Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories

PETER WAKKER
Medical Decision Making Unit, University of Leiden, Leiden, The Netherlands

IDO EREV
Faculty of Industrial Engineering and Management, Technion, Haifa, 32000, Israel

ELKE U. WEBER
Center for Decision Research, Graduate School of Business, University of Chicago.

Abstract

This article compares classical expected utility (EU) with the more general rank-dependent utility (RDU) models. The difference between the independence condition for preferences of EU and its comonotonic generalization in RDU provides the exact demarcation between EU and rank-dependent models. Other axiomatic differences are not essential. An experimental design is described that tests this difference between independence and comonotonic independence in its most basic form and is robust against violations of other assumptions that may confound the results, in particular the reduction principle and transitivity. It is well known that in the classical counterexamples to EU, comonotonic independence performs better than full-force independence. For our more general choice pairs, however, we find that comonotonic independence does not perform better. This is contrary to our prior expectation and suggests that rank-dependent models, in full generality, do not provide a descriptive improvement over EU. For rank-dependent models to have a future, submodels and choice situations need to be identified for which rank-dependence does contribute descriptively.

Key words: rank-dependence, nonexpected utility, comonotonicity, prospect theory, independence

Rank-dependent utility theory, first introduced by Quiggin (1982), is presently perhaps the most popular alternative to expected utility as a model of risky decisions. It permits not only the transformation of outcomes into utilities, as expected utility theory, but also the transformation of probabilities into nonlinear, nonadditive, decision weights. This provides a more natural way of modeling risk attitudes, because probabilities are more directly related to intuitive notions of risk than outcomes. Transformations of probabilities have also been adopted in older theories (e.g., Edwards, 1962), where probabilities for fixed outcomes were transformed rather than "cumulative" probabilities as in rank-dependent utility. These theories, however, had the problem of implying violations of stochastic dominance. Quiggin (1982) developed rank-dependent utility to preserve the idea of transforming probabilities while at the same time satisfying stochastic dominance.

This article describes an experiment that critically tests rank-dependent utility (RDU) against expected utility (EU). We first describe the critical axiomatic difference between the two theories and then demonstrate that it is simple enough for experimental testing. Such a comparison that tests RDU against EU in full generality without confounding it with other assumptions has not been performed before.

Section 2 demonstrates, in an informal manner, that the difference between EU and RDU is exactly the difference between the independence and the comonotonic independence assumption, which are defined below. We show that this is the only essential difference between the two theories, and that other axiomatic differences are nonessential. Formal derivations of these assertions are provided in Wakker (1994b).

Our experiment shows that, at least for the choice alternatives employed, the assumption of comonotonic independence does not describe people's choices better than the (stronger) assumption of independence. That is, this article reports a negative finding. To validate our experimental procedure, we included some classical Allais-type choices. Here our findings fully agreed with the well-established findings in the literature, namely, that comonotonic independence holds where general independence fails. However, the major part of our experiment consisted of gambles with only risky outcomes, which are not subject to the Allais certainty effect (Kahneman and Tversky, 1979), or other special biases. For these gambles, comonotonic independence did not improve upon independence.

Our experiment employed simple choice alternatives with known probabilities. There was no ambiguity concerning probabilities. Our experimental results thus do not speak to decision making in situations of ambiguity. The theoretical results that the difference between independence and comonotonic independence provides the demarcation between EU and RDU theory, do hold for ambiguous as well as risky choices.

Let us briefly comment on prior experimental tests between EU and RDU models in the literature; more extensive comments are provided in section 6. Such tests have usually been conducted on a domain of gambles consisting of all probability distributions over three fixed outcomes (referred to as the "probability triangle"). Observation 3 in section 6 will show that this domain does not provide a suitable structure for testing comonotonic independence. Another complication is that comonotonic independence is best tested in choice pairs that have a "states-of-nature-format," where the outcomes of both alternatives are related to the same set of underlying events (in our case, events with known probabilities), rather than each to a different set of events and probabilities.

Previous experiments usually restrict RDU additionally, for example by assuming special parametric forms for the utility and weighting functions, and test only those restrictions (Camerer, 1989; Battallio, Kagel, and Jiranyakul, 1990; Starmer, 1992). An exception is Starmer and Sugden (1989, Theorem 1), who indeed test an implication of the general RDU model. The tested property, however, is more complicated than comonotonic independence, and is not critical in providing an exact demarcation. Thus, we will not describe it here. In summary, none of the evidence obtained so far can be considered conclusive on the issue of RDU versus EU. When comparing RDU with the other main stream of transitive nonexpected utility models, i.e., the "betweenness" models, Camerer (1992) and Camerer and Ho (1994) strongly favor RDU.[1]

The experiment reported in this article, where gambles were presented in a "states-of-nature-format," is robust against the often found violations of the "reduction principle" (i.e., that gambles can be reduced to probability distributions over outcomes), and against possible violations of transitivity. This will be explained in further detail below. Our results are also not affected by possible "state-dependence" of the utilities, i.e., the phenomenon that an outcome obtained as the result of different events may be appreciated differently. Finally, the states-of-nature format may be more understandable and natural for respondents not used to dealing with numerical probabilities.

1. History, motivation, and preview

Von Neumann and Morgenstern (1947) and Savage (1954) laid the axiomatic foundation of EU by describing the preference conditions that are necessary and sufficient for EU maximization. In decision theory, preferences are taken as observable primitives; thus, axiomatizations describe the testable implications of a model of choice. Von Neumann and Morgenstern considered the case of decision under risk, where the probabilities with which outcomes occur are known and well specified. Savage considered the case of decision under uncertainty, where probabilities need not be known.

In these and all subsequent alternative axiomatizations of EU, the most important axiom is an independence condition. It states that preferences between gambles should be independent of those events for which the gambles give the same outcome. Either independence is imposed explicitly, as in Savage's sure-thing principle (Postulate P2), or it is implied by the other conditions.

Several axiomatizations have been developed for RDU, both for choice under risk and under uncertainty, weakening or modifying one or more of the EU axioms. While independence is implied in all EU models, only a weakened version, "comonotonic independence," is implied by all RDU models. We shall show that this weakening is the essential generalization of EU as provided by RDU. Given this difference, other differences in axioms found in the literature are, in an empirical sense, cosmetic and nonessential. The difference between independence and comonotonic independence is at the heart of the matter and therefore is the object of our experiment.

The empirical test between independence and comonotonic independence, reported below, was designed to be as "pure" as possible. The experiment was devised to avoid all possible extraneous factors that may obscure the test. Only three-outcome gambles, the simplest kind of gambles for testing comonotonic independence, were used. The tests of patterns of indifference curves in the probability triangle, commonly found in the literature, are based on many other assumptions such as transitivity, the reduction principle, and independence of utility from events. Any violations of these conditions, often confirmed, disturb these tests of RDU. Our experiment does not invoke any of these assumptions, hence is more basic.

The experiment was restricted to the simple case where uncertainty was generated by the drawing of a card from a deck, providing both states and probabilities. Therefore our

findings directly apply to both risk, where lack of information is described through probabilities, and uncertainty, where lack of information is described through states. The study of cases where no probabilities are given, or probabilities are vague or ambiguous, is a topic for future research. At present we make no empirical claims about such cases.

2. Independence versus comonotonic independence as the demarcation between EU and RDU

This section describes the EU and RDU theories, and describes in an informal manner the theoretical results derived in Wakker (1994b). In particular, this section shows how EU and RDU imply independence and comonotonic independence, respectively.

A comment on terminology is in order. For decision under uncertainty, the term SEU (subjective expected utility) is often used; EU is then reserved for decision under risk. As decision under risk is a special case of decision under uncertainty, EU is a special case of SEU. Still, the term conventionally used to refer to both is EU, rather than SEU. Similarly, for the generalized utility theories, the term "Choquet expected utility" is often used for decision under uncertainty, and RDU is reserved for decision under risk. However, we shall use RDU in this article as the common term to refer to both cases.

The choices in our experiment were between gambles of the following type: subjects were told that a card would be drawn from a deck of n numbered cards. Gambles yielded different outcomes, i.e., amounts of money, depending on the number of the card selected. We considered only gains to avoid effects of sign-dependence, i.e., the often confirmed finding that risk attitudes are different for gains than for losses. This was done to test rank-dependence in its purest form, with a minimum of other effects that might cause deviations from expected utility. Formalizations of sign-dependence, in addition to rank-dependence, can be found in Luce and Fishburn (1991), and in Tversky and Kahneman's (1992) cumulative prospect theory.

For decision under risk, it is assumed that all cards are equally likely to be drawn. Then each card has an (objective) probability of $1/n$ of being selected, and a given gamble can be described as a probability distribution over the outcomes. For decision under uncertainty the assumption of equal or known probabilities for the selection of cards is no longer imposed. Thus decision under risk is a special case of decision under uncertainty.

In *expected utility* (EU), a gamble is valuated by $\sum_{j=1}^{n} p_j U(x_j)$. For our card-drawing scenario, p_j is the probability that the jth card is selected, x_j is the resulting outcome if card j is selected, and $U(x_j)$ denotes the utility of outcome x_j. For decision under risk the probabilities p_j are known to be $1/n$. For decision under uncertainty they may be different, namely, the subjective probabilities that the decision maker assigns to the events.

Both for decision under uncertainty and for decision under risk, let $P(A)$ denote the probability of a subset $A \subset \{1, \ldots, n\}$, i.e., the probability that the selected card will be an element of A. Subsets A of $\{1, \ldots, n\}$ are called *events*. The gamble considered above can be described by the n-tuple of its possible outcomes, $(x_1, \ldots, x_n)$, or x for short.

To describe the independence condition, suppose that $x \succcurlyeq y$, i.e., gamble x is preferred to gamble y. Under EU, this means that

$$\sum_{j=1}^{n} p_j U(x_j) \geq \sum_{j=1}^{n} p_j U(y_j). \tag{1}$$

Then, if $x_1 = y_1$, the inequality will not be affected if we replace x_1 by x_1' and y_1 by y_1' as long as still $x_1' = y_1'$. In other words, a preference between two gambles is not affected if a common outcome is replaced by another common outcome. By repeated application, preference is also not affected if we replace several common outcomes. This is what *independence* requires.

In the literature on decision under risk, independence is mostly formulated in a way different from ours. In Machina (1989), our version is called replacement separability and the alternative condition is called mixture separability. Fishburn and Wakker (1992) show that replacement separability is somewhat weaker than mixture separability, but in the presence of usual continuity conditions they are equivalent. Chew and Wakker (1993) and Wakker (1994b) use the term comonotonic sure-thing principle instead of our term comonotonic independence.

In the evaluation of gambles, EU uses a probability P that is *additive*. That is, for nonoverlapping sets A, B of cards, $P(A \cup B) = P(A) + P(B)$. This property is generalized in RDU, which uses a nonadditive measure, called *capacity*, and denoted by W. W assigns capacity 0 to the empty set $(W(\emptyset) = 0)$, capacity 1 to the universal event $(W(\{1, \ldots, n\}) = 1)$, and larger capacities to larger events (i.e., if A contains B, then $W(A) \geq W(B)$). While sharing these properties with probabilities, capacities need not be additive.

For the calculation of the RDU value of a gamble x, the rank-ordering of the outcomes of x is relevant. Suppose first that $x_1 \leq \cdots \leq x_n$. Then the RDU value of the gamble is $\sum_{j=1}^{n} \pi_j U(x_j)$, where U denotes utility as in EU and the π_j are *decision weights*, derived from W by $\pi_j = W(j, \ldots, n) - W(j + 1, \ldots, n)$. This expression shows that the decision weight of card j is its marginal capacity contribution to all superior-outcome cards.

