Adam J. Oliver

The internal consistency of the standard gamble : tests after adjusting for prospect theory

Originally published in Journal of health economics, 22 (4). pp. 659-674

The Internal Consistency of the Standard Gamble:
Tests After Adjusting for Prospect Theory

Adam Oliver
LSE Health and Social Care, London School of Economics, London, UK[*]

[*] Correspondence: LSE Health and Social Care, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Tel: +44 (0)20 7955 6471; Fax: +44 (0)20 7955 6803; E-mail: a.j.oliver@lse.ac.uk

Summary

This article reports a study that tests whether the internal consistency of the standard gamble can be improved upon by incorporating loss weighting and probability transformation parameters in the standard gamble valuation procedure. Five alternatives to the standard EU formulation are considered: (1) probability transformation within an EU framework; and, within a prospect theory framework, (2) loss weighting and full probability transformation, (3) no loss weighting and full probability transformation, (4) loss weighting and no probability transformation, and (5) loss weighting and partial probability transformation. Of the five alternatives, only the prospect theory formulation with loss weighting and no probability transformation offers an improvement in internal consistency over the standard EU valuation procedure.

JEL classification: C90, D81, I10

Keywords: Standard gamble, Expected utility theory, Prospect theory, Loss aversion, Probability transformation

1. Introduction

It has been argued that the standard gamble is the gold standard for the elicitation of cardinal health state values (Torrance, 1986; Torrance and Feeny, 1989). The reason for this is that health care decisions invariably involve a degree of risk, and unlike other often used elicitation methods (e.g. the time trade-off and the rating scale), the standard gamble is implied from the axioms of expected utility theory (EU). The standard gamble thus has a firm grounding in the theory of risk and uncertainty.

Unfortunately, the standard gamble has been empirically observed to be internally inconsistent (Bleichrodt, 2001; Llewellyn-Thomas et al., 1982; Morrison, 1996; Rutten-van Molken et al., 1995; Spencer, 1998). To understand the concept of internal consistency, consider the simple standard gamble scenario where an individual is given a choice between remaining in an intermediate health state, and a treatment that offers an unknown probability, p, of attaining a (pre-defined) better health state with a risk of probability (1-p) that the treatment could cause a (pre-defined) worse health state. If the health states are denoted by x_i, the choice is therefore between the certainty of health state x_2 and a treatment that offers $px_1 + (1-p)x_3$ with $x_1 \succ x_2 \succ x_3$, where $\succ$ means 'is better than'. The individual is asked to choose the probability, p, that would render them indifferent between the alternatives. Therefore, $v_{sg}(x_2) = pv_{sg}(x_1) + (1-p)v_{sg}(x_3)$, where $v_{sg}(.)$ is the standard gamble value function. If the values of x_1 and x_3 are respectively normalised at one and zero, then $v_{sg}(x_2) = p$.

Now consider the health state x_2', for which $x_2 \succ x_2' \succ x_3$. The standard gamble value of x_2' is evaluated by asking the individual to choose the probability, q, which would render them indifferent between x_2' and a treatment that offers $qx_1 + (1-q)x_3$. Thus, $v_{sg}(x_2') = q$.

Next, assume that the individual is asked to choose the probability, r, for which they would be indifferent between x_2 and a treatment that offers $rx_1 + (1-r)x_2'$. In this case, $v_{sg}(x_2) = rv_{sg}(x_1) + (1-r)v_{sg}(x_2')$. If the standard gamble is perfectly internally consistent, we ought to be able to substitute into this equation the previously elicited values for x_2 and x_2'. Hence, the individual should state an r commensurate with $p = r + (1-r)q$; i.e. $r = (p-q)/(1-q)$.

Application of this process in testing the internal consistency of the standard gamble has been termed 'chaining' (Rutten-van Molken et al., 1995; Spencer, 1998). The process outlined above chains to the failure outcome, implying that the exercise involves replacing the failure outcome in the gamble with a health state that differs from the original failure outcome (i.e. x_3 was replaced with x_2' when eliciting the individual's probability, r). Alternatively, it is possible to chain to the success outcome, which would mean that an exercise is undertaken where the original success outcome is replaced by an alternative health state.

The treatment arm of a standard gamble usually employs full health and death as the success and failure outcomes, and standard gamble values that are elicited under these circumstances are sometimes called 'basic reference values' (Rutten-van Molken et al., 1995; Spencer, 1998). The values elicited through chaining exercises can be termed 'chained values'. As should by now be clear, if the standard gamble is internally consistent, the basic reference and chained values elicited for any particular health state ought to be identical. The studies that have tested the internal consistency of the standard gamble in the context of health have generally found that although the standard gamble is internally consistent when chaining to the success outcome, chained to failure values often tend to exceed basic reference values (Bleichrodt, 2001; Llewellyn-Thomas et al., 1982; Morrison, 1996; Rutten-van Molken et al., 1995; Spencer, 1998). This finding is problematic, because if the basic reference and chained values for a particular health outcome systematically differ in a particular direction (e.g. in the direction of chained values > basic reference values), we do not know which values (if any) accurately reflect underlying strengths of preference.

