AMARTYA SEN

RATIONALITY AND UNCERTAINTY

1. CONSISTENCY AND INTEREST

There are, it can be argued, two dominant approaches to rational choice extensively used in decision theory and economics:

(1) *Internal consistency*: Rational choice is seen, in this approach, simply in terms of internal consistency of choice.
(2) *Self-interest pursuit*: The rationality of choice is identified here with the unfailing pursuit of self-interest.

The two approaches both have fairly straightforward interpretations in choices with *certainty*. The internal consistency approach has been much used in the theory of 'revealed preference', with various 'axioms' of revealed preference serving as conditions of internal consistency of choice (see Samuelson (1947)).[1] In much of modern economic theory, 'rational choice' is seen as no more – and no less – than consistent choice, and a choice function is taken as 'rationalizable' if and only if it is consistent enough to have a binary representation (or, in a more exacting interpretation, representation by an ordering).

The self-interest approach is crucial to the derivation of certain central results in traditional and modern economic theory, e.g., the Pareto optimality of competitive equilibria.[2] The traditional theory of utility provides a seemingly firm basis for the rationality of pursuing one's utility – defined either in terms of Benthamite hedonism of pleasure calculus, or in terms of various formulations of desire-fulfilment. In fact, ambiguities in the concepts of 'utility' and 'preference' have played quite a substantial part in intermediating between self-interest and choice, giving the appearance of tying rational choice firmly to the pursuit of self-interest.[3]

The self-interest approach is some times confounded with the internal

Theory and Decision **18** (1985) 109–127. 0040–5833/85.10

consistency view, through *defining* interest or utility as the binary relation of 'revealed preference' (i.e., the binary relation that can represent the choice function if it satisfies certain conditions of internal consistency). But, obviously, that definitional trick does not establish a correspondence of choice with any independently defined notion of self-interest. There is a world of difference between the claim that a person is trying to pursue his or her self-interest through choice, and the announcement that whatever the person can be seen as maximizing (if such a binary relation does exist[4]) will be *called* that person's utility (or interest). The internal consistency approach and the self-interest approach are fundamentally different.

I would like to argue that neither approach adequately captures the content of rationality. Consider the internal consistency approach first. Take a choice function $C(.)$, assumed to be 'rationalizable' (i.e., 'binary') and let R be the binary relation representing it.[5] Construct the binary relation R^* from R by 'reversing' every strict preference, and let $C^*(.)$ be the choice function generated by (and 'rationalizable' with respect to) R^*. If a person with unchanged non-choice characteristics (i.e., the same feelings, values, tastes, etc.) were to end up choosing in exactly the 'opposite' way in each case, i.e., according to $C^*(.)$ rather than $C(.)$, it would be hard to claim that his or her choices have remained just as 'rational'. But the 'opposite' choices are exactly as consistent!

Rationality has to include some correspondence of choice with *other* characteristics, and it cannot be fully captured by any notion of *internal* consistency – however exacting it may be. In this sense, the internal consistency approach is too permissive (though it may *also* be too restrictive in other ways, if the consistency conditions turn out to be unduly exacting). The self-interest approach is, in contrast, certainly too restrictive. A person need not be involved in any lapse of reasoning or rationality if he or she decides to pursue some goals other than self-interest.[6] People in real life may or may not be entirely self-seeking, but it would be absurd to claim that anyone who does not pursue what he or she recognizes to be his or her own interest *must be* just irrational!

It is arguable that what goes wrong with these two standard approaches to rationality is their failure to pay adequate and explicit attention to the role of reasoning in distinguishing the rational from the irrational. Reasoning may demand more than consistency.[7] (Also, it need not be seen

as requiring – though this is a more debatable point – that consistency must take a binary form.[8]) There is also no convincing ground for insisting that a person's reasoning must be employed only in the pursuit of his or her self-interest. The internal consistency approach can bring in reasoning only indirectly – only to the extent (and in the form) that is allowed by the nature of the consistency conditions imposed. The self-interest approach refuses to admit reasoned choice in pursuit of any goals other than self-interest. Both approaches sell reasoning very short indeed, in characterising rationality.

The view is often expressed that the notion of rationality is quite 'unproblematic' when the object of attention is choice under certainty, and that the difficulties arise from trying to 'extend' the notion of rationality – obvious in the case of certainty – to cases involving uncertainty. I shall argue that this view is hard to defend, and the enormous difficulties of getting a grip on the notion of rationality under *uncertainty* include a great many problems that also arise in characterising rationality in choices *without* uncertainty.

2. REASONING AND CORRESPONDENCE

Rationality must deal with the correspondence of actual choice with the use of reason. There are two distinct types of failures of rationality. A person can fail to do what he would decide to do if he were to reason and reflect on what is to be done. The failure may arise from one of several causes, e.g., (i) the person has acted 'without thinking', (ii) the person has reasoned lazily about what to do and has not used his faculties properly, (iii) the person has reasoned carefully and decided to do x, but has ended up doing y for, say, the weakness of will (what the Greeks called 'akrasia'). All these cases have one point in common, to wit, the person would reject his own choice on careful reflection – there is, in this sense, a failure of positive correspondence between the person's reasoning and his choice. I shall call this 'correspondence irrationality'.[9]

In contrast with 'correspondence irrationality', a person may fail to be rational because of the limited nature of the reasoning of which he is capable. A person may have reflected as carefully as he can on a choice, but not seen something significant that a sharper reasoning would have

revealed. I shall call this 'reflection irrationality'. In the case of 'correspondence irrationality', the person fails to do the right thing as he himself sees it (or would have seen it if he had carefully reflected on the matter), whereas with 'reflection irrationality' the person fails to see that the objectives he wishes to pursue would have been better served by some other choice (on the basis of the information he has).