In general, if the ordering of outcomes is not $x_1 \leq \cdots \leq x_n$, then we determine the worst outcome $x_{1 : n}$ (where $1{:}n$ denotes the number of the card giving the worst outcome), the second-worst outcome $x_{2:n}, \ldots$, and the best outcome $x_{n:n}$. The ranks of equal outcomes can be chosen arbitrarily. The RDU value of the gamble is

$$\sum_{j=1}^{n} \pi_j U(x_j) \tag{2}$$

where the decision weights π_j are as follows. For $i : n = j$, $\pi_j = W(i : n, \ldots, n : n) - W((i + 1) : n, \ldots, n : n)$. Therefore, again, the decision weight of card j is its marginal capacity contribution to all superior-outcome cards.[2] If W is a probability, i.e., is additive, then π_j reduces to the probability $W(\{j\})$ for selecting the jth card and RDU reduces to EU.

For decision under risk, where objective probabilities $P(A)$ are given for events A, it is assumed that the only relevant aspect of events is their probability. $W(A)$ then depends

only on $P(A)$ and can be written as a function of $P(A)$, i.e., as $w(P(A))$ for $w:[0,1] \rightarrow [0,1]$. Capacities are now transformed objective probabilities. For the gamble x above, with $x_1 \leq \cdots \leq x_n$, $\pi_j = w((n - j + 1)/n) - w((n - j)/n)$ results.

Two gambles that induce the same ranking of outcomes (i.e., generate the same numbers $1:n, \ldots, n:n$), are called *comonotonic*. In the calculation of their RDU value, the same decision weights π_j are used. If comonotonic gambles have a common outcome (i.e., $x_j = y_j$), and this outcome is changed into another common outcome $(x'_j = y'_j)$ without affecting the rank-order of the outcomes, then preference should not be affected, as illustrated in more detail below. This condition is called *comonotonic independence*.

To derive comonotonic independence from RDU, suppose $x \succcurlyeq y$, for two gambles x, y with $x_j = y_j$. Suppose further that the gambles are comonotonic (e.g., $x_1 \leq \cdots \leq x_n$ and $y_1 \leq \cdots \leq y_n$). An RDU evaluation for the preference results in

$$\sum_{j=1}^{n} \pi_j U(x_j) \geq \sum_{j=1}^{n} \pi_j U(y_j). \tag{3}$$

In this expression, the decision weights π_j are the same for both gambles, because the gambles are comonotonic.

In general, if we replace the common outcomes $x_j = y_j$ by other common outcomes $x'_j = y'_j$, the effect of this replacement cannot be predicted. The reason is that the new common outcomes may change the relative rank-order of the outcomes of gamble x (e.g., $x_j < x_{j+1}$, but $x'_j > x_{j+1}$) and may also do so for gamble y. Because W maps A into decision weights using the ranks of outcomes, a change in rank-order changes the decision weights π_j, which may affect the inequality and the preference. (Examples will be given below.) Hence independence need not hold under RDU.

Suppose that the change in common outcomes does not alter the rank-ordering of the outcomes in either x or y. That is, if $x_1 \leq \cdots \leq x_n$, then the same holds with x_j replaced by x'_j, so $x_{j-1} \leq x'_j \leq x_{j+1}$, with corresponding inequalities for y instead of x. Then the same decision weights can be used in the calculation of RDU before and after the change in outcomes. In that case, the terms $\pi_j U(x'_j) = \pi_j U(y'_j)$ cancel, just as the terms $\pi_j U(x_j) = \pi_j U(y_j)$ do in (3), and the other terms are the same both before and after the change of outcome. Thus the inequality is not affected, and neither is the preference. This shows that RDU implies comonotonic independence.

The next observation presents, in a somewhat informal manner, the exact demarcation between EU and RDU. The presently existing axiomatizations, mentioned in the observation, are described in Wakker (1994b); there also a proof is provided.

Observation 1. Suppose that RDU holds, in accordance with one of the presently existing axiomatizations. Then:

(i) Comonotonic independence holds.
(ii) Independence holds if and only if EU holds.   □

3. Experiment

3.1. Subjects

Eighty-four students in an introductory psychology course at the University of North Carolina at Chapel Hill participated in the study. They were randomly assigned to one of the four display conditions described below.

3.2. Task

Subjects' task was to choose the preferred member from each of 64 pairs of gambles. All gambles involved possible profits. To increase the realism of the choices and motivate the subjects, one of the 64 gambles which they chose was randomly selected at the conclusion of the experiment and played for real money. The effectiveness of this incentive method relies on the "isolation effect" (Kahneman and Tversky, 1979, p. 271), the finding that subjects evaluate each choice between gambles in isolation instead of evaluating their choices as subparts of one large joint gamble.

3.3. Procedure

Subjects were seated in front of a personal computer in a private booth. They read instructions and then performed the choice task on the computer. The order of the 64 gamble pairs and the order of the two gambles within each pair were randomly determined.

3.4. Display and instruction

Subjects were instructed to choose that gamble in each presented pair that they would prefer to play.[3] It was emphasized that there were no right or wrong decisions, and that they should simply choose the gamble they preferred; indifference could not be expressed. They were informed that one of the 64 pairs of gambles would be randomly selected at the end of the experiment by selecting one card from a deck of 64 cards. The gamble that they had chosen in that pair would be played by selecting a second card from a deck of 100 cards marked 1 to 100. The number written on that card would determine the outcome which they would receive, depending on their choice of gamble. Subjects were told that they would be paid the money amount won on that trial.

The display of gambles has been shown to affect choice behavior (Keller, 1985; Erev, Bornstein, and Wallstein, 1993). Therefore we used four distinct gamble displays. This manipulation was used to improve the generalizability of the result rather than to test specific hypotheses about the effects of display format. The display condition was a between subject variable to which each subject was randomly assigned.

Figure 1 presents the display of one pair of gambles (the second pair in set 5) under the four display conditions. Twenty-two subjects were assigned to the collapsed display condition, in which events that lead to identical outcomes were collapsed. This condition is similar to the commonly used display of gambles in behavioral decision-making research.
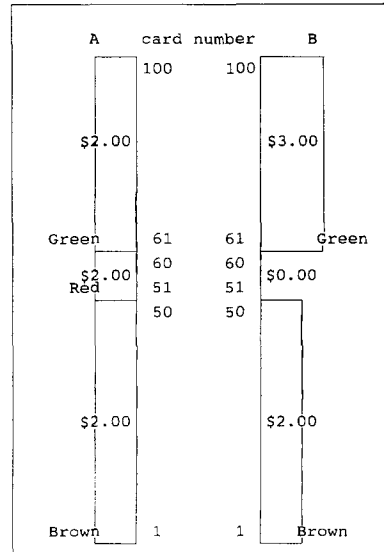
Twenty-one subjects were assigned to the noncollapsed display, which was identical to the collapsed display with the exception that identical outcomes were not collapsed. Twenty subjects were assigned to the verbal display, which was identical to the noncollapsed display, except that the probabilities of the described events were not provided numerically. The remaining 21 subjects were assigned to the graphical display condition, adopted from Camerer (1989).

Figure 1. The four displays of gambles.

3.5. Stimuli

The 64 pairs of gambles included two replications of the 24 experimental pairs described below and 16 "filler" pairs. The fillers were added, since the experimental questions required choices between gambles with similar expected values. A risk-neutral subject could have figured, after a few trials, that random choice would save time without a significant loss in utility. To motivate subjects to examine the choice alternatives carefully, the 16 filler pairs of gambles with clearly distinct expected values were interspersed among the experimental pairs.

The 24 experimental gambles consisted of six sets of four gamble pairs. Each gamble pair consisted of a safer gamble S, and a riskier gamble R. As an example, consider the first gamble pair of set 1, depicted in figure 3.1.

The S-gamble yields \$0.5 if the number of the card drawn is not above 55, \$6.0 if the number is between 56 and 80, and \$7.0 if the number is 81 or higher. The R-gamble yields \$0.5 if the number of the card drawn is not above 55, \$4.5 if the number is between 56 and 80, and \$9.0 if the number is 81 or higher. The general structure of outcomes in a gamble pair is depicted in figure 2. The S-gamble yields either common outcome CO, or S_l or S_h as outcome; these were 0.5, 6.0, and 7.0, respectively, in our example. The R-gamble yields either common outcome CO, or R_l or R_h as outcome; these were 0.5, 4.5, and 9.0, respectively, in the example. The CO outcomes are related to the same card numbers for S and R, and so are the S_l and R_l outcomes and the S_h and R_h outcomes. This was clear to the subjects in all four displays. To avoid triviality of choice due to dominating alternatives, we always have $R_l < S_l \leq S_h < R_h$.

The common outcomes CO in a gamble pair were taken from the regions CO_l, CO_m, or CO_h shown in figure 2, so that the gambles in a given choice pair always have the same rank-ordering of outcomes, i.e., are comonotonic. The other three pairs of set 1 were identical to the first pair, except for the common outcome CO, which took the values \$3.50, \$6.50, and \$9.50, respectively. The gamble pairs of all six sets are illustrated in the upper part of figures 3.1–3.6 (and also in table 2). As described for set 1, the four pairs within each set are identical except for the CO outcome.

The gambles were described by means of a "states-format," generated by the selection of a card. Given that all cards were known to be drawn with equal likelihood, the above states-format notation also provides information about the probabilities of each outcome. In the first two display conditions, this probability was additionally provided explicitly, i.e., numerically (see figure 1, a and b). The first two choice situations in sets 5 and 6 are versions of the Allais paradox, with the S-gamble in the second pair of each set a sure gain.
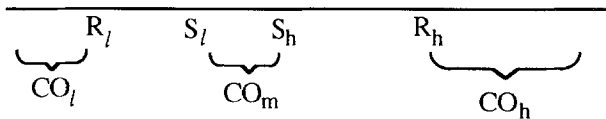

Figure 2. Outcomes of two gambles in a choice pair.

(1st pair)			(2nd pair)			(3d pair)			(4th pair)			cards #
1-55 56-80 81-100			**1-55** 56-80 81-100			56-80 **1-55** 81-100			56-80 81-100**1-55**			
(**0.5**, 6.0, 7.0)			(**3.5**, 6.0, 7.0)			(6.0, **6.5**, 7.0)			(6.0, 7.0, **9.5**)			S-outcomes
≿? / ≾?			≿? / ≾?			≿? / ≾?			≿? / ≾?			
(**0.5**, 4.5, 9.0)			(**3.5**, 4.5, 9.0)			(4.5, **6.5**, 9.0)			(4.5, 9.0, **9.5**)			R-outcomes

0.5	→	**3.5**	→	**6.5**	→	**9.5**	common outcome
	×		×		×		EU
	×		+		+		RDU
	×		S⤬R		R⤬S		RDU, pesm
	×		R⤬S		S⤬R		RDU, optm
	×		S⤬R		S⤬R		CPT

Figure 3.1. Description of first set of gamble pairs, together with model predictions of excluded preference changes for changes in common outcome *CO*. EU (ind.) excludes any preference change if *CO* is changed (×). RDU (Com. ind.) excludes no preference changes if *CO* is changed from 3.5 to 6.5 (+). RDU with pessimism (*w* convex) excludes a change from *S* to *R* if *CO* is changed from 3.5 to 6.5 (but permits a change from *R* to *S*).

(1st pair)			(2nd pair)			(3d pair)			(4th pair)			cards #
1-65 66-85 86-100			**1-65** 66-85 86-100			66-85 **1-65** 86-100			66-85 86-100**1-65**			
(**0.5**, 3.5, 5.5)			(**2.5**, 3.5, 5.5)			(3.5, **4.5**, 5.5)			(3.5, 5.5, **6.5**)			S-outcomes
≿? / ≾?			≿? / ≾?			≿? / ≾?			≿? / ≾?			
(**0.5**, 3.0, 6.0)			(**2.5**, 3.0, 6.0)			(3.0, **4.5**, 6.0)			(3.0, 6.0, **6.5**)			R-outcomes

0.5	→	**2.5**	→	**4.5**	→	**6.5**	common outcome
	×		×		×		EU
	×		+		+		RDU
	×		S⤬R		R⤬S		RDU, pesm
	×		R⤬S		S⤬R		RDU, optm
	×		S⤬R		S⤬R		CPT

Figure 3.2. Description of second set of gamble pairs, together with model predictions. EU (ind.) excludes any preference change if *CO* is changed (×). RDU (Com. ind.) excludes no preference changes if *CO* is changed from 2.5 to 4.5 (+). RDU with pessimism (*w* convex) excludes a change from *S* to *R* if *CO* is changed from 2.5 to 4.5 (but permits a change from *R* to *S*).

