# An Experimental Test of Loss Aversion and Scale Compatibility

Han Bleichrodt, iMTA, Erasmus University, Rotterdam, The Netherlands

Jose Luis Pinto, Universitat Pompeu Fabra, Barcelona, Spain

*Address correspondence to:*

Han Bleichrodt, iMTA, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: bleichrodt@bmg.eur.nl.

*Acknowledgements:*

We are grateful to Peter Wakker for helpful comments on previous drafts of this paper. Han Bleichrodt's research was made possible by a fellowship from the Royal Netherlands Academy of Arts and Sciences. Financial support for the experiments reported in this paper was obtained from the Direccion General de Ciencia y Tecnologia (DGICYT PB 94-0848/95) and from Merck, Sharp and Dohme.

## Abstract

This paper studies two important reasons why people violate procedure invariance, loss aversion and scale compatibility. The paper extends previous research on loss aversion and scale compatibility by studying loss aversion and scale compatibility simultaneously, by looking at a new decision domain, medical decision analysis, and by examining the effect of loss aversion and scale compatibility on "well-contemplated preferences." We find significant evidence both of loss aversion and scale compatibility. However, the sizes of the biases due to loss aversion and scale compatibility vary over trade-offs and most participants do not behave consistently according to loss aversion or scale compatibility. In particular, the effect of loss aversion in medical trade-offs decreases with duration. These findings are encouraging for utility measurement and prescriptive decision analysis. There appear to exist decision contexts in which the effects of loss aversion and scale compatibility can be minimized and utilities can be measured that do not suffer from these distorting factors.

KEY WORDS: *Decision Analysis, Utility Theory, Loss Aversion, Scale Compatibility, Health)*

This paper studies two important explanations of why people's preferences deviate from procedure invariance, loss aversion and scale compatibility. Procedure invariance is the requirement that logically equivalent ways of measuring preferences yield identical results. For example, suppose we ask a client to specify how many years in full health he considers equivalent to living for 40 years with a back injury and he answers 30 years. Then procedure invariance requires that we obtain the same indifference if we ask this client instead to specify the number of years with a back injury that he considers equivalent to living for 30 years in full health. That is, the client's response to the latter question should be 40 years. Procedure invariance is a basic requirement of normative decision analysis. Without procedure invariance, preferences over decision alternatives cannot be measured unambiguously and, in the absence of normative grounds to prefer one method of presenting alternatives over another, the outcomes of decision analyses are equivocal.

Unfortunately, empirical research has displayed many systematic violations of procedure invariance. People's preferences have been shown to depend on the framing of the outcomes (Tversky and Kahneman 1986), the context in which preferences are elicited (Tversky and Simonson 1993), and the response scale used (Slovic 1975, Tversky, et al. 1988, Delquié 1993).

Two important models that can explain violations of procedure invariance are the reference-dependent model (Tversky and Kahneman 1991) and the contingent trade-off model (Tversky et al. 1988). The reference-dependent model posits that people frame outcomes as gains and losses with respect to a given reference point, which is often their current position. Reference-dependence in combination with loss aversion can lead to violations of procedure invariance. The contingent trade-off model explains violations of procedure invariance by scale compatibility: attributes of decision alternatives that are compatible with the response mode are weighted more heavily than those that are not.

The purpose of this paper is to study the impact of loss aversion and scale compatibility in causing violations of procedure invariance. The paper extends previous research on loss aversion and scale compatibility in three ways. First, we study the effects of loss aversion and scale compatibility simultaneously. Previous empirical studies typically focused either on loss aversion or on scale compatibility but did not examine the interaction between the two effects. This may have led to biased conclusions. As we show in Section 3, loss aversion and scale compatibility can interact in trade-offs. More robust tests of loss aversion and scale compatibility are obtained by studying the two effects simultaneously. Performing systematic tests of loss aversion and scale compatibility is important for decision analysis. If loss aversion and/or scale compatibility turn out to follow specific patterns then it may become possible to model these effects and to correct for them in utility measurement. The analogy with the literature on probability transformation may illustrate this point. In response to observed violations of expected utility theory, alternative models of decision under risk have been proposed in which probabilities are weighted through a probability transformation function (Quiggin 1982). The empirical literature on probability transformation has displayed systematic patterns in the way people distort probabilities (Tversky and Kahneman 1992, Tversky and Fox 1995, Wu and Gonzalez 1996). Specific parametric forms have been proposed for the probability transformation function that capture these systematic patterns (Tversky and Kahneman 1992, Lattimore, et al. 1992, Prelec 1998). Finally, it has been shown how these parametric forms can be used to correct for probability transformation in utility measurement (Wakker and Stiggelbout 1995). Regarding loss aversion, several authors have used models in which loss aversion is constant (Tversky and Kahneman 1992, Shalev 1997). Bleichrodt, et al. 2000 have derived how utilities should be corrected for constant loss aversion and they have shown that these corrections reduce the number of violations of procedure invariance.

Second, the paper extends previous research by studying loss aversion and scale compatibility in a new domain, medical decision making. The little evidence that is available on loss aversion and scale compatibility in medical trade-offs is indirect and ambiguous. Two-attribute trade-offs are generally used in health utility measurement and insight into the extent to which these trade-offs are affected by loss aversion and scale compatibility may contribute to the validity of health utility measures.

Third, we test to what extent loss aversion and scale compatibility are present if an attempt is made to elicit well-contemplated preferences. Most previous studies were primarily interested in whether loss aversion and scale compatibility existed and used question formats that were conducive to violations of procedure invariance. For example, most of these studies used matching to elicit indifference. It has been shown that matching is more likely to lead to inconsistencies in preferences than choice-based elicitation procedures (Bostic, et al. 1990). Displaying the presence of violations of procedure invariance is an important research topic. However, for practical decision analysis it is also important to examine to what extent loss aversion and scale compatibility are present under careful preference elicitation. The preferred way to elicit well-contemplated preferences is by the constructive preference approach (Payne, et al. 1999). However, the constructive preference approach is time consuming and requires sophisticated clients and interviewers. Therefore, it is often not feasible in practice. For example, in medical decision making utility elicitations are often performed by medical staff who lack both time and training to elicit preferences by the constructive preference approach. In such cases, it is important to use experimental procedures that are feasible and limit the impact of distorting factors as much as possible. Testing to what extent such experimental procedures are susceptible to loss aversion and scale compatibility is a third aim of this paper.

The paper has the following structure. The next two sections describe the reference-dependent model and the contingent trade-off model. These two models are applied in Section 3 to derive empirical tests of the effect of loss aversion with scale compatibility held constant, of the effect of scale compatibility with loss aversion held constant, and of the joint effect of loss aversion and scale compatibility. The latter test is derived for decision contexts where loss aversion and scale compatibility make conflicting predictions and therefore allows for an assessment of the relative strengths of the two effects. Sections 4 and 5 are empirical and describes the design respectively the results of an experiment aimed to perform the tests derived in Section 3. Section 6 concludes.