The internal inconsistency of the standard gamble points to a failure of the independence axiom of EU, an axiom that is often violated in experimental settings (for a review, see Camerer, 1995). That is to say that the fundamental value that people place on any particular health outcome seems to vary according to the positioning of the outcome in the standard gamble construct. Prospect theory and cumulative prospect theory have been proposed as descriptive theories of choice that can accommodate all of the major violations of EU (Kahneman and Tversky, 1979; Tversky and Kahneman, 1986; 1992). Prospect theory entails the application of probability transformation to the individual probabilities in a gamble. The transformation of individual probabilities has a long history (Edwards, 1955; Preston and Baratta, 1948), but unfortunately this approach allows violations of dominance. In other words, original prospect theory allows unambiguously inferior gambles to be preferred to unambiguously superior gambles. Although Kahneman and Tversky (1979) argued that individuals would reject dominated gambles during an initial 'editing phase', many researchers, both before and since the development of prospect theory, have considered any decision model that allows formal violations of dominance to be fatally flawed (e.g., Diecidue and Wakker, 2001; Fishburn, 1978). Consequently, Tversky and Kahneman (1992) developed the theory so that the transformation applied to cumulative probability, hence the name 'cumulative prospect theory'. This development was borrowed from rank dependent utility theory (RDU). Thus, Tversky and Kahneman formulated cumulative prospect theory by essentially combining original prospect theory with RDU. Nevertheless, with gambles with two outcomes, such as those presented in the standard gamble, prospect theory and cumulative prospect theory yield similar predictions (Tversky and Kahneman, 1992). In this article, the generic 'prospect theory' shall therefore incorporate both theories.

In order to correct for the descriptive violations of EU, Bleichrodt et al. (2001) have advocated adjusting the standard gamble valuation procedure with the corrections proposed by prospect theory. Bleichrodt et al.'s intention was to examine if this modified procedure could account for the discrepancies that have been reported between the values elicited from the probability equivalence and the certainty equivalence versions of the standard gamble. Under EU, the values elicited for a particular health state from these versions of the standard gamble ought to be identical, but it has been demonstrated that the probability equivalence method generally generates higher standard gamble values than the certainty equivalence method, a violation of procedure invariance (Hershey and Schoemaker, 1985; Jansen et al., 1998; Johnson and Schkade, 1989). Bleichrodt et al. (2001) found that adjusting the standard gamble valuation procedure with the predictions of prospect theory removed these discrepancies.

The objective of this article is to test if various variants of the adjustments proposed by prospect theory can improve the internal consistency of the standard gamble. First, however, the major modifications that prospect theory makes to EU are outlined, and an explanation is given as to how these modifications can be applied to the standard gamble valuation procedure.

2. Modifications according to prospect theory

In the interests of clarity regarding the analysis in this article, it is worth explicitly reminding ourselves that under EU the standard gamble value of health state x_2 is evaluated by:

$$v_{sg}(x_2) = pv_{sg}(x_1) + (1\text{-}p)v_{sg}(x_3) \tag{1}$$

where $x_1 \succ x_2 \succ x_3$ and $0 \leq p \leq 1$.

Prospect theory prescribes two major modifications to EU (Tversky and Kahneman, 1992):

(i) The carriers of value are gains and losses, not final assets: this is synonymous with the ‘reference point effect’.

According to prospect theory, people select a reference point and then evaluate gambles according to their expected gains and losses around the reference point. Denote the reference point as ρ. If the health state $x \succ \rho$ then x is a gain. Conversely, if $\rho \succ x$ then x is a loss.

Allowing for the reference point effect permits individuals to demonstrate ‘loss aversion’. Loss aversion is one possible component of the ‘gambling effect’ - i.e. the systematic dislike of risk that cannot entirely be explained by risk aversion through EU (Gafni and Torrance, 1984; Loomes, 1993). There are many possible (interrelated) cognitive components of the gambling effect, including anticipated regret and disappointment, but a discussion of these is beyond the scope of this article. For a brief discussion, see Oliver (2003). Tversky and Kahneman (1991; 1992) reported that individuals will only generally accept ‘mixed’ gambles with a fifty-fifty chance of a gain and a loss if the gain is at least twice as large as the loss (which indicates a strong aversion to losses). However, in ‘positive’ gambles, where no outcome is perceived as a loss, a decrease in the smallest gain is fully compensated for by only a very slightly larger increase in the largest gain, implying that the high sensitivity to the worst outcome in a gamble only applies if the worst outcome is perceived as a loss. The seemingly high magnitude of disutility associated with an outcome when it is perceived as a loss relative to the magnitude of utility when that same outcome is perceived as a gain is not accounted for in the EU model.

Tversky and Kahneman (1992) estimated a loss weighting parameter (or ‘loss aversion coefficient’) by eliciting the factor by which a loss generally has to exceed a gain in order for the loss to balance the gain. For example, if an individual will only just accept a mixed gamble if the loss is precisely twice as large as the gain, then the loss weighting parameter for that individual when answering that particular mixed gamble would equal two. Tversky and Kahneman (1992) estimated a 'general' loss weighting parameter, λ, equal to 2.25, which can be applied to mixed gambles, including the standard gamble (assuming that people adopt an outcome intermediate to the best and worst outcomes in the standard gamble - for example, the certain outcome, x_2 - as their reference point), in order to attempt to control for (or 'factor out') loss aversion.

(ii) Rather than being multiplied by a raw probability, the utility (or value) of each outcome is multiplied by a decision weight, where the decision weights are determined by the non-linear transformation of probability.

