From Uniform Expected Utility to Uniform Rank-Dependent Utility: An Experimental Study

A. Vrijdags, T. Marchant

Ghent University, Dunantlaan 1, 9000 Ghent, Belgium

Abstract

This paper presents an experimental investigation of the Uniform Expected Utility (UEU) criterion, a model for ranking sets of uncertain outcomes. We verified whether the two behavioral axioms characterizing UEU, i.e., Averaging and Restricted Independence, are satisfied in a pairwise choice experiment with monetary gains. Our results show that neither of these axioms holds in general. Averaging in particular, appears to be violated on a large scale. On the basis of the current study and a previous one, we can conclude that none of the models for set ranking that have been axiomatically characterized so far is able to model observed choices between sets of possible outcomes in a satisfactory fashion. In this paper we therefore lay out the foundations for a new descriptive model for set ranking: the Uniform Rank-Dependent Utility (URDU) criterion.

Keywords: Complete uncertainty, uniform expected utility, set ranking

1. Introduction

A far-famed approach to the modelling of uncertainty is the Bayesian one, which claims that, in the absence of objective probabilities, the decision maker should have her own subjective probabilities and these probabilities should guide her decisions. Another approach, not using probabilities, can be found in the literature about set rankings, surveyed by Barberà et al. [4]. In this domain, decisions are quite frugally described by nothing more than the sets of their possible outcomes. Comparing decisions hence reduces to comparing sets of possible outcomes.

The *Min and Max Induced Rankings* (MMIR) form a family of set rankings (Maximin, Maximax, Minmax, Maxmin, etc.) that require preferences over sets to be induced from comparison of the best and/or worst elements within those sets. The Minmax and Maxmin criteria [3, 9], for example, treat the best and worst elements in a lexicographical fashion. According to Minmax, comparison

of the minima will be the primary criterion for ranking sets. In the case where the minima coincide, Minmax prescribes that the decision maker will proceed to comparing the maxima. An indifference will be stated if the minima as well as the maxima of both sets are identical. The Maxmin rule is the dual case in which the decision maker first considers the maxima in the sets to be compared, and when these are identical, she will go on to comparing the minima.

The Uniform Expected Utility (UEU) criterion, axiomatized by Gravel et al. [17], is another type of model for ranking sets of possible outcomes. The UEU criterion shows some similarities to the classical Expected Utility (EU) criterion in that it states that sets are ranked on the basis of the expected utility of their outcomes. In the absence of information about probabilities, however, it is assumed that the decision maker acts as if she considers all the possible outcomes of a decision as equally likely.

Most of the research in the field is pursued by theorists who are mainly concerned with the axiomatic characterizations of models for set ranking. From our point of view, however, it also seems interesting to investigate whether any of these models are capable of describing observed decision behavior.

So far, we know of only two studies that have adopted a descriptive approach in order to study set rankings: Vrijdags [31] investigated whether rankings of sets of monetary consequences obey the transitivity axiom, and in a second paper, the MMIRs were examined empirically [32].

In the current study, UEU will be the model under scrutiny. In specifically designed tests, reported in Vrijdags [32], UEU appears to outperform the MMIRs in predicting the subjects' preferences. Yet, some observations were made that are hard to accommodate within the UEU framework. In Vrijdags [32], subjects are asked to choose between sets of monetary consequences. When asked to choose between $A_1 = \{35, 4, 3\}$[1] and $B_1 = \{35, 3\}$, for example, 46% are estimated to prefer A_1. When confronted with the choice between $A_2 = \{20, 3, 2\}$ and $B_2 = \{20, 1\}$, as much as 78% of all participants opted for A_2. A similar choice is the one between $A_3 = \{20, 2, 1\}$ and $B_3 = \{20, 1\}$, where 38% stated a preference for set A_3. For a considerable share of subjects, it thus appears that they prefer one more outcome in the middle, instead of being constrained to a set with one high and one low outcome, even when the value of this middle outcome is very close to the minimum. This decision behavior might be explained by a positive attitude towards a diversification of uncertainty within the range of the minimum and the maximum of a set. Such choices—where a set with a considerably lower average is preferred—are hard to explain with UEU, unless one assumes an extremely risk averse utility function over the outcomes for all subjects choosing the three-elements sets with the lower arithmetic means over the outcomes. Although this seems rather implausible, it deserves empirical analysis. That is why we devote a large part of this paper to the empirical

---

[1]Set A_1 can be thought of as an occasion to win either €35, €4, or €3 with unknown probabilities. A more detailed account of how the subjects are instructed to conceive of the sets they are presented with can be found in the Method section.

validation of the behavioural axioms, among those characterizing UEU.

The next section presents the model and the axioms that will be tested in the rest of the paper. Sections 3 and 4 will then present the empirical method and the results. In Section 6, we will discuss these results and propose a new promising model, close in spirit to the Rank-Dependent Utility model.

2. The uniform expected utility model

Let X be a non-empty universal set of outcomes, and let $\mathcal{X}$ denote the set of all non-empty, finite subsets of X. We assume the subjects have preferences over $\mathcal{X}$ that can be represented by a weak order (transitive[2] and complete binary relation) $\succsim$ over $\mathcal{X}$. The asymmetric part (strict preference) of $\succsim$ is denoted by $\succ$ while the symmetric part (indifference) is denoted by $\sim$.

We say that $\succsim$ is representable in the UEU model if and only if there exists a real-valued mapping u defined on X such that, for all $A, B \in \mathcal{X}$,

$$A \succsim B \iff \mathrm{UEU}(A) \geq \mathrm{UEU}(B),$$

where

$$\mathrm{UEU}(A) = \frac{1}{\#A} \sum_{a \in A} u(a).$$

Provided the relation $\succsim$ satisfies a richness condition, UEU has been characterized [17] by means of two behavioral axioms (Averaging and Restricted Independence) and a technical condition (Archimedeanness). Sections 3, 4 and 5 focus on these two conditions.

Averaging: for all disjoint sets A and $B \in \mathcal{X}$,

$$A \succsim B \Leftrightarrow A \succsim A \cup B \Leftrightarrow A \cup B \succsim B. \tag{1}$$

The Averaging axiom, first used by [14], ensures that enlarging a set A with a (disjoint) set of outcomes B that is not considered better than A is a worsening of the original set A. On the other hand, the axiom implies that enlarging B with a set A which is considered at least as attractive as B, constitutes an improvement of the original set B. The Averaging axiom is intended to capture an intuitive property satisfied by calculations of "average" in various settings.

Restricted Independence: $\forall A, B, C \in \mathcal{X}$ with $\#A = \#B$ and $A \cap C = B \cap C = \emptyset$,

$$A \succsim B \Leftrightarrow A \cup C \succsim B \cup C. \tag{2}$$

The Restricted Independence axiom is a consistency condition which requires that the ranking of sets with equal cardinality is independent of any elements

---

[2]We know from Vrijdags [31] that transitivity is a reasonable hypothesis in this context.

they may have in common. Hence, adding these common elements to or withdrawing them from the sets should not affect their ranking. A similar condition has been used in [25], although the latter is weaker because it constrains sets A and B to be singletons. Notice also the existence of a condition with a similar appearance in the literature on qualitative probability [e.g., 23, p.204]; there, A and B are events rather than sets of outcomes, and they do not necessarily have the same cardinality; it is used to derive an additive representation of a probability.

3. Method

3.1. Choice stimuli

Our goal is to determine to what extent UEU applies through an investigation of its characterizing behavioral axioms: Averaging and Restricted Independence.

If the subjects' choices do not obey Averaging and/or Restricted Independence, we know that they do not decide according to the UEU model. When devising experiments for axiom tests, one usually tries to "challenge" the axiom under consideration by selecting choice objects or stimuli that are expected to be capable of refuting an axiom if it does not hold in general. Previous research has led us to believe that the choices of people who do not appear to follow UEU might be guided by an inclination towards a diversification of the possible outcomes within the range bounded by the minimum and the maximum of the set. Consequently, where possible, we used this assumption when devising the choice stimuli for this study.

