A Test of Rank-Dependent Utility in the Context of Ambiguity

HEIN FENNEMA
University of Nijmegen, NICI, Nijmegen, The Netherlands

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

Abstract

Experimental investigations of non-expected utility have primarily concentrated on decision under risk ("probability triangles"). The literature suggests, however, that ambiguity is one of the main causes for deviations from expected utility (EU). This article investigates the descriptive performance of rank-dependent utility (RDU) in the context of choice under ambiguity. We use the axiomatic difference between RDU and EU to critically test RDU against EU. Surprisingly, the RDU model does not provide any descriptive improvement over EU. Our data suggest other "framing" factors that do provide descriptive improvements over EU.

Key words: ambiguity, rank dependence, non-expected utility, comonotonicity, presentation effects
JEL code: D81

In the past decades, students of individual choice have demonstrated that preference behavior deviates from Expected Utility theory (EU). In response, various generalizations and alternatives to EU have been put forward. Axioms that are thought to be descriptively invalid have been weakened to accommodate the observed violations. For most theorists, the independence axiom is the major culprit, exemplified by the Allais and Ellsberg paradoxes. In the early 1980s, several new models have been proposed that weaken the independence axiom, thus generalizing EU. Two major kinds of transitive generalizations have emerged: betweenness models and rank-dependent models.

Recently, empirical research is cumulating that tests whether the new proposals are able to explain the experimental facts found thus far (Camerer and Ho, 1994). Most of this research can be summarized using so-called probability triangles. These triangles employ three outcomes that are kept constant. Gambles are constructed by varying the probabilities assigned to these outcomes. Most of the theories considered can be tested by observing preference behavior on selected pairs of gambles from these triangles. The overall picture emerging from this research is that none of the generalized expected utility theories can explain all the systematic violations that have been discovered. The betweenness condition, for example, is found to be systematically violated (Camerer and Ho, 1994). The rank-dependent models (RDU), especially the new version of prospect theory, Cumulative Prospect theory, are at the moment probably the leading contender for best descriptive theory. However, some remarks must be made.

First, several phenomena cannot be explained by the rank-dependent models. For instance, it turns out that the shapes of indifference curves in probability triangles are not independent of the amounts of money that are used, whereas they should be, according to RDU. Second, to reach conclusions, most experiments done thus far assume other axioms, such as transitivity and reduction, axioms that are also necessary for RDU. It is not always evident that the reported violations of EU are due to violations of the independence axiom rather than violations of the assumed axioms. (An exception is Camerer and Ho (1994), who did verify that their observed violations of betweenness were not due to violations of transitivity.)

Third, and perhaps most important, most of the research conducted thus far is not very well-suited for testing rank-dependent models. In probability triangles, outcomes are kept constant and probabilities are varied. Because RDU theories introduce the weighting of probabilities, they are especially well-suited for handling variations in probability, and therefore impose few testable predictions on the probability triangle. Theories that do not assume a probability weighting function are tested more strictly in the triangle. It can be shown (Wakker, Erev, and Weber, 1994, Observation 3) that it is not possible to test RDU conclusively with stimuli that employ only three outcomes, as is done in probability triangles. Another problem with probability triangles is that they can only be used to test theories for decision under risk, not for decision under ambiguity. It is not a priori true that results concerning EU generalizations in the risky case can be extrapolated to the ambiguous case.

Wakker (1996) outlines a test that provides a precise empirical demarcation between RDU and EU. This test avoids the complications described above and will be discussed in detail shortly. Wakker, Erev, and Weber (1994) employed the test in a risky setting and found a negative result. That is, there was no evidence that people were sensitive to rank dependence in the context of risk. Thus, it appears that RDU also faces some serious descriptive shortcomings in the risky case.

This article studies the descriptive performance of RDU for decision under ambiguity. Outcomes are not generated with given probabilities, but depend on events with unknown probabilities generated by irregular dice. Although people rarely encounter decision situations where probabilities are known, nearly all experimental studies of decision making have concentrated on risk. (For a review of exceptions, see Camerer and Weber, 1992.) Therefore, little is known about the descriptive success of generalized EU and RDU under ambiguity. There are, however, several reasons why RDU can be expected to be more useful for ambiguity than for risk. This is explained next.

RDU generalizes EU by permitting decision weights to depend not only on events but also on outcomes. More precisely, the decision weights are affected by the rank ordering of the outcomes. This outcome dependence of decision weights can have more effect under ambiguity than under risk, because, under ambiguity, it will not only affect the weight of given probabilities in decisions but also the preceding assessment of such probabilities or likelihoods.[1] Hence weighting functions can be expected to exhibit more nonadditivity in ambiguity. Weber (1994) describes uncertainty concerning probability as the primary source for outcome dependence. In the normative model of Schmeidler (1989) that initiated the rank-dependent model for ambiguity, it was explicitly assumed

that nonadditivity was only due to ambiguity and not to risk. Further, Tversky and Fox (1995) found more pronounced nonadditivity for ambiguity than for risk in an empirical study. These observations suggest that RDU can be a powerful theory for decision under ambiguity, and they motivate the experiment of this article.