Probability transformation has been observed to differ slightly between the domain of gains and the domain of losses (e.g., Bleichrodt, 2001; Tversky and Kahneman, 1992); the transformed probabilities are therefore denoted $w^+(p)$ and $w^-(p)$ for gains and losses, respectively. For gambles that occur entirely within the domain of gains, the transformed probabilities sum to one. Similarly for gambles that occur entirely within the domain of losses. However, for mixed gambles that involve both possible gains and possible losses the transformed probabilities can sum to greater or less than one (Tversky and Kahneman, 1992). Transformed probabilities equal one when p = 1 and zero when p = 0, and are non-decreasing with probability. The common empirical finding is that individuals transform probabilities in an inverse S-shaped pattern (e.g., Abdellaoui, 2000; Bleichrodt and Pinto, 2000; Bleichrodt et al., 1999; Camerer and Ho, 1994; Gonzalez and Wu, 1999; Kahneman and Tversky, 1979; Lattimore et al., 1992; Tversky and Fox, 1995; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996; 1999), such that people appear to overweight small probabilities, underweight large probabilities, and perceive w(p) as equal to p at approximately 0.4. This general pattern is observed in the domains of both gains and losses (e.g., see Bleichrodt, 2001; Tversky and Kahneman, 1992). A typical pattern of transformed probabilities is presented in Figure 1, where the probabilities are represented by the solid diagonal line and the associated transformed probabilities are represented by the dashed inverse S-shaped curve.

[Insert Figure 1]

Let us consider an example of how the inverse S-shaped probability transformation curve has been estimated. Tversky and Kahneman (1992) presented 25 respondents with a large number of positive and negative gambles. As noted above, a positive gamble is a gamble where no outcome is perceived as a loss; conversely, a negative gamble is a gamble where no outcome is perceived as a gain. The positive gambles took the form p$y + (1-p)$0, and the negative gambles took a similar form in the domain of losses. Probabilities across the entire probability distribution were used.

Tversky and Kahneman elicited the respondents' certainty equivalents, c, for all gambles. For a risk neutral individual, c = py. When faced with positive gambles, risk averse individuals demonstrate $c < py$, and risk seeking individuals demonstrate $c > py$ (the opposite is true of risk averse and risk seeking individuals when faced with negative gambles). Consequently, for positive gambles, risk neutrality implies p = c/y, risk aversion implies $p > c/y$ and risk seeking implies $p < c/y$. Tversky and Kahneman interpreted the transformed probabilities to equal the mean values of c/y, and plotted these values against p. Through this method, they derived the inverse S-shaped pattern of transformed probabilities in both the domain of gains and the domain of losses. The curves imply that their respondents tended to be risk seeking when faced with small probabilities of gains and large probabilities of losses and risk averse when faced with large probabilities of gains and small probabilities of losses.

Tversky and Kahneman (1992) proposed the following functional forms to estimate the transformed probabilities:

$$w^{+}(p) = p^{\delta} / [p^{\delta} + (1-p)^{\delta}]^{1/\delta} \qquad (2)$$

$$w^{-}(p) = (1-p)^{\gamma} / [(1-p)^{\gamma} + p^{\gamma}]^{1/\gamma} \qquad (3)$$

where $\delta = 0.61$ and $\gamma = 0.69$.

In accordance with other studies (e.g., Bleichrodt et al., 2001; Hershey and Schoemaker, 1985), it is assumed here that individuals adopt the certain outcome as their reference point in the standard gamble. Hence, $x_2 = \rho$. It is also assumed that people adopt $w^{+}(p)$ and $w^{-}(p)$ as their decision weights, π^{+} and π^{-}, and respectively apply these weights to the perceived gains and losses in the gamble. Under prospect theory, the standard gamble value of health state x_2 is equated with the value of the perceived gains balanced against the value of the perceived losses, relative to the adopted reference point in the risky (treatment) arm of the gamble. Hence, the value of x_2 is evaluated by (Bleichrodt et al., 2001):

$$v_{pt}(x_2) = v_{pt}(\rho) + \pi^{+}[v_{pt}(x_1) - v_{pt}(\rho)] - \lambda\pi^{-}[v_{pt}(\rho) - v_{pt}(x_3)]$$

$$\Rightarrow v_{pt}(x_2) = v_{pt}(x_2) + \pi^{+}[v_{pt}(x_1) - v_{pt}(x_2)] - \lambda\pi^{-}[v_{pt}(x_2) - v_{pt}(x_3)]$$

$$\Rightarrow v_{pt}(x_2) = v_{pt}(x_2) + \pi^{+}v_{pt}(x_1) - \pi^{+}v_{pt}(x_2) - \lambda\pi^{-}v_{pt}(x_2) + \lambda\pi^{-}v_{pt}(x_3)$$

$$\Rightarrow \pi^{+}v_{pt}(x_2) + \lambda\pi^{-}v_{pt}(x_2) = \pi^{+}v_{pt}(x_1) + \lambda\pi^{-}v_{pt}(x_3)$$

$$\Rightarrow v_{pt}(x_2)(\pi^{+} + \lambda\pi^{-}) = \pi^{+}v_{pt}(x_1) + \lambda\pi^{-}v_{pt}(x_3)$$

$$\Rightarrow v_{pt}(x_2) = [\pi^{+}v_{pt}(x_1) + \lambda\pi^{-}v_{pt}(x_3)] / (\pi^{+} + \lambda\pi^{-}) \qquad (4)$$

where $v_{pt}(.)$ is the prospect theory standard gamble value function.

A study that was designed to test whether the internal consistency of the standard gamble can be improved upon by using the prospect theory valuation formula in Eq. (4) rather than the standard EU formula in Eq. (1) is now reported. Throughout this article, Tversky and Kahneman's (1992) estimates of λ, δ and γ - i.e. 2.25, 0.61 and 0.69, respectively - are used.