Table 1 shows the pairs of sets that were used to investigate the descriptive validity of Averaging and Restricted Independence. These pairs of sets are administered to the subjects in a forced choice experiment. The instructions of the experiment explain that the numerical set elements represent monetary amounts in €. Each set can be conceived of as a lottery in the form of a container holding one hundred tickets. On each of these tickets, one of the amounts in the set is printed. However, the frequency distribution of the different tickets in the container is unknown. For example, a set $\{30, 23\}$ can be thought of as a container holding an unknown number of tickets with "30" printed on them as well as an unknown number of tickets with "23" printed on them, both of which sum to one hundred. In order to play the lottery, one ticket would be drawn at random from the container, and the amount on it would be the prize to be won in €. Hence, in the current experiment, choosing a set comes down to choosing the lottery one would rather play.

In order to test Averaging, three choices need to be made; one between the original sets A and B, and two between each of the original sets and the union of both, A versus $A \cup B$, and $A \cup B$ versus B. Table 1 shows the four tests that were constructed for Averaging, each consisting of three choices. In all four of them, A_1 and A_2 are instances of A in equation 1, B_1 and B_3 were chosen as instances of B in equation 1, and, consequently, B_2 and A_3 represent

4

Test	Choice	Sets	
	No.	First (A)	Second(B)
Averaging 1	4	$A_1 = \{30, 23\}$	$B_1 = \{27, 25\}$
	5	$A_2 = \{30, 23\}$	$B_2 = \{30, 27, 25, 23\}$
	6	$A_3 = \{30, 27, 25, 23\}$	$B_3 = \{27, 25\}$
Averaging 2	7	$A_1 = \{70, 60, 20\}$	$B_1 = \{55, 50, 45\}$
	8	$A_2 = \{70, 60, 20\}$	$B_2 = \{70, 60, 55, 50, 45, 20\}$
	9	$A_3 = \{70, 60, 55, 50, 45, 20\}$	$B_3 = \{55, 50, 45\}$
Averaging 3	10	$A_1 = \{9, 7, 5\}$	$B_1 = \{8, 6\}$
	11	$A_2 = \{9, 7, 5\}$	$B_2 = \{9, 8, 7, 6, 5\}$
	12	$A_3 = \{9, 8, 7, 6, 5\}$	$B_3 = \{8, 6\}$
Averaging 4	13	$A_1 = \{94, 25\}$	$B_1 = \{45, 43, 32\}$
	14	$A_2 = \{94, 25\}$	$B_2 = \{94, 45, 43, 32, 25\}$
	15	$A_3 = \{94, 45, 43, 32, 25\}$	$B_3 = \{45, 43, 32\}$
Restricted Independence 1	19	$A_1 = \{40, 1\}$	$B_1 = \{11, 10\}$
	20	$A_2 = \{40, 37, 25, 22, 1\}$	$B_2 = \{37, 25, 22, 11, 10\}$
Restricted Independence 2	21	$A_1 = \{100, 9\}$	$B_1 = \{29, 27\}$
	22	$A_2 = \{100, 94, 93, 92, 9\}$	$B_2 = \{94, 93, 92, 29, 27\}$
Restricted Independence 3	23	$A_1 = \{65, 5\}$	$B_1 = \{13, 11\}$
	24	$A_2 = \{65, 64, 63, 62, 5\}$	$B_2 = \{64, 63, 62, 13, 11\}$
Restricted Independence 4	25	$A_1 = \{24, 1\}$	$B_1 = \{8, 7\}$
	26	$A_2 = \{24, 15, 13, 1\}$	$B_2 = \{15, 13, 8, 7\}$

Table 1: Choice stimuli used for the examination of the axioms used in the characterization of UEU: Averaging and Restricted Independence.

the union $A \cup B$. Only two of the eight possible response patterns for the sequence of three choices are in line with Averaging: the pattern where the first set is chosen in all three choices of a test, i.e. $A_1 A_2 A_3$, and the pattern where the second set is chosen in all three choices, i.e. $B_1 B_2 B_3$. With the aforementioned inclination towards a diversification of uncertainty in mind, we tried to challenge Averaging by constructing the stimulus pairs in such a way that people would prefer the union of both sets to each of the originals. In that case they will demonstrate patterns $A_1 B_2 A_3$ or $B_1 B_2 A_3$ which constitute violations of Averaging. For the test Averaging 1, we used two binary sets. In this test, $A_1 = \{30, 23\}$ contains the more extreme outcomes, while the two outcomes of $B_1 = \{27, 25\}$ are situated in the middle between those of A_1. Based on the assumed inclination towards diversification, we might expect people to prefer the union of both sets $B_2 = \{30, 27, 25, 23\}$ to $A_2 = \{30, 23\}$. If those people also prefer $A_3 = \{30, 27, 25, 23\}$ to $B_3 = \{27, 25\}$, they are violating Averaging. In the second test, Averaging 2, the sets to be united both have three outcomes. Considering the assumed inclination towards diversification, we anticipate that most people will prefer $B_2 = \{70, 60, 55, 50, 45, 20\}$ to $A_2 = \{70, 60, 20\}$. Maybe, some people will also prefer $A_3 = \{70, 60, 55, 50, 45, 20\}$ to $B_3 = \{55, 50, 45\}$ since A_3 has two extra higher outcomes as opposed to only one extra lower outcome. In the third test, the dispersion of the outcomes is minimal. The last test, Averaging 4, consists of enlarging a rather risky binary set, $A_1 = \{95, 25\}$ with three middle outcomes, $B_1 = \{45, 43, 32\}$.

For the empirical examination of Restricted Independence we need the subjects to make two choices, one between the original sets A and B, and one

between the extended sets $A \cup C$ and $B \cup C$. Again, four tests were constructed, shown in Table 1, each consisting of two choices. In order to comply with Restricted Independence, people should demonstrate pattern $A_1 A_2$ or pattern $B_1 B_2$. In each test, we chose A_1 and B_1 as instances of A and B in equation 2. In the choice between A_2 and B_2, the original sets are enlarged with the same common elements, as prescribed by Restricted Independence. For this axiom it is not possible to use our assumptions about an inclination towards larger sets, since the sets to be compared have, by definition, the same number of outcomes. Therefore, the tests were constructed following a different recipe. In all four tests, the first pair of sets to be compared consisted of one risky binary set A_1 with an extremely high and an extremely low outcome and a considerably higher average over the outcomes than the second, safer set B_1 comprising two relatively low outcomes. In the second choice of each test, three (in the first three tests) or two (in the last test) higher outcomes are added to both original sets. These added outcomes are lower than the maximum of the risky set, but higher than both outcomes in the safe set. Adding these high outcomes might render the highest outcome of the risky set less salient, since it is now flanked by a number of other very high outcomes. The enlarged safe set now also contains a number of very high outcomes, while it does not have the extremely low outcome which is present in the enlarged risky set. This might lead some subjects to switch their preference from the risky set in the first choice to the enlarged safe set in the second choice, which would constitute a violation of Restricted Independence.

3.2. Statistical Model

For the statistical analysis, we will use the "true and error" (TE) model, proposed by Birnbaum and Bahra [7][3]. It assumes that each subject (1) has a true preference relation (that can vary across individuals) and (2) makes random errors in responding. These hypotheses are similar to those made by, a.o., Harless and Camerer [18], Hey and Orme [20]. A particularity of Birnbaum's model is that it does not assume that the error probability is the same for each choice made by the subject. That is why every choice needs to be presented several times in order to unambiguously estimate the error rate for each distinct choice.

This TE model will permit us to estimate, for each test of an axiom, the proportion of subjects with a particular choice pattern. For instance, for the first test of Averaging (see Table 1), we will estimate the proportion of subjects with pattern $A_1 A_2 A_3$. Adding this proportion to the proportion of subjects with pattern $B_1 B_2 B_3$ will yield the proportion of subjects whose choices are compatible with Averaging.

Notice that we will not test the hypothesis that an axiom (say Averaging) holds, because we consider that subjects can be divided into several categories,

---

[3]Birnbaum actually proposed this TE model earlier, in [5], but the way it is presented in [7] will make it easier for the reader to follow our analysis.

according to the way they behave. It is therefore perfectly possible that some subject satisfy Averaging and some others not. That is why we want to estimate, for every axiom, the probability of subjects satisfying it.