Let us point out one restriction concerning the results of our study. One of the main conclusions of recent experimental research in decision theory is that framing and presentation have a large impact on observed decisions. In our experiment, we chose the, in our opinion, most basic binary choice tasks, and our stimuli are relatively simple and transparent. Therefore, the common outcomes in our stimuli were transparent. An interesting question is to what degree the results of our study can be affected by different presentations of the stimuli, in particular, if those presentations conceal common outcomes. This question has been studied in further detail by Weber and Kirsner (1995) and is discussed in section 4 below.

1. Testing RDU: the theoretical design of the experiment

In decision theory, preferences are taken as the observable primitives of the theory, and theories are ultimately characterized by axioms defined on these preferences. The central axiom of decision under ambiguity as formulated by Savage (1954) is the *sure-thing principle*[2]: If two gambles share an event that yields the same outcome, then this common outcome should not matter for the decision. That is, subjects should behave as if they "cancel" the common outcome prior to making their choice. Consequently, the preference between these gambles should not be affected if we change the common outcome.

It can be shown that the essential difference between RDU and EU comes down to a weakening of the sure-thing principle (STP) to the case where the gambles are "comonotonic." All other differences between EU and RDU are, in a way, nonessential (for a precise formulation see Wakker, 1996). Two gambles are *comonotonic* if they induce the same order on the events according to the favorability of the outcomes. This condition will be explained by an example, also introducing the kind of stimuli that were used in the experiment.

Figure 1 presents two gambles. Amounts are Dutch guilders ($1 equals about Fl 1.65).

	Event 1	Event 2	Event 3
Gamble R	2	7	12
Gamble S	3	5	12

Figure 1. Two gambles, R and S. Common outcome equals 12.

The gamble with the larger spread of outcomes is denoted by R (risky), the other one by
S (safe). For gamble R, the best outcome, 12, is obtained under event 3. Event 2 yields the
second best outcome, 7, and Event 1 the worst outcome, 2. The same order of the events
is induced by the rank-order of the outcomes of gamble S, as is easily verified. If we
construct two gambles R' and S' on the same events (figure 2), by changing the common
outcome on event 3 from $12 to $8, then this pair of gambles is again comonotonic.

Moreover, the change of the common outcome does not change the rank order in
comparison to the original gambles. Therefore, any pair of these four gambles is comono-
tonic; they all induce the same rank ordering on the events. In this case, where all gambles
are pairwise comonotonic, the *comonotonic STP* applies. It forbids a preference reversal
for the choices in figure 1 and 2: only the preferences R and R' or the preferences S and
S' are allowed.

Next consider a change of the common outcome from 8 to 4 (figure 3). Now Event 2
has become the best event instead of Event 3. In this case, the gambles R" and S" induce
a different rank ordering on events than the gambles R' and S'. The four gambles are not
pairwise comonotonic, therefore the comonotonic STP does not apply. RDU now allows
each of the four possible patterns of preferences. The choices in figures 2 and 3 provide
a test of the STP that will be called non-comonotonic.

If the axiomatic difference between EU and RDU lies in the difference between the STP
and the comonotonic STP, then the critical empirical test of RDU is to compare the
violation rate of the STP with the violation rate of the comonotonic STP. If RDU has any
additional value to EU, then there must be systematic violations of the non-comonotonic
STP. Consequently, the violation rate of the non-comonotonic STP should then be higher
than the violation rate of the comonotonic STP. The main goal of the experiment is to test
this hypothesis.

An additional part of the design introduces tests of RDU in combination with particular
weighting functions. The explanation of these tests is more formal and is described next.
First we introduce some notation. The *RDU value* of a gamble yielding X_i if event E_i ($i =$
1, 2, 3) occurs can be written as:

$$\pi_1 U(X_1) + \pi_2 U(X_2) + \pi_3 U(X_3), \tag{1}$$

	Event 1	Event 2	Event 3
Gamble R'	2	7	8
Gamble S'	3	5	8

Figure 2. Common outcome equals 8.

	Event 1	Event 2	Event 3
Gamble R''	2	7	4
Gamble S''	3	5	4

Figure 3. Common outcome equals 4.

where the π's are the "decision weights" (and subscripts refer to events). They depend on the rank order of the outcomes. If $X_1 \geq X_2 \geq X_3$, the *decision weights* are defined as follows:

$$\pi_1 = W(E_1), \pi_2 = W(E_1 \cup E_2) - W(E_1) \text{ and } \pi_3 = 1 - W(E_1 \cup E_2);$$

here, W is the *weighting function*. W assigns value 1 to the universal event, value 0 to the impossible event, and satisfies $W(A) \geq W(B)$ whenever A contains B.