## 1. The reference-dependent model

Let X be a set of outcomes. The set of outcomes X is a Cartesian product of the attribute sets $X_1$ and $X_2$. The sets $X_1$ and $X_2$ are connected topological spaces and X is endowed with the product topology. A typical element of X is $(x_1, x_2)$, $x_1 \in X_1$, $x_2 \in X_2$. Let a preference relation Ë be defined over X, where Ë is assumed to be a *continuous weak order*, i.e., it is transitive and complete and the sets {x∈X: x Ë y} and {x∈X: x Ê y} are both closed for all y∈X. The relation Ë is the preference relation adopted by standard choice theory. As usual, É (strict preference) denotes the asymmetric part of Ë and ~ (indifference) the symmetric part. Preference relations over attributes are defined from Ë. Let $x_i\alpha$ denote the outcome that yields $x_i$ on attribute i and $\alpha$ on the other attribute. Then we define for $i\in\{1,2\}$ and $x_i\alpha$, $y_i\alpha \in X$, $x_i$ Ë $y_i$ iff $x_i\alpha$ Ë $y_i\alpha$ Throughout, we assume that the preference relation satisfies *attribute monotonicity*.

**Definition 1 (attribute monotonicity):**

The preference relation Ë satisfies *attribute monotonicity* if for all x, y $\in$ X with $x_j = y_j$, $j\in\{1,2\}$, either

(a) $x_i > y_i$ iff x É y, $i \in\{1,2\}$, $i\neq j$

or

(b) $x_i > y_i$ iff x È y, $i \in\{1,2\}$, $i\neq j$

Reference-dependent theory modifies standard choice theory by making the preference relation dependent on a given reference point. The reference point is often the current position of the individual. Instead of one preference relation Ë, as in standard choice theory, there is a family of indexed preference relations Ë$_r$, where x Ë$_r$ y denotes “x is at least as preferred as y judged from reference point r.” The reference-dependent relations of strict preference and indifference are denoted by É$_r$ and $\sim_r$ respectively. The preference relations Ë$_r$ are continuous weak orders and satisfy attribute monotonicity. The preference relations over single attributes are defined as under the standard choice model. Under attribute monotonicity, thesingle-attribute preference relations are independent of the reference point and we therefore denote them as before by Ë.

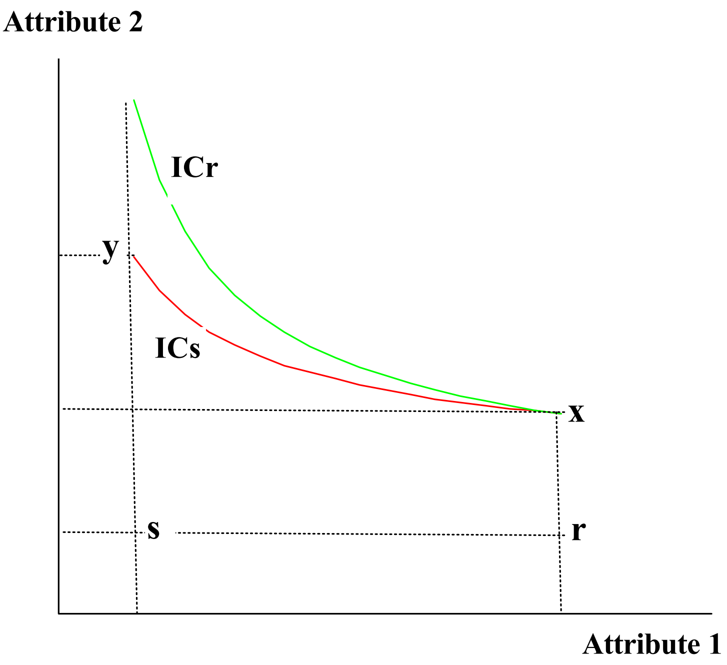
The distinctive predictions of reference-dependent theory follow from the assumptions about the impact of shifts in the reference point. Tversky and Kahneman

(1991) hypothesize that preferences satisfy *loss aversion*, which can be defined as follows.

**Definition 2 (loss aversion):**

Let $i,j\in\{1,2\}$, $i\neq j$. The preference relation satisfies *loss aversion* if for all $r,s,x,y\in X$ such that $x_i = r_i$ É $y_i = s_i$, and $r_j = s_j$, x Ë$_s$ y implies x É$_r$ y.

**Figure 1: An Illustration of Loss Aversion**



The intuition behind loss aversion is that losses loom larger than gains. Because a shift in the reference point can change losses into gains and vice versa, loss aversion can explain violations of procedure invariance. Figure 1 illustrates. Suppose that x and y are equivalent judged from reference point s. That is, the gain $y_2–s_2$ is just sufficient to offset the gains $x_1–s_1$ and $x_2–s_2$. If the reference point shifts from s to r then x yields the reference level and y a loss $r_1–x_1 = –(y_1–s_1)$ on the first attribute. The shift in the reference point does not affect the second attribute. Because losses loom larger than gains and no change occurs on the second dimension, x is now strictly preferred to y. If we draw indifference curves in Figure 1 then loss aversion implies that the indifference curves judged from reference point r, $IC_r$, are steeper than those from reference point s, $IC_s$.

Several empirical studies support loss aversion and closely related concepts as “endowment effects” (Kahneman, et al. 1990) and “status quo bias” (Samuelson and Zeckhauser 1988). Kühberger (1998) gives an overview of the impact of reference-dependence and loss aversion on risky decisions. Illustrations of the influence of reference-dependence and loss aversion on riskless decision making can, for example, be found in Tversky and Kahneman (1991), Bateman et al. (Bateman, et al. 1997), and Herne (Herne 1998).

## 2. The contingent trade-off model

The contingent trade-off model (Tversky et al. 1988) generalizes standard choice theory by making preferences conditional on the elicitation method used. In two-attribute trade-offs, preferences can either be elicited by the first ($x_1$) or by the second ($x_2$) attribute. Let $\text{Ë}_1$ and $\text{Ë}_2$ denote the preference relation when the first respectively the second attribute is used as the response scale. For i=1,2, $\text{É}_i$ is the asymmetric part of $\text{É}_i$ (strict preference) and $\sim_i$ the symmetric part (indifference). It is assumed that the $\text{Ë}_i$, i=1,2, are continuous weak orders and satisfy attribute monotonicity. The preference relations over single attributes are defined as under the standard choice model. Under attribute monotonicity, the single-attribute preference relations are independent of the response scale used and we therefore continue to denote them by Ë.

The distinctive predictions of the contingent trade-off model are a consequence of the effect of changes in the response scale. Tversky et al. (1988) impose *scale compatibility* to explain why preferences depend on changes in the response scale. Scale compatibility posits that an attribute becomes more important if it is used as response scale. Formally, scale compatibility can be expressed as follows.

**Definition 3 (scale compatibility)**

If x, y ∈ X are such that for i,j∈{1,2}, i≠j, $x_i$ É $y_i$ and $x_j$ È $y_j$ then x Ë$_j$ y implies x É$_i$ y.