From Observation 3 in section 6 and its proof in appendix A, it can be inferred that our experiment tested comonotonic independence in the simplest possible way. Further criteria for the selection of the stimuli used in our experiment were:

• Amounts should be nonnegative (to avoid sign-dependence effects, as well as the possibility of losses for subjects).
• Amounts should be high enough to motivate subjects and low enough to be payable.

(1st pair)	(2nd pair)	(3d pair)	(4th pair)	
1-40 41-80 81-100	41-80 **1-40** 81-100	41-80 **1-40** 81-100	41-80 81-100 **1-40**	cards #
(**0.5**, 2.5, 6.0)	(2.5, **3.0**, 6.0)	(2.5, **5.5**, 6.0)	(2.5, 6.0, **8.0**)	S-outcomes
≽?	≽?	≽?	≽?	
(**0.5**, 1.5, 7.5)	(1.5, **3.0**, 7.5)	(1.5, **5.5**, 7.5)	(1.5, 7.5, **8.0**)	R-outcomes

0.5	→	**3.0**	→	**5.5**	→	**8.0**	common outcome
		×		×		×	EU
		+		×		+	RDU
		S⤬R		×		R⤬S	RDU, pesm
		R⤬S		×		S⤬R	RDU, optm
		S⤬R		×		S⤬R	CPT

Figure 3.3. Description of third set of gamble pairs, together with model predictions. EU (ind.) excludes any preference change if *CO* is changed (×). RDU (Com. ind.) excludes no preference changes if *CO* is changed from 0.5 to 3.0 (+). RDU with pessimism (*w* convex) excludes a change from *S* to *R* if *CO* is changed from 0.5 to 3.0 (but permits a change from *R* to *S*).

(1st pair)	(2nd pair)	(3d pair)	(4th pair)	
1-70 71-80 81-100	71-80 **1-70** 81-100	71-80 **1-70** 81-100	71-80 81-100 **1-70**	cards #
(**2.5**, 5.5, 10.5)	(5.5, **6.0**, 10.5)	(5.5, **9.5**, 10.5)	(5.5, 10.5, **13.0**)	S-outcomes
≽?	≽?	≽?	≽?	
(**2.5**, 3.5, 12.5)	(3.5, **6.0**, 12.5)	(3.5, **9.5**, 12.5)	(3.5, 12.5, **13.0**)	R-outcomes

2.5	· →	**6.0**	→	**9.5**	→	**13.0**	common outcome
		×		×		×	EU
		+		×		+	RDU
		S⤬R		×		R⤬S	RDU, pesm
		R⤬S		×		S⤬R	RDU, optm
		S⤬R		×		S⤬R	CPT

Figure 3.4. Description of fourth set of gamble pairs, together with model predictions. EU (ind.) excludes any preference change if *CO* is changed (×). RDU (Com. ind.) excludes no preference changes if *CO* is changed from 2.5 to 6.0 (+). RDU with pessimism (*w* convex) excludes a change from *S* to *R* if *CO* is changed from 2.5 to 6.0 (but permits a change from *R* to *S*).

- Amounts should be multiples of 50 cents (to enhance simplicity and thus reduce errors).
- The gambles in experimental choice pairs should be close in preference to allow for preferences to be switched after a change in common outcome. If the gambles are too far apart in preference, choices are not likely to be reversed when the common outcome is changed, even if people are sensitive to changes in rank order. Then only few violations of independence, either comonotonic or noncomonotonic, would result, and the statistical power would be low.

(1st pair)	(2nd pair)	(3d pair)	(4th pair)	cards #
1-50 51-60 61-100	51-60 **1-50** 61-100	51-60 61-100 **1-50**	51-60 61-100 **1-50**	
(**0.0**, 2.0, 2.0)	(2.0, **2.0**, 2.0)	(2.0, 2.0, **4.0**)	(2.0, 2.0, **6.0**)	S-outcomes
≿?	≾?	≿?	≿?	
(**0.0**, 0.0, 3.0)	(0.0, **2.0**, 3.0)	(0.0, 3.0, **4.0**)	(0.0, 3.0, **6.0**)	R-outcomes

0.0 →	2.0 →	4.0 →	6.0	common outcome
×	×	×		EU
+	+	×		RDU
~~S→R~~	~~R→S~~	×		RDU, pesm
~~R→S~~	~~S→R~~	×		RDU, optm
~~S→R~~	~~S→R~~	×		CPT

Figure 3.5. Description of fifth set of gamble pairs, together with model predictions. EU (ind.) excludes any preference change if *CO* is changed (×). RDU (Com. ind.) excludes no preference changes if *CO* is changed from 0.0 to 2.0 (+). RDU with pessimism (*w* convex) excludes a change from *S* to *R* if *CO* is changed from 0.0 to 2.0 (but permits a change from *R* to *S*).

(1st pair)	(2nd pair)	(3d pair)	(4th pair)	cards #
1-50 51-60 61-100	51-60 **1-50** 61-100	51-60 61-100 **1-50**	51-60 61-100 **1-50**	
(**2.0**, 4.0, 4.0)	(4.0, **4.0**, 4.0)	(4.0, 4.0, **6.0**)	(4.0, 4.0, **8.0**)	S-outcomes
≿?	≾?	≿?	≿?	
(**2.0**, 2.0, 5.0)	(2.0, **4.0**, 5.0)	(2.0, 5.0, **6.0**)	(2.0, 5.0, **8.0**)	R-outcomes

2.0 →	4.0 →	6.0 →	8.0	common outcome
×	×	×		EU
+	+	×		RDU
~~S→R~~	~~R→S~~	×		RDU, pesm
~~R→S~~	~~S→R~~	×		RDU, optm
~~S→R~~	~~S→R~~	×		CPT

Figure 3.6. Description of sixth set of gamble pairs, together with model predictions. EU (ind.) excludes any preference change if *CO* is changed (×). RDU (Com. ind.) excludes no preference changes if *CO* is changed from 2.0 to 4.0 (+). RDU with pessimism (*w* convex) excludes a change from *S* to *R* if *CO* is changed from 2.0 to 4.0 (but permits a change from *R* to *S*).

- The gambles in "filler" choice pairs should differ strongly in preferability, to promote deliberate decision making.
- The expected value of the riskier gambles should not always be higher than that of the safer gambles.
- The change in the magnitude of common outcomes from choice pair to choice pair should be the same, so that, in this respect, the tests of comonotonic independence and general independence do not differ.

- Comonotonic independence should be tested for increases in common outcome when the common outcome is the lowest outcome, when it is the middle outcome, as well as when it is the highest outcome.
- Some safer gambles should have a more widespread distribution of outcomes, whereas others should have a narrower range of outcomes.
- The probabilities should be such that both the common outcomes and the noncommon outcomes have impact.

4. Predictions

This section derives the predictions that EU and different versions of the RDU model make for the patterns of choices encountered in our experiment. As above, U denotes the utility function for outcomes. We only consider decision under risk, i.e., we assume that all cards are equally likely to be selected, with probability 1/100. For RDU, $w : [0, 1] \rightarrow [0, 1]$ denotes the probability transformation function.[4] However, all theoretical predictions below can be extended to decision under uncertainty without any difficulty.

We analyze in detail the patterns of choices possible under the different utility theories for the first set in our design. For this set, shown in figure 3.1, the three events were described by cards numbered between 1–55, 56–80, and 81–100, respectively, with corresponding probabilities of .55, .25, .20. The corresponding outcomes were

$$R = (\$CO, \$4.5, \$9)$$

and

$$S = (\$CO, \$6, \$7),$$

in the riskier (R) and safer (S) gambles, respectively. The common outcome CO takes the values of $0.50, $3.50, $6.50, and $9.50, respectively, in the four choice pairs of the set.

Under EU, independence should be satisfied, i.e., preferences should be independent of the value of CO. If the independence assumption holds, we should either find four preferences for the R-gamble or four preferences for the S-gamble.

Next we present the predictions under RDU. For each choice pair, the two gambles have the same rank-order of outcomes (i.e., are comonotonic), hence the decision weights π_j are the same for the two gambles in their RDU evaluation. Leaving these decision weights unspecified for the moment, we write

$$RDU(R) = \pi_1 U(CO) + \pi_2 U(4.5) + \pi_3 U(9)$$
$$RDU(S) = \pi_1 U(CO) + \pi_2 U(6) + \pi_3 U(7).$$

Cancelling the common term $\pi_1 U(CO)$, we have RDU(R) $\geq$ RDU(S) if and only if

$$\pi_3(U(9) - U(7)) \geq \pi_2(U(6) - U(4.5)). \tag{4}$$

In other words, for R to be preferred to S, the "R-gain" of receiving $9 instead of $7 must outweigh the "$S$-gain" of receiving $6 instead of $4.5.

Under RDU, the independence requirement should only be satisfied between pairs of gambles that are mutually comonotonic. Changing the common outcome CO from $0.50 to $3.50 does not affect the rank-order of the outcomes. That is, the .20 probability outcome is still the best, the .25 probability outcome the second best, and the CO outcome the worst, for the new gambles. In other words, the two pairs of gambles with $CO = $0.50 and with $CO = $3.50 are mutually comonotonic. Thus, comonotonic independence applies, and the choices should be the same for both pairs, i.e., either two S or two R choices. In terms of (4), comonotonicity of the four gambles implies that the decision weights π_2, π_3 do not change as CO is changed from $0.50 to $3.50, thus resulting in the same choices.

When the common outcome is changed from $3.50 to $6.50, on the other hand, RDU makes a different prediction because the change in CO results in a change in the rank-ordering of the outcomes. Now CO is preferred to the .25 probability outcomes, whereas before it was not. The new gambles are not comonotonic with the old gambles, and RDU no longer predicts the same preference. In terms of Formula (4), changing CO now has the potential to change the decision weights π_2 and π_3, which may affect the inequality.

The nature of the change in preference depends on the shape of the weighting function w. A convex w function[5] corresponds to *pessimism*. Predictions for this special case of RDU are also shown in the lower part of figure 3.1. The increase of CO from $3.50 to $6.50 changes the rank of the .25 probability outcomes from middle to worse. For a pessimist, this means, intuitively, that the .25 outcomes of both gambles now have more importance to the subject. Mathematically, due to the convexity of w, the new decision weight ($\pi_2 = w(1) - w(.75)$) associated with these outcomes is larger than the old decision weight $\pi_2 = w(.45) - w(.20)$, so that these outcomes play a more important role in the choice. As the .25 probability outcome for the S-gamble, i.e., $6, is greater than the corresponding $4.50 outcome for the R-gamble, this means that the utility of the S-gamble will increase more than the utility of the R-gamble. The decision weight π_3 of the .20 outcomes has not been altered by the increase of CO. These outcomes have remained the best throughout, and their decision weight has continued to be $w(0.20)$. Therefore, for a pessimist, the change of CO from $3.50 to $6.50 enhances the safer choice. A choice of R when $CO = $3.50 may be accompanied by a choice of S when $CO = $6.50, but not the other way around: A choice of S when $CO = $3.50, must imply a choice S when $CO = $6.50.

Analogously, if w is concave, then the increase of the common outcome from $3.50 to $6.50 enhances a risky choice. In other words, a preference reversal $S \rightarrow R$ is permitted and a preference reversal $R \rightarrow S$ is excluded. Concavity of w can be interpreted as *optimism*.

