together. If his utility represents the degree to which a person wants an outcome, then according to decision theory his expected utility will represent the degree to which he wants an action (or would want it if he was properly informed). Both utility and expected utility, then, stand for degrees of desire. But there is still a vital difference between them. Expected utility – the desire for an action – is the desire for a means, whereas utility – the desire for an outcome – is the desire for an end, for something wanted for its own sake. This difference is vital because it means that whereas utility may be intrinsically valuable – it may be intrinsically valuable that someone should have something he desires for its own sake – expected utility is not. We value expected utility, but only because promoting it is a way (on average) of promoting utility. Now, we may have reasons to favour an equal distribution of a good worth having for its own sake, such as utility. But there is no reason to think these reasons will carry over to something intrinsically valueless, such as expected utility.

Of course, if we have reasons to promote equality in utilities those are also reasons to promote whatever leads to equality in utilities. Consequently we shall certainly have reasons to promote the expected equality of utilities (the mathematical expectation of the degree of equality of utilities). This is a way (on average) to promote equality in utilities. But whether we should promote the equality of expected utilities is quite another matter.<sup>1</sup> In Diamond's example equalising expected utilities by choosing action *B* does nothing towards equalising utilities.

If equality in utilities is valuable it does not follow that equality in expected utilities is valuable. We have not yet found a reason for preferring action *B* to *A* in Diamond's example. So I turn to Diamond's second suggested reason, that *B* embodies a fairer process of choice.

<sup>1</sup> As a referee pointed out to me, the difference arises because an index of equality, as a function of utility or expected utility, is bound to be non-linear.

## II. RANDOM SELECTION AND WELFARE

If Diamond thinks that action *B* is fairer than *A* because of the process of choice, that can only be because he takes *B* to incorporate some random process for deciding who is to get the utility. Random selection does sometimes seem to be a way of achieving some sort of fairness when, as in the example, inequality cannot in the end be avoided. At any rate, in this Section I shall take that for granted and show that, even so, it gives us no reason to value equality in expected utilities, nor to doubt the sure-thing principle.

Diamond assumes that action *B* contains a random process of selection. But he does not actually describe the process, and Table 1 by itself is not enough to show that the process in *B* is really more random than in *A*. Suppose the story is this:

(a) Both people need haemodialysis, but only one machine is available. It has been decided to choose by lot who is to be treated. The lottery has been held, and Michael won. Now two courses of action are being considered. *A* is to give Michael the treatment as arranged. *B* is to cancel the arrangement and instead treat whichever of Maggie and Michael turns out after a genealogical study to be the least Jewish.

In this example the fairness achieved by random selection belongs to action *A* rather than *B*. It would certainly be absurd to prefer *B* to *A* because even if we free *B* from its taint of racism there could still be no reason for preferring it to *A*. Suppose we make the story this:

(b) Everything is as in (a) except that action *B* is to ignore the first lottery and hold another to decide whom to treat.

If you were now to prefer *B* to *A* on grounds of randomness, then every time a random selection was made you would want to ignore its result and have another. This is absurd.<sup>1</sup> Indeed, action *A* is actually strictly better than *B* even in the softened version (b), because ignoring the results of the first lottery is tampering with the results of a fair random process.

These examples show that even if we grant that random selection is sometimes fair, that fairness does not necessarily attach to the action that equalises expected utilities. What makes random selection fair, when and if it is, I have considered in a companion paper to this one (Broome 1984). These examples illustrate one of the conclusions I reached there: a necessary condition for random selection to be fair is that the candidates should have roughly equal claims to the good that is to be distributed. To know a candidate's claim you need to know more than the sort of information about utilities and expected utilities that is contained in Table 1. For instance, in example (b)

<sup>1</sup> A referee has pointed out to me that it is not obviously absurd to want to delay the lottery as long as possible. If you were in favour of equalising expected utilities that is presumably what you would want to do. The question addressed by this paper is whether you could actually have a good reason for it. (Remember this must not be seen, in the manner of Deschamps and Gevers described in Section I, as a covert means towards equalising actual utilities.) The question addressed by this section is whether, in particular, the value we attach to random choice can give us a reason for it. As a step towards answering this I am asking now whether the value we attach to random choice can give us a reason to ignore a random choice already made. I claim it is absurd to answer yes to *this* question.

once Michael has won the first lottery he has a perfect claim to the treatment, and Maggie has none. Consequently it is not at all fair to hold a second lottery. But none of that appears in the table.

Table 1, then, is not enough to show which action is the fairer one. And that means the table is incomplete. The theory of choice under uncertainty requires an action to be treated simply as a mathematical function from states of nature to outcomes (Savage 1954, p. 14). Any feature of an action that has significance or value must be written into the description of the outcome. Otherwise (it is easy to see) the sure-thing principle is a lost cause from the start. So to express Diamond's intentions properly – to make sure that *B* really is the fairer action – we must use Table 2.