**Figure 2: An Illustration of Scale Compatibility**

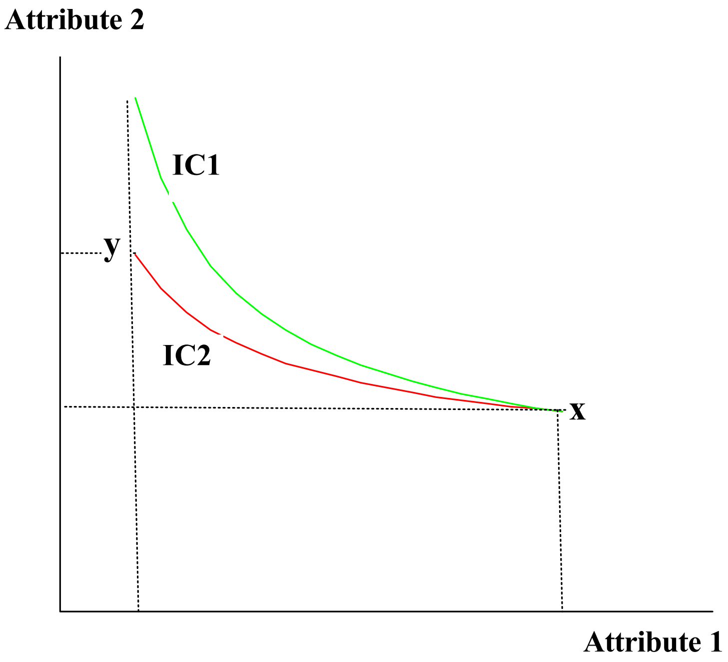


Figure 2 illustrates scale compatibility. The two preference relations Ë$_1$ and Ë$_2$ give rise to different sets of indifference curves $IC_1$ and $IC_2$. The indifference curves corresponding to Ë$_1$, $IC_1$, are steeper, reflecting that the first attribute gets more weight when it is used as the response scale. Figure 2 shows that if x and y lie on the same indifference curve when the second attribute is used as the response scale then x, which yields a strictly preferred level on the first attribute, lies on a higher, i.e. more preferred, indifference curve when the first attribute is used as the response scale.

Several studies have provided empirical support for scale compatibility. Delquié (1993) gives a comprehensive overview of the impact of scale compatibility on both decision under risk and decision under certainty. His results provide strong support for scale compatibility.

Two of the abovementioned studies provide some insight into the relative contribution of loss aversion and scale compatibility on trade-offs. The two studies yield conflicting results. Delquié (1993), who focused on scale compatibility, argues that the effect of scale compatibility is stronger than the effect of loss aversion. Bateman et al. (1997), whose aim was to test the influence of loss aversion, conclude that loss aversion is more important than scale compatibility.

## 3. Empirical tests

We used a linked equivalence design to derive empirical tests of loss aversion with scale compatibility constant, of scale compatibility with loss aversion constant, and of the joint effect of loss aversion and scale compatibility. Consider two outcomes $x = (x_1, x_2)$ and $y = (y_1, y_2)$, suppose that for both attributes higher levels are preferred to lower, and that $x_2 < y_2$. In the first stage of the linked-equivalence design, three of the parameters $x_1$, $x_2$, $y_1$, and $y_2$ are fixed and participants were asked to establish indifference between x and y by specifying the value of the remaining parameter. Suppose that $x_1$ is used to elicit indifference in the first stage and denote the first-stage response by $x_1$¢. It follows by attribute monotonicity that $x_1$¢ $> y_1$. In the second stage, $x_1$¢ is substituted and one of $x_1$, $x_2$, and $y_2$ is used to establish indifference. Standard choice theory predicts that the second-stage response should always be equal to the first-stage stimulus value. This follows immediately from transitivity of the indifference relation and attribute monotonicity. The second-stage response predicted by either the reference-dependent model or the contingent trade-off model can differ from the first-stage stimulus value depending on which parameter is used to elicit indifference. Table 1 provides an overview of the various possibiilities. A formal derivation of these predictions is given in Appendix A. Let us state for completeness that inequalities corresponding to an attribute reverse if that attribute is such that lower levels are preferred to higher levels. For example, if lower levels are preferred to higher levels on the first attribute then the reference-dependent model predicts that $y_1 > y_1¢¢$.

**Table 1: Predictions of the reference-dependent model (RDM) and the contingent trade-off model (CTO)**

| Parameter used in the second-stage | Prediction RDM with loss aversion | Prediction CTO with scale compatibility |
|---|---|---|
| $y_1$ | $y_1¢¢ > y_1$ | $y_1¢¢ = y_1$ |
| $x_2$ | $x_2¢¢ = x_2$ | $x_2¢¢ > x_2$ |
| $y_2$ | $y_2¢¢ > x_2$ | $y_2¢¢ < x_2$ |

NOTE**:** Second-stage responses are indicated by the symbol ″.

Table 1 displays how two-attribute trade-offs can be used to test loss aversion and scale compatibility. The use of $y_1$ to elicit indifference in the second stage of the linked equivalence questions yields a “pure” test of loss aversion. In this case, the contingent trade-off model predicts that the effect of scale compatibility is held constant. A pure test of scale compatibility is obtained if $x_2$ is used to elicit second-stage responses. The reference-dependent model predicts that this test will not be confounded by changes in loss aversion. Finally, the joint impact of scale compatibility and loss aversion is tested if $y_2$ is used to elicit indifference in the second stage. We study trade-offs where scale compatibility and loss aversion make conflicting predictions. This allows tests of the relative size of the two effects.

## 4. Experiment

### *Participants*

Fifty-one economics students at the University Pompeu Fabra, aged between 20 and 25 participated in the experiment. Participants were paid five thousand Pesetas (approximately $30 at the time of the experiment). The experiment was carried out in two personal interview sessions. The two sessions were separated by two weeks. Before the actual experiment was administered, we tested the questionnaire in several pilot sessions.

### *Questions*

The experiment consisted of three groups of questions. We describe each group of questions in turn.

#### *Group 1: Back pain questions*

In the first group of questions, health status was qualitative and participants were asked to make trade-offs between years with back pain and years in full health. Questions in which health status is qualitative are commonly used in health utility measurement. In Section 3, it was assumed that both attributes in the decision problem can be used to elicit indifference. This implies that both attributes must satisfy certain "richness conditions" (connectedness of the topological spaces) to ensure that such indifference values can always be found. The richness conditions are unlikely to be satisfied for qualitative health states. In health utility assessment utilities are therefore elicited by varying only the quantitative attribute, generally duration. The impact of scale compatibility can then not be tested, because the tests for scale compatibility require shifts in the response scale used. Therefore, the first group of questions, to which we refer as the "back pain questions" only tested for loss aversion. The back pain questions were included because of the common use in health utility measurement of questions with qualitative health status.

**Table 2: The description of back pain**

| |
|---|
| Unable to perform some tasks at home and/or at work |
| Able to perform all self care activities (eating, bathing, dressing) albeit with some difficulties |
| Unable to participate in many types of leisure activities |
| Often moderate to severe pain and/or other complaints |

Back pain was described to participants as level of functioning on four dimensions: daily activities, self care, leisure activities, and pain. Table 2 shows the description of back pain. This description was taken from the Maastricht Utility Measurement Questionnaire, a widely used instrument to describe health states in medical research (Rutten-van Mölken, et al. 1995). We selected the health state back pain, because it is a familiar condition and participants were likely to know people suffering from it. Full health was defined as no limitations on any of the four dimensions.

