Paper Presented at HESG, Nottingham, July 2000

DRAFT COPY – PLEASE DO NOT QUOTE

Testing Rank-Dependent Utility Theory in the Context of Health Care

Adam J. Oliver
Office of Health Economics*

Abstract

Systematic violations of expected utility theory (EU) in the context of health care have been reported, potentially undermining the validity of the standard gamble (SG). Rank-dependent expected utility theory (RDU) is currently the most influential descriptive alternative to EU. This paper reports a test of the exact demarcation between EU and RDU – that of the weakening of the independence axiom to comonotonic independence – in the context of health care. The results report an insignificant number of violations of independence. Moreover, there were more violations of comonotonic independence than of non-comonotonic independence, though the difference was not significant. Therefore, the alternative hypothesis that RDU offers a descriptive improvement over EU is rejected on the basis of the experiment reported in this paper.

* Office of Health Economics, 12 Whitehall, London SW1A 2DY, UK; Tel: +44 (0)20 7930 9203 x1467; Fax: +44 (0)20 7747 1419; E-mail: aoliver@ohe.org

1. Introduction

The standard gamble (SG) is widely accepted among health economists as the gold standard for health state utility elicitation. The reason for this is that health care decisions invariably involve a degree of risk; unlike other often used methods for utility elicitation, SG is implied from the axioms of expected utility theory (EU) and is thus firmly based on a theory of risk and uncertainty.

EU was developed during the Second World War (von Neumann and Morgenstern, 1944) and has subsequently been accepted into mainstream economics. However, the descriptive validity of its axioms has been subjected to increasing criticism (e.g., Allais, 1953; Slovic and Tversky, 1974; Kahneman and Tversky, 1979; MacCrimmon and Larsson, 1979; Machina, 1987; Conlisk, 1989; Camerer and Weber, 1992; Tversky and Kahneman, 1992; Camerer and Ho, 1994). The most controversial aspect of EU is the independence axiom. In the context of health care, independence requires that the patient should make their choices as if all common outcomes with the same probability of occurring across two or more treatments are irrelevant. Formally:

For all $x_1, x_2, x_3 \in X$, $p \in (0,1)$: if $x_1 \geq x_2$ then $px_1 + (1-p)x_3 \geq px_2 + (1-p)x_3$ (1)

where x_i are health states attainable from the treatment option, X, and $\geq$ is the preference relation, 'preferred or indifferent to'.

Most of the direct experimental testing of the EU axioms has been undertaken in the context of monetary outcomes. An exception that incorporated health care outcomes is reported in Oliver (2000). In that study, which was a sub-study undertaken at the same time and with the same subjects as the study that will be reported in this paper, 38 subjects were asked to imagine they had an illness from which, without treatment, they would die immediately. The subjects were then presented with two contexts (one after the other), in which there were two hypothetical treatment options in each context. In each context, the subjects were asked for their ordinal rating of the two available treatment options. The two contexts were designed to mirror the classic Allais paradox scenario, and were essentially composed of the treatment options presented in Table 1.

[Insert Table 1]

In the table, treatment (a) in context 1 offers a 1% chance of immediate death, a 10% chance of living for 18 years in full health and then death and an 89% chance of living for 12 years in full health and then death. The outcomes and associated probabilities of all other treatments in Table 1 can be read similarly. According to the independence axiom, subjects should rate (a) against (b) in context 1 (and (a') against (b') in context 2) as if all common outcomes across treatment options are irrelevant. Thus, under independence, subjects should act as if they disregard the 89% chance of living for 12 years in context 1, and the 89% chance of immediate death in context 2. If this is done, contexts 1 and 2 are identical. Therefore, an individual who prefers treatment (a) ((b)) should prefer treatment (a') ((b')).

However, in the scenario on which the contexts in Table 1 are based, Allais (1953) employed monetary outcomes, and argued that individuals will systematically violate independence. Transferring Allais' proposition to the contexts in Table 1, it is hypothesised that individuals will tend to prefer treatments (b) and (a'). This is the specific proposition that is now known as the Allais paradox (or the 'common consequence effect'), and, in the context of monetary outcomes, has been repeatedly supported in empirical analyses (e.g., Morrison, 1967; Moskowitz, 1974; Slovic and Tversky, 1974; Kahneman and Tversky, 1979; MacCrimmon and Larsson, 1979; Conlisk, 1989).[1] Of the 38 subjects in the experiment reported by Oliver, 20 (52.6%) violated independence, with 14 of these (77.8% of those violating independence) conforming with the Allais hypothesis. Thus, significant and systematic violations of independence (and hence EU) in the context of health care outcomes were found. This places a question mark against the validity of using SG for eliciting cardinal health state utility measures.