To illustrate, take the case of Buridan's ass, which died of starvation dithering between two haystacks both of which looked alluring. Was the ass irrational? We can't, of course, know whether it was or not without knowing more about the story. Perhaps it was an extremely noble and 'do-gooder' ass, committing suicide to leave the haystacks for other asses, and pretending to dither to avoid embarrassing the other asses? If so, Buridan's ass may have been far from irrational (even though members of the 'self-interest' school of rationality would not see this).

Let us assume, however, that the ass did indeed wish to live and was not intending to bequeath the haystacks to other asses. Why didn't it choose one of the haystacks, then? Did it fail to see that touching neither haystack and dying of starvation was the worst of the three alternatives, no matter how it ranked the haystacks? If it saw this and was still paralysed (say, by greed), or – alternatively – *would have* seen it if it had reflected carefully but did not so reflect (say, because of nervousness), then this is a case of 'correspondence irrationality'. Another possibility is that the ass would not have been *able* to figure this out at all (i.e., to see that even if it could not decide which of the two haystacks was the larger, it was still sensible to choose either of then rather than neither).[10] If this was the case, then this exemplifies 'reflection irrationality'. Perhaps the ass had read too much 'revealed preference' theory, and felt unable to choose x when y was available without being sure that x was superior to (or even at least as good as) y, and – relevantly for the 'weak axiom' – without being sure that it would never choose y in the presence of x.

Both these issues of rationality are deeply problematic in the sense that it is not easy to find simple criteria that will diagnose rationality or irrationality of either type in a decisive way. 'Correspondence rationality' involves the use of counterfactuals (what the person would have decided on careful reflection). While social science is hard to do without counterfactuals,[11] the no-nonsense operationalist dreads the excursion into 'what

would have happened if . . . '. Similarly, it is not easy to be sure how much reasoning to demand in diagnosing 'reflection irrationality'. For example, is a choice 'reflection irrational' if the person chose wrongly because he was unable to figure out (relevantly for his choice of action) a hard mathematical puzzle the solution of which was 'contained' – analytically – in the problem itself. Where do we draw the line?

I should make it absolutely clear that I do not regard it as at all embarrassing to the approach I am presenting here that decidability is a problem for both 'correspondence irrationality' and 'reflection irrationality'. Quite the contrary. My claims include: that the notion of rationality involves inherent ambiguities; that the decidability problems of correspondence rationality and reflection rationality merely make these ambiguities clear; that many of the sources of ambiguities are present with or without uncertainty; that the standard approaches to rationality avoid these ambiguities (insofar as they do avoid them) by misspecifying the problem of rationality. I would also argue that to try to jettison all the ambiguities of rationality and to aim at a sure-fire test that will work in every case would tend to take us away from the reasons that make rationality an important concept. The partial undecidabilities of rationality are, in fact, part and parcel of my thesis.

Decidability problems do not make a concept useless. The indentification of many unambiguous cases may well be both easy and useful. Indeed, the belief – often implicit – that a satisfactory criterion must be a 'complete' one has done, it can be argued, a good deal of harm in the social sciences by forcing us to choose between groundless defeatism and arbitrary completion.

I have tried to argue the case for systematically accommodating 'incompleteness' in such contexts as interpersonal comparison of utilities, measurement of inequality, making real income comparison, quantifying poverty, and measuring capital.[12] A similar approach may be useful in dealing with rationality. There will be clear cases of 'correspondence irrationality', where the person himself accepts unambiguously that he would have chosen differently had he bothered to think at all on the matter. There are clear cases also when 'correspondence irrationality' is caused by the 'weakness of will' despite the person having made a reasoned decision to do something else.

Similarly, though there may be doubts about how much reasoning to incorporate in the standards of 'reflection irrationality', some cases are clear enough. It is known that people learn techniques of decision making with practice. Indeed, one major objective of decision theory has been to improve people's ability to reason about decisions.[13] There may be great difficulties on drawing an exact line, but it may be easy enough to agree that some cases involve obvious reasoning failures of an uncomplicated kind, and which can very easily be avoided with just a little training.

3. UNCERTAINTY AND REASONING

Having outlined an approach to the problem of assessing rationality of choice, I should make a few remarks on the contrasts with other approaches. The differences with the approaches of 'internal consistency' and 'self-interest' in their pure forms must be obvious enough. But some approaches are more complex.

John Harsanyi (1978) presents his "rational-choice models of social behaviour" by noting that his theory "is a normative (*prescriptive*) theory" and that "formally and explicitly it deals with the question of how each player *should* act in order to promote his own interests most effectively" (p. 16). One obvious difference between our approach and Harsanyi's lies in his apparent concentration on the person promoting 'his own interests' (rather than any other goals that he may have). But this may not be a major problem here, since much of Harsanyi's analysis can be reinterpreted in terms of pursuit of general goals – subject to certain formal restrictions – rather than only the particular goal of self-interest maximization.