Table 3 displays the five trade-offs between years with back pain and years in full health. The first attribute is duration and the second health status (BP stands for back pain and FH for full health). A possible problem in these types of questions is that people always respond in round numbers, e.g. multiples of five. To reduce this problem, we did not use multiples of five as first-stage stimulus values.  We only used durations less than forty years to avoid so-called “ceiling effects.” We learnt from the pilot sessions that participants find it hard to perceive living for very long durations, say sixty additional years. Such perception problems act against loss aversion which predicts that participants’ second-stage response should exceed the first-stage stimulus value.

The final columns of Table 3 display the predicted responses according to the reference dependent model with loss aversion (RDM) and according to the contingent trade-off model (CTO). These predictions are derived in a similar way as Table 1.

**Table 3: The Back Pain Questions**

| Question | First stage | Second stage | Prediction<br>RDM | Prediction<br>CTO |
|---|---|---|---|---|
| 1 | [13,BP] vs [$x_1$¢,FH] | [$y_1$¢¢,BP] vs [$x_1$¢,FH] | $y_1$¢¢ > 13 | $y_1$¢¢ = 13 |
| 2 | [19,BP] vs [$x_1$¢,,FH] | [$y_1$¢¢,BP] vs [$x_1$¢,FH] | $y_1$¢¢ > 19 | $y_1$¢¢ = 19 |
| 3 | [24,BP] vs [$x_1$¢,,FH] | [$y_1$¢¢,BP] vs [$x_1$¢,FH] | $y_1$¢¢ > 24 | $y_1$¢¢ = 24 |
| 4 | [31,BP] vs [$x_1$¢,,FH] | [$y_1$¢¢,BP] vs [$x_1$¢,FH] | $y_1$¢¢ > 31 | $y_1$¢¢ = 31 |
| 5 | [38,BP] vs [$x_1$¢,,FH] | [$y_1$¢¢,BP] vs [$x_1$¢,FH] | $y_1$¢¢ > 38 | $y_1$¢¢ = 38 |

*Group 2: Migraine questions*

In the second group of questions, participants were asked to make trade-offs between duration and the number of days per month they suffer from migraine. Hence, health status was quantitative and both tests of loss aversion and scale compatibility were possible. Table 4 gives the description of migraine, for which we again used the Maastricht Utility Measurement Questionnaire. Migraine was selected as health status, because it is a relatively common disease and participants were likely to know people suffering from it.

**Table 4: The description of migraine**

| |
|---|
| Unable to perform normal tasks at home and/or at work |
| Able to perform all self care activities (eating, bathing, dressing) |
| Unable to participate in any type of leisure activity |
| Severe headache |

Table 5 displays the migraine questions asked. The first attribute denotes duration and the second the number of days with migraine per month. Again, we avoided round numbers in the first stage and durations longer than 40 years. Six trade-offs were asked in the first experimental session. Three questions used duration as the response scale and three questions used days with migraine as the response scale. In the second session, ten trade-offs were asked. Questions 6 to 11 test loss aversion with scale compatibility constant. Questions 12 and 13 test scale compatibility with loss aversion constant. Questions 14 and 15 test the joint effect of scale compatibility and loss aversion in trade-offs where they make opposite predictions. Questions 12-15 used the first-stage responses from questions 6-11. The predictions according to the reference dependent model with loss aversion and the contingent trade-off model are displayed in the final two columns of the table.

**Table 5: The Migraine Questions**

| Question | First stage | Second stage | Prediction RDM | Prediction CTO |
|---|---|---|---|---|
| 6 | [16,3] vs $[x_1',8]$ | $[y_1'',3]$ vs $[x_1',8]$ | $y_1'' > 16$ | $y_1'' = 16$ |
| 7 | [19,8] vs $[x_1',4]$ | $[y_1'',8]$ vs $[x_1',4]$ | $y_1'' > 19$ | $y_1'' = 19$ |
| 8 | [34,13] vs $[x_1',4]$ | $[y_1'',13]$ vs $[x_1',4]$ | $y_1'' > 34$ | $y_1'' = 34$ |
| 9 | [22,4] vs $[28, x_2']$ | $[22, y_2'']$ vs $[28, x_2']$ | $y_2'' < 4$ | $y_2'' = 4$ |
| 10 | [26,8] vs $[17, x_2']$ | $[26, y_2'']$ vs $[17, x_2']$ | $y_2'' < 8$ | $y_2'' = 8$ |
| 11 | [32,8] vs $[20, x_2']$ | $[32, y_2'']$ vs $[20, x_2']$ | $y_2'' < 8$ | $y_2'' = 8$ |
| 12 | [34,13] vs $[x_1',4]$ | [34,13] vs $[x_1', x_2'']$ | $x_2'' = 4$ | $4 < x_2'' < 13$ |
| 13 | [22,4] vs $[28,x_2']$ | [22,4] vs $[x_1'', x_2']$ | $x_1'' = 28$ | $22 < x_1'' < 28$ |
| 14 | [16,3] vs $[x_1',8]$ | $[16, y_2'']$ vs $[x_1',8]$ | $y_2'' < 3$ | $3 < y_2'' < 8$ |
| 15 | [26,8] vs $[17, x_2']$ | $[ y_1'',8]$ vs $[17, x_2']$ | $y_1'' > 26$ | $17 < y_1'' < 26$ |

*Group 3: Car accident questions*

In the third group of questions, health status was also quantitative. Participants were told that they had experienced a car accident as a result of which they are temporarily unable to walk. To restore their ability to walk, participants have to undergo rehabilitation therapy for some time. Rehabilitation sessions last 2 hours per day and result in moderate to severe pain for a couple of hours after the rehabilitation sessions. Participants were asked to elicit indifference between two types of therapy, described as intensive and less intensive therapy, that differ in the time that elapses until participants are able to walk again and the number of hours of pain after the rehabilitation sessions.

Table 6 displays the car accident questions. The first attribute denotes years until being able to walk again and the second the number of hours of pain after the rehabilitation sessions. Four trade-offs were asked in the first experimental session. In the second session, eight trade-offs were asked. Questions 16 to 19 test loss aversion while holding scale compatibility constant. Questions 20 and 21 test scale compatibility while holding loss aversion constant. Questions 22 and 23 test the joint effect of scale compatibility and loss aversion in trade-offs where they make opposite predictions. Questions 20-23 used the first-stage responses from questions 16-19. The predictions according to the reference dependent model with loss aversion and the contingent trade-off model are displayed in the final two columns of the table.