Finally, consider the effects of a change in CO from \$6.50 to \$9.50. Again, the change affects the rank-ordering of the outcomes, moving the ranks of the .20 probability outcomes from best to second best. Thus, again, RDU no longer predicts invariance of preference. In the case of pessimism, i.e., convexity of w, the increase of CO will increase the decision weight of the .20 probability outcome (π_3 in (4)) and thus will enhance a risky choice. So now a preference reversal $S \rightarrow R$ can occur, but a preference reversal $R \rightarrow S$ is excluded. The converse holds if w is concave.

We also derive the predictions of an S-shaped weighting function. This form was suggested by several authors, e.g., by Quiggin (1982), Karni and Safra (1990), and by Tversky and Kahneman (1992) in their "cumulative prospect theory" (CPT). We shall refer to it as the CPT prediction. Models very similar to CPT were introduced by Starmer and Sugden (1989, appendix) and Luce and Fishburn (1991); the latter differs mainly in its axiomatization, invoking conditions on a "joint receipt" operation. These models generalize RDU by incorporating "sign-dependence," which allows the weighting function for probabilities associated with gains to be different from the weighting function for probabilities associated with losses.[6] As discussed above, our experiment was designed to avoid effects of sign-dependence by using gambles involving only gains. In that case, the general form of CPT coincides with RDU; therefore, our experiment is also a test of CPT. CPT assumes a weighting function w for gains that is convex for large probabilities, leading to the overweighting of certainty, and thus explaining phenomena such as the Allais paradox. At the same time, the weighting function is assumed to be concave for small probabilities, leading to the "long-shot" effect, where a person overappreciates a gamble that, with a small probability, gives a large gain, thus explaining the buying of lotteries. In short, this special case of w, and thus of RDU, assumed by CPT predicts that more attention and decision weight is attributed to extreme outcomes.

The change in common outcome from \$3.50 to \$6.50 in set 1 makes the .25 probability outcomes more extreme for both gambles by making them the lowest outcomes after the change. According to CPT, this will increase the associated decision weight π_2. Mathematically, the weight changes from $w(.45) - w(.20)$ to $w(1) - w(.75)$, resulting in an increase if w is convex for high probabilities, as assumed by CPT. (This is also implied by dual subadditivity of w, as predicted by Tversky and Kahneman, 1992. Dual subadditivity means $w(p) - w(p - r) \leq w(1) - w(1 - r)$ and is somewhat less restrictive than convexity of w for high probabilities.) Increasing π_2 leads to greater preference for the S-gamble. The predictions of CPT coincide in this case with those of "pessimistic" RDU with its convex weighting function. Shifts from R choices to S choices are predicted; shifts from S choices to R choices constitute a violation of CPT.

The change in common outcome from \$6.50 to \$9.50 makes the .20 probability outcomes less extreme, thus decreasing their decision weight π_3 and enhancing the safer choice. This prediction coincides with that of optimistic RDU with its concave weighting function. Shifts from R choices to S choices are predicted in this case, and shifts from S choices to R choices are violations of CPT.

For sets 2 to 6, predictions can be derived in the same way as just outlined for set 1. In each set, there are two changes in the common outcome CO that make the resulting

two pairs of choice situations not mutually comonotonic. RDU in general allows changes in preference between these pairs. Special cases of RDU that assume particular shapes for the weighting function w make more specific predictions: For changes of CO that move it from lowest to middle-ranked outcome, CPT and pessimistic RDU enhance the safer choice and exclude a preference reversal $S \rightarrow R$, whereas optimistic RDU excludes $R \rightarrow S$. For changes of CO that move it from middle to highest ranked outcome, CPT and optimistic RDU enhance the safer choice and exclude $S \rightarrow R$; whereas, pessimistic RDU excludes $R \rightarrow S$. *All* RDU theories exclude violations of comonotonic independence. These predictions are detailed in the lower parts of figures 3.1–3.6. It is important to note that these predictions hold true regardless of the nature of the transformation of outcomes; that is, no assumptions about the shape of the utility have been made.

Summarizing, pessimistic weighting in RDU (convex w) means that more attention is paid to events as they yield outcomes that are low in the rank-order of possible outcomes, optimistic weighting in RDU (concave w) means that more attention is paid to events as they yield outcomes that are high in the distribution of possible outcomes, and CPT (S-shaped w with concave lower portion and convex upper portion) means that more attention is paid to events as they yield outcomes that are relatively more extreme.

4.1. Expectations prior to the experiment

Prior to the experiment, our expectation was as follows: There are many causes for deviations from EU and the independence condition. One of these many causes is rank-dependence, i.e., the phenomenon that people let the importance attributed to events depend on the rank in relative favorability. Thus we expected that, ceteris paribus, the independence condition would be less frequently violated for comonotonic gambles than for noncomonotonic ones. This is to be expected for the Allais-type choices in sets 5 and 6, as many experiments have demonstrated. We expected to find the effect also for the other sets, but less pronounced, because several studies have found violations of independence primarily when nonlinearity of the weighting function near 1 or 0 is involved (Cohen and Jaffray, 1988). In studies of indifference curves in the probability triangle, nonlinearities are also mainly found near the borders (Camerer, 1992, stylized fact 1). Estimations of the probability weighting function for RDU usually find that the function is relatively linear in the middle of its region (e.g., Tversky and Kahneman, 1992).

5. Experimental results

5.1. Consistency

Each subject chose twice among the gambles of each experimental pair, allowing for an examination of the consistency of repeated choices. Over the 24 experimental gambles and across display formats, the average subject made consistent choices (two R's or two

S's) for only 67% of the pairs. The graphical display produced the highest consistency (69%), and the collapsed display the lowest (61%), with no significant difference in consistency between conditions. This proportion of consistent choices against a chance level of 50% may seem disappointingly low and may seem to argue against the assumption that choice is deterministic. However, other explanations are also possible; this is discussed further below.

The results are similar to those of Camerer (1989) and Starmer and Sugden (1989), where the estimates of inconsistencies were based on one observation per subject, allowing only for an overall estimate across subjects. In our experiment, consistency could be examined as an individual difference characteristic that may vary from individual to individual. Low consistency could be due to some individuals who choose at random, while others are highly consistent. Figure 4 shows the distribution of inconsistent choice proportions in our sample of 84 subjects. While there is some variation, over 90% of all subjects were inconsistent at least 25% of the time. The observed degree of inconsistency is smaller than would be expected if subjects had simply chosen at random ($z = 2.93$, p $< .005$). An analysis of variance showed that consistency also depends on the set of gamble pairs ($F[5,3] = 8.93; p < .0001$); it is considerably higher in set 5 than in the other sets.

Part of the observed inconsistencies can be explained through indifference, or weak preference: If two gambles are truly indifferent to the subject, or very close in value, then it is rational that the subject will choose randomly between these gambles; repeated choices can then be mutually different. As pointed out at the end of section 3, the experimental gambles were designed to be close in preference, so true indifference is quite plausible. We found a high correlation between inconsistency and a measure for
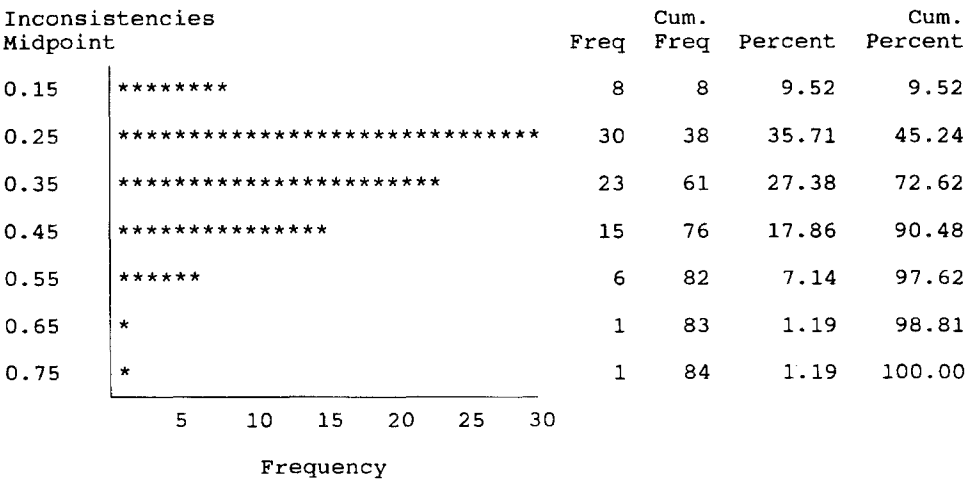
```
Inconsistencies                                      Cum.                  Cum.
Midpoint                                 Freq   Freq   Percent   Percent

0.15    |********                          8      8      9.52      9.52

0.25    |******************************   30     38     35.71     45.24

0.35    |**********************           23     61     27.38     72.62

0.45    |***************                  15     76     17.86     90.48

0.55    |******                            6     82      7.14     97.62

0.65    |*                                 1     83      1.19     98.81

0.75    |*                                 1     84      1.19    100.00
        |_____
             5    10   15   20   25   30

              Frequency
```

Figure 4. The distribution of inconsistent choice proportions in the sample of 84 subjects.

strength of preference,[7] i.e., $\rho = -.78, p < .0001$. The higher consistency rate for the filler gamble choices, i.e., 84%, further supports the indifference explanation for the inconsistencies in choice for the experimental pairs, which makes the inconsistencies less problematic.

5.2. Violations of independence; descriptive analysis within subjects

Our null hypothesis, H_0, assumes that violations of independence will occur equally likely in all pairwise comparisons of choices in this study. EU predicts that, without inconsistencies and errors in choice, independence is satisfied for all comparisons. We first discuss the data descriptively, without committing to any specific error theory. All that we assume for now is that, under EU, factors that contribute to violations of independence are just as likely to operate when the two pairs of gambles are mutually comonotonic as when they are not. Later we shall assume a specific error theory and then present a statistical analysis.

Each set of choices in our study provides one test of comonotonic independence (for the two CO values in the same region in figure 2) and two tests of noncomonotonic independence. H_0 predicts that the proportion of violations of comonotonic independence will be equal to the proportion of violations of noncomonotonic independence. RDU, on the other hand, predicts that the proportion of violations of noncomonotonic independence will be greater.

The choices made by a subject for a given set of four gamble pairs can be described by a quadruple of numbers between 0 and 2 that describe the number of R choices out of a total of two possible ones for each pair (i.e., for the two replications of each choice). For example, a quadruple (2,1,0,2) means that the subject, consistently, chose R both times from the first choice pair (with the lowest value of CO), chose R once and S once from the second choice pair (i.e., was inconsistent), chose S both times from the third choice pair, and R both times in the final choice.

Independence of preference from the value of the common outcome is assessed by comparing pairs of choices within the set. Suppose we observe, in set 1, a quadruple (2,0,0,0). This is taken as one violation of independence, observed in the comparison of the first two choices, and two verifications of independence, one in the comparison of second and third choice, the other in the comparison of third and fourth choice. Comparisons of the first and third, and first and fourth choice also show evidence for violations of independence; however, these violations follow from the one already counted in the comparison of first and second choice. Furthermore, they concern larger increases in the common outcome and thus should not be compared with changes in preference between pairs with adjacent levels of CO. Therefore, to avoid double counting and to compare violations of comonotonic versus noncomonotonic independence under conditions that are as similar as possible, we only compare independence violations for choice pairs with adjacent levels in the common outcome.

For quadruples with inconsistent choices, for example, (2,2,2,1), the scoring of independence violations was as follows. Assuming that the inconsistency in choice for the

fourth choice pair reflects greater indifference between the S and the R alternative than the consistent choice of two R selections in the previous three choice pairs, this pattern provides some evidence for a violation of independence, namely a preference reversal from strict preference to indifference. Since this is a weaker violation than a reversal from strict preference to strict dispreference, we refer to the following choice patterns in adjacent pairs: (0,1), (2,1), (1,2), and (1,0), as *weak violations of independence*, and to (2,0) and (0,2) choice patterns in adjacent pairs as *strong violations of independence*.