A second difference, which is more fundamental, arises from Harsanyi's firmly prescriptive motivation, and this relates ultimately to seeing decision-theoretic recommendations as consistency conditions that any person *must* obey to make sense of his practice. In contrast 'correspondence rationality' is not prescriptive, and 'reflection rationality' is only conditionally prescriptive.

To illustrate the contrast – at the risk of being a little *ad hominem* – both Allais' own response to the choice in the paradox that bears his name and Savage's well-known first-blush response (in the same lines as Allais')

are simply 'irrational' in Harsanyi's framework since they violated the condition of 'strong independence' which is seen as a 'prescriptive requirement of rationality'. In contrast, in our framework of 'correspondence rationality', Allais' choices were *not* 'correspondence irrational'; he did defend his choice by reasoned reflection and has continued to do so.[14] On the other hand, Savage's choices were clearly 'correspondence irrational', and he did in fact reject his first-blush choices after reasoned reflection about the implications of his choices.

Regarding 'reflection irrationality', there is more of a problem of decidability. But if anyone does claim that Allais' reasoning regarding these choices are 'erroneous', he has to show why the apparent justification is not 'really' acceptable. The issue of reflection rationality in this case may well be an important one to pursue, but that is a very different exercise from simply insisting on strong independence as a consistency condition. I shall take up that question for a closer examination in the next section.

The 'internal consistency' approach has been used powerfully, in analysing decision making under uncertainty, in many rational decision models.[15] Some – like the von Neumann-Morgenstern utility model – have been successful both in raising important questions about rational behaviour under uncertainty and also – as Harsanyi (1978) notes – in "explaining or predicting real-life human behaviour" (p. 16).[16] The latter question – that of explanation or prediction of *actual* behaviour – involves a somewhat different issue from that of rationality – a distinction that is especially important in the context of interpreting various 'obviously irrational' psychological responses found in experimental research by Kahneman, Slovik, Tversky and others.[17]

As far as rationality is concerned, the difficulties with the internal consistency approach in the case of decisions under uncertainty are not radically different from those in the case of certainty. A person can be internally consistent and still be doing the opposite of the things he should obviously do to pursue his own goals. As was discussed earlier – no test of internal consistency, however stringent, can deal with this problem. Also, on reasoned reflection a person might revise his choices substantially, even though the first-blush responses had satisfied all the conditions of internal consistency. It should be clear that whether or not these consistency conditions are necessary for rationality, they can scarcely be *sufficient* for it.

The issue of *necessity* raises problems similar to those faced in the context of choice under certainty, but with greater force. 'Why binary choice?' has now to be supplemented by such questions as 'Why strong independence?' These are certainly matters for reasoning *for* and *against*. This, in turn, leads to possible applications of the concepts of 'correspondence rationality' (involving 'self-policing') and 'reflection rationality' (involving a host of issues from decision-theoretic training to 'agreeing to disagree').

4. INDEPENDENCE AND RATIONALITY

The rationality axiom for choice under uncertainty that has caused the most controversy is almost certainly the condition of strong independence. One of several versions of this condition demands that a lottery L^1 will be preferred to a lottery L^2 if and only if, for any lottery L^3, the combined lottery $(pL^1, (1 - p)L^3)$ will be preferred to the combined lottery $(pL^2, (1 - p)L^3)$ for all probability numbers p. Mixing each lottery with a third one – in the same proportion in the two cases – does not change the ranking. It was this axiom that was clearly violated by Allais' famous counterexample, and it has been the subject of several other interesting counterexamples as well.

The strong independence axiom is indeed crucial to the expected utility approach. Given this axiom, the linear form of evaluation is pretty much unavoidable in choosing between lotteries, since the other axioms needed (including conditions of complete ordering and a mild condition of continuity) are not particularly exacting.[18] The battle of expected utility has been largely fought on the field of independence. While strong independence has appeared to some to be self-evidently a necesary condition of rationality – indeed of internal consistency – it certainly does need a detailed defence. Violating it is not obviously silly in the way in which the behaviour of Buridan's ass clearly is. If an 'error' is being made, it is a less immediate one, and more must be said on this than asserting that strong independence is self-evidently essential for reasoned choice.

One approach, among others, in defence of expected utility (including strong independence) that has persuasive features is Peter Hammond's (1982) derivation of expected utility from what he calls – taking a little liberty – 'consequentialism'. In Hammond's characterisation,

'consequentialism' requires that acts be chosen exclusively on the basis of choosing from the 'feasible set of contingent consequences' – and these reflect 'prizes' with the overall uncertainties specified. Adding some continuity, Hammond gets home to expected utility on the basis of 'probability consequentialism', with the uncertainty specified in terms of probabilities. The operative choices here are confined to 'consequence lotteries' and the choice of acts follow from that.