**Table 6: The Car Accident Questions**

| | | | Prediction | Prediction |
|---|---|---|---|---|
| Question | First stage | Second stage | RDM | CTO |
| 16 | [3,5] vs [$x_1$¢,2] | [$y_1$¢¢,5] vs [$x_1$¢,2] | $y_1$¢¢ < 3 | $y_1$¢¢ = 3 |
| 17 | [4,2] vs [$x_1$¢,3.5] | [$y_1$¢¢,2] vs [$x_1$¢,3.5] | $y_1$¢¢ < 4 | $y_1$¢¢ = 4 |
| 18 | [3,6] vs [7, $x_2$¢] | [3, $y_2$¢¢] vs [7, $x_2$¢] | $y_2$¢¢ < 6 | $y_2$¢¢ = 6 |
| 19 | [5,2] vs [1.5, $x_2$¢] | [5, $y_2$¢¢] vs [1.5, $x_2$¢] | $y_2$¢¢ < 2 | $y_2$¢¢ = 2 |
| 20 | [3,5] vs [$x_1$¢,2] | [3,5] vs [$x_1$¢, $x_2$¢¢] | $x_2$¢¢ = 5 | 2 < $x_2$¢¢ < 5 |
| 21 | [5,2] vs [1.5,$x_2$¢] | [5,2] vs [$x_1$¢¢, $x_2$¢] | $x_1$¢¢ = 1.5 | 1.5 < $x_1$¢¢ < 5 |
| 22 | [4,2] vs [$x_1$¢,3.5] | [4, $y_2$¢¢] vs [$x_1$¢,3.5] | $y_2$¢¢ < 2 | 2 < $y_2$¢¢ < 3.5 |

| | | | | |
|---|---|---|---|---|
| 23 | [3,6] vs [7, $x_2$¢] | [ $y_1$¢¢,6] vs [7, $x_2$¢] | $y_1$¢¢ $<$ 3 | 3 $<$ $y_1$¢¢ $<$ 7 |

### *Methods*

To avoid order effects, we varied the order in which the three groups of questions were administered. Similarly, within each group the order of the questions was varied.

Recruitment of participants took place one week before the actual experiment started. After recruitment, we gave participants three practice questions, one from each group. Participants were asked to answer these practice questions at home before coming to the experiment. This procedure was selected to familiarize participants with the questions. Prior to each group of questions, participants were asked whether they had experienced any problems in answering the practice question that corresponded to that group. Participants were then asked to explain their answer to the practice question. This allowed a test of whether participants had understood the questions. In case we were not convinced that participants had understood the question, we went over the task again until we were convinced participants understood the task.

Indifferences were elicited by a choice bracketing procedure. Participants reported their answers by filling in a table, as shown in Appendix B. At any time during the interview, participants were allowed to check earlier responses and to adjust these if desired. It is crucial for our test of the reference-dependent model that participants interpret the option in which both parameters are given as their reference point. We took special care to  formulate the questions in such a way that ambiguities about the reference point were ruled out. We consistently referred to the option in which both parameters were given as the participant’s current health state and to the option in which a parameter had to be specified as the health state to which the participant could change to. The choice-bracketing procedure used three answer categories: "I prefer my current health state," "I want to change to the other health state," and "I am indifferent between my current health state and a change to the other health state." After each question the participants were asked to confirm the elicited indifference value. The final comparison was displayed once again and participants were asked whether they agreed that the two options were equivalent. Appendix B shows the formulation of the "back pain questions". A similar format was used for the "migraine questions" and the "car accident questions."

The trade-offs used in this study are hypothetical. We do not believe that the hypothetical nature of the outcomes is problematic. Several studies have shown that people's responses to hypothetical and real tasks do not differ in a systematic way (Tversky and Kahneman 1992, Camerer 1995, Beattie and Loomes 1997, Camerer and Hogarth 1999). More specifically, previous studies on loss aversion demonstrated its effect on preferences both in hypothetical (Tversky and Kahneman 1986, Samuelson and Zeckhauser 1988, McDaniels 1992, Jones-Lee, et al. 1995) and in real tasks (Tversky and Kahneman 1991, Bateman et al. 1997).

In the introduction we mentioned that one of the aims of our study is to elicit "well-contemplated preferences." Let us summarize the steps we took to try and elicit such well-contemplated preferences: (i) several pilot sessions were organized in an attempt to develop an appropriate questionnaire; (ii) familiar health states were used; (iii) participants received practice questions to take home and had to explain their answers to these practice questions; (iv) personal interviews were used; (v) choice bracketing was used to elicit indifference; (vi) an indifference confirmation procedure was used.

## Results

One participant was excluded from the analyses because he did not answer all the questions. Because the tests of loss aversion and scale compatibility require attribute monotonicity, those participants who violated attribute monotonicity were excluded in each of the groups of questions. This left 42, 46, and 38 participants in the analyses of the back pain, migraine, and car accident questions respectively.[1]

### *Tests of loss aversion*

*Back pain questions*

**Figure 3: Loss Aversion in the Back Pain Questions**

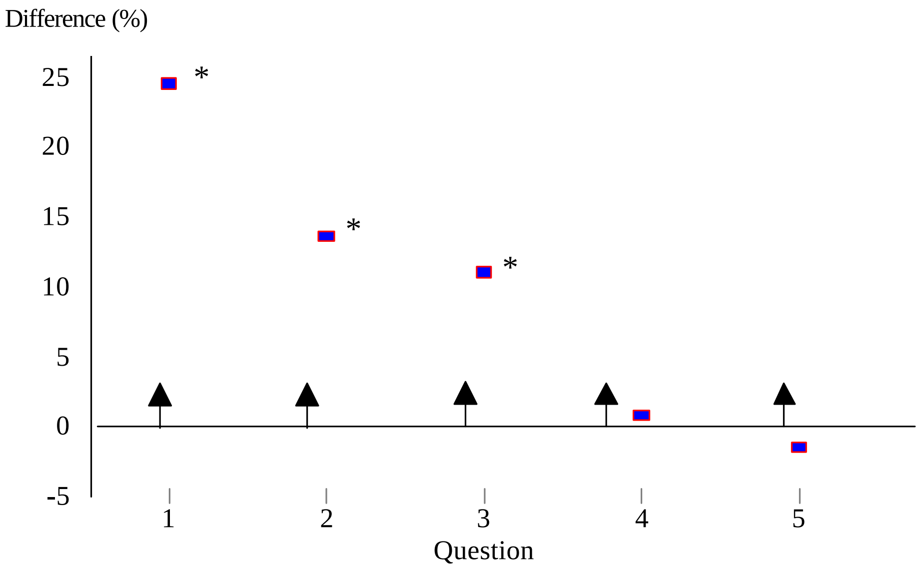


Figure 3 shows the difference between the second stage response and the first stage stimulus value expressed as a proportion of the first stage stimulus value for the five back pain questions. The solid arrows show the direction of the difference that is predicted by the reference-dependent model with loss aversion. The figure shows

---

[1] Thirty participants satisfied attribute monotonicity in all three groups of questions. The data were also analyzed using only these 30 participants. The results were qualitatively similar.

significant evidence of loss aversion. The figure also shows that the impact of loss aversion is not constant but decreases with duration. Loss aversion is significant in the first three back pain questions, but not in the final two questions. An explanation of why loss aversion decreases with duration may be that the substitutability of health status and duration increases with duration. Empirical studies have shown that loss aversion decreases with increases in substitutability (Ortona and Scacciati 1992, Chapman 1998). Mc Neil et al. (1981) found that the substitutability between health status and duration increases with duration. They observed that people are unwilling to trade duration for improved health status if duration is low. If duration increases, however, people become more prepared to trade-off duration and health status.

At the individual level, we find that most participants do not behave consistently according to the reference dependent model with loss aversion. Thirteen participants are *uniformly loss averse*, i.e., they behaved according to the predictions of the reference-dependent model with loss aversion in each question. One participant is *uniformly gain seeking*, i.e., he behaved contrary to the predictions of the reference-dependent model with loss aversion in each question. This preference pattern implies that the participant gives more weight to gains than to losses of the same size, hence the term gain seeking.