As a further example, the quadruple choice pattern (1,0,2,1) for set 1 shows two weak violations and one strong violation of independence. In set 1 the first two gamble pairs are comonotonic, hence this quadruple gives one weak violation of comonotonic independence; there is also a weak violation of noncomonotonic independence and one strong violation of noncomonotonic independence. The quadruple (1,1,2,0) for set 1 similarly reveals one strong violation of noncomonotonic independence, one weak violation of noncomonotonic independence, and no violation of comonotonic independence. For set 5, where the four gambles of the last two choice situations are comonotonic, the quadruple (1,0,0,2) reveals one strong violation of comonotonic independence and one weak violation of noncomonotonic independence.

Table 1 lists the number of observed violations of independence. For 84 subjects, 6 sets, and 3 tests of independence per set, there were altogether $84 \times 6 \times 3 = 1512$ different tests; one-third of them, 504, concerned comonotonic independence. There were 95 strong violations of (EU-) independence and 613 weak violations. Under H_0, one expects 95/3, i.e., 31.66, strong violations of comonotonic independence, and 613/3 = 204.33 weak violations of independence. The observed numbers, 33 and 205 under the general RDU column, do not differ significantly from those expected under H_0. We see that comonotonic independence is not significantly less violated than non-comonotonic independence. If anything, comonotonic independence was slightly more frequently violated than noncomonotonic independence! Thus, across all choice pairs, there was no evidence that a general version of RDU accommodated the data better than EU.

For the Allais-type choices, however, comonotonic independence performed better, in agreement with previous observations in the literature. Let us give results for the collapsed presentation, the commonly used format in the literature, concerning sets 5 and 6. In these sets, the first increase in common outcome gives the Allais-type test of independence. Here there were, altogether, 9 strong violations of independence, leading

Table 1. Number of observed violations versus number of violations predicted by H_0.

	RDU		$RDU_{w\ cvx}$|pess		$RDU_{w\ ccv}$|optm		$CPT_{w\ s-sh}$		EU
	obsd	pred. H_0	obsd	pred. H_0	obsd	pred. H_0	obsd	pred. H_0	
STRONG	33	31.66	64	63.33	64	63.33	60	63.33	95
WEAK	205	204.33	398	408.66	420	408.66	394	406.66	613

to a proportion of $9/44 = .205$; the overall proportion of strong violations of independence was $95/1512 = .063$. Only 1 (H_0 predicts $9/3 = 3$) concerned a violation of comonotonic independence, and 7 (H_0 predicts 1.5) were in the Allais tests and in the direction predicted by the Allais paradox. The weak violations give less pronounced results: out of 55 weak violations, 17 (H_0 predicts 18.33) concerned comonotonic independence and 16 (H_0 predicts 9.17) were in the Allais tests and in the predicted direction.

Let us next consider the special cases of RDU described in section 4. Just as the general version of RDU, the special cases of RDU prohibit any violations of comonotonic independence; in addition, they also prohibit some of the violations of noncomonotonic independence (see figures 3.1–3.6). For example, as can be inferred from the lower half of figure 3.1, a quadruple (0,0,2,2) in set 1 means a strong violation of RDU with pessimism, and so does the quadruple (2,2,2,0); quadruples (2,2,0,0) and (0,0,0,2) do not violate RDU with pessimism. Thus, under H_0, two-thirds of the 95 strong violations of independence could be expected to constitute violations of RDU with pessimism. A similar reasoning applies to the special cases of RDU with an optimistic (concave) weighting function w and to CPT (RDU with an S-shaped weighting function).

The data in table 1 show that, for strong violations, neither RDU with optimism nor RDU with pessimism perform better than the null hypothesis would predict. For the weak violations, RDU with pessimism performs a little better than predicted by the null hypothesis (only 398 violations instead of 408.66), whereas RDU with optimism performs a little worse. CPT with its S-shaped w-function provides the best description of the choice patterns observed in this experiment. However, the improvement in prediction does not reach statistical significance.

The Allais-type choices, i.e., the first two in sets 5 and 6, make it possible to compare our experimental design to other studies in the literature. Allais-paradox experiments in the literature mostly presented the gamble pairs in a display format similar to the collapsed format of our experiment. In other studies, choices are usually not repeated. To compare our repeated choices with single choices, it seems reasonable to count scores for our repeated choices as follows:

- (0,0) and (2,2) each count as no violation of independence.
- (2,0) and (0,2) each count as two "single" violations of independence.
- (0,1), (1,2), (2,1), and (1,0) each count as one violation of independence.
- (1,1) also counts as one violation of independence.

Only the (1,1) score may need some elucidation; here the subject once chose R and once chose S, both in the first and second choice of the set. These choices can either be grouped as choices (R,R) and (S,S), giving no single violation of independence, or as choices (R,S) and (S,R), giving two violations. Therefore it seems reasonable, when comparing our repeated choices with single choices, to count also a (1,1) score as one single violation. (Only two of the 22 subjects gave a (1,1) score on these questions.)

With 22 subjects in the collapsed display format, one test in set 5 and one in set 6, and all choices made twice by each subject, there are 88 possibilities for "single-choice-violations" of independence in an Allais setting. By the above way of counting, we found 39 violations of independence, i.e., in 44% of the cases. This is quite comparable to other results in the literature (MacCrimmon and Larsson, 1979; Keller, 1985; Conlisk, 1989). Previous studies that examined conditions in which outcomes were not collapsed generally found fewer violations (around 20% in Keller, 1985).

5.3. Statistical analysis within subjects

We formulate now RDU hypotheses concerning the degree of deviation from EU in the true preferences. We show that even weak deviations from EU can be rejected on the basis of our data. The hypotheses that we consider only specify the *degree* of deviation from EU towards RDU, and do not consider specifications of RDU such as CPT or RDU with pessimism.[8]

For the statistical analysis, we assume that, for each choice, a subject has a "true preference" that is strict and remains constant throughout the decision. In addition, we assume that there is a probability p_e that the subject makes an error and chooses contrary to her true preference. This probability is assumed to be the same for all subjects and gamble pairs, and each error is assumed to be independent of everything else in the experiment. Obviously, in reality, the probability of error does depend at least partly on strength of preference, as shown in our data, but our assumption is an approximation that makes the analysis tractable.

With an inconsistency rate of 1/3, the probability p_e for a choice error is determined by the equality $1/3 = 2p_e(1 - p_e)$ (an inconsistency occurs if there is a choice error on the first choice and not on the second, or vice versa). From this, $p_e = .2113$ is derived ($p_e = .7887$ is excluded because $p_e < 1/2$). This error probability p_e will be assumed henceforth. The analysis below uses the number X (62 in our experiment) of observed strong violations of noncomonotonic independence as statistic.[9]

There are a total of 12 tests per subject $\times$ 84 subjects $= 1008$ tests of noncomonotonic independence in our experiment. We consider hypotheses that claim that there are n violations of noncomonotonic independence in the true preferences, and $1008 - n$ tests that conform to EU, for varying values of n; such hypotheses are denoted by H_n. H_0 is the hypothesis that EU holds; H_{1008} claims that all individuals violate noncomonotonic independence on every possible occasion.

Tversky and Kahneman (1992) suggest the weighting function

$$w(p) = \frac{p^{.61}}{[p^{.61} + (1-p)^{.61}]^{1/.61}}$$

as a plausible weighting function for the "average" subject. For the utility function they suggest $\alpha \to \alpha^{.88}$. (For all choices in our experiment, the same predictions would follow if we take utility linear.) With these parameters for CPT, one violation of comonotonic

independence is predicted in each of the six sets in our experiment; to be precise, the predicted choice patterns are $RRSS, RRSS, RSSS, RSSS, RSSS$, and $RSSS$ for the six sets. Hence, if all subjects were perfectly described by this representation, then H_{504} would hold true.

"Configural weighting theory" is an early developed theory by Birnbaum (1974) that contains most of the ideas of RDU. With the parameters derived from the experiment for pricing judgments[10] in Birnbaum et al. (1992), the predicted patterns are $RRSS$, $SSSS, SSSS, RSSS, RSSS$, and $RSSS$. Hence it predicts fewer violations of noncomonotonic independence than CPT. If all subjects were perfectly described by this representation, then H_{336} would be implied. Let us repeat here that the present analysis only tests the strength of deviation from EU predicted by RDU theories, and does not test specific predictions about the places where violations of noncomonotonic independence are to be expected. For such specific RDU theories, other tests statistics than X should be used, and these theories may actually perform better than RDU-in-general.

Appendix B explains that the number of strong violations of noncomonotonic independence, X, is approximately normally distributed, and that:

Lemma 2. Under H_n, the expectation of X is

$$E(X) = n/3 + 55.944,$$

and an upper bound for the variance of X is given by

$$VAR(X) \le .2595n + 63.1680.$$

Our observation $X = 62$ leads to an estimate of $n = 18$. That is, given an inconsistency rate of 1/3 for choices among the same pair of gambles, we can estimate that 62 observed strong violations of noncomonotonic independence are indicative of only 18 true violations of noncomonotonic independence among the 1008 tests. In the calculations below, we shall use the upper bound for the variance as given above; therefore, our results are actually slightly stronger than what is claimed below.

Statistically, we can clearly reject H_{504} in favor of the one-sided alternative hypothesis that there are fewer than 504 true violations of noncomonotonic independence. Under H_{504}, the observed value of $X = 62$ has a z-statistic of

$$z = \frac{62 - (504/3 + 55.944)}{\sqrt{.2595(504) + 63.168}} = -11.6282, \qquad p < .0001.$$

Similarly, we can reject H_{336}. Under H_{336}, the observed value of $X = 62$ has a z-statistic of

$$z = \frac{62 - (336/3 + 55.944)}{\sqrt{.2595(336) + 63.168}} = -8.6399, \qquad p < .0001.$$

Even the hypothesis of an average of at least one violation of noncomonotonic indepen-
dence per subject among the 12 tests, H_{84}, can be rejected at the .05 level. Then

$$z = \frac{62 - (84/3 + 55.944)}{\sqrt{.2595(84) + 63.168}} = -2.3806, \qquad p < .01.$$

The strongest prediction not rejected at the .05 level by our data is the prediction of 62
or more violations across subjects and choice sets (H_{62}), i.e., an average of slightly less
than one violation per subject.

Conclusion: The hypothesis that there are at least 63 violations of noncomonotonic
independence among the true preferences is rejected at the .05 level to the favor of the
hypothesis that there are fewer deviations from EU.   □

Our data can also be analyzed by Bayesian methods; these assume that all statistical
information is captured in the likelihood function. Appendix B demonstrates that the
likelihood ratio of H_0 against H_{63}, with the observed value $X = 62$, is about 3.41. For $n >$
63, the likelihood ratio is larger than 3.41. The first value for which it is larger than 2 is n
$= 54$, and for $n = 84$ it is 14.8. Therefore, on the basis of our data, prior odds for EU
(H_0, $n = 0$) against the hypothesis of an average of at least one violation of non-
comonotonic independence per individual ($n \geq 84$) should be increased by at least a
factor 14.8.

5.4. An aggregated preference change score

Next we present an alternative analysis, which allows us to incorporate weak violations of
independence.
   The "preference change" columns in table 2 summarize the proportions of reversals in
preferences for each increase in the common outcome. A preference reversal score was
calculated for each subject and each increase in the common outcome as the mean
absolute difference in the number of R choices. For example, if a subject made two R
choices in the first pair of the set and only one R choice in the second, her preference
reversal score was $|2 - 1|/2 = .5$. This score vector does not distinguish between
preference reversals from R to S and from S to R, so it does not provide information for
or against the special cases of RDU that were described above, such as CPT (such
information is provided in table 3). Table 2 lists these scores (normalized to be 1 for
maximal deviations from EU).
   The increases in common outcome that do not affect the rank-ordering of outcomes
and thus give a test of comonotonic independence, are marked in table 2 by a line
connecting the two relevant common outcomes. These increases yield a considerable
amount of preference reversals, an average of .268 over gambles and display conditions,
which was not significantly smaller than the average over the noncomonotonic increases

Table 2. Aggregated preference change scores.