Hammond's argument is interesting and important, but it is not adequate (nor is it claimed to be so) for establishing exclusive reasonableness of expected utility. Part of the difficulty arises from limitations of consequentialist reasoning that have received much attention in recent years in moral philosophy (see Williams (1973, 1982), Nagel (1980), Parfit (1984)). But the property defined by Hammond is, in some important respects, even more demanding than traditional consequentialism. The main 'consequentialist' approach in moral philosophy has been based on the utilitarian view, which has involved restricting attention to the 'utilities' of the persons in question in the consequent states of affairs.[19] In Hammond's formulation, these mental attitudes do not figure at all, and true to the tradition of von Neumann-Morgenstern-type 'expected utility', 'utilities' are determined *by the choices* over lotteries rather than the other way round. This has, of course, been a bone of contention between Allais and his followers on one side, who have preferred to start with a psychological cardinal utility that influences choice over the lotteries.[20] The issue is of decisive importance since the consideration of 'could have been' outcomes can influence the contingent choice over lotteries through affecting the person's happiness and other psychological features.

This is, of course, the door that opens on to old arguments on such subjects as the relevance of 'regret' (e.g. 'minimax regret', or newer theories due to Bell (1982), Loomes and Sugden (1982), and others), which the 'expected utility' theorist tends to see as red herrings. There is some scope for genuine confusion about two distinct issues related to such matters as 'regret'. The question of rationality of 'regretting' has to be distinguished from the question of the rationality of taking note of regret if it were to occur. Even if it is the case that it is irrational for me to regret something that cannot be changed, if nevertheless I am willy nilly doomed to regretting the thing in question, then I must take note of that *fact* of regretting.[21]

Aside from the psychological problems involved in this issue, there are further considerations that question the entire consequentialist perspective, e.g., the relevance of agency (*who* took *what* decision). Information on this is lost in the 'consequence lotteries', which do not distinguish between the path through a 'decision node' as opposed to a 'chance node' so long as the consequences are the same. There is a more information-preserving way of characterising 'consequential reasoning' that will permit such considerations to be included in 'consequential reasoning' (see Sen 1982b, 1983), but for that we must go beyond 'consequential lotteries'.

I would argue that the condition of strong independence is deeply questionable from either of these two perspectives: (1) psychology sensitivity, and (2) agency sensitivity. To this we can add a third, viz., (3) information sensitivity. The information that a person gathers about prizes and uncertainty does, of course, get reflected in the specifications of 'consequence lotteries', but the valuation that a person attaches to the consequences may well depend on things about which a person learns more by considering what lotteries he is, in fact, facing. There is an odd asymmetry in the traditional 'expected utility' story whereby the observer (such as the decision analyst) learns about the chooser's preferences by observing his decisions, but the chooser does not use the *nature of the lotteries* that he faces to learn about the nature of the world, which may affect his valuation of consequences and thus his choices. To be sure, there is no formal restriction on such learning, but once such learning is systematically admitted, some of the axioms of expected utility (including 'strong independence') becomes difficult to sustain. As lotteries are combined with others, the determinants of the person's valuation of the states and acts can change, even within a broadly consequentialist framework.

Some of the 'counterexamples' to expected utility and its axioms (including 'strong independence') that have been offered in the literature (e.g., Allais' (1953), Machina's (1981), Tversky's (1975)) can be seen in the light of these three conditions, in particular the first two (psychology sensitivity and agency sensitivity).[22]

I suggest three other 'counterexamples' below.

Case I. *The No-letter Response*

You come home after the day's work and check your mail. You may

possibly have won a prize in the national draw (you think with probability p), in which case you would find a letter waiting for you. If no letter, you would choose to do something useful like painting the garbage can, which needs doing some time. In another case, there is the possibility (you think with probability p) of your finding a court summons for a motoring incident – the policeman was vague and the last day for summoning you will pass tonight. If you find no letter, you would like to open a bottle of champagne and enjoy yourself, rather than painting the garbage can. The significance of the no-letter situation depends on what *could have been*, but hasn't (cash prize in one case, court summons in the other).
So your preferences are the following:

p, win cash, no summons; $1-p$, no win, no summons, paint garbage can	*preferred to*	p, win cash, no summons; $1-p$, no win, no summons, drink champagne

and

p, no cash win, summons received; $1-p$, no win, no summons, drink champagne	*preferred to*	p, no cash win, summons received $1-p$, no win, no summons, paint garbage can

The same case of 'no letter' – implying 'no win, no summons' – is read differently depending on whether the alternative expectation was for a cash win, or for getting a summons (depending on the nature of the lottery with which the decision regarding drinking champagne and painting garbage can is combined).

You have violated strong independence all right, and you must prepare to face the 'expected utility' lot.[23] But if you don't change your mind on further reflection (showing no sign of 'correspondence irrationality'), you will not get the big stick of 'reflection irrationality' from us.