We can now demonstrate formally that RDU cannot explain a change of preference if the common outcome changes from 8 to 12, as discussed above. The change from 8 to 12 does not affect the rank ordering of outcomes; hence the decision weights do not change. One easily verifies that preferences cannot then reverse.[3] This proves that RDU implies the comonotonic STP. Next we consider tests of the non-comonotonic STP. Here the general RDU model does not impose restrictions, for decision weights can change and no conclusions can be drawn.

If we specify characteristics of the weighting function, however, additional restrictions are implied. We derive these restrictions using again the example from the experiment. First we discuss the change of common outcome from 0 to 4, next from 4 to 8. Suppose a subject prefers R''' to S''' in the choice question depicted in figure 4. This implies that the RDU value of R''' exceeds the RDU value of S''':

$$\pi_1 U(2) + \pi_2 U(7) + \pi_3 U(0) > \pi_1 U(3) + \pi_2 U(5) + \pi_3 U(0),$$

	Event 1	Event 2	Event 3
Gamble R'''	2	7	0
Gamble S'''	3	5	0

Figure 4. Common outcome equals 0.

where the decision weights π_1, π_2, and π_3 are next explained. For R''' and S''' , the highest outcome is obtained under event E_2; therefore its decision weight π_2 is $W(E_2)$. Event E_1 yields the second highest outcome; therefore π_1 is $W(E_1 \cup E_2) - W(E_2)$. Finally, event E_3 yields the worst outcome; therefore π_3 is $1 - W(E_1 \cup E_2)$. The above formula can be rewritten as:

$$\pi_2[U(7) - U(5)] > \pi_1[U(3) - U(2)]. \tag{2}$$

If we change the outcome from 0 to 4 (figure 3), Event 1 becomes the worst event. Suppose that, in this case, the same subject prefers S'' to R'', implying a violation of the non-comonotonic STP. Now the RDU value of S'' exceeds the RDU value of R'':

$$\pi_2'[U(7) - U(5)] < \pi_1'[U(3) - U(2)], \tag{3}$$

where $\pi_1' = 1 - W(E_2 \cup E_3)$ and $\pi_2' = W(E_2)$ ($=\pi_2$). This pattern of preferences does not constitute a violation of RDU, for it can be explained if π_1 in (2) is smaller than π_1' in (3), so if

$$W(E_1 \cup E_2) - W(E_2) < 1 - W(E_2 \cup E_3).$$

A weighting function W is *concave* if for all disjoint events A,B,D,

$$W(A \cup B) - W(B) \geq W(A \cup B \cup D) - W(B \cup D).$$

That is, the marginal weight contribution of Event B (i.e., its decision weight) becomes smaller if Event A is enlarged to A∪D. A *convex* function is defined similarly, with the reversed inequality; so then the marginal weight contribution of Event B to Event A is increasing in A. If we substitute E_1 for A, E_2 for B, and E_3 for D, then it follows that a concave weighting function cannot explain the stated inequality. Thus, RDU in combination with a concave weighting function cannot explain the above preferences. RDU combined with a convex weighting function can explain these preferences, for it predicts that π_1 in (2) is smaller than or equal to π_1' in (3). Convexity of the weighting function can be interpreted as modeling pessimism. A pessimistic subject predominantly looks at the worst outcome. In figure 4, the common outcome of 0 is the lowest outcome, so Event 3 receives much attention. If we change the common outcome from 0 to 4, Event 1 becomes the worst event and receives more attention than before. Under Event 1, the risky gamble yields 2; whereas the subject receives a better outcome, 3, if he chooses the safe gamble. Therefore, the change of the common outcome from 0 to 4 enhances the safe choice, in agreement with the predictions of RDU with a convex weighting function.

If we change the common outcome from 4 to 8 (figure 3 and figure 2), the common outcome becomes the best outcome. If W is convex, Event 2 will receive a larger decision weight than before. In this case, the risky choice is enhanced, for the risky gamble yields a larger outcome on Event 2 than does the safe gamble. Thus, RDU with a convex weighting function cannot model a preference for the gamble R'' in figure 3 combined

with a preference for the gamble S' after the change of the common outcome to 8. RDU with a concave weighting function precludes the reversed change of preference.