3.3. Data

The data for this paper were gathered by means of a forced choice experiment presented as an Internet questionnaire. Participants were recruited with an e-mail, inviting them to complete a questionnaire regarding decision behavior. The questionnaire comprised a total of 66 choices between sets. The participants were asked to state their preference by clicking a "radio button" besides the preferred set. There were 33 different pairs of sets (the experimental choices in Table 1 and some warm-up and filler choices). Each of the pairs was presented twice, since the proportion of preference reversals between replications of the same choice is needed to estimate the error rate for that choice.

In the first series, the pairs were presented a first time. Consequently, the whole series was repeated. Between the first and the second presentation, the order of the sets within each pair was reversed in order to counterbalance any potential effect of this order. Within each series, the order of presentation of the choices was randomized. It was repeatedly stressed in the instructions that the proportions of the different numbers in the containers, and thus the probabilities of winning the respective monetary prizes, were unknown. Participants were also informed that ten of them would be selected at random to play one of their chosen lotteries for real money.

A total of 193 people participated in the experiment. Most of the participants were students in the Faculty of Psychology and Educational Sciences of Ghent University, 75% were female and 81% were between 18 and 24 years of age.

4. Results

4.1. Averaging

The frequencies of each response pattern for the complete sequence of three choices in the test Averaging 1 are tabulated in Table 2. Since the three choices were presented twice, there are $2^6 = 64$ possible response patterns in total. Most of these 64 patterns have zero frequencies. Therefore, the data are pooled into 16 cells as in [7]. For each of the eight preference patterns in Table 2, the number of times it was shown on both replicates was counted as well as the number of times it was shown on either the first or the second replicate, but not both, divided by two. For example, in the first row of Table 2 it is shown that 24 out of 193 participants demonstrated pattern $A_1 A_2 A_3$ (i.e., $A_1 \succ B_1, A_2 \succ B_2$, and $A_3 \succ B_3$) on the first replicate, and 20 did so on the second replicate. Out of these, nine people showed pattern $A_1 A_2 A_3$ on both replicates. The number of times $A_1 A_2 A_3$ was shown on either the first or the second replication, but *not* both is then given by $(24 - 9) + (20 - 9) = 26$. In order to avoid that responses with differing patterns in the first and second replicate are counted

twice (once for the pattern in the first replicate and once for the pattern in the second replicate), this number is divided by two, which yields 13. Grouping the 64 frequencies of the complete patterns over the two replicates in the way described above yields the 16 mutually exclusive frequencies in the fourth and fifth column of Table 2.

Response	Observed frequencies				Estimated true
pattern	Rep 1	Rep 2	Both	One not both	probability
$A_1A_2A_3$	24	20	9	13	**0.110**
$A_1A_2B_3$	7	11	2	7	0.039
$A_1B_2A_3$	75	62	37	31.5	0.508
$A_1B_2B_3$	9	20	2	12.5	0.002
$B_1A_2A_3$	9	6	1	6.5	0.000
$B_1A_2B_3$	6	12	0	9	0.022
$B_1B_2A_3$	39	49	19	25	0.221
$B_1B_2B_3$	24	13	6	12.5	**0.099**
Total	193	193	76	117	1.000

Table 2: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns in the first test of Averaging, comprised by the sequence of choices 4, 5 and 6. Entries in bold are the probabilities of the true patterns that comply with Averaging. Estimated error rates are 0.238, 0.135, and 0.156 for Choices 4, 5, and 6, respectively. Evaluation of the TE model yields $\chi^2(5) = 7.76$, $p = 0.170$.

The TE model was fit to the 16 frequencies in the "Both" and "One not both" columns in Table 2, as in [7]. From these 16 frequencies that have 15 degrees of freedom (they sum to the total number of participants), there are three error terms and eight true probabilities to be estimated. Since the probabilities of the eight possible patterns sum to one, this leaves $15 - 3 - 7 = 5$ degrees of freedom to test the fit of the TE model. The column labelled "Estimated true probability" in Table 2 shows the estimated true probabilities of each pattern. For the analysis of the first test of Averaging, Averaging 1, in Table 2, $\chi^2(5)$ equals 7.76 ($p = 0.17$), which is not significant (with $\alpha = 0.05$), suggesting that the general TE model can be retained. With this model, merely 20.9% ($\hat{p}(A_1A_2A_3) + \hat{p}(B_1B_2B_3)$) of all participants are estimated to have a true pattern that complies with Averaging. As much as 72.9% ($\hat{p}(A_1B_2A_3) + \hat{p}(B_1B_2A_3)$) is violating Averaging by choosing the union in choices 4 and 5. However, 6.1% ($\hat{p}(A_1A_2B_3) + \hat{p}(B_1A_2B_3)$) of the violations is produced by people who have a true preference for the smaller set in choices 4 and 5.

Table 3 shows the analysis of the response patterns for Averaging 2, which reveals that more than half of the participants, i.e. 54.1%, have a true pattern that complies with Averaging. As anticipated when constructing these stimuli, the largest portion of violations is again accounted for by patterns $A_1B_2A_3$ and $B_1B_2A_3$, where the union is preferred to the original sets in choices 8 and 9.

As shown in Table 4, the third test of Averaging produced the highest number of violations. Of all participants, 55.6% violated averaging by opting for the largest set in all three choices. Surprisingly, 18.1% truly preferred the smallest

| Response | Observed frequencies | | | | Estimated true |
pattern	Rep 1	Rep 2	Both	One not both	probability
$A_1A_2A_3$	12	13	4	8.5	**0.049**
$A_1A_2B_3$	1	7	0	4	0.000
$A_1B_2A_3$	37	27	18	14	0.213
$A_1B_2B_3$	5	10	1	6.5	0.000
$B_1A_2A_3$	8	9	1	7.5	0.004
$B_1A_2B_3$	20	14	6	11	0.060
$B_1B_2A_3$	43	43	14	29	0.182
$B_1B_2B_3$	67	70	38	30.5	**0.492**
Total	193	193	82	111	1.000

Table 3: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns in the second test of Averaging, choices 7, 8, and 9. Estimated error rates are 0.124, 0.138, and 0.187 for choices 7, 8, and 9, respectively. Evaluation of the TE model yields $\chi^2(5) = 6.94$, $p = 0.225$, an acceptable fit. Entries in bold are the probabilities of the true patterns that comply with Averaging.

| Response | Observed frequencies | | | | Estimated true |
pattern	Rep 1	Rep 2	Both	One not both	probability
$A_1A_2A_3$	14	17	5	10.5	**0.048**
$A_1A_2B_3$	18	17	3	14.5	0.053
$A_1B_2A_3$	62	60	33	28	0.556
$A_1B_2B_3$	23	19	3	18	0.005
$B_1A_2A_3$	5	11	0	8	0.000
$B_1A_2B_3$	26	24	8	17	0.181
$B_1B_2A_3$	19	26	4	18.5	0.021
$B_1B_2B_3$	26	19	7	15.5	**0.137**
Total	193	193	63	130	1.000

Table 4: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns in the third test of Averaging with choices 10, 11, and 12. Estimated error rates are 0.210, 0.177, and 0.214 for choices 10, 11, and 12 respectively. Evaluation of the TE model yields $\chi^2(5) = 6.14$, $p = 0.293$, an acceptable fit. Entries in bold are the probabilities of the true patterns that comply with Averaging.

set in all three choices of this test.

Table 5 presents the the results of the fourth test. For this test as well, the bulk of the violations is generated by subjects who opted for the union instead of the original sets in choices 14 and 15. The responses of only 22.1% of all subjects are estimated to comply with Averaging.