Many alternatives to EU have been proposed in the experimental economics literature. For example, theories have been developed that specifically remove the independence axiom, such as generalised expected utility theory (Machina, 1982) and quadratic expected utility theory (Chew et al., 1991). Theories have also been proposed that weaken the independence axiom; for example, weighted expected utility theory (Chew and MacCrimmon, 1979; Chew, 1983), implicit expected utility theory (Dekel, 1986) and disappointment aversion theory (Gul, 1991). Moreover, there are (descriptive) theories that were not developed with the specific objective of removing or weakening the axioms of EU, but are instead based upon certain psychological notions that people may incorporate into a risky or uncertain decision making context. Theories of this kind include prospect theory (Kahneman and Tversky, 1979), regret theory (Loomes and Sugden, 1982; 1987a; b) and disappointment theory (Loomes and Sugden, 1986).

It is often claimed, however, that rank-dependent expected utility theory (RDU) is currently the most influential and/or popular alternative to EU (e.g., Wakker et al., 1994; Fennema and Wakker, 1996; Weber and Kirsner, 1997; Bleichrodt, 1999). RDU was first developed in the 1980s (Quiggin, 1982; Yaari, 1987) and has subsequently been incorporated into a cumulative version of prospect theory (Tversky and Kahneman, 1992).[2] Wakker and Stiggelbout (1995) have gone so far as to propose a modified version of SG based on RDU. Essentially, RDU generalises EU by weakening the independence axiom. However, experiments that have thus far been undertaken to test the performance of RDU have found that the theory is not significantly better than EU at encapsulating people's preferences (Wakker et al., 1994; Fennema and Wakker, 1996). The previous tests of RDU have incorporated monetary outcomes and its axioms have yet to be tested in the context of health care. Until such tests have been undertaken, proposing a modified version of SG based on RDU is premature.

The objective of the experiment reported in this paper is to quantitatively test the descriptive performance of RDU against EU in the context of health care outcomes. First, however, the distinction between EU and RDU will be detailed.

---

[1] The independence violation rate reported in these experiments typically ranges between 30-60%.
[2] The original prospect theory (Kahneman and Tversky, 1979) allowed violations of stochastic dominance, which most experimental economists deemed unacceptable.

2. The demarcation between EU and RDU

2.1. The distinction between independence and comonotonic independence

The only essential axiomatic difference between EU and RDU is a weakening of the independence axiom to comonotonic independence (Wakker, 1996). The reader will perhaps best understand the concept of comonotonic independence with the aid of an example that has been developed from Fennema and Wakker (1996). Consider Table 2.

[Insert Table 2]

In the table, p_1, p_2 and p_3 are the probabilities attached to the occurrence of the outcomes of treatments (a) and (b). Probabilities $p_1 + p_2 + p_3 = 1$. The outcomes are denoted as survival years in full health. Utility is assumed to be an increasing function of life years, and c years is a common outcome across both treatments. Assume, for the moment, that c is equal to zero. Under both treatments, the worst outcome occurs with probability p_1, the intermediate outcome occurs with probability p_2, and the best outcome occurs with probability p_3. Therefore, treatments (a) and (b) are 'comonotonic'.

Now assume that c is equal to four rather than zero. The rank ordering of outcomes for both treatments is still worst, intermediate and best with probabilities p_1, p_2 and p_3, respectively. Therefore, both treatments remain comonotonic. Moreover, since the rank orderings of outcomes associated with the probabilities presented are identical for treatments (a) and (b) when c = 0 and when c = 4, any pair of the four treatment options described thus far (i.e. treatments (a) and (b) when c = 0 and c = 4) are comonotonic. Thus, the change in the common outcome from zero to four is a comonotonic change, which implies that comonotonic independence should apply. Under both EU and RDU, an individual who prefers treatment (a) (treatment (b)) when c = 0 should prefer treatment (a) (treatment (b)) when c = 4.