In cumulative prospect theory, the weighting function is S shaped. For decision weights, this implies that, for good outcomes and for bad outcomes, decision weights in cumulative prospect theory are relatively large. Middle outcomes receive relatively small decision weights. Recapitulating, we can say that, for a convex weighting function, the low outcomes receive more attention (pessimism); for a concave weighting function, the high outcomes receive more attention (optimism); and for cummulative prospect theory, the extreme outcomes receive more attention. Table 1 summarizes all additional restrictions that RDU combined with a particular weighting function imposes on our data.

2. The experiment

The experiment was conducted with 60 undergraduate students at the University of Nijmegen, most of whom were majoring in psychology. They were paid according to the random lottery method (see Starmer and Sugden, 1991, for comments on its reliability): one of the choice questions was randomly selected at the end, and the choice of the subjects was played out. Besides the lottery, yielding maximally Fl 15 (about $9) and minimally nothing, the subjects all received Fl 3 for their participation. All preferences were administrated by computer, including the response times.

The ambiguous events were generated by three different irregular dice. Each die had six different sides, which were grouped into three events by different colors. Colors are used to label and identify events, i.e., "red" denotes the event that the die lands red side *down*. (This unusual way of "reading" the event from a die simplified the estimation of probabilities: larger sides are more probable than smaller ones.) Subjects received no frequentistic information about the dice: they were encouraged to examine the dice, but they were not allowed to throw them. We ran an independent experiment ($n = 10$), to get an impression of the perceived likelihoods of the events generated by the dice. The results are reported in table 2.

Two gambles were constructed on the events, called S and R in this article, shorthand for safe and risky. The subjects were asked to state which of the two gambles they would prefer to play if they were given the opportunity to play one of them. No indifferences were allowed.

Table 1. Patterns of violations of the non-comonotonic STP (i.e., changes of preference) that *contradict* RDU combined with a weighting function as specified in the columns. Opposite changes can be explained.

rank order of common outcome event changes from:	Convex	Concave	cumulative prospect theory
worst → middle	S → R	R → S	S → R
middle → best	R → S	S → R	S → R

Table 2. Means and standard deviations (SD) of the perceived likelihoods.

	Die I			Die II			Die III		
	White	Yellow	Grey	Green	Blue	Red	Purple	Green	Yellow
Mean	0.099	0.242	0.659	0.135	0.186	0.679	0.311	0.304	0.384
SD	0.029	0.071	0.075	0.071	0.072	0.113	0.031	0.031	0.041

The test consisted of 36 choice questions. The order of these choice questions was randomized. Three choice questions were replicated in the experiment. These were randomly selected and varied from one subject to the other. The replications were added to provide some indication of the consistency of the subjects. We also included nine filler items, where one gamble was clearly superior to the other (for any reasonable estimate of the likelihoods of the events). The filler items were included in order to motivate subjects to consider their choices carefully. Also, these fillers helped to prevent recognition, and successfully so, according to the pilot study (21 subjects, written evaluation). The test was preceded by twelve training questions to avoid learning effects during the test.

The stimuli were chosen according to a number of criteria. The gambles had to be close in preference, because we wanted to leave room for changes in preference due to changes in the common outcome. For all choice problems, we included a small risk premium. All outcomes were positive to exclude sign effects (Tversky and Kahneman, 1992).

We constructed nine quadruples of choice questions that were the same except for a change of a common outcome. The quadruples are called sets henceforth, and are denoted by A, $\cdots$, I (see tables 3–5). For example, the safe gamble in the first gamble pair of set A, denoted by (2, White; 2, Yellow; 0, Grey), yields 2 if die I lands white down, 2 if die I lands yellow down, and 0 if die I lands grey down. The risky gamble in the first gamble pair of set A is (0, White; 3, Yellow; 0, Grey). The other three gamble pairs of set A are the same, except that the common outcome for grey is replaced by another common outcome (2,4, and 6, respectively). The other sets in tables 3,4, and 5 are defined similarly, but table 4 referes to die II, and table 5 to die III. Set G of table 5 was already illustrated in figures 1–4.

For comparability, we always changed the common outcome by the same amount. One change of the common outcome yielded a test of the comonotonic STP, the other *two* changes provided tests of the non-comonotonic STP.

Table 3. Sets of gambles generated by die 1. The underlined common outcomes yield gamble pairs that provide a test of the comonotonic STP.

Set of gambles	S/R	White	Yellow	Grey (Common outcome)			
A	Safe	2	2	0	2	4	6
	Risky	0	3				
B	Safe	10	10	0	5	10	15
	Risky	5	12				
C	Safe	4	9	0	4	8	12
	Risky	0	11				

Table 4. Sets of gambles generated by die II.

Set of gambles	S/R	Green	Blue	Red (Common outcome)			
D	Safe	5	5	1	3	5	7
	Risky	7	4				
E	Safe	2	8	0	4	8	12
	Risky	0	12				
F	Safe	2	3	0	3	6	9
	Risky	0	6				

Table 5. Sets of gambles generated by die III.