Case 2: *The Doctor's Dilemma*

Dr. Chang faces the problem that he is in a remote rural area, facing two critically ill persons, and with just one unit of the medicine that can possibly help cure each. If administered to Hao, there is – Dr. Chang believes – a 90 per cent chance of cure for Hao. If given to Lin there is,

Dr. Chang believes, an even higher chance of cure – he thinks around 95 per cent. If the medicine is divided between the two neither will be cured. Faced with the need for an unequivocal choice between the two ('please say who'), Dr. Chang would decide to give the medicine to Lin. But when he is given the option of a 50–50 chance mechanism (either directly or indirectly through the choices of other doctors), he opts for that lottery over either of the two certain strategies. That is, he chooses trivial lottery L^1 = (0, Hao; 1, Lin) over trivial lottery L^2 = (1, Hao; 0, Lin), but chooses (0.5, Hao; 0.5, Lin) over (0, Hao: 1, Lin), which is equivalent to (0.5, L^1; 0.5, L^2) being chosen over (0.5, L^1; 0.5, L^1).

The violation of strong independence and expected utility may be due to a sense of fairness in the treatment of Hao and Lin (not ignoring Hao just because he has a somewhat lower probability of cure, though it is very high anyway).[24] But it may also be due to Dr. Chang's dislike of having to make the choice himself between Hao and Lin, 'condemning' – as it were – one of them to death. Dr. Chang may, in fact, even prefer that the lottery be won by Lin, who has a somewhat higher probability of cure, but nevertheless prefer to have the genuine lottery over simply giving the medicine to Lin, ignoring Hao's claims altogether. The agency of the actual choice – whether Dr. Chang has to *name* one of two to be saved (and the other left to die) – may make a difference to him. Whether Dr. Chang is morally right to prefer the lottery is, of course, a debatable issue (there are arguments on both sides), but certainly it is very hard to claim that Dr. Chang is being straightforwardly irrational in being 'agency sensitive'.

Case 3: *Deportation Information*

Ayesha – an immigrant to the United Kingdom – is wondering whether to become a civil rights lawyer or a commercial lawyer in her choice of career. Given that simple choice, she would be inclined towards the latter, i.e., commercial law practice. But she learns that since there were some minor technical irregularities in her immigration papers (and since she comes from what is politely called the 'new' Commonwealth countries, as opposed to white ones), she has about a 50 per cent chance of being simply deported from the U.K. rather than doing either kind of practice there. She decides that if these are the prospects *and* if – in the event – she is

not deported, then she will prefer after all to be a civil rights lawyer. However, everything in the real world (except in her mind) will be exactly the same if she is not deported as it would have been if that issue had not been raised at all. Is she being irrational in violating strong independence?

Ayesha's choices can be given reasoned support in line with 'psychology sensitivity', rather like in the case of 'the No-letter Response'. She could also believe that she has some 'responsibility' now to concentrate on civil rights issues having become involved in one herself, at the receiving end. But I don't want to pursue either of these lines here. (I assume that Ayesha is psychologically unaffected and also does not accept any special moral responsibility by virtue of facing the prospects of her own deportation.) But the very fact of her facing the probability of deportation herself may give her more *knowledge* of the issue of immigration and of the problems faced by immigrants. The world is no different, but her understanding of it is not unaffected by the uncertainty she herself faces regarding deportation. Her contingent preference reflects her greater understanding of the realities of the U.K. immigration policy and of the nature of the civil rights problem, if she faces the prospect of deportation herself.

If the nature of the uncertainties faced affects a person's knowledge and if this affects the person's *valuation* of the outcomes (without changing the "outcomes", as they are defined in this literature), then the axiomatic requirements of expected utility models may well be seriously compromised.

5. CONCLUDING REMARKS

Some of the main points of this paper can be briefly put together.

(1) The two standard approaches to 'rational choice', viz., 'internal consistency' and 'self-interest pursuit', are both deeply defective.

(2) The view that the problem of rationality is 'unproblematic' for choice under certainty, with difficulties arising only with uncertainty, is mistaken. Many serious difficulties are present whether or not uncertainty is faced by the chooser.

(3) The problem of rational choice can be split into two different types of problems, which are respectively called here 'correspondence rationality' and 'reflection rationality'.

(4) 'Correspondence irrationality' is a matter of failure of correspondence between the person's reasoned reflection and his actual choices. The failure can arise from various causes, e.g., (i) acting 'without thinking', (ii) 'lazy' reflection, and (iii) 'weakness of will'.

(5) 'Reflection irrationality' is a matter of failure of careful reflection. Despite reflecting carefully, connections may be missed and relevant considerations ignored because of intellectual limitations, possibly due to lack of training on decision problems.

(6) Both 'correspondence rationality' and 'reflection rationality' have serious decidability problems. This is no embarrassment to the approach to rationality suggested in this paper. The notion of rationality involves inherent ambiguities, and the decidability problems of 'correspondence rationality' and 'reflection rationality' relate to these basic ambiguities. Sensible criteria of checking a property cannot lead to complete and clear-cut answers when the property itself includes ambiguities. There is a strong case for systematically admitting incompleteness in rationality judgements, separating out clear cases of irrationality (of either type) from others.

(7) The approach of 'expected utility' raises interesting issues of 'reflection rationality'. The axioms used (including 'strong independence') and the demands of 'probability consequentialism' both help to bring out the main contentious issues in the 'expected utility' approach. While the approach has much appeal, there are serious arguments *against* as well. The problem of 'reflection rationality' has genuine ambiguities in dealing with violations of strong independence and probability consequentialism.