Now assume that c is equal to ten rather than four. The rank ordering of outcomes for both treatments has now changed. For both treatments, the best outcome still occurs with probability p_3, but the intermediate outcome now occurs with probability p_1 and the worst outcome occurs with probability p_2. Therefore, although both treatments remain comonotonic with each other when c = 10, the differing rank ordering of outcomes for (a) and (b) when c = 10 than when c = 4 implies that there has been a non-comonotonic change in the common outcome. Under EU, independence should hold regardless of whether the change in the common outcome is comonotonic or non-comonotonic. However, under RDU, independence is not a requirement for non-comonotonic changes in the common outcome. Thus, under RDU, a person who prefers treatment (a) (treatment (b)) when c = 4 is allowed to prefer treatment (b) (treatment (a)) when c = 10.

2.2. Explaining the distinction in terms of the decision weights

Under EU, a treatment is evaluated by:

$$\sum^{n}_{i=1} p_i U(x_i) \quad (2)$$

where p_i is the probability of health state x_i occurring, and U(.) is the health state utility index.

RDU, on the other hand, incorporates a probability weighting function that determines 'capacity', and is denoted by W(.). W(.) assigns a value of one when p_i = 1 and zero when $p_i = 0$, and is increasing in the probabilities. Assume that $x_1 \leq \ldots \leq x_n$. Under RDU, a treatment is evaluated by:

$$\sum^{n}_{i=1} \pi_i U(x_i) \quad (3)$$

where π_i is the decision weight associated with health state x_i,[3] and U(.) is the same health state utility index as that denoted under EU.

The decision weights are derived from W(.) by:

$$\pi_i = W(p_i, \ldots, p_n) - W(p_{i+1}, \ldots, p_n) \quad (4)$$

Therefore, the decision weight attached to health state i is its marginal capacity contribution to all superior outcomes (Wakker et al., 1994). For example, consider again Table 2, and assume that c = 0. Further assume that an individual prefers treatment (a) over treatment (b). For this individual, under RDU:

$$\pi_1 U(0 \text{ yrs}) + \pi_2 U(5 \text{ yrs}) + \pi_3 U(15 \text{ yrs}) > \pi_1 U(0 \text{ yrs}) + \pi_2 U(8 \text{ yrs}) + \pi_3 U(12 \text{ yrs}), \text{ or}$$

$$\pi_3[U(15 \text{ yrs}) - U(12 \text{ yrs})] > \pi_2[U(8 \text{ yrs}) - U(5 \text{ yrs})] \quad (5)$$

Treatments (a) and (b) with c = 0 give $x_1 < x_2 < x_3$. Thus, eliciting from Eq. (4),

$$\pi_1 = 1 - W(p_2 + p_3)$$
$$\pi_2 = W(p_2 + p_3) - W(p_3)$$
$$\pi_3 = W(p_3)$$

Now assume that c = 4. The rank ordering of outcomes remains identical to that when c = 0. Therefore, under RDU, the decision weights are calculated in the same way as π_1, π_2, and π_3 above. Thus, the individual who preferred treatment (a) when c = 0 should also prefer treatment (a) when c = 4.

Now assume that c = 10. The rank ordering of the outcomes has changed from when c = 0 or c = 4 and is given by $x_2 < x_1 < x_3$. The decision weights when c = 10 are derived from:

$$\pi'_1 = W(p_1 + p_3) - W(p_3)$$
$$\pi'_2 = 1 - W(p_1 + p_3)$$
$$\pi'_3 = W(p_3)$$

[3] If $\pi_i = p_i$, RDU reduces to EU.

Substitute π'_2 and π'_3 for π_2 and π_3 in Eq. 5. Since the decision weight π'_2 is derived differently from π_2, the inequality given in Eq. (5) does not necessarily hold when c = 10. More specifically, if π_2 is of a lesser magnitude than π'_2 in this example, then RDU allows the individual to reverse their preference following the non-comonotonic change in the common outcome.

To summarise, the essential axiomatic difference between EU and RDU is that RDU weakens the independence axiom to that of comonotonic independence. Therefore, if RDU is a better descriptive theory than EU, non-comonotonic changes in the common outcome will be expected to generate significantly more violations of independence than will comonotonic changes.

3. Methods

3.1. Subjects

Subjects were recruited in June 1999 on a voluntary basis from the staff of a large health care-related organisation situated in London. By informing the subjects that their participation in the experiment may, in some small way, contribute towards health care science, it was hoped that there would be a sufficient incentive for them to express their preferences to the best of their abilities. No monetary payment was offered to the subjects.