Table 2  
*Diamond's Example Redescribed*

Actions	States of nature	
	1	2
<i>A</i>	(1, 0) the selection not having been fair	(1, 0) the selection not having been fair
<i>B</i>	(1, 0) the selection having been fair	(0, 1) the selection having been fair

This table shows that *B*'s results, being fairer, are better than *A*'s in both states of nature. And action *B* is better than action *A*, so there is no breach of the sure-thing principle.

There is, however, a breach of welfarism (Sen 1979). Welfarism is the idea that states of affairs are to be judged entirely by the utilities of the people in them. I have just said that in Table 2 *B*'s results are better than *A*'s in both states of nature. Yet in state 1 both people's utilities are exactly the same under both actions.

Do not suppose that the reason *B*'s outcome in state 1 is better than *A*'s can only be because Maggie feels better until the state of nature is revealed. That, as I explained in Section I, is what Deschamps and Gevers supposed, and it negates the point of the example. Any feelings of Michael's and Maggie's must have been taken account of in the utilities when the table was drawn up. The easiest thing is to imagine that the process of selection is kept secret from the candidates.

If we deny welfarism here we do not have to claim that there is a sort of good or bad that is not a good or bad for a person. We may claim instead that there is a sort of good or bad for a person that is not included in his utility. The outcome of *B* in state 1 might be better *for Maggie* than the outcome of *A*, simply because she is more fairly treated. She might be more fairly treated even though her utility is not improved and even though (as I suggested we imagine) she does not know of her fairer treatment. It is by no means absurd to suggest that good or bad may be done a person without his knowing it and without its altering his utility. For instance, if you ridicule a friend of yours behind his back

you are presumably doing him a wrong even if he never finds out and his utility is not reduced (compare Nagel 1979). (Of course it would be possible to alter the meaning of 'utility' so as to include the non-welfarist good of fairness; I have followed this course in another paper (Broome 1983).)

In the argument of this section I have conceded to Diamond his implicit assumption that there are occasions when it is fairest to choose between people randomly. Both John Harsanyi (1975) and J. A. Mirrlees (1982, p. 82), on the other hand, deny this assumption as their way of countering the threat posed by Diamond to their respective utilitarian theories. Mirrlees agrees that there may be times when a government ought to choose randomly, but only if it cannot be trusted, because of partiality or corruption or some other reason, to make the best non-random choice. Making the best non-random choice would be better than choosing randomly, so that an ideally fair government would have no use for random choice. I do not know if this is right; that is the subject of my companion paper (Broome 1984) but in that paper I was not able to reach a definite conclusion. If it is right, then of course random selection does not provide even a *prima facie* case against the sure-thing principle or in favour of equalising expected utilities. It also, notice, does not then provide a case against welfarism; the argument presented above is premised on the assumption that it is sometimes fairest to choose randomly.

### III. FURTHER EXAMPLES

Consider this example, which is adapted from one of John Taurek's (1977). A volcanic island is about to explode. People are waiting to be rescued, some at the North end and some at the South. There are more of them at the North. You have the only boat near enough to help but your boat is too small to pick up even everybody at the South. You are now at sea approaching the island. You have to decide whether to go to the North or the South. When you arrive at either end time will be so short that the people you rescue will have to be selected by lot. Everyone, assume, has an equal claim to be rescued. Assume, too, that the populations at either end are similar in character, so that your contributions to total utility and to equality in utility can be expected to be about the same whichever end you go to. Is there any moral merit in going to the North? If you decide to go to the South because, say, it is cheaper on fuel, have you done anything wrong? Going to the North is the way to spread expected utilities more equally, so if it is true that this is a worthwhile object then that is where you ought to go. But actually I cannot persuade myself that it makes any moral difference which end you go to. If anyone believes that it is valuable to equalise expected utilities, then they must believe that going to the South would be morally blameworthy, and I leave that as a matter between them and their moral intuitions.

I find this artificial example persuasive enough to stand as a counterexample to the claim that equalising expected utility is valuable.<sup>1</sup> My next example, on the other hand, is controversial and contains some hidden complications. I offer

<sup>1</sup> It is discussed more fully in Broome (1984).



it, not as a counterexample, but as a practical *application* of what I have been saying.

Suppose some radioactive waste leaks into the sea and is absorbed by fish. As a result a number of people die of cancer caused by eating the fish. Compare two cases. In one the fish are widely distributed to shops through the country, so the deaths are scattered throughout the population. In the other the deaths are confined to people living in a small area, each of whom is therefore exposed to quite a big risk. The number of deaths is the same in each case. Is one or the other case preferable? The National Radiological Protection Board (Clark *et al.* 1981, particularly pp. 35–8) strongly prefers the deaths to be widely distributed. Indeed it would prefer a larger number of deaths widely distributed to a smaller number narrowly confined. Once more, I cannot persuade myself that this is right. If we are to sacrifice lives to the goal of equalising expected utilities then we should have pretty good reasons for valuing that goal. But I believe we have no good reasons at all.<sup>1</sup>

#### IV. SUMMARY

I have argued that we should not value equality in expected utilities. I showed, first, that valuing equality in utilities gives us no reason to do so. Then I showed that, even if it is sometimes right to choose randomly between candidates for a good, that too is no reason to value equality in expected utilities despite first appearances to the contrary; occasions where random choice may be appropriate are not always those where it equalises expected utilities. During the course of the paper I gave several counterexamples to the view I disagree with, and I ended with a direct practical application of the case I have been arguing.