(8) Three different arguments for violating strong independence were identified and distinguished, viz., (1) psychology sensitivity, (2) agency sensitivity, and (3) information sensitivity. These arguements can be used to explain reasoned violations of the axioms of expected utility in some of the counterexamples that have been presented in the literature.

(9) Three counterexamples to the reasonableness of strong independence were presented, called respectively, (1) 'The No-letter Response', (2) 'The Doctor's Dilemma', and (3) 'Deportation Information'. The first illustrates 'psychology sensitivity', and the second 'agency sensitivity'; whereas the third can be seen as exemplifying either 'psychology sensitivity' or 'information sensitivity'.

(10) Finally, rational choice is a matter of the correspondence of choice to the person's reasoning and of the quality of that reasoning. While both questions are hard to deal with, they have to be explicitly faced. To try to avoid these questions either by externally imposing specific objectives and substantive rules (e.g., self-interest maximization), or by imposing conditions of internal consistency (e.g., binariness, strong independence), amounts to losing important dimensions of the problem of rationality of choice. No set of internal consistency conditions – however stringent – can be *sufficient* for the rationality of choice. Nor – it appears – can the usual consistency conditions be seen as *necessary*. Rationality deserves a less mechanical approach.

NOTES

[1] See also Arrow (1959), Richter (1971), Sen (1971), Herzberger (1973).

[2] See Arrow (1951b), Debreu (1959), Arrow and Hahn (1971). These results require *actual* behaviour to be self-interest maximizing, and this involves the further assumption that actual behaviour is also 'rational' (seen as self-interest maximization).

[3] See Sen (1973) for a critique; also Sen (1982a).

[4] See Arrow (1959), Sen (1971), Herzberger (1973).

[5] See Richter (1971), Sen (1971), Suzumura (1976).

[6] See Nagel (1969), Sen (1973, 1977a), Hirschman (1982), Margolis (1982), Akerlof (1983), Schelling (1984), and Schick (1984).

[7] In an illuminating review article, Mark Machina (1981) remarks: "It is not irrational, for example, to hate asparagus." It certainly isn't (though what rotten luck!). However, it would be difficult to take as rational the person who hates asparagus but continues eating it nevertheless, without being able to provide any convincing reason for choosing what he hates (e.g., seeking some particular vitamins present in asparagus, or facing a threat of being murdered by an asparagus-maniac gang if he does not eat 'the good vegetable'). As formulated here, the issue of rationality of choice is connected with the correspondence of choice with reasoning and the quality of that reasoning. In the context of certainty, Machina sees rationality as 'transitivity' of the person's preference.

[8] The reasonableness of choices being 'binary' has been differently assessed in Arrow (1951a), Sen (1970a, 1977a), Schwartz (1972), Fishburn (1973), Herzberger (1973), Plott (1973), Kanger (1976), Campbell (1975), Suzumura (1983), Sugden (1985).

[9] I have discussed the motivational issues underlying 'correspondence rationality' in Sen (1984b).

[10] An alternative reading – perhaps even the most frequent reading – of the problem of Buridan's ass makes it *indifferent* between the two haystacks (rather than seeing it as unable to decide which one was preferable). In this case the ass should have even less problem in choosing either haystack (with a guarantee of maximization no matter which of the two haystacks it chose).

[11] See Elster (1978).

[12] In various papers, reproduced in two selections, Sen (1982a, 1984a).
[13] See Raiffa (1968) and Keeney and Raiffa (1976).
[14] See also Allais and Hagen (1979) and Stigum and Wenstøp (1983).
[15] For an illuminating review, see Fishburn (1981).
[16] See also Arrow (1970).
[17] See especially Kahneman, Slovik and Tversky (1983). For a challenging defence of the rationality of some of the alleged irrationalities of observed psychology, see Cohen (1983). See also Jeffrey (1965), Levi (1974, 1982), Arrow (1982, 1983), Gärdenfors and Sahlin (1982), Machina (1983), McClennen (1983), among other contributions.
[18] The independence condition is strictly necessary for global linearity (i.e., fixed utilities), but can be dispensed with for more permissive 'expected utility analysis' with 'local utilities' (locally linear coefficients for weighting the probabilities); see Machina (1982).
[19] I have tried to argue that even with consequentialism, this concentration on 'utility consequences' *only* is a further severe limitation of the utilitarian approach; see Sen (1979).
[20] For an illuminating analysis of the distinction between the 'actual psychological reality' of a person's feelings about the choices (e.g., Allais'), and the 'psychological values' assigned by the expected utility procedure, see Machina (1981).
[21] Only an upper-class Englishman properly brought up by a strict nanny can believe that if a person decides that some psychological attitude is not sensible, then it certainly can be prevented from occurring.
[22] See also MacCrimmon (1968), Drèze (1974), Allais and Hagen (1979), McClennen (1983), Stigum and Wenstøp (1983).
[23] An alternative way of dealing with the case is to allow your 'disappointment' (at not getting the cash prize) or 'relief' (not 'getting' the summons) to enter the description of the states of affairs or outcomes, but this goes against the approach of 'expected utility' and also makes 'strong independence' a rather vacuous restriction. A third possibility is to assume that the person does not *know* what the alternative outcomes might be (i.e., does not known whether a cash prize is expected, or a summons may be coming). However, to combine this ignorance with rational decision making over lotteries, we would have to assume that the person forgets what the nature of the lotteries (and the prizes) are, *after* taking his decisions. Independence cannot be easily rescued by any of these 'cunning' tricks.
[24] Cf. Diamond (1967), Sen (1970a), and Broome (1984), for a somewhat different case with symmetric individual positions.