Set		Basic gambles			CO	R choices C	N	V	G	M	Preference Change C	N	V	G	M
1	P	.55	.25	.20	0.5	.45	.62	.67	.74	.62					
											.32	.36	.38	.29	.33
	S	CO	6.0	7.0	3.5	.50	.40*	.60	.60	.52					
											.27	.43	.38	.36	.36
	R	CO	4.5	9.0	6.5	.50	.55	.63	.62	.57					
											.27	.36	.22	.26	.28
					9.5	.41	.57	.65	.64	.57					
2	P	.65	.20	.15	0.5	.50	.45	.47	.57	.50					
											.41	.19	.38	.43	.35
	S	CO	3.5	5.5	2.5	.64	.55	.60	.52	.58					
											.27	.24	.32	.33	.29
	R	CO	3.0	6.0	4.5	.59	. 55	.52	.57	.56					
											.30	.33	.27	.33	.31
					6.5	.43	.50	.70*	.57	.55					
3	P	.40	.40	.20	0.5	.43	.29	.47	.38	.39					
											.25	.26	.27	.29	.27
	S	CO	2.5	6.0	3.0	.45	.31	.35	.52	.41					
											.16	.21	.30	.26	.23
	R	CO	1.5	7.5	5.5	.43	.19	.25	.50	.35					
											.30	.19	.32	.17	.24
					8.0	.36	.14	.22	.43	.29					
4	P	.70	.10	.20	2.5	.70	.81	.63	.81	.74					
											.18	.21	.35	.21	.24
	S	CO	5.5	10.5	6.0	.57*	.74	.47	.79	.64*					
											.23	.29	.30	.19	.25
	R	CO	3.5	12.5	9.5	.70	.64	.52	.83	.68					
											.14	.29	.30	.14	.21
					13.0	.66	.79	.63	.83	.73					
5	P	.50	.10	.40	0.0	.57	.12	.02	.10	.21					
											.41	.24	.13	.17	.24
	S	CO	2.0	2.0	2.0	.34*	.21	.10	.17	.21					
											.16	.17	.07	.14	.14
	R	CO	0.0	3.0	4.0	.23	.24	.07	.21	.19					
											.20	.12	.10	.24	.17
					6.0	.20	.26	.07	.12	.17					
6	P	.50	.10	.40	2.0	.66	.57	.38	.60	.55					
											.43	.33	.22	.29	.32
	S	CO	4.0	4.0	4.0	.36*	.48	.25	.50	.40*					
											.23	.29	.27	.29	.27
	R	CO	2.0	5.0	6.0	.23	.43	.38	.45	.37					
											.23	.24	.30	.36	.28
					8.0	.32	.43	.38	.48	.40					

**p < .01, *p < .05

The basic gambles column presents the probabilities (P) of the different outcomes and the basic structure of gambles S and R. The value of the common outcome is presented in the CO column. Common outcomes with identical ranks are connected by a line. The R choices columns present the proportions of R choices in the condition collapsed (C), not collapsed (N), verbal (V), graphical (G), and mean over conditions (M). The preference change columns show the within-subject preference changes in the same format.

in common outcomes, i.e., .264 ($t[83] = .36, p = .72$). Comonotonic independence again does not provide for a better prediction of the response patterns.

5.5. A between-subjects analysis

In a between-subjects analysis, choices are aggregated across individuals. The greater the difference is between the number of R choices for one common outcome and for another common outcome, the more evidence this provides against independence. While this evidence is indirect, it has the advantage that it is less sensitive to random error operating at the individual level (Tversky and Kahneman, 1992, section 2.2). It is also closer to economic observations, where data are often only available at the macro-level.

The five columns below "R choices" in table 2 present the proportion of risky (R) choices made in each choice situation under the four display conditions, as well as averaged over all four conditions (column M). The display condition did not have an overall significant effect on risk taking ($F[3,80] = 2.16, p = .098$), nor did the set ($F[5,76] = 1.1, p = .37$). We examined the effect of each of the increases in the common outcome on the proportion of R choices in each condition. Of the 72 tests (3 increases per set × 6 sets × 4 display conditions), only one was significant at the .01 level. The significant increase, marked by two stars, occurred in the first increase of set 6 under the collapsed condition. Note that this is one of the two "Allais-type" increases, in agreement with the prediction of RDU with pessimism and of CPT, and contrary to the prediction of RDU with optimism. Five increases were significant at the .05 level: the two collapsed Allais problems and three of the other 70 increases.

These results, again, fail to reject the null hypothesis, under which one can expect one rejection at the .01 level among 70 tests and 3.5 rejections at the .05 level. Of the three significant non-Allais increases, one involved comonotonic gambles. Given that one-third of all the increases involves comonotonic gambles, this observed proportion is again consistent with the null hypothesis.

Of the four significant results for noncomonotonic independence, the three significant results in sets 4, 5, and 6 all concern the first increase in common outcome and all show a decrease in risk-taking. This agrees with pessimism-RDU and with CPT. The significant result in set 2 gives an increase in risk-taking for the last increase of the common outcome. This is in agreement with pessimism-RDU and contrary to CPT's prediction; however, when this result is taken together with the results from other sets concerning the last increase in common outcome, CPT performs slightly better than RDU with pessimism, as we shall see below, although this is not significant.

Aggregated over the four display conditions, only the first increase in set 6 (the Allais problem) is significant at the .01 level. The first increase in set 4 is significant at the .05 level. Both of these results are consistent with RDU with pessimism and with CPT. Again, for Allais-type choices, our finding is consistent with the findings in the literature. Other than for the Allais-type choices, we find no positive results for RDU; for the Allais choices there exist, of course, many explanations other than rank-dependence.

The effect of pessimism when the common outcomes have low values, is more pronounced than the optimism effect when the common outcomes have higher values. Over conditions, in all six sets the first increase in common outcome led to more preference reversals than the last increase. This linear trend is significant ($p < .02$ in a sign-test using set as a unit of analysis, and in a traditional t-test with subject as a unit of analysis). There was no significant interaction between the magnitude of this trend and the display condition.

Table 3 presents preference change scores similar to table 2, that, however, do distinguish between different specifications of RDU, the specifications being pessimism (indicated in columns Cp ... Mp), optimism (indicated in columns Co ... Mo), and CPT (indicated by underlining). This shows how the specific theories performed in particular sets of gambles (high scores indicate many violations, therefore bad performance). For the score of a specific model, only those preference changes contribute that do violate the specific model. For example, for the first increase of the CO outcome in set 1, any preference change constitutes a violation of all the RDU theories, so that here the scores in table 2 coincide with all the scores in table 3. For the second increase of the CO outcome in set 1, each preference change is either a violation of RDU with optimism, or a violation of RDU with pessimism; hence, the score in table 2 is here the sum of the scores of RDU with optimism and RDU with pessimism in table 3. We see from this table that CPT's better performance occurs mainly in the collapsed presentations, and mainly in the Allais-type choices of sets 5 and 6.

6. Comparison with other tests in the literature

RDU has usually been tested in the domain of decision under risk, where gambles are probability distributions over outcomes. An underlying assumption of such tests, known under various names such as the "reduction principle," is that different gambles that induce the same probability distribution, are indifferent. For the context where multistage gambles are used, this also invokes the "reduction of compound lotteries" assumption. Usually properties of indifference curves in the probability triangle are derived and tested. The analysis of indifference curves also assumes some transitivity conditions. Sometimes conditions for derivatives of indifference curves have been considered, and predictions are valid only in the direct neighborhood of the edges of the probability triangle. Conlisk (1989), Camerer (1992), and Harless (1992) found that phenomena are quite different for gambles inside the triangle than for gambles on the edges, with the pronounced violations of independence occurring mainly near the edges.

It is well known that the above assumptions are empirically problematic; see Kahneman and Tversky (1979), Luce and Fishburn (1991) and Bernasconi and Loomes (1992) for critical discussions of the reduction of the reduction principle. The reduction principle for single-stage gambles and transitivity have been abandoned in regret theory, as is discussed further below. Reduction also becomes problematic if utility is event-dependent.

A test of independence versus comonotonic independence with these additional assumptions is more problematic. In this case, the actual test is between {independence +

Table 3. Aggregated preference change scores for specifications of RDU.

Set	CO	Cp Co	Np No	Vp Vo	Gp Go	Mp Mo
	Violations					
1	0.5					
		.32 .32	.36 .36	.38 .38	.29 .29	.33 .33
	3.5					
		.14 .14	.29 .14	.20 .18	.19 .17	.20 .15
	6.5					
		.18 .09	.17 .19	.10 .13	.12 .14	.14 .14
	9.5					
2	0.5					
		.41 .41	.19 .19	.38 .38	.43 .43	.35 .35
	2.5					
		.11 .16	.12 .12	.13 .20	.19 .14	.14 .15
	4.5					
		.23 .07	.19 .14	.05 .23	.17 .17	.16 .15
	6.5					
3	0.5					
		.14 .11	.14 .12	.08 .20	.21 .07	.14 .13
	3.0					
		.16 .16	.21 .21	.30 .30	.26 .26	.23 .23
	5.5					
		.18 .11	.12 .07	.18 .15	.12 .05	.15 .10
	8.0					
4	2.5					
		.02 .16	.07 .14	.10 .25	.10 .12	.07 .17
	6.0					
		.23 .23	.29 .29	.30 .30	.19 .19	.25 .25
	9.5					
		.09 .05	.07 .21	.10 .20	.07 .07	.08 .13
	13.0					
5	0.0					
		.09 .32	.17 .07	.10 .03	.12 .05	.12 .12
	2.0					
		.14 .02	.07 .10	.05 .03	.05 .10	.08 .06
	4.0					
		.20 .20	.12 .12	.10 .10	.24 .24	.17 .17
	6.0					
6	2.0					
		.07 .36	.12 .21	.05 .18	.10 .19	.08 .24
	4.0					
		.18 .05	.17 .12	.08 .20	.17 .12	.15 .12
	6.0					
		.23 .23	.24 .24	.30 .30	.36 .36	.28 .28
	8.0					

The headings *Set* and *CO* and the first letter in the violation columns names, C, N, V, G, and M, keep table 2's format. The second letter stands for RDU with pessimism (ρ) or RDU with optimism (o). The underlined scores measure violations of CPT.

reduction principle + transitivity} as opposed to {comonotonic independence + reduction principle + transitivity}, and the obtained results may be disturbed by violations of the reduction or transitivity principle. Camerer and Ho (1994) discuss the desirability of disentangling principles that are to be tested from the confounding influence of assumptions such as the reduction principle and transitivity.

Our notion of independence can be tested in the common consequence version of the Allais paradox, when formulated in a state-format. It is not tested directly in the common ratio version of the Allais paradox; the latter involves more assumptions (including a kind of state-independence of utility). The common ratio test provides a direct test of *mixture independence* for decision under risk ($P \succcurlyeq Q \Leftrightarrow \lambda P + (1 - \lambda)R \succcurlyeq \lambda Q + (1 - \lambda)R$). Independence as defined and tested in this article is more basic than mixture independence for decision under risk.

It can be seen that our test, for the context of uncertainty, is also robust against state-dependence of utility which in the context of rank-dependence is studied in Chew and Wakker (1993).

The most well-known theory that does not assume the reduction principle or transitivity is regret theory. However, regret theory still requires independence in the event-contingent presentation of gambles studied in this article. Thus, all violations of independence observed in our experiment also constitute violations of regret theory. The explanation of regret theory for violations of independence in decision under risk, which assumes that subjects perceive different gambles as statistically independent, does not apply to our experiment. In our experiment, the two gambles of each choice pair were related to the same events; thus, it was clear that they are correlated and hence not statistically independent. Starmer (1992), who found strong regret effects, also devised an experiment that was robust against regret effects. He tested implications of RDU with pessimism, as well as some other theories, and found that none of these theories described his findings well. None of his tests of independence, however, included tests of comonotonic independence.