### *Migraine questions*

**Figure 4: Loss Aversion in the Migraine Questions**

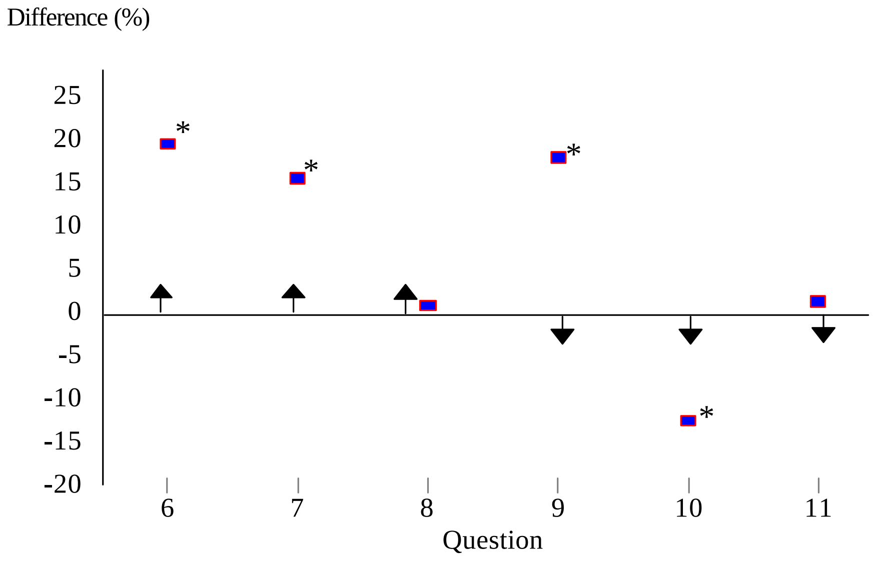


Figure 4 shows the results for the migraine questions. Note from Table 5 that duration increases in questions 6-8. Hence, we observe again that the degree of loss aversion decreases with duration. The pattern observed in Questions 9-11 is mixed. In question 10 there is significant loss aversion, in questions 9 there is a significant deviation in the opposite direction and in question 12 no significant difference exists between second stage response and first stage stimulus value. An explanation for the mixed evidence in questions 9-11 can be that participants were confused that the response scale in these questions (number of days with migraine) was such that lower values were preferred whereas on the other response scale (duration) higher values were preferred. Against this explanation pleads, however, that the number of violations of attribute monotonicity, a symptom of confusion in our opinion, is lowest in the migraine questions.

The mixed evidence with respect to loss aversion is reflected in the individual data. Only two participants are uniformly loss averse. No participant is uniformly gain seeking. Hence, we observe again that for most participants the effect of loss aversion is not constant but depends on the trade-off.

*Car accident questions*

**Figure 5: Loss Aversion in the Car Accident Questions**

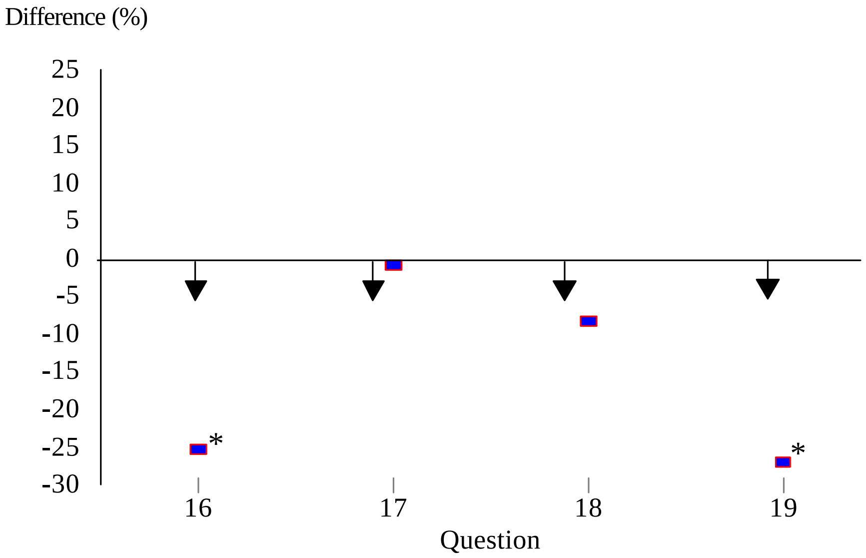
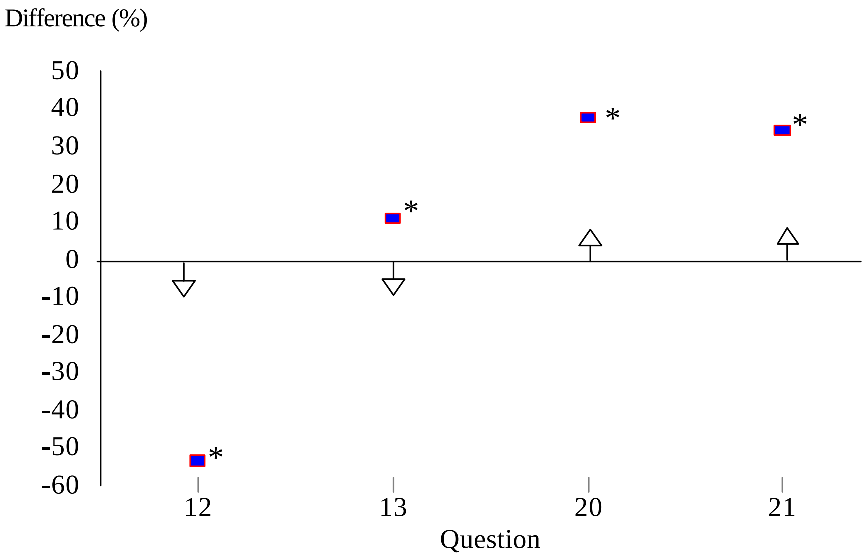


Figure 5 shows the results for the car accident questions. The reference-dependent model with loss aversion predicts a negative proportion in each question, as indicated by the solid arrows. This prediction is confirmed, but the difference between second-stage response and first stage stimulus value is only significant in two questions. Again, we observe at the individual level that most participants do not behave consistently according to the reference dependent model with loss aversion, but that there are trade-offs in which participants are loss averse and trade-offs in which participants are gain seeking. Six participants are uniformly loss averse. No participant is uniformly gain seeking.

### *Tests of scale compatibility*

**Figure 6: Scale Compatibility in the Migraine and Car Accident Questions**



Both the migraine questions and the car accident questions contained two tests of scale compatibility with loss aversion constant. Figure 6 displays the difference between second stage response and first stage stimulus value as a proportion of the first stage stimulus value. The open arrows indicate the direction of the difference predicted by the contingent trade-off model with scale compatibility. Three out of four tests support the contingent trade-off model with scale compatibility. In Question 13, however, the bias is in the opposite direction. Hence, we observe mixed results on scale compatibility in the migraine questions. Again, this may be due to the different "sign" of the reponse scales in the migraine questions.

There is only one participant behaves uniformly according to the contingent trade-off model with scale compatibility, i.e. his responses are in the direction predicted by the model in all four questions. No participant behaves opposite to the model in all four questions. Hence, the effect of scale compatibility depends on the trade-off for almost all participants.