4.2. Restricted Independence

With respect to the choices in Table 1, people should demonstrate either pattern A_1A_2 (i.e., A is preferred over B in the first choice of the test as well as in the second choice), or pattern B_1B_2 (if B is preferred over A in both

| Response | Observed frequencies | | | | Estimated true |
pattern	Rep 1	Rep 2	Both	One not both	probability
$A_1A_2A_3$	30	30	12	18	**0.116**
$A_1A_2B_3$	4	3	1	2.5	0.007
$A_1B_2A_3$	93	84	54	34.5	0.661
$A_1B_2B_3$	9	8	1	7.5	0.000
$B_1A_2A_3$	6	2	0	4	0.000
$B_1A_2B_3$	2	1	0	1.5	0.000
$B_1B_2A_3$	26	49	9	28.5	0.111
$B_1B_2B_3$	23	16	9	10.5	**0.105**
Total	193	193	86	107	1.000

Table 5: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns in the fourth test of Averaging composed of choices 13, 14, and 15. Estimated error rates are 0.153, 0.119, and 0.121 for choices 13, 14, and 15, respectively. Entries in bold are the estimated probabilities of the true patterns that comply with Averaging. Evaluation of the TE model yields an only marginally acceptable fit ($\chi^2(5) = 9.483$, $p = 0.091$), implying that the estimated probabilities need to be considered with the necessary caution.

choices of the test) in order to comply with the requirements of Restricted
270 Independence. For each test of the axiom, there are now four possible patterns
of choice (A_1A_2, A_1B_2, B_1A_2 or B_1B_2) instead of eight. The data and results can
therefore be presented in a single table (Table 6). The estimated probabilities
of someone's true preferences corresponding to a specific pattern are displayed
in columns four to seven of Table 6, i.e., the "Parameter Estimates" columns.
275 The results in Table 6 confirm rather than contravene Restricted Independence.
A large portion of the people who opted for A_1 in the first choice of each test,
also preferred A_2 in the second choice. Of the people who chose B_1 in the first
choice, on the other hand, almost everyone chose B_2 as well. This results in
violation percentages ($\hat{p}(A_1B_2) + \hat{p}(B_1A_2)$) varying from 18.7% (in the fourth
280 test) to 34.7% (in the second test), so we can hardly conclude that this axiom
is systematically violated with our stimuli.

| | | Parameter Estimates | | | | | |
Test	Choice	$\hat{e}$	$\hat{p}(A_1A_2)$	$\hat{p}(A_1B_2)$	$\hat{p}(B_1A_2)$	$\hat{p}(B_1B_2)$	p-value
Restr. Ind. 1	19	0.126	**0.193**	0.318	0.023	**0.466**	0.066
	20	0.093					
Restr. Ind. 2	21	0.095	**0.269**	0.341	0.006	**0.383**	0.241
	22	0.169					
Restr. Ind. 3	23	0.074	**0.613**	0.269	0.000	**0.117**	0.227
	24	0.244					
Restr. Ind. 4	25	0.105	**0.419**	0.147	0.040	**0.395**	0.638
	26	0.169					

Table 6: Estimated true probabilities of each response pattern in the four tests of Restricted Independence. The p-values resulting from the tests of the TE model are shown in the last column. All five of them show acceptable fits ($\alpha = .05$). Entries in bold are the probabilities of the true patterns that comply with Restricted Independence.

Nevertheless, one has to consider the possibility that some of the subjects who comply with Restricted Independence in one of the tests, violate the axiom in one or more of the other tests. In order to truly know the compliance rate for Restricted Independence axiom, we should examine how many people comply with the axiom in all four tests.

Since each test of Restricted Independence consists of two choices, we would need to analyze response patterns for a sequence of eight choices if we want to obtain the percentage of people complying with all four tests. Unfortunately, we do not have enough data to estimate the excessive number of parameters needed to fit such a model (256 possible true patterns of which the probabilities sum to one, plus eight error rate estimates yields 256 - 1 + 8 parameters). With the amount of data that we have, it seems feasible to analyze sequences of up to four choices with the TE model. This way, we can estimate for every possible couple of tests how many people comply with both of them.

Tests	Estimated percentage of people who showed		
	2 violations	2 compliances	1 viol. + 1 compl.
Restricted Independence 1 & 2	0.275	0.555	0.169
Restricted Independence 1 & 3	0.155	0.497	0.347
Restricted Independence 1 & 4	0.115	0.587	0.299
Restricted Independence 2 & 3	0.196	0.509	0.295
Restricted Independence 2 & 4	0.073	0.539	0.388
Restricted Independence 3 & 4	0.051	0.561	0.389

Table 7: Estimated true probabilities of the proportions of violations, compliances, and mixed responses for each couple of tests of Restricted Independence.

The results of these analyses are shown in Table 7. The two by two analysis of the four tests of Restricted independence yields 6 couples of tests. Each line in Table 7 presents the results for one such couple of tests. To estimate the fraction of respondents complying with both tests in a couple, we had to analyze sequences of four choices, resulting in $2^4 = 16$ possible true patterns. In order to present the results in an orderly and concise manner, the estimated proportions were summed for all patterns complying with Restricted Independence in both tests[4]— remember that a compliance consists of either choosing set A or set B in both choices of a test—patterns with two violations[5] and the remaining patterns with one compliance and one violation.

In the first row of table 7, it is shown that 55.5% of all subjects is estimated to truly comply with Restricted Independence in the first two tests. Since for both tests the compliance rate was approximately 65% (see Table 6), this means

---

[4]There are four patterns with two compliances: $AAAA$ if one consistently chooses the first set in both tests, $AABB$ if the first set is chosen twice in the first test and the second set is chosen twice in the second test, $BBAA$ for the reverse pattern, and finally $BBBB$ if one consistently chooses the second set in both tests.

[5]There are four patterns with two violations: $ABAB$, $ABBA$, $BAAB$, and $BABA$

that most people who chose according to Restricted Independence in one test,
also complied with the axiom in the other test. We can see in Table 7 that
for each couple of tests, there is an "overlap" of at least 49.7% of the subjects
complying with the axiom in both tests.

From this, we can conclude that the proportion of subjects complying with
the axiom in all four tests is at most 49.7%. A lower bound for this proportion
is provided by the proportion of subjects complying with the axiom in all four
tests without making any "error", i.e., those complying with Restricted Inde-
pendence in the first as well as in the second presentation of all four tests. This
proportion is obtained by simply counting the number of such subjects (without
any statistical analysis) and dividing this number by the sample size (193). We
obtain 9.8%. Note that this is a very severe lower bound because it neglects all
subjects that made at least one "error".

In summary, the proportion of subjects satisfying Restricted Independence
is estimated between 10 and 50%.

5. Discussion of the experiments

The empirical investigation of the axiomatic characterizations of models for
set ranking in Vrijdags [31, 32] has revealed that choices between sets of possi-
ble monetary outcomes appear to be transitive. Furthermore, peoples' choices
obey axioms that prevent rankings to be based on total-goodness, as well as
monotonicity axioms which ensure that replacing a set element with a better
one results in a better set. Additionally, the strength of preference between
monetary outcomes seems to be taken into account by the subjects, and ax-
ioms that prevent rankings to be based on average-goodness were violated on a
large scale.[6] All of these findings are compatible with UEU and thus constitute
evidence in favour of the criterion. Yet, we also found that under particular
circumstances, people appear to prefer sets that are obviously less attractive in
terms of the arithmetic mean over the outcomes, which can only be explained
by UEU if an extremely risk averse utility function is assumed. This rather
surprising finding was attributed to an inclination towards a diversification of
uncertainty. The latter observations indicate that UEU might not be an all-
encompassing model which is able to explain observed set rankings in different
kinds of situations. In the current study, we sought to bring more clarity to
the situation by testing the axioms characterizing UEU, i.e., Averaging and
Restricted Independence, in a pairwise choice experiment.

In order to construct empirical tests for the axioms, we attempted, where
possible, to devise stimuli in such a way that we believe they are most likely to
yield violations. Averaging was challenged by taking into account the assumed
inclination towards diversification. A strategy which appears to have worked
well since our stimuli gave rise to a rather strong refutation of Averaging, with

---

[6]For a more detailed account of these findings, the interested reader is referred to Vrijdags
[32].