Thirty-eight people agreed to participate in the experiment. Subjects were recruited from all grades within the organisation (i.e. from the general office staff to the directors) and 55% (21/38) were women. Subjects undertook the experiment in nine groups of between two and six people during July and August 1999.

3.2. Design

The subjects who undertook the experiment were presented with a total of 23 questions. Three of these were practice questions that the subjects completed at the beginning of the experiment, and two questions comprised a test of the classic Allais paradox, the full results of which are reported elsewhere (Oliver, 2000). The order in which the 20 questions in the main experiment were presented was randomised across subjects. The practice questions were included to ensure that the subjects understood their task before they embarked on the main experiment. The subjects were allowed to ask questions during the practice session. All of the subjects indicated that they understood their task before they began the main experiment, which they were required to complete without conferring. During the experiment, the subjects were free to return to previous questions in order to revise their answers.

Each question was presented as a health care context comprising two treatment options. In order to gain and maintain the full understanding and interest of the subjects, an effort was made to present each context with as much clarity as possible. To facilitate this, the options in each context were presented in pie chart format. An example of the pie chart format in a typical context is given in Figure 1.

[Insert Figure 1]

The subjects were asked to imagine that they have an illness from which, without treatment, they would die almost immediately. They were also asked to imagine that their doctor tells them that there are two alternative available treatments for their illness. The construct of the two pies in each context illustrates the chances of certain outcomes from the two available treatments. For example, in the context of Figure 1, those subjects who preferred treatment (a) were faced with a 35% chance of living for 9 years in full health and then death, a 25% chance of living for 16 years in full health and then death, and a 40% chance of immediate death. Alternatively, if they preferred treatment (b) they would definitely live for 12 years in full health and then die.

For each context the subject was informed that there is a new treatment being developed for their illness. If taken, the new treatment would definitely give 30 years in full health followed by death. However, the subjects were told that the new treatment would not be available for them.

Each subject was then asked to rate (a) and (b) on a scale. The outcome of the unavailable treatment (30 years in full health) was marked at the top of the scale, and the outcome of no treatment (immediate death) was marked at the bottom of the scale. The subjects were consequently required to rate their preferred treatment relatively higher up the scale. The subjects could indicate indifference between (a) and (b) by rating both treatment options at the same point on the scale. It was emphasised that this was an ordinal rating. No effort was made to measure cardinality in this experiment. The subjects were informed that there are no right or wrong answers to any of the questions.

Subjects were asked to rate (a) and (b) on a scale because the nature of the experiment required many contexts to have (a) and (b) sharing a common outcome. If the subjects had been asked for a direct choice between (a) and (b), they may have soon learned that many contexts contain alternatives with a common outcome. This could have caused the subjects to search for and cancel any common outcomes before evaluating the available options with the appropriate level of care and consideration. Such an occurrence would represent an immediate focus on a particular outcome and may introduce a cognitive process that would hardly, if ever, be induced in real world settings. It was thus considered important to reduce the possibility of an unnatural and immediate focus upon any particular outcome. By asking the subjects to rank each treatment on a scale with endpoints marked by other outcomes, they may have been more likely to consider all outcomes in each treatment option before indicating a preference. Additionally, the experiment contained a large number of contexts in which the alternatives did not share a common outcome. A partial function of these contexts was to serve as ‘fillers’, used to reduce the possibility that the subjects would learn to search for common outcomes.

3.3. Tests

The contexts which form the basis of this paper are summarised in Table 3.

[Insert Table 3]

In the table, the treatment alternatives in each context are given under (a) and (b). Let $(x_i, p_i;\ldots;x_n, p_n)$ denote a treatment that has outcome x_i years of life in full health and then death with probability p_i. For example, in context 3, treatment (a) gives a 35% chance of living for 9 years in full health and then death, a 25% chance of living for 16 years in full health and then death, and a 40% chance of immediate death. All other treatment options summarised in Table 3 can be read similarly. There are four tests of independence in the experiment. Two of the tests are of comonotonic independence and two are of non-comonotonic independence.