3. Methods

3.1. Respondents

Thirty respondents, recruited via the London School of Economics e-mail network, participated. Therefore the respondents did not comprise a representative subgroup of society, and their answers should not be taken as indicative of societal preferences as a whole. However, it is possible that many people would find standard gamble questions quite difficult to answer, and it was thought that that the answers given by a relatively educated subgroup *might* provide a more insightful starting point in testing whether the internal consistency of the instrument can be improved. Thus, in an initial methodological exercise (such as the study reported in this article), the respondents are not inappropriate. However, if any of the alternatives to the standard EU valuation procedure tested in this study were observed to offer greater internal consistency, it is recognised that larger, more representative samples, both in studies that aim to corroborate the findings reported here and in studies that aim to use any promising procedures to elicit values for practical decision making purposes, might be more appropriate.

The results reported here were taken from one interview per respondent, for which each respondent was paid £12. All of the respondents were either undergraduates (77%) or graduates, 63% were female, 83% were aged 16-30 (the remainder were aged 31-45), 77% had a social science background (with 7%, 7% and 10% having a humanities, science and ‘other’ background, respectively), and 77% were unfamiliar with decision theory.

3.2. Design

The interview lasted between 15 and 30 minutes and the respondents attended their interviews individually. The interview was divided into two parts and comprised a total of 17 questions. The analysis in this article is based on the answers to the first part of the interview, which contained 11 questions, one of which was a practice question that the respondents were required to answer at the beginning of the interview.

The respondents were taken through a pre-answered question and the practice question, and were free to ask questions during the practice session. After completing the practice question, the order in which the ten remaining questions were presented was randomised across respondents, and the respondents were required to answer these without asking any questions. The respondents were informed that they were free to return to previous questions in order to revise their answers if they wished. Each question was presented as a health care context in the standard gamble format. A typical context is reported as Figure 2.

[Insert Figure 2]

The respondents were asked to imagine that they themselves face the choices given in the questions. The three outcomes in each question, x_1, x_2 and x_3 (where, in Figure 2, x_2 = 10 years, x_1 = 40 years and x_3 = 0 years), were defined by length of life. This contrasts with the previous tests of the internal consistency of the standard gamble where health outcomes were defined by the severity of the health state (Bleichrodt, 2001; Llewellyn-Thomas et al., 1982; Morrison, 1996; Rutten-van Molken et al., 1995; Spencer, 1998). However, because some respondents may have difficulty in comprehending health states of varying severity, longevity is used so as to simplify the elicitation process and thus identify the extent to which anomalies in the standard gamble occur when it is stripped to its basic elements. It was assumed that all of the respondents would prefer more life in full health to less life in full health. Therefore, x_1, x_2 and x_3 could be set unambiguously at $x_1 \succ x_2 \succ x_3$. The maximum length of life on offer throughout the questionnaire was 40 years in good health. All of the respondents could reasonably expect to live for at least 40 more years.

In the standard gamble, it is usual practice to attempt to elicit respondents' indifference probabilities for p in Eq. (1). However, when piloting the questions, it was found that people sometimes have difficulty in comprehending the concept of indifference in this experimental design. Therefore, a respondent's stated minimum required chance of treatment success is used to approximate the probability that would render him or her indifferent between accepting the certainty, x_2, or the treatment.

A copy of the full questionnaire is available from the author on request.

3.3. Tests

The outcomes used in each of the 11 questions are summarised in Table 1.

[Insert Table 1]

Since 40 years in full health and immediate death represent the best and worst outcomes throughout the questionnaire, these outcomes are respectively normalised with values of one and zero. Following Eq. (1), the respondents' basic reference EU standard gamble values for living for 10, 20 and 30 years in full health equate to their stated probabilities in qs. 1-3. Note that the basic 'reference' values should not be confused with the 'reference' point in the standard gamble, which, as noted earlier, is assumed in this article to be the certain outcome, x_2. Also in accordance with Eq. (1), these basic reference values can be used with the stated probabilities given in the remaining questions to derive chained EU standard gamble values for 10 years in full health (qs. 4 and 5), 20 years in full health (qs. 6-8), and 30 years in full health (qs. 9 and 10). As stated earlier, under EU the elicitations for each particular life expectancy, whether they be derived from basic reference or chained gambles, should not significantly or systematically differ from one another. A comparison of the basic reference and chained EU standard gamble values elicited from the answers given to qs. 1-10 facilitate several tests of the internal consistency of this method.

As an alternative to the EU standard gamble, we can again apply Eq. (1) but with the raw probabilities replaced with the decision weights, π^+ and π^- (i.e. the transformed probabilities, $w^+(p)$ and $w^-(p)$). This specification also implicitly assumes that final 'assets' are the carriers of value, but that individuals unconsciously transform the probabilities when evaluating a risky treatment. Given the abundant empirical evidence that purports to demonstrate that individuals transform probabilities (cited earlier), we may reasonably expect this method to offer an improvement over the EU standard gamble. Wakker and Stiggelbout (1995) similarly recommended that standard gamble values be measured by transformed rather than raw probabilities. Indeed, Bleichrodt (2001) has applied this recommendation with use of the inverse S-shaped transformation function, though he actually found this process to exacerbate internal inconsistency in the standard gamble. The study reported in this article will therefore either substantiate or contradict the results of Bleichrodt's application of Wakker and Stiggelbout's recommendation. This specification is defined as the 'WS standard gamble', and its internal consistency is tested through a procedure identical to that outlined for the EU standard gamble.