12

violation rates ranging from 46% to 81.5%. As a result we can conclude that most people probably do not perform any averaging operation as required by UEU when validating sets of possible monetary outcomes.

In order to design stimuli for the test of Restricted Independence a different strategy was applied, which yielded violations for a significant part of the subjects. Still, with our stimuli, the choices of roughly 10% to 50% of the sample met the requirements of Restricted Independence in all four tests. Consequently, if one intends to model actual choice behavior for a significant portion of people, it seems expedient to take Restricted Independence into account. However—as always when rather small numbers of violations are found—the question might be asked whether Restricted Independence would continue to hold up in studies with other recipes for devising the stimuli.

6. A new model

In the current and previous studies, we have tested all models for ranking sets of uncertain outcomes that have been axiomatically characterized so far. None of them could be retained as being descriptively valid, since for each model, one or more of the characterizing axioms was violated on a large scale. Our previously obtained results clearly show that the MMIR, which prescribe that attention is restricted to merely the extremes of the set, are not adequate for describing observed choices between sets of monetary consequences. The UEU criterion, which supposes that the decision maker's attention is divided equally over all possible outcomes, has none of the shortcomings of the MMIR. Nevertheless, both its characterizing behavioral axioms were violated by a non-negligible portion of people, although Restricted Independence continues to hold true for at least 10% of all subjects.

Because UEU looked so promising at first, we wondered whether UEU could be extended in some way so that it would perform better in empirical tests. Since Averaging was profusely violated, and Restricted independence much less so, we tried to find a model satisfying some kind of weakened version of Restricted Independence, that would not be refuted by the data. Through examination of the manner in which violations were evoked for Restricted Independence, the idea arose that the ranks of the outcomes in the sets might play a crucial role in establishing the violating patterns. Note that, in each test, the set C to be added to both sets A and B consists of outcomes that are more attractive than all outcomes in B and less attractive than $\max(A)$. These reflections led us to a model for set ranking that resembles Rank-Dependent Utility (RDU) for risky decision making.

6.1. Rank-Dependent Utility

In the literature on risky decision making, a variety of RDU models have been proposed as solutions to different theoretical and empirical problems with

13

EU-models and their extensions. They were introduced by Quiggin [26] who attempted to explain the Allais paradox without producing violations of Stochastic Dominance as implied by other non-EU models such as Edwards' Subjective Expected Value model [13] and Prospect Theory (PT) [21]. The central idea of RDU, i.e., rank-dependent weighting, was then incorporated into PT, and the resulting model is known as Cumulative Prospect Theory (CPT) (Tversky & Kahneman, 1992). The intuition of rank-dependence entails that the attention given to an outcome does not only depend on the probability of the outcome, but also on how attractive the outcome is in comparison to the other possible outcomes of a prospect.

For a lottery $L = (x_1, p_1; x_2, p_2; \ldots x_n, p_n)$, with possible outcomes x_i and respective probabilities p_i, RDU can be formally represented as follows:

$$\text{RDU}(L) = \sum_{i=1}^{n} \pi_i u(x_{(i)}), \tag{3}$$

with the decision weights, π_i, defined by

$$\pi_i = w\left(\sum_{j=1}^{i} p_j\right) - w\left(\sum_{j=1}^{i-1} p_j\right).$$

In this formulation, $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ constitute a reordering of the outcomes such that $x_{(i)}$ is the ith largest outcome in L, and w is the probability weighting function which is strictly increasing and assigns $w(0) = 0$ and $w(1) = 1$.

Several processes may lead people to evaluate utility in a rank-dependent fashion. Tversky and Kahneman [30], for example, suggest perceptual biases as a possible cause for the non-linear weighting of outcomes. According to CPT, overweighting occurs mainly for the most and least favorable outcomes in a lottery. Being positioned at the ends of the distribution, these outcomes are assumed to be perceptually more salient. One can also think of motivational processes to account for rank-dependence. For some reason, the decision maker may think that an unfavorable (favorable) outcome is especially important in decision making and therefore should receive more attention than a favorable (unfavorable) one with the same probability. The reason therefore might be a pessimistic attitude that causes an irrational belief that unfavorable events tend to happen more often, leading to an unrealistic overweighting of unfavorable likelihoods [12].

Several researchers have provided axiomatizations of RDU models [1, 11, 22, 24, 27, 28, 33], and, according to Wakker [34, p. 352], they are said to be the best-performing and most-confirmed empirical models for decision under risk and uncertainty to date. Because of their extensive use, however, numerous empirical violations of these models have been documented as well [e.g., 6, 16, 19, 29, 35].

6.2. Uniform Rank-Dependent Utility

When the notion of RDU is applied to the context of set ranking, it gives rise to an extension of UEU, which we will refer to as Uniform Rank-Dependent Utility (URDU). The URDU criterion can be formally represented by

$$\text{URDU}(A) = \sum_{i=1}^{\#A} w_i^{\#A} u(a_{(i)}), \tag{4}$$

where $a_{(i)}$ is the ith largest element in set A with respect to $\succsim$, and where each set has an associated collection of decision weights, $\mathbf{w}^{\#A} = (w_1^{\#A}, ..., w_{\#A}^{\#A})$, lying in the unit interval and summing to one. As shown in (4), the weights associated with each of the ranks are conditional on the number of elements in the set. Consequently, for one person, sets with an equal cardinality have the same weights assigned to each of the ranks.

In the absence of any probabilities, we assume that the decision maker acts as if all outcomes in a set are equally probable; the decision weights in URDU will therefore solely depend on the ranks of the outcomes. Hence, the attention given to each of the outcomes in a set is not necessarily equally distributed as assumed by UEU, but will depend on how good or attractive the outcome is in comparison to the other possible outcomes in the same set. To exemplify this notion of rank-dependence in the context of set ranking, consider an optimistic decision maker evaluating the set $\{50, 30, 15\}$. As she is assumed to have an optimistic attitude, more than one third of her attention will be devoted to the best outcome in the set, €50. Suppose that the decision weight for this outcome, i.e., w_1^3, equals $1/2$. Consequently, the other two outcomes will receive less attention ($w_2^3 + w_3^3 = 1/2$). Since the decision maker is said to be an optimist, the majority of her remaining attention will be devoted to the middle outcome, which entails that $w_2^3 > 1/4$, for example $w_2^3 = 1/3$. Accordingly, the decision weight pertaining to the lowest outcome, €15 will be relatively small ($w_3^3 = \frac{1}{6}$). Now, replacing the middle outcome, €30, by €55 yields set $\{55, 50, 15\}$ where €50 is no longer the most attractive outcome. Therefore, our optimistic decision maker will pay less attention to it than in the previous set.

Note that URDU can be conceived of as a special case of RDU, that is, the RDU of a lottery where all outcomes have an equal probability of occurrence, with $p_j = 1/\#L$.

Moreover, URDU also bears strong similarities with the Ordered Weighted Averaging (OWA) aggregation operator [36].[7] In fact, the OWA operator is the special case of URDU where $u(x) = x$.

6.3. Plausibility of Uniform Rank-Dependent Utility

To this point, we have gathered empirical data on every single behavioral axiom that has been used in characterizations of models for set ranking. In order

---

[7]For more information about the OWA operator and its potential as a valuation method in decision making under uncertainty, the interested reader is referred to Yager [36, for instance].

for URDU to be qualified as a descriptive model, it should satisfy the five axioms that we were unable to refute in [31] and [32], i.e., Transitivity, Dominance, Simple Top Monotonicity, Simple Bottom Monotonicity, Simple Monotonicity, and potentially a weakened version of Restricted Independence. At the same time, URDU should violate the remaining axioms, namely the ones that did not prove valid.

In [31], it was shown that most people's choices between sets of possible monetary outcomes obey Transitivity: $\forall A, B, C \in \mathcal{X}, A \succsim B$ and $B \succsim C \Rightarrow A \succsim B$. It is evident that URDU also satisfies Transitivity.

Since Vrijdags [32] failed to elicit violations for Dominance, we consider that URDU should also satisfy Dominance in order to be descriptive.