### *Joint tests of loss aversion and scale compatibility*

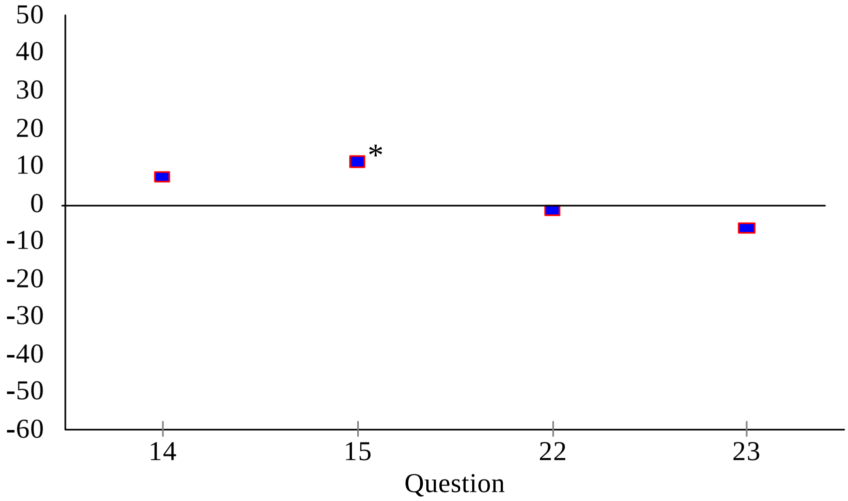
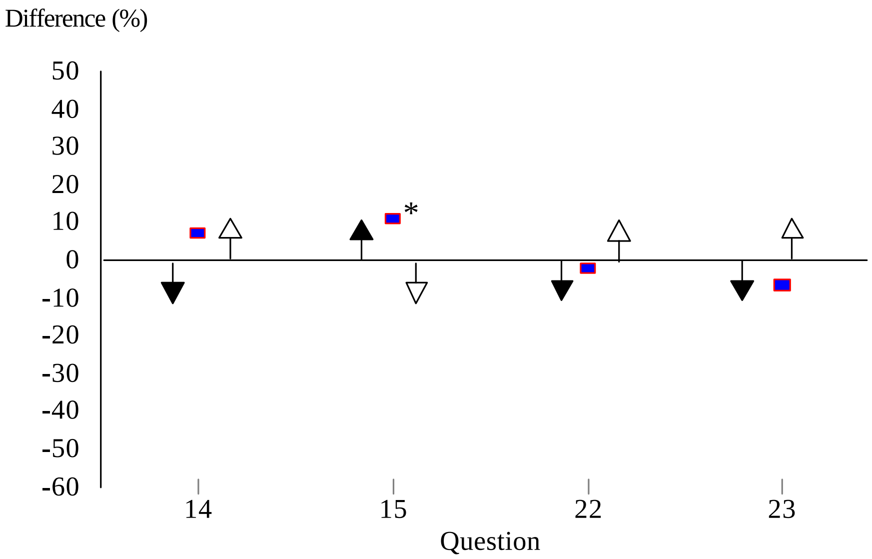


Figure 7: Loss Aversion and Scale Compatibility in the Migraine and Car Accident Questions



Four tests, two in the migraine questions and two in the car accident questions, test the joint impact of loss aversion and scale compatibility. Figure 7 shows the results. As before, solid arrows indicate the predictions of the reference-dependent model with loss aversion and open arrows indicate the predictions of the contingent trade-off model with scale compatibility. Figure 7 shows that the relative sizes of the biases due to loss aversion and scale compatibility depend on the trade-off. In three tests the bias is in the direction of loss aversion, in one in the direction of scale compatibility. However, the difference is only significant in Question 15. In particular in Question 22 the effects of loss aversion and scale compatibility more or less compensate each other.

The individual data confirm that the relative strengths of the biases due to loss aversion and scale compatibility are not constant but vary over trade-offs. There are only 2 participants for whom the effect of loss aversion dominates the effect of scale compatibility in all 4 questions. There is no participant for whom the effect of scale compatibility dominates the effect of loss aversion in all 4 questions.

## 5. Conclusion

The overall message of the paper is positive for prescriptive decision analysis. The finding that loss aversion is not constant and that people are not consistently loss averse suggests that it is possible to identify decision contexts that are hardly affected by loss aversion. The identification of such decision contexts may allow the measurement of utilities without the distorting impact of loss aversion. For example, our results suggest that health utility measurement should rely on trade-offs between health status and duration where duration is relatively long, because the effect of loss aversion decreases with duration. The results on the joint effect of loss aversion and scale compatibility show that there exist trade-offs in which loss aversion and scale compatibility approximately outweigh each other. Future research should examine to what extent our preliminary findings on the joint efect of loss aversion and scale compatibility can be generalized. By identifying situations in which the two effects cancel out, it may become possible to obtain utilities that are approximately free of the distorting influences of loss aversion and scale compatibility.

## Appendix A: Derivation of the empirical tests

Throughout this appendix superscripts $'$ and $''$ denote first-stage respectively second-stage responses. We consider three cases depending on which parameter is used to elicit indifference in the second stage. Recall that it is assumed that on both attributes higher levels are preferred to lower levels and that $x_2 < y_2$.

CASE 1: $y_1$ is used to elicit indifference in the second stage.

In this case the first attribute is still used as the response scale, but the outcome in which both parameters are held fixed changes from x to y. The contingent trade-off model predicts that $y_1'' = y_1$, because the response scale does not change and, thus, the second stage yields $(x_1', x_2) \sim_1 (y_1'', y_2)$. $y_1'' = y_1$ then follows from transitivity of $\sim_1$ and attribute monotonicity.

According to the reference-dependent model, the reference point shifts from y to x and the second stage elicits $(x_1', x_2) \sim_x (y_1'', y_2)$. Let z denote the point $(x_1', x_2)$. By loss aversion, the first stage indifference $(x_1', x_2) \sim_y (y_1, y_2)$ implies $(x_1', x_2) \succ_z (y_1, y_2)$. Let $(x_1', x_2) \sim_z (z_1, y_2)$. Such a $z_1$ exists, because $X_1$ is a connected topological space and $\succsim$ is continous. By attribute monotonicity $z_1 > y_1$. Loss aversion also implies that if $(y_1'', y_2) \sim_x (x_1', x_2)$ then $(y_1'', y_2) \succ_z (x_1', x_2)$. By transitivity, $(y_1'', y_2) \succ_z (z_1, y_2)$ and by attribute monotonicity $y_1'' > z_1 > y_1$.

CASE 2: $x_2$ is used to elicit indifference in the second stage.

In this case, y is still the outcome in which both parameters are given, but the response scale changes from the first to the second attribute. The reference-dependent model predicts that $x_2'' = x_2$. This follows straightforwardly from transitivity of $\sim_y$, attribute monotonicity, and the second-stage indifference $(x_1', x_2'') \sim_y (y_1, y_2)$.

According to the contingent trade-off model, the second stage elicits $(x_1', x_2'') \sim_2 (y_1, y_2)$. By attribute monotonicity, $y_2 > x_2''$ and thus by scale compatibility $(x_1', x_2'') \succ_1 (y_1, y_2)$. The first stage yielded $(x_1', x_2) \sim_1 (y_1, y_2)$ and hence $(x_1', x_2'') \succ_1 (x_1', x_2)$ by transitivity of $\succsim_1$. Attribute monotonicity implies $x_2'' > x_2$.