By using Eq. (4) rather than Eq. (1), with full application of probability transformation and weighting against the possible influence of loss aversion, a similar procedure is undertaken to estimate the level of internal consistency in the 'prospect theory standard gamble'. Full application of Eq. (4) is based on the assumption that the carriers of value are gains and losses (rather than final assets), and that losses loom larger than gains.

There are many possible explanations for the most persistent violations of EU, and CCthere has been some debate (e.g., Cohen and Jaffray, 1988) concerning whether the violations are largely the result of the certainty effect (of which loss aversion is a possible contributory factor) or of non-linear probability weighting. In an attempt to ascertain the respective isolated impact of both loss aversion and probability transformation on the internal consistency of the standard gamble, Eq. (4) is further applied both in the absence of the loss weighting parameter (defined as the 'prospect theory π standard gamble'), and in the absence of probability transformation (defined as the 'prospect theory λ standard gamble').

Finally, since prospect theory predicts significant non-linear weighting of only small and large probabilities, the internal consistency of the prospect theory standard gamble is also tested by incorporating probability transformation in only those cases where a respondent expresses a probability ≤ 0.1 or ≥ 0.9 (defined as the 'prospect theory partial-π standard gamble'). Therefore, whenever a respondent expresses a probability in the range $0.1 < p < 0.9$, the prospect theory partial-π standard gamble is modified only with the loss weighting parameter.

All tests of significance are calculated with Wilcoxon's matched-pairs signed-ranks test. The Wilcoxon test is the distribution-free analogue to the parametric t-test for related samples; unlike the t-test, it does not require the samples to be normally distributed. This is convenient, because the Lilliefors test showed very strong evidence against normality for the EU standard gamble values elicited in each of the questions in this study (the Lilliefors test for normality can be found at http://ubmail.ubalt.edu/~harsham/Business-stat/otherapplets/LillforsTest.htm). A further advantage of using the Wilcoxon test is that it is not affected by extreme scores or outliers. This is a particularly useful characteristic in testing data sets where outliers may or may not be the result of respondent and/or recording errors. The use of this test removes the temptation to exclude outliers from the analysis. The tests are two-tailed, and the 5% and 1% levels of significance are used.

4. Results

A summary of the results is given in Table 2.

[Insert Table 2]

To recap, qs. 1, 2 and 3 are basic reference gambles, qs. 4, 5 and 6 are gambles chained to the success outcome and qs. 7, 8, 9 and 10 are gambles chained to the failure outcome (strictly speaking, q.7 is chained to both the success and failure outcomes, but is categorised in this article as a 'chained to failure' question). With the exception of q.3 versus q.10, changing the failure outcome in the EU standard gamble caused significant and systematic internal inconsistency, with the chained values tending to be significantly greater than the basic reference values.

The WS standard gamble was, overall, even more internally inconsistent than the EU standard gamble, a finding that complies with those reported by Bleichrodt (2001). The prospect theory standard gamble performed better than the WS standard gamble, but there was no overall discernible improvement in internal consistency compared to the EU standard gamble. The prospect theory π standard gamble performed as poorly as the WS standard gamble. *However, for all of the tests bar one, the hypothesis that the prospect theory λ standard gamble values were internally consistent could not be rejected.* In this specification for valuing the health outcomes, the only test where internal inconsistency was still observed to a statistically significant level was for q.2 versus q.8 (with chained values tending to be greater than basic reference values). Finally, with the exception of q.2 versus q.7, the prospect theory partial-π standard gamble was observed to be internally inconsistent in all of the tests.

Of the alternatives considered, only the prospect theory λ standard gamble appears to offer an improvement in internal consistency over the standard EU valuation procedure. Due to the study's small sample size and the non-normal distribution of the values, little attention ought to be paid to the 'population' values elicited via these two valuation procedures, particularly with respect to the internal consistency of the values. Nonetheless, Table 3 lists the median value and range of values elicited from each of the ten main questions, and from these descriptive statistics two general points can be noted. First, controlling for loss aversion will (substantially) decrease the median standard gamble value for each outcome. Second, an EU standard gamble value equal to one will remain at one when applying the prospect theory λ standard gamble, which will mean than the range of the latter valuation procedure will almost inevitably be larger.

[Insert Table 3]

5. Discussion

The results indicate that the internal consistency of the EU standard gamble is compromised following changes in the failure outcome in the gamble, a finding consistent with those previously reported in the literature (Bleichrodt, 2001; Llewellyn-Thomas et al., 1982; Morrison, 1996; Rutten-van Molken et al., 1995; Spencer, 1998). If the source of this internal inconsistency stems from the same cognitive processes that underlie the often observed violations of EU, and if prospect theory can indeed accommodate most of the major violations of EU, then there is reason to expect the prospect theory-adjusted standard gamble to remedy the problem of internal inconsistency.

Five alternatives to the normal EU method of calculating standard gamble values have been considered in this article. One of the alternatives is based on the continued application of Eq. (1); the remaining four alternatives are based on the prospect theory standard gamble valuation formula given as Eq. (4). The latter four alternatives implicitly assume that individuals adopt the certain outcome as their reference point and then evaluate the risky treatment arm of the gamble in terms of the possible movements away from the reference point. The only alternative that appeared to offer an improvement in the internal consistency of the standard gamble was the prospect theory formulation that incorporated the loss weighting parameter but did not internalise probability transformation: i.e. the prospect theory λ standard gamble. The exact specification of this formula is:

$$v_\lambda(x_2) = [pv_\lambda(x_1) + \lambda(1-p)v_\lambda(x_3)] / [p + \lambda(1-p)] \qquad (5)$$

where $v_\lambda(.)$ is the prospect theory λ standard gamble value function.