Set of gambles	S/R	Purple	Green	Yellow (Common outcomes)			
G	Safe	3	5	0	4	8	12
	Risky	2	7				
H	Safe	8	3	0	4	8	12
	Risky	12	0				
I	Safe	6	6	0	3	6	9
	Risky	9	4				

Subjects were randomly assigned to one of four different conditions concerning the presentation of the items. These conditions were included, because we wanted to control for artefacts due to presentation effects.

The matrix condition presented the gambles in rows and the events in columns (figure 5). The graphic condition presented colored two-dimensional "snapshots" of the die with the outcomes plotted on the sides (figure 6).

The other two conditions were almost identical, both using a verbal description of the gambles. In the verbal condition, each gamble had a column that first stated the event followed by its outcome (figure 7a). The verbal-collapsed condition was the same, except that if two outcomes were the same, the events were grouped, for example, "White and Grey Side: Fl 4" (figure 7b). For all conditions, the positioning (top/bottom) of the R and S gamble was randomized over subjects.

Die I		White	Yellow	Grey
	Option A	ʃ 4	ʃ 9	ʃ 4
	Option B	ʃ 0	ʃ 11	ʃ 4

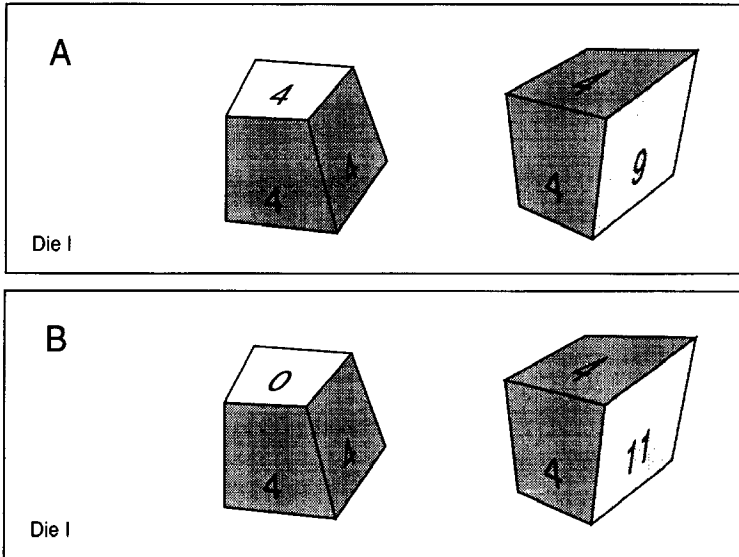*Figure 5.* The matrix condition.

Figure 6. The graphic condition.

3.Results

Our main hypothesis states that if we change the common outcome, subjects will violate the STP less often if the change is comonotonic, then if it is non-comonotonic. The null hypothesis assumes, to the contrary, that comonotonicity has no effect, i.e., violations of the STP are as likely for non-comonotonic cases as for comonotonic cases. As all sets of

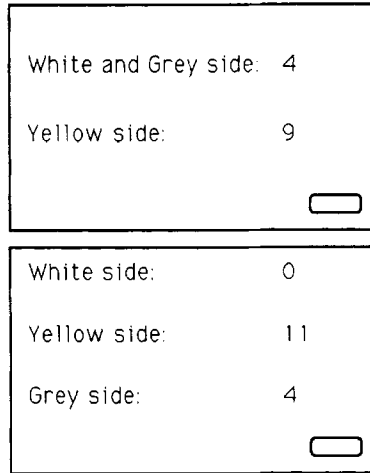*Figure 7a.* The verbal condition.

Figure 7b. The verbal-collapsed condition.

items (A-I, tables 3–5) offered one comonotonic change and two non-comonotonic changes of the common outcome, the null hypothesis predicts that in our data there will be twice as many violations of the non-comonotonic STP as of the comonotonic STP.

The left half of table 6 shows that our expectation was invalidated. We found no effect of comonotonicity ($p = .41$, binomial).

Some violations of the STP will be due to inconsistency on behalf of the subjects. A comparison with the inconsistency rate of replications gives an indication if there are also systematic violations of EU. The null hypothesis states that all violations are due to inconsistencies. If we take the variability between subjects into account, the null hypothesis cannot be rejected: the inconsistency rate of 0.283, and the violation rate of 0.264 are essentially equal in a paired comparison ($t(59) = -.57, p = .714$, one-tailed). It seems that observed violations of the STP should be explained by random inconsistencies of the subjects. The inconsistency rate of about 30% is not uncommon in choice experiments (Camerer, 1989). While strict inconsistency is a form of irrationality, in the case of our experiment there may be a more rational explanation for part of the observed inconsis-

Table 6. Violations of the STP and violations of RDU, combined with a particular weighting function.