Other than containing a different common outcome across treatments (a) and (b), contexts 3, 6, 9, 12 and 15 are identical. When comparing contexts 3 and 6, the common outcome remains the worst outcome under both (a) and (b). Similarly, when comparing contexts 9 and 12, the common outcome remains the intermediate outcome. Therefore, contexts 3 and 6, and contexts 9 and 12, are mutually comonotonic; comparing context 3 with 6 and 9 with 12 offer two tests of comonotonic independence. On the other hand, comparing context 6 with 9, and 12 with 15, reveals that the common outcome across treatment options has changed from the worst to the intermediate outcome, and from the intermediate to the best outcome, respectively. Therefore, the ranking of the common outcome has changed across these contexts, offering two tests of non-comonotonic independence.

Under EU, independence should hold regardless of whether the change in the common outcome is comonotonic or non-comonotonic. Moreover, if subjects appear to conform with and violate independence at random (i.e. if the frequency of preference patterns conforming with independence is not significantly greater than fifty-fifty), a question mark can be placed against the descriptive validity of EU within the context of this experiment. For RDU to be a better descriptive theory than EU, any violations of independence would be expected to occur significantly more frequently following non-comonotonic, as opposed to comonotonic, changes in the common outcome.

4. Results

4.1. Quantitative results

The results from the independence tests are given in Table 4.

[Insert Table 4]

There were four general tests of independence on 38 individuals, giving a total of 152 individual tests of this axiom. There were 76 individual tests of both comonotonic and non-comonotonic independence. Of the 152 individual tests of independence, 134 responses conformed with this axiom. There were 18 violations. Using this data set, the hypothesis that the subjects will randomly conform with independence can be categorically rejected at the 5% level of significance ($\chi^2 = 88.53 > \chi^2_{.05}(1) = 3.84$).

In terms of comonotonic independence, there were 65 conforming responses and 11 violations. The comparable results for non-comonotonic independence were 69 and 7, respectively. Thus, there were more violations of comonotonic than non-comonotonic independence, though this difference was not significant at 5% ($\chi^2 = 0.89 < \chi^2_{.05}(1) = 3.84$).[4] These results are similar to those reported in the context of monetary outcomes (Wakker et al., 1994). Since, under RDU, it would be expected that there be significantly more violations of non-comonotonic than comonotonic independence, the results of this experiment appear to refute the hypothesis that RDU offers a descriptive improvement over EU.

5. Discussion

In the context of health care, the results of the experiment reported in this paper do not offer any evidence that RDU is a better descriptive theory than EU. In fact, the finding that comonotonic independence is violated more often than non-comonotonic independence implies that the descriptive performance of RDU was worse than that of EU, though this difference did not reach statistical significance. This finding is consistent with experiments that have been reported in the context of monetary outcomes (Wakker et al., 1994; Fennema and Wakker, 1996). In answering the contexts reported in this paper, the subjects conformed very closely with general independence, which, at least on a superficial level, implies that there may be very little wrong with this axiom.

However, before concluding that EU is a perfectly acceptable descriptive theory in the context of health care, it should be remembered that the 38 subjects who answered the contexts reported here also, within the same one hour period, significantly and systematically violated independence in Allais paradox-type health care contexts (Oliver, 2000). The Allais paradox scenario contains contexts where a subject is asked to choose between certainty and risk in one context, and risk and risk in a second context (see Table 1). The best known proposition that has been put forward to explain why subjects often violate independence over these type of contexts is the 'certainty effect', where it is suggested that some people overweight outcomes that are considered certain relative to outcomes that are merely probable (Kahneman and Tversky, 1979). The qualitative results reported in Oliver (2000) imply that those subjects who conformed with the Allais paradox, when asked to rate an option that gave a positive outcome with certainty against an option that had a small probability of immediate death (Table 1, context 1), often focussed upon the small probability of immediate death, inducing risk averse behaviour. However, with an identical difference in the percentage chance of immediate death between the two treatment options, but with a large chance of death in both options (Table 1, context 2), the subjects often attached less emphasis on the difference in the probability of death over treatments, and were more likely to base their preference on the best possible outcome, which resulted in behaviour that was consistent with risk seeking. Thus Oliver's results imply that individuals often use different cognitive processes to formulate their decisions over Allais paradox-type contexts.