REFERENCES

Akerlof, G. (1983). 'Loyalty Filters,' *American Economic Review*, 73.
Allais, M. (1953). 'Le Comportement de l'Homme Rational devant le Risque: Critique de Postulates et Axiomes de l'Ecole Américaine,' *Econometrica*, 21.
Allais, M., and Hagen, O., eds, (1979). *Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of Decisions under Uncertainty with Allais' Rejoinder* (Dordrecht: Reidel).
Arrow, K. J. (1951a). *Social Choice and Individual Values* (New York: Wiley, 2nd edition, 1963).
Arrow, K. J. (1951b). 'An Extension of the Basic Theorems of Welfare Economics,' in J. Neyman, ed., *Proceedings of the 2nd Berkeley Sumposium of Mathematical Statistics* (Berkeley, Calif.: University of California Press).

Arrow, K. J. (1959). 'Rational Choice Functions and Orderings,' *Economica*, 26.
Arrow, K. J. (1970). *Essays in the Theory of Risk-Bearing* (Amsterdam: North-Holland).
Arrow, K. J. (1982). 'Risk Perception in Psychology and Economics,' *Economic Inquiry*, 20.
Arrow, K. J. (1983). 'Behaviour under Uncertainty and Its Implications for Policy,' in Stigum and Wenstøp (1983).
Arrow, K. J. and Hahn, F. H. (1971). *General Competitive Analysis* (San Francisco: Holden-day; republished North-Holland, Amsterdam, 1979).
Bell, D. E. (1982). 'Regret in Decision Making under Uncertainty', *Operations Research*, **30**.
Borch, K., and Mossin, J. (1968). *Risk and Uncertainty* (London: Macmillan).
Broome, J. (1984). 'Uncertainty and Fairness', *Economic Journal*, **94**.
Campbell, D. E. (1976). 'Democratic Preference Functions,' *Journal of Economic Theory*, 12.
Chipman, J. S., Hurwicz, L., Richter, M. K., and Sonnenschein, H. F., eds., *Preference, Utility and Demand* (New York: Harcourt).
Cohen, J. (1982). 'Are People Programme to Commit Fallacies? Further Thoughts about Interpretation of Experimental Data on Probability Judgement,' *Journal of the Theory of Social Behaviour*.
Davidson, D., Suppes, P., and Siegel, S. (1957). *Decision Making: An Experimental Approach* (Standford: Stanford University Press).
Debreu, G. (1959). *A Theory of Value* (New York: Wiley).
Diamond, P. (1967). 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility: A Comment,' *Journal of Political Economy*, 75.
Drèze, J. H. (1974). 'Axiomatic Theories of Choice, Cardinal Utility and Subjective Probability: A Review,' in J. H. Drèze, ed., *Allocation under Uncertainty: Equilibrium and Optimality* (London: Macmillan).
Edwards, W., and Tversky, A., eds. (1967). *Decision Making* (Harmondsworth: Penguin Books).
Elster, J. (1978). *Logic and Society* (New York: Wiley).
Fishburn, P. C. (1973). *The Theory of Social Choice* (Princeton, N. J.: Princeton University Press).
Fishburn, P. C. (1981). 'Subjective Expected Utility: a Review of Normative Theories,' *Theory and Decision*, 13.
Gärdenfors, P., and Sahlin, N. -E. (1982). 'Unreliable Probabilities, Risk Taking and Decision Making,' *Synthese*, 53.
Hammond, P. J. (1976). 'Changing Tastes and Coherent Dynamic Choice,' *Review of Economic Studies*, 43.
Hammond, P. J. (1982). 'Consequentialism and Rationality in Dynamic Choice under Uncertainty,' Technical Report 387, Institute for Mathematical Studies in the Social Sciences, Stanford University.
Harsanyi, J. C. (1966). 'A General Theory of Rational Behaviour in Game Situations,' *Econometrica*, 34.
Harsanyi, J. C. (1977). *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations* (Cambridge: Cambridge University Press).
Herzberger, H. G. (1973). 'Ordinal Preference and Rational Choice,' *Econometrica*, 41.
Hirschman, A. O. (1982). *Shifting Involvements* (Princeton: Princeton University Press).
Jeffrey, R. C. (1965). *The Logic of Decision* (New York: McGraw-Hill).
Kahneman, D., and Tversky, A. (1979). 'Prospect Theory: An Analysis of Decisions under Risk,' *Econometrica*, 47.