*University of Bristol*

*Date of receipt of final typescript: February 1984*

#### REFERENCES

- Broome, J. (1978). 'Trying to value a life.' *Journal of Public Economics*, vol. 9, pp. 91–100.  
 — (1982). 'Uncertainty in welfare economics, and the value of life.' In Jones-Lee (1982), pp. 201–16.  
 — (1983). 'Utilitarianism and separability.' Discussion Paper 139/83 of the Department of Economics, University of Bristol.  
 — (1984). 'Selecting people randomly'. *Ethics*, vol. 95, (October).  
 Clark, M. J., Fleishman, A. B. and Webb, G. A. M. (1981). 'Optimisation of the radiological protection of the public.' National Radiological Protection Board, NRPB-R120.

<sup>1</sup> The NRPB's own reason for its conclusion has nothing to do with the arguments about equality in expected utility discussed in this paper, and it seems to me plainly mistaken despite the support it has received from M. W. Jones-Lee (1981) and Joanne Linnerooth (1982). The NRPB's report (Clark *et al.* 1981, p. 32) bases its case on earlier work of Jones-Lee's (1976). Jones-Lee shows that if a person is exposed to greater and greater risks of dying, the money he would require to compensate him for the risk increases more than in proportion to the risk. (As the danger approaches certainty the compensation may well approach infinity.) Consequently, exposing a lot of people to a small risk requires less compensation than exposing a few people to a large one. However, it is plain from Jones-Lee's analysis that the reason for this phenomenon is simply the diminishing marginal utility of money (and the bounded utility of money). As a person is exposed to a greater risk of dying, nothing suggests his *life* becomes more precious; it is just that the medium it is valued in becomes less precious. If the valuation were properly adjusted for the marginal utility of money there would be no temptation to follow the NRPB's line. The extreme end of the NRPB's way of thinking is to count the death of an identified person for infinitely much more than a 'statistical' death, a mistake I have inveighed against before (Broome 1978, 1982).

- Currie, D., Peel, D. and Peters, W. (eds.) (1981). *Microeconomic Analysis*. Croom Helm.
- Deschamps, R. and Gevers, L. (1979). 'Separability, risk-bearing and social welfare judgements.' In Laffont (1979), pp. 145-60.
- Diamond, P. A. (1967). 'Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment.' *Journal of Political Economy*, vol. 75, pp. 765-6.
- Hammond, P. J. (1982). 'Utilitarianism, uncertainty and information.' In Sen and Williams (1982), pp. 85-102.
- (1983). 'Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty.' In Pattanaik and Salles (1983).
- Harsanyi, J. (1955). 'Cardinal welfare, individualistic ethics and interpersonal comparisons of utility.' *Journal of Political Economy*, vol. 63, pp. 309-21.
- (1975). 'Nonlinear social welfare functions: do welfare economists have a special exemption from Bayesian rationality?' *Theory and Decision*, vol. 6, pp. 311-32.
- Jones-Lee, M. W. (1976). *The Value of Life: An Economic Analysis*. Martin Robertson.
- (1981). 'The value of non-marginal changes in physical risk.' In Currie *et al.* (1981), pp. 233-68.
- (ed.) (1982). *The Value of Life and Safety*. North-Holland.
- Laffont, J.-J. (ed.) (1979). *Aggregation and Revelation of Preferences*. North-Holland.
- Linnerooth, J. (1982). 'Murdering statistical lives...?' In Jones-Lee (1982), pp. 229-61.
- Mirrlees, J. A. (1982). 'The economic uses of utilitarianism.' In Sen and Williams (1982), pp. 63-84.
- Myerson, R. B. (1981). 'Utilitarianism, egalitarianism, and the timing effect in social choice problems.' *Econometrica*, vol. 49, pp. 883-97.
- Nagel, T. (1979). 'Death.' In his *Mortal Questions*, Cambridge University Press, pp. 1-10.
- Pattanaik, P. K. and Salles, M. (eds.) (1983). *Social Choice and Welfare*. North-Holland.
- Savage, L. J. (1954). *The Foundation of Statistics*. Wiley, and Chapman and Hall.
- Sen, A. K. (1979). 'Utilitarianism and welfarism.' *The Journal of Philosophy*, vol. 76, pp. 463-89.
- and Williams, B. A. O. (1982). *Utilitarianism and Beyond*. Cambridge University Press.
- Taurek, J. M. (1977). 'Should the numbers count?' *Philosophy and Public Affairs*, vol. 6, pp. 293-316.