Dominance: For all $A \in \mathcal{X}$, for all $x, y \in X$,

$$[\{x\} \succ \{y\} \text{ for all } y \in A] \quad \Rightarrow \quad A \cup \{x\} \succ A, \tag{5}$$

$$[\{y\} \succ \{x\} \text{ for all } y \in A] \quad \Rightarrow \quad A \succ A \cup \{x\}. \tag{6}$$

If one allows the weights to be chosen freely, URDU does not satisfy Dominance. Suppose, for example, $A = \{10, 1\}$, and $\{x\} = 11$. If the decision weights are the following: $\mathbf{w}^2 = (0.8, 0.2)$ and $\mathbf{w}^3 = (0.1, 0.2, 0.7)$, calculating URDU will result in a violation of Dominance. In this case, $\text{URDU}(A) = 0.8 \times 10 + 0.2 \times 1 = 8.2$ will be larger than $\text{URDU}(A \cup \{x\}) = 0.1 \times 11 + 0.2 \times 10 + 0.7 \times 1 = 3.8$, which contravenes the implications of Dominance. Consequently, some restrictions on the weights are needed. They are provided in the following result.

Proposition 1. *Suppose $X = \mathbb{R}$. A preference relation representable by URDU with u increasing and continuous satisfies Dominance if and only if the weight vectors satisfy*

$$\sum_{i=1}^{j} w_i^n \leq \sum_{i=1}^{j+1} w_i^{n+1}, \quad \forall j \in \{1, \ldots, n-1\} \tag{7}$$

and

$$\sum_{i=j}^{n} w_i^n \leq \sum_{i=j}^{n+1} w_i^{n+1}, \quad \forall j \in \{2, \ldots, n\}. \tag{8}$$

The proof[8] of all propositions can be found in the appendix.

For the comparison of a four element set with a five element set, for example, Restrictions (7) and (8) entail:

$$w_1^4 + w_2^4 + w_3^4 \leq w_1^5 + w_2^5 + w_3^5 + w_4^5,$$
$$w_1^4 + w_2^4 \leq w_1^5 + w_2^5 + w_3^5,$$
$$w_1^4 \leq w_1^5 + w_2^5,$$

---

[8]The authors thank Jean-Luc Marichal for his assistance in formulating and proving these restrictions.

and

$$w_2^4 + w_3^4 + w_4^4 \leq w_2^5 + w_3^5 + w_4^5 + w_5^5,$$
$$w_3^4 + w_4^4 \leq w_3^5 + w_4^5 + w_5^5,$$
$$w_4^4 \leq w_4^5 + w_5^5.$$

These restrictions do not give the impression of being overly far-fetched, and they still allow sufficient variability to accommodate individual and/or situational differences.

In order to be of merit in a descriptive sense, URDU must also satisfy Simple Top Monotonicity and Simple Bottom Monotonicity, since, for these axioms, no violations at all were recorded [32].

Simple Top Monotonicity: For all $x, y, z \in X$,

$$\{x\} \succ \{y\} \succ \{z\} \quad \Rightarrow \quad \{x, z\} \succ \{y, z\}. \tag{9}$$

Simple Bottom Monotonicity: For all $x, y, z \in X$,

$$\{x\} \succ \{y\} \succ \{z\} \quad \Rightarrow \quad \{x, y\} \succ \{x, z\}.$$

We easily obtain:

Proposition 2. *URDU satisfies Simple Top Monotonicity and Simple Bottom Monotonicity.*

Finally, there is one more axiom that could not be refuted in Vrijdags [32]: Simple Monotonicity. Consequently, this axioms also needs to be satisfied by URDU.

Simple Monotonicity: For all $x, y \in X$,

$$\{x\} \succ \{y\} \quad \Rightarrow \quad \{x\} \succ \{x, y\} \succ \{y\}.$$

The next result shows that URDU satisfies Simple Monotonicity provided a hardly restrictive condition is met.

Proposition 3. *URDU satisfies Simple Monotonicity iff* $0 < w_1^2 < 1$.

URDU does not only satisfy the axioms that appeared to hold empirically, but with certain values for the decision weights this model is also capable of violating all axioms for which substantial numbers of violations were found.

In the remainder of this paper, we will use the following weighting scheme:

$$w_i^{\#A} = k w_{i+1}^{\#A}, \qquad \text{with } k > 0 \text{ and } \sum_{i=1}^{\#A} w_i^{\#A} = 1. \tag{10}$$

One can easily verify that this condition implies

$$w_i^n = \frac{k^{n-i}}{\sum_{i=0}^{n-1} k^i} \tag{11}$$

17

and that these weights satisfy (7) and (8).

Accordingly, $k > 1$ entails that the most favorable outcomes, with the low
indices, will be assigned the largest weights. With $k < 1$, the less favorable outcomes will receive more weight, and $k = 1$ indicates that there is no weighting; all $w_i^n = 1/n$. Although this scheme seems very rigid, it is more tractable than (7) and (8) and it effectively serves our purpose of proving that URDU does not satisfy the axioms that were violated on a large scale. Moreover, we will show that URDU with restriction (10) is capable of predicting the most frequently occurring choice patterns recorded in Vrijdags [31, 32] and the current paper.

Consider, for instance, the data for the fourth test of Averaging in Table 5. According to this test, 66% of all subjects truly violate averaging by preferring $A = \{94, 25\}$ over $B = \{45, 43, 32\}$, and also choosing $A \cup B = \{94, 45, 43, 32, 25\}$ over both A and B. The UEU criterion cannot account for this choice pattern. However, if we assume that the subjects assign larger weights to the most attractive outcomes ($k = 3$), and that they use the simplest possible utility function $u(x) = x$, URDU with weight restriction (10) can predict the observed modal choice pattern. Using (11), we find that the weights for set A are $w_1^2 = k/(1 + k) = 3/4$ and $w_2^2 = 1/(1 + k) = 1/4$. Set B contains three elements, so the weights are $w_1^3 = k^2/(1 + k + k^2) = 9/13$, $w_2^3 = k/(1 + k + k^2) = 3/13$, and $w_3^3 = 1/(1 + k + k^2) = 1/13$. For the five elements of $A \cup B$, the weights are $w_1^5 = 81/121$, $w_2^5 = 27/121$, $w_3^5 = 9/121$, $w_4^5 = 3/121$, and $w_5^5 = 1/121$, respectively. URDU(A) is obtained by summing the products of the utilities of the respective outcomes and their weights as follows $w_1^2 u(a_{(1)})) + w_2^2 u(a_{(2)})) = 3/4 \times 94 + 1/4 \times 25 = 76.75$. In the same way one can calculate URDU(B) = 43.54, and URDU($A \cup B$) = 77.17. Hence, decision makers who rank sets according to URDU with $k = 3$ and $u(x) = x$ should show the most frequently observed pattern for this choice sequence, i.e., $A \cup B \succ A \succ B$, a violation of Averaging. Obviously, other combinations of weights and utility functions exist that would yield the same preference pattern. This result indicates that the observed preference for larger sets in the tests of Averaging, which was attributed by the authors to an inclination towards a diversification of uncertainty, can be explained by URDU with an overweighting of the more favorable outcomes in the sets.

Table 6.3 lists the other axiom tests along with examples of k-values that will bring about violations. For all tests, the simplest possible utility function, $u(x) = x$, is assumed. Where $k = 1$, no weighting is needed to cause a violation. Given that $u(x) = x$, URDU then comes down to simply calculating the arithmetic mean over the outcomes in the sets.