Kahneman, D., Slovik, P., and Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press).
Kanger, S. (1976). 'Choice Based on Preference,' mimeographed, Uppsala University.
Keeney, R. L., and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (New York: Wiley).
Levi, I. (1974). 'On Indeterminate Probabilities,' *Journal of Philosophy*, 71.
Levi, I. (1982). 'Ignorance, Probability and Rational Choice,' *Synthese*, 53.
Loomes, G., and Sugden, R. (1982). 'Regret Theory: An Alternative Theory of Rational Choice', *Economic Journal*, **92**.
Luce, R. D., and Raiffa, H. (1957). *Games and Decisions* (New York: Wiley).
MacCrimmon, K. R. (1968). 'Descriptive and Normative Implications of Decision Theory Postulates,' in Borch and Mossin (1968).
Machina, M. (1981). '"Rational' Decision Making vs. 'Rational' Decision Modelling?" *Journal of Mathematical Psychology*, 24.
Machina, M. (1982). '"Expected Utility' Analysis without the Independence Axiom," *Econometrica*, 50.
Machina, M. (1983). 'Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom,' in Stigum and Wenstøp (1983).
McClennen, E. F. (1983). 'Sure-Thing Doubts,' in Stigum and Wenstøp (1983).
Margolis, H. (1982). *Selfishness, Altruism and Rationality* (Cambridge: Cambridge University Press).
Nagel, T. (1970). *The Possibility of Altruism* (Oxford: Claredcn Press).
Nagel. T. (1980). 'The Limits of Objectivity,' in S. McMurrin, ed., *Tanner Lectures on Human Values* (Cambridge: Cambridge University Press).
Parfit, D. (1984). *Reasons and Persons* (Oxford: Claredon Press).
Plott, C. (1973). 'Path Independence, Rationality and Social Choice,' *Econometrica*, 41.
Raiffa, H. (1968). *Decision Analysis* (Reading, Mass.: Addison-Wesley).
Ramsey, F. P. (1931). 'Truth and Probability,' in F. P. Ramsey, *The Foundations of Mathematics and other Logical Essays* (London: Kegan Paul).
Richter, M. K. (1971). 'Rational Choice,' in Chipman, Hurwicz, Richter and Sonnenschein (1971).
Savage, L. J. (1954). *The Foundations of Statistics* (New York: Wiley).
Samuelson, P. (1947). *The Foundations of Economic Analysis* (Cambridge, Mass.: Harvard University Press).
Schelling, T. C. (1984). 'Self-Command in Practice, in Policy, and in a Theory of Rational Choice,' *American Economic Review*, 74, Papers and Proceedings.
Schick, F. (1984). *Having Reasons: An Essay on Rationality and Sociality* (Princeton: Princeton University Press).
Schwartz, T. (1972). 'Rationality and the Myth of the Maximum,' *Nous*, 7.
Sen, A. K. (1970a). *Collective Choice and Social Welfare* (San Francisco: Holden-Day; republished by North-Holland, Amsterdam, 1979).
Sen, A. K. (1970b). 'Interpersonal Aggregation and Partial Comparability,' *Econometrica*, 38; "A Correction," *Econometrica*, 40 (1972).
Sen, A. K. (1971). 'Choice Functions and Revealed Preference,' *Review of Economic Studies*, 38.
Sen, A. K. (1973). 'Behaviour and the Concept of Preference,' *Economica*, **40**.
Sen, A. K. (1977a). 'Social Choice Theory: A Re-examination,' *Econometrica*, 45.

Sen, A. K. (1977b). 'Rational Fools: A Critique of the Behavioural Foundations of Economic Theory,' *Philosophy and Public Affairs*, 6.
Sen, A. K. (1979). 'Utilitarianism and Welfarism,' *Journal of Philosphy*, 76.
Sen, A. K. (1982a). *Choice, Welfare and Measurement* (Oxford: Blackwell, and Cambridge, Mass.: Harvard University Press).
Sen, A. K. (1982b). 'Rights and Agency,' *Philosphy and Public Affairs*, 11.
Sen, A. K. (1983). 'Evaluator Relativity and Consequential Evaluation,' *Philosophy and Public Affairs*, 12.
Sen, A. K. (1984a). *Resources, Values and Development* (Oxford: Blackwell, and Cambridge, Mass.: Harvard University Press).
Sen, A. K. (1984b). 'Rationality, Interest and Identity,' written for a festschrift for A. O. Hirschman.
Simon, H. A. (1957). *Models of Man* (New York: Wiley).
Sugden, R. (1985). 'Why Be Consistent? A Critical Analysis of Consistency Requirements in Choice Theory,' *Economica*, **52**.
Suzumura, K. (1976). 'Rational Choice and Revealed Preference,' *Review of Economic Studies*, 43.
Suzumura, K. (1983). *Rational Choice, Collective Decisions and Social Welfare* (Cambridge: Cambridge University Press).
Stigum, B. P., and Wenstøp, F., eds. (1983). *Foundations of Utility and Risk Theory with Applications* (Dordrecht: Reidel).
Tversky, A. (1975). 'A Critique of Expected Utility Theory: Descriptive and Normative Considerations,' *Erkenntnis*, 9.
Tversky, A., and Kahneman, D. (1974). 'Judgement under Uncertainty: Heuristics and Biases,' *Science*, 185.
von Neumann, J., and Morgenstern, O. (1947). *Theory of Games and Economic Behaviour* (Princeton: Princeton University Press).
Williams, B. (1973). 'A Critique of Utilitarianism,' in J. Smart and B. Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press).
Williams, B. (1982). *Moral Luck* (Cambridge: Cambridge University Press).

All Souls College,
Oxford,
England