Condition	Comono- tonic STP	Non-comono- tonic STP	STP	Convex	Concave	Cumulative prospect theory
Matrix	24	55	79	27	28	28
Verbal	36	70	106	42	28	39
Graphic	30	73	103	37	36	39
Collapsed	50	90	140	54	36	41
Total	140	288	428	160	128	147

tencies: The gambles were designed to be close in preference, so it could well be that subjects were more or less indifferent between several gambles and would then choose randomly.

The design of the experiment also allowed us to investigate which functional forms of the weighting function performed best: convex, concave, or S-shaped. The results are presented in the right half of the table 6. RDU with a concave weighting function performed best. This is quite remarkable, as it deviates from common findings reported in other experimental papers (Tversky and Kahneman, 1992; Wakker, Erev and Weber, 1994; Wu, 1994) and is also contrary to our expectations that subjects would be ambiguity-averse for the kind of events which we employed. However, the difference between violation rates of concave and convex RDU are not significant if we take between-subject variability into account (t(59) = 1.51, p = .135). Besides, there is also reason to believe that the success of the concave functional can be partly explained by a side effect found in the verbal-collapsed condition, to which we will turn later. The patterns of preferences which we have found here also fall perfectly well within the range of what may be expected if violations of the STP are only random inconsistencies.

We expected that there would be no large differences between the conditions, for all the conditions gave the same information in a reasonably transparent manner. However, there were some significant differences, summarized in table 7.

There seems to be a significant and straightforward relation between the transparency of the presentation and the coherence and response time of the choices. The matrix condition, and, to a smaller degree, the verbal and graphic conditions, promote cancelling of the common outcome. Hence, we observe a much smaller rate of violations of the STP here than in the verbal-collapsed condition, where common outcomes need not be transparently displayed. The inconsistency rate and response time are also much larger in the verbal-collapsed condition, indicating that the verbal-collapsed condition is more difficult, yielding more inconsistencies.

The high rate of violations of the STP in the verbal-collapsed condition can also be partly explained by an interesting phenomenon that is due to collapsing. The collapsing effect asserts (in the context of gains) that if the common outcome collapses with an outcome of one of the gambles, then this gamble become *less* attractive. This biases violations of the STP, for, if there is a choice question where one gamble has an outcome that collapses with the common outcome, the collapsing effect predicts that subjects will be biased in favor of the gamble that has no collapse (Tversky and Kahneman, 1986;

Table 7. Comparison of mean values.

Condition	Violations of STP (Out of a possible 27)	Inconsistencies (Out of a possible 3)	Mean time (sec)	% Risky
Matrix	5.27†	0.60	16.62	40.9
Verbal	7.07	0.40†	18.66	41.5
Graphic	6.87	1.13*	16.92	41.3
Verbal-Collapsed	9.33*	1.27*	21.97	46.5

Within columns, † is significantly different from *, using Tukey's HSD at the .05 level.

Starmer and Sugden, 1989). The intuitive rationale for this effect seems to be that, because there are fewer distinct outcomes when there is a collapse, the gamble appears to be less attractive *visually* (see figure 6). A similar phenomenon is described in support theory (Tversky and Koehler, 1994), where "unpacking" an event enhances its perceived likelihood.

If a gamble has only one outcome besides the common outcome (as happens in sets A,B,D, and I), then collapsing introduces a certain outcome, and a certainty effect (Kahneman and Tversky, 1979) can be induced. In such cases, we cannot test for collapsing properly, because the certainty effect interferes. Therefore, collapses that introduce a certain outcome have been ignored in the analysis.

We found significant collapsing effects in the verbal-collapsed condition. The collapsing effect predicts that there will be additional violations of the STP, because a gamble with a collapsed common outcome is less attractive. For each individual, the number of violations of the STP agreeing with the collapsing effect, and the number of violations disagreeing, were determined. A paired sample t-test showed a significant effect ($t(14) = 3.29$, $p < .01$). The hypothesis that there is no difference between the direction of violations predicted by the collapsing effect between the verbal-collapsed condition and the other conditions could also be rejected ($F(1,58) = 8.70$, $p < .01$). The finding of a collapsing effect suggests that subjects are more sensitive to seemingly innocuous and, according to EU, irrelevant variations in presentation than they are to comonotonicity.

The program that administrated the choices also registered the response time. Although the experiment was not designed to test explicit hypotheses concerning response times, some points can be made. For instance, we observed that there were significant differences between mean response times on items constructed on the three different dice.[4] This indicates that subjects were sensitive to differences in the source of ambiguity.