The pairwise choice tasks used in this paper and in Vrijdags [32] can be considered fairly neutral in their situational motivation for differential weighting of outcomes as a function of their rank, as well as in the perceptual emphasis on certain outcomes. Therefore, one can assume that the subjects' weighting schemes will be mainly determined by their personal dispositions which will be supposed to remain constant throughout the course of the experiments. Hence, if URDU really is a descriptively valid choice model, we should be able to find for each subject a unique weighting scheme and utility function with which we can

Axiom	k-value	Axiom	k-value
Neutrality*	1	Restricted Ind. 4	0.6
Bottom Ind.*	1	Simple Uncertainty Appeal*	1
Top Ind.*	1	Simple Uncertainty Aversion*	1
Disjoint Ind.*	1	Monotone Consistency*	1
Robustness*	1	Type 1 Dominance*	1
Averaging 1	1.3	Type 2 Dominance*	1
Averaging 2	1.1	Type 1 Extension Principle*	0.6
Averaging 3	1.1	Type 2 Extension Principle*	4.1
Averaging 4	3	Type 1 Monotonicity*	0.6
Restricted Ind. 1	0.5	Type 2 Monotonicity*	0.9
Restricted Ind. 2	0.28	Extension Independence*	1
Restricted Ind. 3	0.12		

Table 8: List of axioms and corresponding k-values that yield violations according to URDU with weight restriction $w_i^{\#A} = k w_{i+1}^{\#A}$ and $u(x) = x$. For axioms marked with an asterisk, the definition as well as the stimuli with which they were tested can be found in Vrijdags [32].

predict all of the true preferences demonstrated by that subject. It is especially important to show that URDU, with a unique set of weights and utility function, can explain the preferences of individuals who do not choose according to the MMIR, nor according to UEU. Unfortunately, the nature of our dataset does not allow such analyses per person. Subjects participated either in one of both MMIR experiments [32] or in the UEU experiment. Furthermore, experimental choices were presented only twice, which makes it impossible to determine the true preferences for each person separately.

As an approximation of fitting a weighting scheme and utility function at the individual level, we managed to find a combination of weights and a utility function that can predict the grand majority of the modal choice patterns produced by all three experiments in [32] and the current paper. If weight restriction (10) is adopted with $k = 3$, and if one assumes an exponential utility function $u(x) = 1 - e^{-ax}$ with $a = 0.1$, URDU can predict the modal choice pattern for 20 out of the 23 tests in Table 6.3. Only for Averaging 2, Averaging 4 and Restricted Independence 4, it does not predict the modal choice pattern, but even in these cases it still explains the correct pattern for 39%, 11% and 19% of the subjects, respectively. Evidently, URDU also makes correct predictions for the tests of axioms that were not violated.

It goes without saying that these results are extremely encouraging. Remember that we used the same, fairly rigid, weight restriction and utility function for all individuals although they might differ. Rank-dependent weighting allows us to explain why people seem to choose according to UEU in Vrijdags [32] whilst violating Averaging and, to a lesser degree, Restricted Independence the way they do in the current paper. Furthermore, this new model is also capable of explaining the puzzling asymmetries found in the tests of some of the MMIR

axioms.[9] Allowing for individual differences would of course give us even more flexibility for the predictions, but, as mentioned above, analysis at the individual level is not possible with the available data.

In the literature concerning RDU, there is a remarkable degree of convergence between studies regarding the functional form of the weighting function; the key ingredient for accurate predictions appears to be an inverted S-shaped weighting function, where the lowest and (to a lesser degree) the highest outcomes are simultaneously overweighted at the expense of the middle ones [e.g., 2, 8, 10, 15, 30]. In our study, however, we are able to explain the majority of the results with an overweighting of the most favorable outcomes ($k > 1$), which seems atypical if compared with the findings for risky decision making. Of course, one has to keep in mind that restriction (10) constrains the weights to be either ascending or descending, which prevents overweighting of the lowest and the highest outcomes simultaneously. Yet, it is not impossible that with the less stringent restrictions (7) and (8), combined with some utility function we would find a weight set that is able to explain our data with an overweighting of the lowest as well as the highest outcomes. Nevertheless, some caution is required in order to avoid thoughtlessly generalizing results from one area of research to another. It may well be that decision makers employ different weighting schemes for set ranking than for deciding between risky prospects, where the weights depend on the probabilities as well as on the ranks.

7. Conclusion

We have tested the foundations of the UEU model and we have laid out the foundations of URDU, a new descriptive model for set ranking, by proving that it satisfies, with certain restrictions, the axioms we were unable to refute in previous research, and that it is also capable of violating the axioms for which substantial numbers of violations were found. Furthermore, it was shown that with fixed parameter values, this model is capable of predicting the most frequently observed preference patterns that have emerged from experimental research in this domain. Still, a lot of questions remain unanswered and constitute therefore ideal avenues for future research. It would, for example, be particularly advantageous to have a full axiomatization of the URDU model. That way we would have a guideline for testing its specific properties without having to estimate any utility or weighting function.

---

[9]Considering the tests of Type 1 and Type 2 Dominance, Type 1 and Type 2 Extension Principle, and Type 1 and Type 2 Monotonicity, Vrijdags [32] found for each couple of dual axioms a very large violation rate for one of the two counterparts, while for the other counterpart the grand majority of people complied with the axiom. The author concluded that the underlying behavioral properties represented by each couple of dual axioms must not hold and that the lack of violations for one of the two counterparts was most likely due to a rather unfortunate choice of stimuli. Results in the current paper, however, show that these asymmetries follow naturally if one employs a certain weighting scheme.

Acknowledgement

[1] Abdellaoui, M., 2002. A genuine rank-dependent generalization of the von Neumann-Morgenstern expected utility theorem. Econometrica 70, 717–736.

[2] Abdellaoui, M., Bleichrodt, H., L'Haridon, O., 2008. A tractabe method to measure utility and loss aversion under prospect theory. Journal of Risk and Uncertainty 36, 245–266.

[3] Arlegi, R., 2003. A note on Bossert, Pattanaik and Xu's "Choice under complete uncertainty: Axiomatic characterizations of some decision rules". Economic Theory 22, 219 – 225.

[4] Barberà, S., Bossert, W., Pattanaik, P. K., 2004. Ranking sets of objects. In: Barberà, S., Hammond, P., Seidl, C. (Eds.), Handbook of Utility Theory. Volume 2 Extensions. Kluwer Academic Publishers, Dordrecht, pp. 893–977.

[5] Birnbaum, M. H., 2004. Test of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. Organizational Behaviour and Human Decision Processes 95, 40–65.

[6] Birnbaum, M. H., 2008. New paradoxes of risky decision making. Pychological Review 115, 463–501.

[7] Birnbaum, M. H., Bahra, J., 2007. Gain-loss separability and coalescing in risky decision making. Management Science 53, 1016–1028.

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

[9] Bossert, W., Pattanaik, P. K., Xu, Y., 2000. Choice under complete uncertainty: Axiomatic characterization of some decision rules. Economic Theory 16, 295–312.

[10] Bruhin, A., Fehr-Duda, H., Epper, T., 2010. Risk and rationality: Uncovering heterogeneity in probability distortion. Econometrica 78, 1345–1412.

[11] Chateauneuf, A., 1999. Comonotonicity axioms and rank-dependent expected utility theory for arbitrary consequences. Journal of Mathematical Economics 32, 21–45.

[12] Diecidue, E., Wakker, P. P., 2002. On the intuition of rank-dependent utility. Journal of Risk and Uncertainty 23, 281–298.

[13] Edwards, W., 1962. Subjective probabilities inferred from decisions. Psychological Review 69, 109–135.

[14] Fishburn, P. C., 1972. Even-chance lotteries in social choice theory. Theory and Decision 3, 18–40.

[15] Gonzalez, R., Wu, G., 1999. On the shape of the probability weighting function. Cognitive Psychology 38, 129–166.

[16] Gonzalez-Vallejo, C. C., Reid, A. A., Schiltz, J., 2003. Context effects: The proportional difference model and the reflection of preference. Journal of Experimental Psychology: Learning, Memory, and Cognition 29, 942–953.

[17] Gravel, N., Marchant, T., Sen, A., 2012. Uniform expected utility criteria for decision making under ignorance or objective ambiguity. Journal of Mathematical Psychology 56, 297–315.

[18] Harless, D. W., Camerer, C. F., 1994. The predictive utility of generalized expected utility theories. Econometrica 62, 1251–1290.

[19] Hertwig, R., Barron, G., Weber, E. U., Erev, I., 2003. Decisions from experience and the effect of rare events in risky choice. Psychological Science 15, 534–539.