CASE 3: $y_2$ is used as the second-stage response scale.

In this case, there is both a shift in the reference point from y to x and a change in the attribute that is used as the response scale. We show that the reference-dependent model and the contingent trade-off model yield conflicting predictions. The reference-dependent model predicts that $y_2'' > y_2$. Let z denote the point $(x_1', y_2)$. By loss aversion, the first stage indifference $(x_1', x_2) \sim_y (y_1, y_2)$ implies $(x_1', x_2) \succ_z (y_1, y_2)$. Let $z_2$ be such that $(x_1', x_2) \sim_z (y_1, z_2)$. $z_2$ exists by conectedness of $X_2$ and continuity of $\succsim$. $z_2 > y_2$ by attribute monotonicity. Loss aversion implies that if $(x_1', x_2) \sim_x (y_1, y_2'')$ then $(y_1, y_2'') \succ_z (x_1', x_2)$. Because $(x_1', x_2) \sim_z (y_1, z_2)$, attribute monotonicity implies that $y_2'' > z_2$. Hence, $y_2'' > y_2$.

According to the contingent trade-off model, the second stage elicits $(y_1, y_2'') \sim_2 (x_1', x_2)$. $y_2'' > x_2$ by attribute monotonicity. By scale compatibility, $(x_1', x_2) \succ_1 (y_1, y_2'')$. $(x_1', x_2) \succ_1 (y_1, y_2'')$ and the first stage indifference $(x_1', x_2) \sim_1 (y_1, y_2)$ imply $(y_1, y_2) \succ_1 (y_1, y_2'')$ by transitivity of $\succsim_1$. Hence, $y_2'' > y_2$ by attribute monotonicity.

## Appendix B: Formulation of the back pain question

Suppose that you have 13 more years to live with back pain. In this question you are asked to state the number of years in full health that you consider equivalent to living for 13 more years with back pain. That is, you have to determine the number Y that makes the following two options equivalent:

1. Living for 13 years with back pain. After these 13 years you die.
2. Living for Y years in full health. After these Y years you die.

Use the following table to answer this question.

| | **Your current situation is 1** | **You can change to situation 2** | **DECISION** | | |
|---|---|---|---|---|---|
| | **Situation 1** | **Situation 2** | | | |
| Step | Years with back pain | Years in full health | I remain in 1 | I am indifferent between 1 and 2 | I change to 2 |
| 1 | 13 | 13 | | | |
| 2 | 13 | 0 | | | |
| 3 | 13 | 11 | | | |
| 4 | 13 | 2 | | | |
| 5 | 13 | 9 | | | |
| 6 | 13 | 4 | | | |
| 7 | 13 | 7 | | | |
| 8 | 13 | 5 | | | |

## References:

I. Bateman, A. Munro, B. Rhodes, S. C. and R. Sugden (1997). A Test of the Theory of Reference-Dependent Preferences. *Quarterly Journal of Economics,* 62, 479-505.

J. Beattie and G. Loomes (1997). The Impact of Incentives Upon Risky Choice Experiments. *Journal of Risk and Uncertainty,* 14, 155-168.

H. Bleichrodt, J.-L. Pinto and P. P. Wakker (2000). Using Descriptive Findings of Prospect Theory to Improve Prescriptive Applications of Expected Utility. Mimeo.

R. Bostic, R. J. Herrnstein and R. D. Luce (1990). The Effect on the Preference Reversal of Using Choice Indifferences. *Journal of Economic Behavior and Organization,* 13, 193-212.

C.F. Camerer (1995). Individual Decision Making. In: J. Kagel and A. Roth, *The Handbook of Experimental Economics* (pp. 587-703). Princeton, NJ: Princeton University Press.

C. F. Camerer and R. M. Hogarth (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty,* 19, 7-42.

G. B. Chapman (1998). Similarity and Reluctance to Trade. *Journal of Behavioral Decision Making,* 11, 47-58.

P. Delquié (1993). Inconsistent Trade-Offs between Attributes: New Evidence in Preference Assessment Biases. *Management Science,* 39, 1382-1395.

K. Herne (1998). Testing the Reference-Dependent Model: An Experiment on Asymmetrically Dominated Reference Points. *Journal of Behavioral Decision Making,* 11, 181-192.

M. W. Jones-Lee, G. Loomes and P. R. Philips (1995). Valuing the Prevention of Non-Fatal Road Injuries: Contingent Valuation vs. Standard Gambles. *Oxford Economic Papers,* 47, 676-695.

D. Kahneman, J. L. Knetsch and R. H. Thaler (1990). Experimental Test of the Endowment Effect and the Coase Theorem. *Journal of Political Economy,* 98, 1325-1348.

P. M. Lattimore, J. R. Baker and A. D. Witte (1992). The Influence of Probability on Risky

Choice. *Journal of Economic Behavior and Organization,* 17, 377-400.

T. L. McDaniels (1992). Reference Points, Loss Aversion, and Contingent Values for Auto Safety. *Journal of Risk and Uncertainty,* 5, 187-200.

G. Ortona and F. Scacciati (1992). New Experiments of the Endowment Effect. *Journal of Economic Psychology,* 13, 277-296.

J. Payne, J. R. Bettman and D. A. Schkade (1999). Measuring Constructed Preferences: Towards a Building Code. *Journal of Risk and Uncertainty,* 19, 243-270.

D. Prelec (1998). The Probability Weighting Function. *Econometrica,* 66, 497-528.

J. Quiggin (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization,* 3, 323-343.

M. P. Rutten-van Mölken, C. H. Bakker, E. K. A. van Doorslaer and S. van der Linden (1995). Methodological Issues of Patient Utility Measurement. Experience from Two Clinical Trials. *Medical Care,* 33, 922-937.

W. Samuelson and R. Zeckhauser (1988). Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty,* 1, 7-59.

J. Shalev (1997). Loss Aversion in a Multi-Period Model. *Mathematical Social Sciences,* 33, 203-226.

P. Slovic (1975). Choice Between Equally Valued Alternatives. *Journal of Experimental Psychology: Human Perception and Performance,* 1, 280-287.

A. Tversky and C. Fox (1995). Weighing Risk and Uncertainty. *Psychological Review,* 102, 269-283.

A. Tversky and D. Kahneman (1986). Rational Choice and the Framing of Decisions. *Journal of Business,* 59, 251-278.

A. Tversky and D. Kahneman (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics,* 56, 1039-1061.

A. Tversky and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty,* 5, 297-323.

A. Tversky, S. Sattath and P. Slovic (1988). Contingent Weighting in Judgment and Choice.

*Psychological Review,* 95, 371-384.

A. Tversky and I. Simonson (1993). Context-Dependent Preferences. *Management Science,* 39, 1179-1189.

P. P. Wakker and A. M. Stiggelbout (1995). Explaining Distortions in Utility Elicitation Through the Rank-Dependent Model for Risky Choices. *Medical Decision Making,* 15, 180-186.

G. Wu and R. Gonzalez (1996). Curvature of the Probability Weighting Function. *Management Science,* 42, 1676-1690.