The tests of independence reported in this paper always involved contexts which contained two risky treatment options; i.e. at no time in contexts 3, 6, 9, 12 and 15 was the subject asked to rate a risky treatment against a treatment that gave an

[4] There was also a statistically insignificant difference in the number of people (34 versus 31) who conformed with independence across the two comonotonic tests. The same result holds across the two non-comonotonic tests.

outcome with certainty. Moreover, the probabilities of the outcomes from each treatment lay relatively close to the middle of the probability distribution; i.e. whilst outcomes with a 0.01 and 0.89 chance of occurring are incorporated in the Allais paradox, the probabilities employed in the tests of independence in this experiment were 0.25, 0.35 and 0.45 (Table 3). That independence was generally conformed with in this experiment therefore suggests that it is most often violated when probabilities at the ends of the distribution are used, a supposition that is widely held within the experimental economics literature in the context of monetary outcomes. By specifying that people will apply particular weights to small and large probabilities of outcomes, RDU significantly deviates from EU for only those treatments that involve small or large probabilities of outcomes. Therefore, future research might try to test if the independence / comonotonic independence demarcation between EU and RDU performs better by incorporating probabilities that lie closer to zero and one.

If it is (or even if it is not) discovered that RDU (and the other generalisations of EU) offer no descriptive improvement over EU, should experimental and health economists accept that EU is the appropriate theory to adopt in the context of health care? Given that EU is significantly and systematically violated in certain circumstances, there is a danger that the SG method for health state utility elicitation yields non-cardinal values. This implies that the results of cost-utility analyses that employ the SG method may recommend an allocation of health care resources that is inconsistent with peoples' true underlying preferences (Oliver, 2000).

Those who have attacked EU have traditionally searched for violations of the EU axioms, and if violations have not been discovered, have assumed that individuals are behaving in accordance with EU. For economists, consistency with the predictions of the theory has sufficed, irrespective of whether or not the decision heuristics that people actually employ are consistent with the axioms that form the basis of the theory. For example, people may often employ a decision heuristic that leads them to behave in a way that is consistent with the predictions of independence, though the heuristic itself, if isolated by the researcher, may be used to predict much wider preference behaviour, and may account for violations such as the Allais paradox.

Along this line of argument, Machina (1999) has recently offered a challenge to those who criticise EU: first, he states, describe empirical findings in their own right, not as flaws in the EU model, and second, search for connections between the findings, and for underlying principles. This will be difficult, and presents a massive research agenda, but it is perhaps time for economists, psychologists and behavioural scientists to abandon EU as the reference point (or, in the case of health economists, accepting EU as given) when analysing decision making behaviour under conditions of risk and uncertainty.

Acknowledgements

The author is grateful to Graham Loomes, Mike Jones-Lee and Han Bleichrodt for ideas and inspiration, and to the volunteers from the Association of the British Pharmaceutical industry who agreed to sit the experiment. Financial support came in the form of ESRC award number R00429834596. The usual disclaimer applies.

References

Allais, M., 1953, Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole Américaine. Econometrica 21, 503-546.

Bleichrodt, H. and J. Quiggin, 1999, Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis. Journal of Health Economics 18, 681-708.

Camerer, C. and M. Weber, 1992, Recent developments in modelling preferences: uncertainty and ambiguity. Journal of Risk and Uncertainty 5, 325-370.

Camerer, C. and T. Ho, 1994, Violations of the betweenness axiom and non-linearity in probability. Journal of Risk and Uncertainty 8, 167-196.

Chew, S.H., 1983, A generalisation of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. Econometrica 51, 1065-1092.

Chew, S.H. and K.R. MacCrimmon, 1979, Alpha-nu choice theory: an axiomatization of expected utility. Working paper 669, University of British Columbia, British Columbia.

Chew, S.H., L.G. Epstein and U. Segal, 1991, Mixture symmetry and quadratic utility. Econometrica 59, 139-163.

Conlisk, J., 1989, Three variants on the Allais example. American Economic Review 79, 392-407.

Dekel, E., 1986, An axiomatic characterization of preferences under uncertainty: weakening the independence axiom. Journal of Economic Theory 40, 304-318.

Fennema, H. and P. Wakker, 1996, A test of rank-dependent utility in the context of ambiguity. Journal of Risk and Uncertainty 13, 19-35.

Gul, F., 1991, A theory of disappointment in decision making under uncertainty. Econometrica 59, 667-686.

Kahneman, D. and A. Tversky, 1979, Prospect theory: an analysis of decision under risk. Econometrica 47, 263-291.

Loomes, G.C. and R. Sugden, 1982, Regret theory: an alternative theory of rational choice under uncertainty. Economic Journal 92, 805-824.