Nakamura (1992) provides a joint generalization of RDU and regret theory. It permits violations of noncomonotonic independence, but still implies comonotonic independence, without invoking the reduction principle or transitivity. Our experiment actually tested this nontransitive version of comonotonic independence and found that for our choice pairs the condition was not of descriptive value.

Most empirical studies in the literature have simultaneously tested several non-EU theories, and questions have not been tailored to RDU as they have been in our experiment. Usually, the stimuli are gambles for which the probabilities are not related to underlying events; for such stimuli, an elementary test of comonotonic independence is not feasible. In addition, most tests have almost exclusively studied the probability triangle, which is not a suitable domain for testing RDU for the following two reasons. First, if outcomes are fixed and probabilities are varied, then RDU allows for much freedom. This is contrary to betweenness theories; the latter, for instance, imply linearity of indifference curves in the probability triangle, and thus can well be tested there. Second, the probability triangle considers no more than three fixed outcomes, whereas any test of comonotonic independence requires four or more distinct outcomes. Let us state this

formally; a proof is given in appendix A. "Nontrivial" means that the choice should not be determined by dominance.

Observation 3. A nontrivial test of comonotonic independence requires at least four distinct outcomes.   □

RDU provides predictions across different probability triangles, i.e., indifference curves should retain certain patterns. They were, however, found not to (Starmer and Sugden, 1989; Camerer, 1989, p. 94; Camerer, 1992, Stylized Fact 5). Since all preferences considered in these references were between gambles from within the same triangle, a direct test of comonotonic independence was not possible.

Tests of RDU have also been restricted mostly to testing the pessimistic or optimistic version, and have concerned themselves with specific pessimism or optimism implications.[11] Camerer and Weber (1992, end of section 4) suggested the testing of comonotonic independence. Our experiment adds to the evidence that neither the pessimistic, nor the optimistic, version of RDU is empirically convincing. The mix of the two, as suggested by Tversky and Kahneman (1992), seems most promising. A mix of the two can also agree with the first proposal of RDU, by Quiggin (1982), that assumed $w(1/2) = 1/2$.

Wu (1993) tested "ordinal independence," a condition for decision under risk that under the reduction principle is somewhat weaker than comonotonic independence, but under continuity is equivalent (Chew and Wakker, 1993, Remark A1.1). For outcomes $x > y > z$, and probabilities p, q, r, s summing to one, Wu found, in choices between gambles $R = (CO, p; y, q; 0, r; 0, s)$ and $S = (CO, p; z, q; z, r; 0, s)$, that for $CO = x$ subjects would more often choose R, but for $CO = y$ they would more often prefer S. This constitutes a systematic violation of RDU, thus casting further doubt on the theory. Wu used a collapsed presentation of gambles, and explained his results through a hypothesis about the editing of transparently common outcomes. Note that in our experiment no collapsing of outcomes occurred outside the "Allais-type" questions in sets five and six.

7. Conclusion

This article has shown that the critical test between RDU and EU is the validity of comonotonic independence of preference against full-force independence. This test is the most basic one possible and is not affected by violations of the reduction principle, transitivity, or independence of utility from events. The experiment reported here provides the first pure test of RDU in its most general form.

The finding of the experiment is negative. Comonotonic independence does not perform better than full-force independence; i.e., RDU does not improve upon EU. Strong degrees of deviation of EU into the direction of RDU can be rejected by our data. Given the current popularity of RDU-like theories as descriptive alternatives to EU, our negative finding is noteworthy. We have not yet investigated the performance of comonotonic

independence for decision under uncertainty; this will be tested in Fennema and Wakker (in preparation).

The negative empirical evidence about better performance of the general RDU form in our experiment does not, of course, invalidate the theory. RDU can still provide a convenient and flexible mathematical form, special cases of which may yet provide significant improvements over EU. This may be the future of RDU: the discovery of special forms and special cases where it is of value. Our results suggest that most of the descriptive effectiveness of current rank-dependent utility models and extensions thereof in the experimental investigations may derive from features other than rank-dependence as such.

Appendices

Appendix A proves Observation 3. Appendix B gives probability calculations for a statistical analysis in section 5.

Appendix A. Comonotonic independence and the probability triangle

The proof of Observation 3 provided here shows that a nontrivial test of comonotonic independence is not possible in the probability triangle, which contains only three distinct outcomes.

For a nontrivial choice between two gambles, it is necessary that, for one event, one gamble is strictly better, and for another event, the other gamble. For a test of comonotonic independence, there must yet be another event for which the two gambles have a common outcome. In the simplest case, no other outcomes are involved. Therefore, suppressing events in the notation, we consider choices between gambles (R_h, R_l, CO) and (S_h, S_l, CO) and between gambles (R_h, R_l, CO') and (S_h, S_l, CO'). For nontriviality, $R_h > S_h, S_l > R_l$, and $CO' \neq CO$ and for comonotonicity $R_h > S_h \geq S_l > R_l$. Further, for comonotonicity, either $\{CO', CO\} \geq R_h$ or $S_h \geq \{CO', CO\} \geq S_l$ or $R_l \geq \{CO', CO\}$. These comparisons provide the simplest tests of comonotonic independence.

If there are only three distinct outcomes in $\{R_h, S_h, S_l, R_l\}$, i.e., if $S_h = S_l$, then at least one of $\{CO, CO'\}$ is not contained in $\{R_h, S_h, S_l, R_l\}$. Therefore, a nontrivial test of comonotonic independence always requires at least four distinct outcomes.

Appendix B. Statistical analysis based on an error theory

This appendix gives the probability calculus needed for the within-subjects statistical analysis of section 5.3. Recall that p_e denotes the probability for a choice error. Table 4 below gives the probabilities for all possible outcomes with two replications of a test of independence. A test of independence is defined as the choices of a given subject for two subsequent gamble pairs from a set. These gamble pairs differ only in the common

Table 4. Probability of observations, given state of true preferences, in pair of subsequent gambles in a set.

	Probability of observations in one test			
	strong viol.	weak viol.	(1,1)	strong verif.
true prefs verify independence	.0555	.4444	.1111	.3889
true prefs violate independence	.3889	.4444	.1111	.0555

Strong violations correspond to (2,0) and (0,2) scores, weak violations to scores (1,2), (1,0), (2,1), (0,1), and strong verifications to scores (0,0) and (2,2).

outcome value CO. Suppose first that the true preferences of the subject do not violate independence, i.e., for both gamble pairs, either the riskier gamble is preferred, or the safer gamble is preferred. Then a strong verification of independence, i.e., the observed pattern (2,2) or (0,0), occurs if, for the choices in both gamble pairs and for the two replications of both choices, the subject made no choice error, which occurs with probability $(1 - p_e)^4 = .3869$, or if the subject erred on all four choices, which occurs with probability $p_e^4 = .0020$. Altogether, the probability for a strong verification is thus .3889. The probability for a strong violation of independence (2,0 or 0,2), given that the true preferences do not violate independence, is $p_e^2(1 - p_e)^2 + (1 - p_e)^2 p_e^2 = .0555$.

If the true preferences do violate independence, i.e., in the first gamble pair R is preferred, and in the second gamble pair S or vice versa, then the probabilities of a strong violation and a strong verification are reversed.

Finally, the probability for a weak violation of independence is the probability for one inconsistency and one consistency, i.e., it is $\frac{1}{3} \times \frac{2}{3} + \frac{2}{3} \times \frac{1}{3} = 4/9 = .4444$, and the probability for a (1,1) score is the probability for two inconsistencies, i.e., $\frac{1}{3} \times \frac{1}{3} = .1111$. It may seem remarkable that weak violations are as probable if true preferences violate independence, as when they satisfy independence, so that an observed weak violation would not provide evidence against independence. In practice, weak violations do provide evidence against independence, because they suggest a change from strict preference to indifference (or from stronger to weaker preference strength), a point ignored in the present analysis where only strict preference and inconsistency are assumed.

Neither EU nor RDU predicts any true-preference violations for the comonotonic tests of independence. Hence the expected number of observed strong violations in the 504 comonotonic tests is $.0555 \times 504 = 27.97$, which does not deviate significantly from the 33 observed strong violations ($p = .29$). Since both theories predict the same for the comonotonic tests of independence, and the observed value does not deviate significantly from this prediction, the statistical analysis in the text only considered the 1008 noncomonotonic tests of independence.

The probability distribution of X, the total number of observed strong violations of noncomonotonic independence, depends of course on the total number n of true-preference violations. It, however, depends also on the number k of ("individual") sets[12] that contain two true-preference violations of noncomonotonic independence (leaving n

$- 2k$ sets that contain exactly one true-preference violation of independence). We denote by $H_{n, k}$ the hypothesis that k sets contain two true-preference violations of noncomonotonic independence, and $n - 2k$ sets contain exactly one true-preference violation of noncomonotonic independence. $H_{n, k}$ is a subhypothesis of H_n. We shall derive upper bounds for variances, given H_n, that hold true for each hypothesis $H_{n, k}$, so that our conclusions, given H_n, in the main text indeed hold true independently of the value of k.[13] We assume that choices differ from the true preferences only because of errors, and that errors in different choices are mutually independent. The contributions of the 504 different individual sets to X are then also mutually independent; further, the variances of these 504 contributions do not differ widely. Therefore, X is approximately normally distributed.

Under H_n, the expectation of X is independent of k, and it is

$$E(X) = .3889 \times n + .0555 \times (1008 - n)$$
$$= .3334 \times n + .0555 \times 1008 = n/3 + 55.944.$$

Next we calculate an upper bound for the variance of X under H_n, i.e., an upper bound for all hypotheses $H_{n,k}$. Note that, for sets 3 and 4, the first test of noncomonotonic independence concerns the first two gamble pairs, and the second test concerns the third and fourth gamble pairs. These two tests concern different gambles, so are statistically independent. For sets 1 and 2, the two tests involve one same gamble pair, i.e., the third pair, and thus are not independent. Similarly, the two tests for sets 5 and 6 both involve the second gamble pair, and thus are also not independent.

If a test of noncomonotonic independence shows a strong violation, then it contributes one count to the total number X of observed strong violations of independence, otherwise 0. Let us denote this contribution by Y; so $Y = 1$ or $Y = 0$. If the underlying true preferences satisfy independence, the variance of Y is $.0555(1 - .0555) = .0524$. If the underlying true preferences violate independence, the variance of Y is $.3889(1 - .3889) = .2377$. These values are listed in tables 5 and 6, as individual variance components. Let us now turn to a detailed derivation and explanation of the two tables.

We determine now the variance of the contribution to X of two tests of independence for a given subject in a set, when the two tests contain one gamble pair in common (e.g., in set 1, where the first test involves the second and third gamble pair, and the second test involves the third and fourth gamble pair). The contribution of the first test to X is denoted by Y, the contribution of the second test by Z.

Table 5. Probability calculations for sets 1,2,5,6.

	Var(Y), Var(Z)	P(2 strong viols)	Covar.	Var(Y + Z)
ind. verified in true prefs	.0524, .0524	.0185	.0154	.1356
1 viol. ind. in true prefs	.2377, .0524	.0185	− .0031	.2839
2 viols. ind. in true prefs	.2377, .2377	.2408	.0896	.6546

Table 6. Probability calculations for sets 3, 4.