It is important to emphasise that the certain outcome is only one of many possible outcomes that individuals may adopt as their reference point. For example, the respondents in this study were generally quite young, and it is possible that some of them *may* have perceived all of the outcomes in each of the questions (including the outcome of 40 more years of full health) as losses. It is also possible that some respondents may choose a different reference point across different questions, and even within questions (i.e. between the treatment on no treatment arms of the standard gamble). Such a broad array of possible reference points clearly has implications concerning the acceptance of Eq. (5) as a general valuation procedure, and an exploration into this concern might present a useful topic for further research. Nonetheless, in order for a valuation procedure to be practical, we have to assume that all individuals adopt the same general decision rule when processing the standard gamble, and if the respondents are successful in believing that they are actually faced with the standard gamble questions presented to them, the assumption that they anchor upon their current *certain* outcome when evaluating the risky treatment arm of the gamble carries intuitive appeal. The results presented in this article do suggest that Eq. (5), with the implicit assumption that people adopt the certain outcome as the reference point, offers a promising and parsimonious method by which to reduce the levels of internal inconsistency in the standard gamble.

In terms of the apparent 'failure' of probability transformation in this study, it is possible that the choice of outcome domain (e.g. money or life expectancy) may influence the probability transformation function (Bleichrodt and Pinto, 2000; Rottenstreich and Hsee, 1999; Wakker and Deneffe, 1996). Thus, that none of the standard gamble formulations that incorporated probability transformation were able to offer an improvement in internal consistency on the EU standard gamble may possibly be explained by a context-specific probability weighting function; i.e. Tversky and Kahneman's (1992) probability transformation functions, which were estimated in the context of money, may not be appropriate for the respondents and/or contexts reported in this article. Alternatively, probability transformation simply may not be a cause of bias in the standard gamble. Even if probability transformation is an important source of bias in the standard gamble, eliciting probability transformation functions is no easy task. Therefore, it may be advisable to exclude consideration of probability transformation in the practical elicitation of health state values, particularly if relatively simple modifications with the loss weighting parameter prove sufficient for eradicating most of the internal inconsistency.

An obvious caveat of the study is that it used quite a small sample size, and we should thus be cautious against over-interpreting the findings. Larger studies need to be undertaken to test whether the results presented in this article are generalisable, and if they are found to be generalisable, efforts ought to be made to elicit the most appropriate general value for λ. Tversky and Kahneman (1992) did, after all, estimate $\lambda = 2.25$ from gambles that incorporated money outcomes. In health contexts, where it is plausible that health outcomes that are perceived as losses will be associated with quite substantial negative utility, a higher λ might be more appropriate.

Although the actual elicitation of values for health outcomes is a descriptive exercise, health care decision making invariably involves an element of risk, and therefore a strong case can be made to argue that the process of eliciting values *ought* to internalise respondents' attitudes towards risk. If this normative proposition is accepted (as it is in this article), then a risk-based method has to be adopted for eliciting these values. Since the standard gamble, the most commonly accepted risk-based method, suffers from internal bias, there is a need to either develop the underlying methodology of this method so as to reduce the bias, or search for an alternative risk-based technique that is free of (or at least, less restricted by) the bias. This article serves as a contribution to how the standard gamble *might* be developed so as to reduce its level of internal inconsistency. The article has provided some evidence to suggest that, when processing the standard gamble, people focus upon the certain outcome, and view the success outcome in the gamble as a gain and the failure outcome as a loss. That is, the carriers of value are gains and losses, not final 'assets'. The evidence suggests that people place a 'disproportionate' weight on the failure outcome, because correcting for this hypothesised effect improved the internal consistency of the gamble. No evidence was found in support of probability transformation. Therefore, in terms of the standard gamble scenarios presented in this article, the results allow us to tentatively conclude that the application of only one of the two main prospect theory modifications to EU may be appropriate, and that a risk-based standard gamble that is free from the potentially biasing influence of loss aversion merits further investigation.

Acknowledgements

I indebted to Mike Jones-Lee, Peter Wakker and Anne Spencer for comments received on previous drafts of this article. I am also grateful to participants of the first annual meeting of the Preference Elicitation Group held at the LSE on December 11th 2002, in particular John Brazier, who discussed a draft of this article at that meeting. The useful comments of two anonymous referees are also gratefully acknowledged. Financial support for the study came from the Office of Health Economics and the ESRC (award number R00429834596). I alone remain responsible for all opinions and errors.

References

Abdellaoui, M., 2000, Parameter-free elicitation of utilities and probability weighting functions. Management Science 46, 1497-1512.

Bleichrodt, H., 2001, Probability weighting in choice under risk: an empirical test. Journal of Risk and Uncertainty 23, 185-198.

Bleichrodt, H. and J.L. Pinto, 2000, A parameter-free elicitation of the probability weighting function in medical decision analysis. Management Science 46, 1485-1496.

Bleichrodt, H., J.L. Pinto and P.P. Wakker, 2001, Using descriptive findings of prospect theory to improve prescriptive applications of expected utility. Management Science 47, 1498-1514.