[20] Hey, J. D., Orme, C., 1994. Investigating generalizations of expected utility theory using experimental data. Econometrica 62, 1291–1326.

[21] Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–293.

[22] Köbberling, V., Wakker, P. P., 2003. Preference foundations for nonexpected utility: A generalized and simplified technique. Mathematics of Operations Research 28, 395–423.

[23] Krantz, D., Luce, R. D., Suppes, P., Tversky, A., 1971. Foundations of Measurement, vol. 1. Academic Press, New York.

[24] Luce, R. D., 2000. Utility of gains and losses: Measurement-theoretical and experimental approaches. Lawrence Erlbaum Associates, Mahwah, NJ.

[25] Nehring, K., Puppe, C., 1996. Continous extension of an order on a set to the power set. Journal of Economic Theory 68, 456-479.

[26] Quiggin, J., 1982. A theory of anticipated utility. Journal of Economic Behavior and Organization 3, 323–343.

[27] Quiggin, J., 1993. Generalized expected utility theory: The rank-dependent model. Kluwer Academic, Boston.

[28] Segal, U., 1989. Anticipated utility: A measure representation approach. Annals of Operation Research 19, 359–373.

[29] Starmer, C., 1999. Cycling with rules of thumb: An experimental test for a new form of non-transitive behavior. Theory and Decision 46, 141–158.

[30] Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

[31] Vrijdags, A., 2010. An experimental investigation of transitivity in set ranking. Theory and Decision 68, 213–232.

[32] Vrijdags, A., 2013. Min- and max-induced rankings: an experimental study. Theory and Decision 75, 233–266.

[33] Wakker, P. P., 1994. Separating marginal utility and probabilistic risk aversion. Theory and Decision 36, 1–44.

[34] Wakker, P. P., 2010. Prospect Theory for risk and ambiguity. Cambridge University Press, Cambridge.

[35] Wu, G., 1994. An empirical test of ordinal independence. Journal of Risk and Uncertainty 9, 39–60.

[36] Yager, R. R., 1988. On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190.

Appendix A. Proofs

Proof of Proposition 1. $(5) \Rightarrow (7)$. Since $X = \mathbb{R}$, $\succsim$ is representable by URDU and $u : \mathbb{R} \to \mathbb{R}$ is increasing, (5) can be rewritten as

$$\forall x \in \mathbb{R}, \quad x > y_{(1)} \Rightarrow w_1^{n+1} u(x) + \sum_{i=1}^{n} w_{i+1}^{n+1} u(y_{(i)}) > \sum_{i=1}^{n} w_i^n u(y_{(i)}).$$

Because of the continuity of u, this amounts to

$$w_1^{n+1} u(y_{(1)}) + \sum_{i=1}^{n} w_{i+1}^{n+1} u(y_{(i)}) \geq \sum_{i=1}^{n} w_i^n u(y_{(i)}). \qquad (A.1)$$

Let j be an integer in $\{1, \ldots, n-1\}$ and let us choose $A = \{z + j\epsilon, z + (j-1)\epsilon, \ldots, z + \epsilon, w + (n-j)\epsilon, w + (n-j-1)\epsilon, \ldots, w + \epsilon\}$, with $z > w$. If $\epsilon \to 0$, then, because of the continuity of u, (A.1) implies

$$w_1^{n+1} u(z) + \sum_{i=1}^{j} w_{i+1}^{n+1} u(z) + \sum_{i=j+1}^{n} w_{i+1}^{n+1} u(w) \geq \sum_{i=1}^{j} w_i^n u(z) + \sum_{i=j+1}^{n} w_i^n u(w),$$
$$(A.2)$$

23

for any $j \in \{1, \dots, n-1\}$. Since $\sum_{i=1}^{n+1} w_i^{n+1} = 1 = \sum_{i=1}^{n} w_i^n$, we also have $\sum_{i=1}^{n+1} w_i^{n+1} u(w) = \sum_{i=1}^{n} w_i^n u(w)$. This, combined with (A.2) yields

$$w_1^{n+1}(u(z) - u(w)) + \sum_{i=1}^{j} w_{i+1}^{n+1}(u(z) - u(w)) + \sum_{i=j+1}^{n} w_{i+1}^{n+1}(u(w) - u(w))$$

$$\geq \sum_{i=1}^{j} w_i^n(u(z) - u(w)) + \sum_{i=j+1}^{n} w_i^n(u(w) - u(w))$$

for any $j \in \{1, \dots, n-1\}$. Therefore

$$w_1^{n+1}(u(z) - u(w)) + \sum_{i=1}^{j} w_{i+1}^{n+1}(u(z) - u(w)) \geq \sum_{i=1}^{j} w_i^n(u(z) - u(w))$$

and

$$\sum_{i=1}^{j+1} w_i^{n+1} \geq \sum_{i=1}^{j} w_i^n, \quad \forall j \in \{1, \dots, n-1\}.$$

Notice that this is the same condition as (7).

(7) $\Rightarrow$ (5) . Let us fix $A = \{y_{(1)}, \dots, y_{(n)}\}$ and define $a_1 = u(y_{(1)})$ and $a_i = u(y_{(i)}) - u(y_{(i-1)})$ for all $i \in \{2, \dots, n\}$. Notice that $a_i \leq 0$ for all $i \in \{2, \dots, n\}$ and $u(y_{(i)}) = \sum_{j=1}^{i} a_j$ for all $i \in \{1, \dots, n\}$. We have

$$\sum_{i=1}^{n} (w_i^n - w_{i+1}^{n+1}) u(y_{(i)}) = \sum_{i=1}^{n} (w_i^n - w_{i+1}^{n+1}) \sum_{j=1}^{i} a_j = \sum_{j=1}^{n} a_j \sum_{i=j}^{n} (w_i^n - w_{i+1}^{n+1})$$

$$= \sum_{j=1}^{n} a_j \left(\sum_{i=j}^{n} w_i^n - \sum_{i=j}^{n} w_{i+1}^{n+1} \right)$$

$$= \sum_{j=1}^{n} a_j \left(1 - \sum_{i=1}^{j-1} w_i^n - 1 + \sum_{i=1}^{j} w_i^{n+1} \right)$$

$$= \sum_{j=1}^{n} a_j \left(\sum_{i=1}^{j} w_i^{n+1} - \sum_{i=1}^{j-1} w_i^n \right)$$

$$= w_1^{n+1} u(y_{(1)}) + \sum_{j=2}^{n} \underbrace{a_j}_{\leq 0} \underbrace{\left(\sum_{i=1}^{j} w_i^{n+1} - \sum_{i=1}^{j-1} w_i^n \right)}_{\geq 0 \text{ (by (7))}}$$

$$\leq w_1^{n+1} u(y_{(1)}).$$

So, $\sum_{i=1}^{n} (w_i^n - w_{i+1}^{n+1}) u(y_{(i)}) \leq w_1^{n+1} u(y_{(1)})$, which is equivalent to (A.1). Going from (A.1) to (5) is easy and left to the reader.

A similar reasoning shows that $\succsim$ satisfies (6) iff (8) holds. $\qquad \square$

Proof of Proposition 2. We prove only Simple Top Monotonicity. By definition, any monotonic utility function prescribes

$$\{x\} \succ \{y\} \succ \{z\} \quad \Leftrightarrow \quad u(x) > u(y) > u(z). \qquad (A.3)$$

Suppose URDU does not satisfy (9). Then, $\text{URDU}(\{x, z\}) \leq \text{URDU}(\{y, z\})$. Hence, by (4), $w_1^2 u(x) + w_2^2 u(z) \leq w_1^2 u(y) + w_2^2 u(z)$, or equivalently, $w_1^2 u(x) \leq w_1^2 u(y)$. Cancellation of w_1^2 yields $u(x) \leq u(y)$, which contradicts (A.3) and thus proves that Simple Top Monotonicity holds. □

Proof of Proposition 3. $\{x\} \succ \{y\}$ implies $u(x) > u(y)$. This, in turn, implies $\text{URDU}(\{x\}) > \text{URDU}(\{x, y\}) > \text{URDU}(\{y\}) \iff 0 < w_1^2 < 1$. □