	Var(Y), Var(Z)	P(2 strong viols)	Covar.	Var($Y + Z$)
ind. verified in true prefs	.0524, .0524	.0031	0	.1048
1 viol. ind. in true prefs	.2377, .0524	.0216	0	.2901
2 viols. ind. in true prefs	.2377, .2377	.1512	0	.4754

First, suppose that the true preferences satisfy independence, e.g., the subject prefers R in all three choices. Then the probability that Y and Z are both 1 (i.e., that the observed scores for the last three choice pairs of set 1 are either (2,0,2) or (0,2,0)) is $(1 - .2113)^2$ $.2113^2(1 - .2113)^2 + .2113^2(1 - .2113)^2.2113^2 = .0185$. The covariance of Y and Z is the expectation of their product (which is the probability that both are 1) minus the product of their expectations, i.e., it is $.0185 - (.0555)^2 = .0154$. The variance of $Y + Z$ is the sum of the variances of Y and Z plus twice their covariance, i.e., it is .1356.

Next, suppose that the true preferences contain one violation of independence. Say the true preferences are RSS for the last three choices of set 1 (the other cases, SRR, RRS, and SSR, are similar, and give the same results, except that in the latter two cases the individual variances of Y and Z are interchanged). Then the probability that Y and Z are both 1 (i.e., that two strong violations of independence are observed) is again $(1 - .2113)^4.2113^2 + .2113^4(1 - .2113)^2 = .0185$. The covariance of Y and Z is $.0185 - (.3889 \times .0555) = -.0031$. The variance of $Y + Z$ is the sum of the individual variances plus twice the covariance, i.e., it is $.2377 + .0524 - .0062 = .2839$.

Finally, suppose that the true preferences contain two violations of independence. Say the true preferences are RSR for the last three choices of set 1. Then the probability that Y and Z are both 1 (i.e., that two strong violations of independence are observed) is $(1 - .2113)^6 + .2113^6 = .2408$. The covariance of Y and Z is $.2408 - (.3889^2) = .0896$. The variance of $Y + Z$ is the sum of the individual variances plus twice the covariance, i.e., it is $.2377 + .2377 + 2(.0896) = .6546$.

For sets 3 and 4, the two tests of noncomonotonic independence are statistically independent. Thus, the variance of $Y + Z$ is computed as above, except that the covariance terms are zero. The variance of $Y + Z$ is equal to the variance of Y plus the variance of Z, i.e., it is .1048 if the true preferences satisfy independence, .2901 if the true preferences contain one violation of independence, and .4754 if the true preferences contain two violations of independence.

To find an upperbound for the variance under $H_{n,k}$, we shall maximize the difference of this variance with the variance under $H_{0,0}$ (no true-preference violations of independence). In sets 1, 2, 5, or 6, an individual set with two true-preference violations of independence generates an increase in variance (as compared to $H_{0,0}$) of $.6546 - .1356$ $= .5190$; in sets 3 or 4 the increase in variance is $.4754 - .1048 = .3706$ (that is, the generated increase in variance is highest in sets 1,2,5, or 6).

In sets 1, 2, 5, and 6, an individual set with one true-preference violation of independence generates an increase in variance (as compared to $H_{0,0}$) of $.2839 - .1356 = .1483$; in sets 3,4 the increase in variance is $.2901 - .1048 = .1853$ (hence, this is highest in sets

3,4). Therefore, under $H_{n,k}$, the maximal variance results if all k individual sets containing two true-preference violations of independence are in the sets 1, 2, 5, 6, and all $n - 2k$ individual sets containing one true-preference violation of independence are in sets 3, 4. Given that setup, we maximize with respect to k. If k is increased by one, then the increase in variance as compared to $H_{0,0}$ of sets 1, 2, 5, 6 is increased by another .5190, and the increase in variance as compared to $H_{0,0}$ of sets 3, 4 decreases by $2(.1853) = .3706$. Thus the maximal variance under H_n results if k is maximal, i.e., $k = n/2$; in this case, all violations of noncomonotonic independence are maximally grouped together in same individual sets, and further they all occur in sets 1, 2, 5, or 6. (For the moderate values of n, k that are relevant for our analysis, that configuration can always be obtained.) Note that this configuration is quite implausible, so that in reality the variance will be lower.

Given the described configuration of true-preference violations (and $4 \times 84 = 336$ individual sets belonging to sets 1, 2, 5, or 6, and $2 \times 84 = 168$ to sets 3 or 4), the variance of X is $.6546n/2 + .1356(336 - n/2) + .1048(168) = .2595n + 63.168$. We conclude that the variance of X is less than or equal to

$$.2595n + 63.168.$$

Next we determine the likelihood function over the parameter values of n, given the observation $X = 62$. For the variance we take again the upper bound derived above. The true variance is somewhat smaller, which means that the likelihood ratios are in reality even farther removed from 1, and our results are actually slightly stronger than our claims. With the variance as described, the likelihood function (multipiled by $\sqrt{2\pi}$) is

$$\frac{1}{\sqrt{.2595n + 63.168}} \times \exp\left(-\frac{(6.056 - n/3)^2}{2(.2595n + 63.168)}\right).$$

Table 7 gives the likelihood function for some values of n, as well as the likelihood ratio of EU ($n = 0$) against those values of n.

Acknowledgments

Table 7. The likelihood function, and likelihood ratios, for some values of n.

	$n = 0$	$n = 54$	$n = 63$	$n = 84$
likelihood	.094	.045	.028	.0064
likelihood ratio EU against H_n	1	2.08	3.42	14.76

Notes

1. Wu (1993), in a study of which we were only recently informed, tested "ordinal independence," a condition for decision under risk that under continuity is equivalent to comonotonic independence. He found systematic violations of ordinal independence, and thus of RDU.
2. In the literature on decision under risk, a dual approach is often used, where outcomes are processed in a reversed order. By some elementary manipulations, this approach can be shown to be data-equivalent to our approach; convexity of the weighting function in our approach is equivalent to concavity in the dual approach, and vice versa.
3. Copies of the instructions and the computer program to run the experiment are available upon request.
4. There are a variety of psychological interpretations of the RDU transformation that go beyond transformations from objective to subjective probabilities (Weber, in press). However, the present experiment was designed to demonstrate the existence of the phenomenon rather than its interpretation.
5. Remember that, in the literature for decision under risk, a way of integration is often used that is dual to our way. Convexity of our weighting function corresponds to concavity in this dual approach.
6. Weber, Anderson, and Birnbaum (1992) found strong evidence for differential weighting of probabilities associated with positive, negative, or zero outcomes for judgments of attractiveness as well as risk.
7. Absolute difference between .5 and the proportion of risky choices across subjects.
8. The latter two would perform somewhat better than RDU-in-general, the model tested here. Exact probability calculations would, however, be complicated, because a symmetry between riskier and safer choices that has greatly simplified the following analysis, does not hold true for specific RDU models such as CPT.
9. As explained in appendix B, the strong violations of comonotonic independence, and all weak violations of independence, do not provide statistical information under the present assumptions.
10. Given by Birnbaum (1992), personal communication.
11. Wakker (1994a) showed that, given RDU, pessimism is equivalent to quasiconvexity of preferences with respect to probabilistic mixtures, and optimism is equivalent to quasiconcavity. The derivation of this result is not elementary, and it has not been known before in the literature. It suggests that tests of RDU with pessimism were rather tests of quasiconvexity.
12. We use the term "set" both in a generic sense, and to describe a set-per-individual. When confusion might arise, we say "individual set" for set-per-individual.
13. The probability distribution of X depends, given n and k in $H_{n,k}$, on one more thing, i.e., the numbers n' and k' of these violations in sets 3 and 4. (Thus, the remaining numbers of these violations in sets 1,2,5,6 are $n - n'$ and $k - k'$.) The upper bounds derived below will cover all possible configurations.

References

Battalio, R.C., J.H. Kagel, and K. Jiranyakul. (1990). "Testing between Alternative Models of Choice under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty* 3, 25–50.

Bernasconi, M., and G. Loomes. (1992). "Failures of the Reduction Principle in an Ellsberg-Type Problem," *Theory and Decision* 32, 77–100.

Birnbaum, M.H. (1974). "The Nonadditivity of Personality Impressions," *Journal of Experimental Psychology* 102, 543–561.

Birnbaum, M.H., G. Coffey, B.A. Mellers, and R. Weiss (1992). "Utility Measurement: Configural-Weight Theory and the Judge's Point of View," *Journal of Experimental Psychology: Human Perception and Performance* 18, 331–346.

Camerer, C.F. (1989). "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty* 2, 61–104.

Camerer, C.F. (1992). "Recent Tests of Generalizations of Expected Utility Theory." In W. Edwards (ed.), *Utility Theories: Measurement and Applications*. Dordrecht: Kluwer Academic Publishers, pp. 207–251.

Camerer, C.F., and T.-H. Ho. (1994). "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty* 8, 167–196.

Camerer, C.F., and M. Weber. (1992). "Recent Developments in Modelling Preferences: Uncertainty and Ambiguity." *Journal of Risk and Uncertainty* 5, 325–370.

Chew, S.H., and P.P. Wakker. (1993). "Generalizing Choquet Expected Utility by Weakening Savage's Sure-Thing Principle." University of California, Irvine Research Unit in Mathematical Behavioral Sciences.

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.

Conlisk, J. (1989). "Three Variants on the Allais Example," *American Economic Review* 79, 392–407.

Edwards, W. (1962). "Subjective Probabilities Inferred from Decisions," *Psychological Review* 69, 109–135.

Erev, I., G. Bornstein, and T.S. Wallsten. (1993). "The Negative Effect of Probability Assessments on Decision Quality," *Organizational Behavior and Human Decision Processes* 55, 78–94.

Fishburn, P.C., and P.P. Wakker. (1992). "The Invention of the Independence Condition," *Management Science*, forthcoming.

Harless, D.W. (1992)."Predictions about Unit Triangle Indifference Curves inside the Unit Triangle: A Test of Competing Decision Theories," *Journal of Economic Behavior and Organization*, forthcoming.

Kahneman, D., and A. Tversky. (1979). "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263–291.

Karni, E., and Z. Safra. (1990). "Rank-Dependent Probabilities," *Economic Journal* 100, 487–495.

Keller, L.R. (1985). "Testing of the 'Reduction of Compound Alternatives' Principle," *Omega* 13, 349–358.

Luce, R.D., and P.C. Fishburn. (1991). "Rank- and Sign-Dependent Linear Utility Models for Finite First-Order Gambles," *Journal of Risk and Uncertainty* 4, 29–59.

Machina, M.J. (1989). "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty," *Journal of Economic Literature* 27, 1622–1688.

MacCrimmon, K.R., and S. Larsson. (1979). "Utility Theory: Axioms versus 'Paradoxes'." In M. Allais and O. Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox*. Dordrecht: Reidel, pp. 27–145.

Nakamura, Y. (1992). "A Generalization of Subjective Expected Utility without Transitivity and Additivity." Paper presented at Sixth FUR Conference, Cachan, France.

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

Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley. (Second edition: 1972, New York: Dover.)

Slovic, P., and A. Tversky. (1974). "Who Accepts Savage's Axiom?" *Behavioral Science* 19, 368–373.

Starmer, C. (1992). "Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect," *Review of Economic Studies* 59, 813–830.

Starmer, C., and R. Sugden. (1989). "Violations of the Independence Axiom in Common Ratio Problems: An Experimental Test of Some Competing Hypotheses," *Annals of Operations Research* 19, 79–101.

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

von Neumann, J., and O. Morgenstern. (1944, 1947, 1953). *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.

Wakker, P.P. (1994a). "Separating Marginal Utility and Probabilistic Risk Aversion," *Theory and Decision*, 1–44.

Wakker, P.P. (1994b). "Independence and Comonotonic Independence: An Axiomatic Analysis." Medical Decision Making Unit, University of Leiden, The Netherlands.

Weber, E.U. (1993). "From Subjective Probabilities to Decision Weights: The Effects of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events," *Psychological Bulletin*, in press.

Weber, E.U., C.J. Anderson, and M.H. Birnbaum. (1992). "A Theory of Perceived Risk and Attractiveness," *Organizational Behavior and Human Decision Processes* 52, 492–523.

Wu, G. (1993). "Editing and Prospect Theory: Ordinal Independence and Outcome Independence." Working paper, Harvard Business School, Boston, Mass.