Bleichrodt, H., J. van Rijn and M. Johannesson, 1999, Probability weighting and utility curvature in QALY based decision making. Journal of Mathematical Psychology 43, 238-260.

Camerer, C., 1995, Individual decision making. In: J.H. Kagel and A.E. Roth, eds., The handbook of experimental economics. Princeton University Press, Princeton, 587-703.

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

Cohen, M. and J.-Y. Jaffray, 1988, Certainty effect versus probability distortion: an experimental analysis of decision making under risk. Journal of Experimental Psychology: Human Perception and Performance 14, 554-560.

Diecidue, E. and P.P. Wakker, 2001, On the intuition of rank-dependent utility. Journal of Risk and Uncertainty 23, 281-298.

Edwards, W., 1955, The prediction of decisions among bets. Journal of Experimental Psychology 50, 201-214.

Fishburn, P.C., 1978, On Handa's 'new theory of cardinal utility' and the maximization of expected return. Journal of Political Economy 86, 321-324.

Gafni, A. and G.W. Torrance, 1984, Risk attitude and time preference in health. Management Science 30, 440-451.

Gonzalez, R. and G. Wu, 1999, On the form of the probability weighting function. Cognitive Psychology 38, 129-166.

Hershey, J.C. and P.J.H. Schoemaker, 1985, Probability versus certainty equivalence methods in utility measurement: are they equivalent? Management Science 31, 1213-1231.

Jansen, S.J.T, A.M. Stiggelbout, P.P. Wakker, T.P.M. Vliet Vlieland, J-W.H. Leer, M.A. Nooy and K. Kievit, 1998, Patient expected utilities for cancer treatments: A study on the feasibility of a chained procedure for the standard gamble and the time trade-off. Medical Decision Making 18, 391-399.

Johnson, E.J. and D.A. Schkade, 1989, Bias in utility assessments: further evidence and explanations. Management Science 35, 406-424.

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

Lattimore, P.M., J.R. Baker and A.D. Witte, 1992, The influence of probability on risky choice. Journal of Economic Behavior and Organization 17, 377-400.

Llewellyn-Thomas, H., H.J. Sutherland, R. Tibshirani, A. Ciampi, J..E. Till and N.F. Boyd, 1982, The measurement of patients' values in medicine. Medical Decision Making 2, 449-462.

Loomes, G., 1993, Disparities between health state measures: an explanation and some implications. In: W. Gerrard, ed., The economics of rationality. Routledge, London, 149-178.

Morrison, G.C., 1996, An assessment and comparison of selected methods for eliciting quality of life valuations from the public. PhD Thesis, University of York, York.

Oliver, A., 2003, A quantitative and qualitative test of the Allais paradox using health outcomes. Journal of Economic Psychology. In press.

Preston, M.G. and P. Baratta, 1948, An experimental study of the auction value of an uncertain outcome. American Journal of Psychology 61, 183-193.

Rottenstreich, Y. and C.K. Hsee, 1999, Money, kisses and electric shocks: on the affective psychology of risk. Working Paper, University of Chicago, Chicago.

Rutten-van Molken, M.P.M.H., C.H. Bakker, E.K.A. van Doorslaer and S. van der Linden, 1995, Methodological issues of patient utility measurement. Medical Care 33, 922-937.

Spencer, A.E., 1998, Value elicitation techniques: methodological issues involved in two-staged procedures. PhD Thesis, University of York, York, UK.

Torrance, G.W., 1986, Measurement of health state utilities for economic appraisal. Journal of Health Economics 5, 1-30.

Torrance, G.W. and D. Feeny, 1989, Utilities and quality-adjusted life years. International Journal of Technology Assessment in Health Care 5, 559-575.

Tversky, A. and C.R. Fox, 1995, Weighing risk and uncertainty. Psychological Review 102, 269-283.

Tversky, A. and D, Kahneman, 1986, Rational choice and the framing of decisions. Journal of Business 59, S251-S278.

Tversky, A. and D. Kahneman, 1991, Loss aversion in riskless choice: a reference dependent model. Quarterly Journal of Economics 107, 1039-1061.

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

Wakker, P.P. and D. Deneffe, 1996, Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. Management Science 42, 1131-1150.

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.

Wu, G. and R. Gonzalez, 1996, Curvature of the probability weighting function. Management Science 42, 1676-1690.

Wu, G. and R. Gonzalez, 1999, Nonlinear decision weights in choice under uncertainty. Management Science 45, 74-85.

Table 1. Summary of outcomes

	Years in full health		
	x_2	x_1	x_3
Practice 1	10	40	0
Question 1	10	40	0
Question 2	20	40	0
Question 3	30	40	0
Question 4	10	20	0
Question 5	10	30	0
Question 6	20	30	0
Question 7	20	30	10
Question 8	20	40	10
Question 9	30	40	10
Question 10	30	40	20

Table 2. Summary of results

Eliciting according to Eq. (1)

	The EU standard gamble	The WS standard gamble
10 years		
q.1 v q.4:	ns	**
q.1 v q.5:	ns	*
20 years		
q.2 v q.6:	ns	ns
q.2 v q.7:	*	ns
q.2 v q.8:	**	**
30 years		
q.3 v q.9:	**	**
q.3 v q.10:	ns	*

Eliciting according to Eq. (4)

	The prospect theory standard gamble	The prospect theory π standard gamble
10 years		
q.1 v q.4:	**	**
q.1 v q.5:	ns	ns
20 years		
q.2 v q.6:	ns	ns
q.2 v q.7:	ns	*
q.2 v q.8:	**	**
30 years		
q.3 v q.9:	*	**
q.3 v q.10:	ns	*

	The prospect theory λ standard gamble	The prospect theory partial-π standard gamble
10 years		
q.1 v q.4:	ns	**
q.1 v q.5:	ns	*
20 years		
q.2 v q.6:	ns	*
q.2 v q.7:	ns	ns
q.2 v q.8:	**	**
30 years		
q.3 v q.9:	ns	**
q.3 v q.10:	ns	*

Note: ns indicates that there is not a significant difference between the elicited basic reference and chained values; * indicates a significant difference at 5%; ** indicates a continued significant difference at 1%.