It is plausible to suppose a connection between strength of preference and response time. If the subject has a strong preference, he will find the choice problem easy and will answer quickly. If the subject is more or less indifferent between the two gambles, he will need more time to consider all relevant aspects of the available gambles. This is our main working hypothesis concerning response times. Given this hypothesis, we can investigate the explanation that part of the inconsistencies and thus part of the violations of the STP are caused by indifference. If subjects are indifferent between two gambles, the chance at violations of the STP and at violations of consistency on replications will increase. Hence, these chances will be positively correlated with response time. Table 8 shows that subjects indeed used significantly more response time for the inconsistent choices.[5]

Table 8. Mean response times (in seconds) and the probability of obtaining these differences (or larger), on the assumption that mean times are equal.

	Violating pairs	Verifying pairs	Significance
Test of consistency	21.74	18.18	.045
Test of STP	21.57	17.93	.000

We also investigated which factors determine the response times of individual items. Most of the variance (25%) could be explained by differences between subjects (predicting a response time by the mean of the response times of the subject that gave the response). There are large differences between overall speed: the average time to make a choice was 18.5 seconds, with a large standard deviation of 6.4 seconds. Some of this variability (16%) can be explained by risk attitude. Subjects who are relatively risk-averse (more then 50% safe choices) are also relatively quick. This could point to the use of quick heuristics leading to safe choices.

If we look at differences between items, we can also obtain predictive power concerning response time. The most predictive component, additional to within-subject differences, was the order of the item in the sequence (multiple regression, additional $R^2 = .065, p < .01$). At the beginning, subjects took much more time to answer a question than later on. In spite of a training session of twelve questions, we still found a strong learning effect. The second item characteristic that yields predictive power is, interestingly, the choice itself, i.e., whether the choice is risky or safe (additional $R^2 = .0065$, p $< .01$). Given our working hypothesis, we might expect a curved relationship with its peak at the .5 risk percentage. We found, however, that, on the average, subjects take more time to give a risky than a safe choice.[6] This may be explained by a justification process: it is easier to justify a safe then a risky choice.

4. Discussion

The test that we employed provides a critical test of rank-dependent utility. The experiment suggests that the general RDU model does not provide descriptive improvement over EU. We found no evidence that subjects are sensitive to comonotonicity for decision under ambiguity. The same was found for decision under risk by Wakker, Erev, and Weber (1994).

Even if the general RDU model does not provide a descriptive improvement, more restrictive submodels may still be found that are of descriptive use. In that case, there will be relatively few violations of the non-comonotonic STP that are excluded by such a submodel (and relatively many that are permitted). In other words, the loss in accuracy is then outweighed by the gain in parsimony. Previous studies have suggested that RDU with an S-shaped weighting function can be of interest (Tversky and Kahneman, 1992; Wu and Gonzalez, 1994; Wakker, Erev, and Weber, 1994), and that the S-shape will be more pronounced for ambiguity than for risk (Tversky and Fox, 1995). Our experiment, however, does not confirm these results. It seems that the data of this experiment could best be described as produced by a noisy EU-maximizer (Hey and Orme, 1994).

Some dependencies on presentation were apparent in our study. The different presentation modes that we used (in particular, the verbal-collapsed presentation) affected the rate of violations of the sure-thing principle to a significant degree. Not very surprisingly, consistency, and therefore violation rates of the STP, are affected by the complexity of presentation. The effect size that we found with seemingly trivial variation in presentation indicates a dependence of choice on presentation that is worrisome. Also, we found a

significant collapsing effect. That is, by presenting events with the same outcomes in a collapsed manner, the gamble becomes less attractive. This effect is not related to likelihood or outcomes, but to presentation only, and is therefore difficult to accommodate in the present EU-like theories.

By incorporating additional stages in decision theory that depend on the presentation of the gambles and precede an EU- or RDU-like evaluation, we might be able to give a better description of preferences. First, a subadditive evaluation of collapsed events, decreasing the decision weight of such an event, can explain the collapse effect. This process is very similar to the subadditive evaluation of "packed" hypotheses formalized in support theory (Tversky and Koehler, 1994). Second, we might introduce an editing operation where transparently coded common outcomes are cancelled (Kahneman and Tversky, 1979). (Wu (1994) finds positive evidence for such an operation, while Li (1994) reports negative findings). This cancellation operation sheds a different light on our negative findings. If subjects always cancel common outcomes, RDU "cum cancellation" is untested by the current design. Additional stages in the theory, however, come at a price. They yield more descriptive qualities but weaken the elegance and the predictive power of the theory.

The effects of variations in presentation modes on the performance of rank-dependent utility is a topic for further research. Weber and Kirsner (1995) addressed the question to what degree rank dependence can be enhanced by presentation. By highlighting the highest or the lowest outcome of the gambles, they enhanced optimistic and pessimistic choice making, respectively. Such modes of evaluation induce violations of the non-comonotonic STP. They also considered buying and selling tasks, where subjects can face losses so that loss-aversion effects can occur. This increases pessimism and therefore also induces rank dependency. Weber and Kirsner showed that these representational manipulations can generate significant rank dependence.