Loomes, G.C. and R. Sugden, 1986, Disappointment and dynamic consistency in choice under uncertainty. Review of Economic Studies 53, 271-282.

Loomes, G.C. and R. Sugden, 1987a, Some implications of a more general form of regret theory. Journal of Economic Literature 41, 270-287.

Loomes, G.C. and R. Sugden, 1987b, Testing for regret and disappointment in choice under uncertainty. Economic Journal 97, 118-129.

MacCrimmon, K.R. and S. Larsson, 1979, Utility theory: axioms versus paradoxes. In: M. Allais and O. Hagen, eds., Expected utility hypotheses and the Allais paradox. Reidel, Dordrecht, 27-145.

Machina, M.J., 1982, 'Expected utility' analysis without the independence axiom. Econometrica 50, 227-323.

Machina, M.J., 1987, Choice under uncertainty: problems solved and unsolved. Economic Perspectives 1, 121-154.

Machina, M.J., 1999, A challenge to the "econoclasts": A commentary on "Rationality for economists?" Journal of Risk and Uncertainty 1-3, 107-108.

Morrison, D.G., 1967, On the consistency of preferences in Allais' paradox. Behavioral Science 12, 373-383.

Moskowitz, H., 1974, Effects of problem representation and feedback on rational behavior in Allais and Morlat-type problems. Decision Sciences 5, 225-242.

Oliver, A.J., 2000, Testing the Allais paradox in the context of health care. Paper presented at the Health Economics Study Group Meeting, Newcastle, UK.

Quiggin, J., 1982, A theory of anticipated utility. Journal of Economic Behavior and Organisation 3, 323-343.

Slovic, P. and A. Tversky, 1974, Who accepts Savage's axiom? Behavioral Science 19, 368-373.

Tversky, A. and D. Kahneman, 1992, Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297-323.

von Neumann, J. and O. Morgenstern, 1944, Theory of games and economic behavior. Princeton University Press, Princeton.

Wakker, P.P., 1996, The sure-thing principle and the comonotonic sure-thing principle: an axiomatic analysis. Journal of Mathematical Economics 25, 213-227.

Wakker, P., I. Erev and E.U. Weber, 1994, Comonotonic independence: the critical test between classical and rank-dependent utility theories. Journal of Risk and Uncertainty 9, 195-230.

Wakker, P. and A. Stiggelbout, 1995, Explaining distortions in utility elicitation through the rank-dependent model for risky choices. Medical Decision Making 15, 180-186.

Weber, E.U. and B. Kirsner, 1997, Reasons for rank-dependent utility evaluation. Journal of Risk and Uncertainty 14, 41-61.

Yaari, M.E., 1987, The dual theory of choice under risk. Econometrica 55, 99-115.

Table 1
Allais paradox-type contexts reported in Oliver (2000)

	Probability		
	0.01	0.10	0.89
Context 1			
Treatment (a)	0 yrs	18 yrs	12 yrs
Treatment (b)	12 yrs	12 yrs	12 yrs
Context 2			
Treatment (a')	0 yrs	18 yrs	0 yrs
Treatment (b')	12 yrs	12 yrs	0 yrs

Table 2
Distinction between independence and comonotonic independence

	p_1	p_2	p_3
Treatment (a)	c yrs	5 yrs	15 yrs
Treatment (b)	c yrs	8 yrs	12 yrs

Table 3

Tests of comonotonic and non-comonotonic independence

	(a)	(b)
Context 3	(9, 0.35; 16, 0.25; 0, 0.40)	(6, 0.35; 19, 0.25; 0, 0.40)
Context 6	(9, 0.35; 16, 0.25; 5, 0.40)	(6, 0.35; 19, 0.25; 5, 0.40)
Context 9	(9, 0.35; 16, 0.25; 10, 0.40)	(6, 0.35; 19, 0.25; 10, 0.40)
Context 12	(9, 0.35; 16, 0.25; 15, 0.40)	(6, 0.35; 19, 0.25; 15, 0.40)
Context 15	(9, 0.35; 16, 0.25; 20, 0.40)	(6, 0.35; 19, 0.25; 20, 0.40)

Table 4
Quantitative results from the tests of independence

	Conform	Violate
Comonotonic		
Contexts 3/6	34	4
Contexts 9/12	31	7
Total	65	11
Non-comonotonic		
Contexts 6/9	33	5
Contexts 12/15	36	2
Total	69	7

Figure 1
The context design