Table 3. Median and range values for the EU and the prospect theory λ standard gambles

	EU standard gamble		Prospect theory λ standard gamble	
	Median	Range	Median	Range
Basic ref. value				
q.1:	0.75	0.65 (0.35-1.00)	0.57	0.81 (0.19-1.00)
q.2:	0.90	0.30 (0.70-1.00)	0.79	0.49 (0.51-1.00)
q.3:	0.98	0.18 (0.82-1.00)	0.96	0.33 (0.67-1.00)
10 years				
q.4:	0.77	0.62 (0.38-1.00)	0.57	0.83 (0.17-1.00)
q.5:	0.77	0.43 (0.57-1.00)	0.57	0.67 (0.33-1.00)
20 years				
q.6:	0.92	0.51 (0.49-1.00)	0.83	0.72 (0.28-1.00)
q.7:	0.94	0.32 (0.68-1.00)	0.85	0.55 (0.45-1.00)
q.8:	0.96	0.38 (0.62-1.00)	0.88	0.61 (0.39-1.00)
30 years				
q.9:	0.99	0.18 (0.82-1.00)	0.96	0.39 (0.61-1.00)
q.10:	0.99	0.23 (0.77-1.00)	0.95	0.43 (0.57-1.00)

Note: All values rounded to 2 decimal places

Figure 1. The inverse S-shaped probability transformation curve

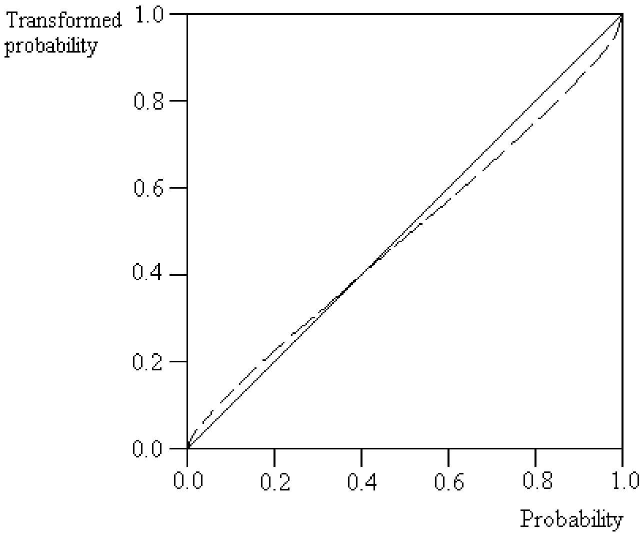


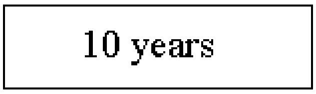
Figure 2. A typical context

Imagine that you go to your doctor for a routine medical examination. To your surprise, your doctor informs you that you have an unusual health condition. In this condition your doctor informs you that, without treatment, you will live for 10 more years in good health, and then you will die.

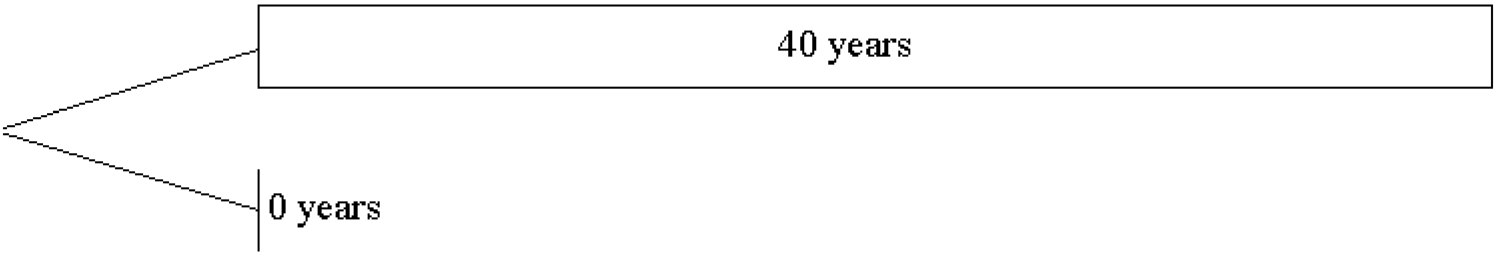
However, your doctor also informs you that there is a treatment for your condition, which, if taken, would give you a chance of living for 40 years in good health before death. However, there is also a chance that the treatment would kill you immediately.

So, your two options are:

1. Do not take the treatment and live for 10 years.



2. Take the treatment for a chance of living for 40 years and risk the chance of immediate death.



So, if the treatment is successful, you will live for 40 years.

Your doctor tells you that the size of the chance that the treatment will succeed is not known.

Please indicate on the scale below **the minimum chance of success** you would require for you to accept the treatment.