Overall, our findings are negative. Taking EU and RDU as they have been axiomatized, we do not find evidence for rank dependency. We do think that RDU can be of descriptive value in specific domains of decision making. In our opinion, RDU yields an elegant manner for modeling preferences if the certainty or possibility effect applies. Also, multi-outcomes gambles can easily be too complicated to be evaluated in an EU-like style. It seems natural that in such cases people need to group outcomes to evaluate these gambles more easily (Lopes, 1987). A very likely grouping in best, worst, and intermediate outcome groups is essentially rank-dependent, providing high hopes for the descriptive success of RDU. The identification of special domains of application of RDU is a topic for future research.

Notes

1. This interpretation is based in a two-stage approach to ambiguity. In the first stage, probabilities are estimated; in the second stage, they are transformed into decision weights. Both stages can generate non-additivity (two-factor model, Tversky and Fox, 1994, Tversky and Wakker, 1995).
2. Combined with usual continuity assumptions, the STP is equivalent to independence for the context of risk (Fishburn and Wakker, 1995).
3. Cancel common terms $\pi_3 U(8)$ and $\pi_3 U(12)$, respectively, in the RDU formula.
4. Mean response times were 16.04 sec for die I, 19.68 sec for die II, and 21.21 sec for die III. The null hypothesis stating that mean times are equal could be rejected ($F(2,58) = 35.73, p < .001$)).
5. There is a risk of contamination here, due to the influence of base rates. Safe choices were more frequent, so there will be more consistent pairs of choices that are both safe. Mean response times for safe choices were also less than for risky choices. This implies that the response times of consistent pairs of choices will be less than for inconsistent pairs. However, the large base rate of safe choices cannot explain the magnitude of the differences reported. Based on the mean times and base rates for safe and risky choices, a mean time for consistent pairs of 18.88 sec is predicted, and a mean time of 19.33 sec for inconsistent pairs.
6. Because the mean time of subjects was already entered as a predictor, the predictive value of the answer does not reflect differences between subjects but purely differences within subjects.

References

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

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.

Fishburn, P.C. and P.P. Wakker. (1995). "The Invention of the Independence Condition," *Management Science* 41, 1130–1144.

Hey, J.D., and C. Orme. (1994). "Investigating Parsimonious Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica* 62, 1291–1326.

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

Li, S. (1994). "What is the Role of Transparency in Cancellation?" *Organizational Behavior and Human Decision Processes* 60, 353–366.

Lopes, L.L. (1987). "Between Hope and Fear: The Psychology of Risk," *Advances in Experimental Psychology* 20, 255–295.

Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley; 2nd ed. (1972), New York: Dover.

Schmeidler, D. (1989) "Subjective Probability and Expected Utility without Additivity," *Econometrica* 57, 571–587.

Starmer, C. and R. Sugden (1989). "Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect," *Journal of Risk and Uncertainty* 2, 140–151.

Starmer, C. and R. Sugden. (1991). "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review* 81, 971–978.

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

Tversky, A. and D. Kahneman. (1986). "Rational Choice and the Framing of Decision," *Journal of Business* 59, pt. 2, 251–278.

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

Tversky, A. and D. Koehler. (1994). "Support Theory: A Nonextensional Representation of Subjective Prob-
ability," *Psychological Review* 101, 547–567.

Tversky, A. and P.P. Wakker. (1995). "Risk Attitudes and Decision Weights," *Econometrica* 63, 1255–1280.

Wakker, P.P. (1996). "The Sure-Thing Principle and the Comonotonic Sure-Thing Principle: An Axiomatic
Analysis," *Journal of Mathematical Economics* 25, 213–227.

Wakker, P.P., I. Erev, and E. Weber. (1994). "Comonotonic Independence: The Critical Test between Classical and
Rank-Dependent Utility Theories," *Journal of Risk and Uncertainty* 9, 195–230.

Weber, E.U. (1994). "From Subjective Probabilities to Decision Weights: The Effects of Asymmetric Loss
Functions on the Evaluation of Uncertain Outcomes and Events," *Psychological Bulletin* 115, 228–242.

Weber, E.U. and B. Kirsner. (1995). *Reasons for Rank-Dependent Utility Evaluation.* New York: Center for
Decision Research.

Wu, G. and R. Gonzalez. (1994). "Curvature of the Probability Weighting Function," *Management Science*,
forthcoming.

Wu, G. (1994). "An Empirical test of Ordinal Independence," *Journal of Risk and Uncertainty* 9, 39–60.