Min and Max induced rankings: An experimental study.

Amélie Vrijdags[1]
Ghent University

Abstract

The current paper is the first to report an experimental study of "Min and Max induced rankings", i.e. a family of set rankings that require preferences over sets to be induced from comparison of the best and/or worst elements within those sets. When pitted against the Uniform Expected Utility criterion (UEU), the Min and Max induced rankings perform particularly worse in predicting real-life decision makers' preferences. The latter finding prompted us to investigate their axiomatic underpinnings by means of pairwise choice experiments. From this investigation, some important conclusions can be drawn: Axioms that prevent rankings to be based on total-goodness, as well as monotonicity conditions (ensuring that replacing a set element with a better one results in a better set) cannot be refuted. Axioms that rule out any utilization of the relative difference in position of the elements and axioms that prevent rankings to be based on average-goodness are all systematically violated. The UEU criterion seems to meet the apparent shortcomings of the Min and Max induced rankings. Some frequently occurring preference patterns, however, suggest that a significant portion of the participants uses neither a Min or Max induced ranking, nor UEU, but some other unspecified decision rule, possibly characterized by the tendency to prefer a diversification of uncertainty.

[1]Department of Data Analysis, Ghent University, Henri Dunantlaan 1, 9000 Ghent, Belgium. Tel: +32 9 264 63 87, Fax: +32 9 264 64 87, e-mail: Amelie.Vrijdags@UGent.be.

The author would like to thank Thierry Marchant and Denis Bouyssou for their helpful comments, as well as Michael Birnbaum for his help with the implementation of the true and error model.

Introduction

One can distinguish among different models for decision making by the quantity of information they require. In models for decision under risk, a unique and objectively known probability distribution over the possible outcomes is available to guide one's decision. In this respect, von Neumann & Morgenstern (1947) formulated the first classical axiomatization of Expected Utility, the most widely known model for choosing among objective probability distributions. Models for decision making under uncertainty, on the other hand, assume that objective probabilities do not necessarily exist. Savage (1956) argued that in such situations the decision maker acts as if she assigns her own subjective probabilities to the states of nature associated with the various possible outcomes. As demonstrated by the Ellsberg paradox, however, situations exist in which people do not act upon the usual probability rules. Other theorists investigated the concept of complete uncertainty or ignorance, which describes situations where the decision maker has no information about the probabilities of the outcomes, nor about their likelihood ranking. Usually, the resulting models follow the classical Savagian approach by defining a decision or an action as a mapping from the set of possible states of the world to the set of outcomes (e.g., Arrow & Hurwicz, 1972; Cohen & Jaffray, 1980; Maskin, 1979; Milnor, 1954; Luce & Raiffa, 1957). Nonetheless, scenarios exist for which the idea of states of the world is pointless because the decision maker does not have any notion at all of which particular state might lead to a given outcome, or because the number of possible states of the world is extremely large. For these scenarios, the previously described families of models are not adequate; they call for the development of models that associate each decision directly with a set of possible outcomes. As such, the problem of ranking decisions according to one's preferences will be reduced to the ranking of sets of outcomes. Therefore, we will refer to these kinds of decisions as "set rankings". Models within this domain are surveyed by Barberà et al. (2004). All of these models were axiomatically characterized, that is, axioms which take into account the decision maker's preferences over outcomes are imposed on the relation over the set of nonempty subsets of the universal space of outcomes. These axioms typically refer to reasonable attitudes towards uncertainty by the decision maker, as well as particular consistency conditions. An important observation in the domain of set ranking is that various combinations of plausible axioms imply that people rank sets by considering exclusively their best and worst elements. Indeed, in most rankings that have emerged in the literature, such as Maximin, Maximax, Minmax, Maxmin and their lexicographic extensions, an essential role is granted to the maximum and minimum of the sets to be compared. We will refer to these decision models as "Min and Max induced rankings".

For example, people who choose according to the *Maximin* decision rule, will base their decision on comparison of the worst elements only. If the worst elements of the sets to be compared are equal, the decision maker using Maximin will be indifferent. The *Maximax* rule entails a comparison of the best elements only.

Pattanaik & Peleg (1984) provided characterizations of *Leximin* and *Leximax*. Like Maximin, the Leximin rule starts with comparing the worst elements or minima when deciding between two sets. However, if both minima are equal, the decision maker is not indifferent but she will eliminate the worst elements from both sets and compare the reduced sets by their minima. If once again these reduced sets have the same minimum, the procedure will be repeated until a situation is reached in which the reduced sets each have a different minimum and a preference can be stated, or until one set is exhausted before the other in which case the nonempty set will be preferred. The Leximax rule, on the other hand, is the dual of the former in the sense that it starts with comparing the maxima after which it successively eliminates identical maxima. If by using Leximax, one set is eliminated completely while some elements of the other remain, the empty set will be preferred. A decision maker using either Leximin or Leximax will only be indifferent between two sets if they are completely identical.

The *Minmax* and *Maxmin*[2] rules (Arlegi, 2003) 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, a decision maker following Minmax will proceed to comparing the maxima. An indifference will be stated when 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.

Bossert et al. (2000) axiomatically characterized the *Lexicographic Minmax* rule where the decision maker initially proceeds as under Minmax. However, instead of

[2]Note that the Maxmin rule is not the same as the Max**i**min rule described in one of the previous paragraphs.

immediately imposing indifference in case of a ‘tie’, she will eliminate the best and worst elements of both sets and apply the Minmax rule to the reduced sets. This procedure will be repeated until a preference can be stated. Should one of the sets be exhausted in this process of repeated elimination whereas the other is not, then the nonempty set will be declared better than the empty set. The *Lexicographic Maxmin* rule is defined analogously.

As mentioned, all of these models were characterized by means of various plausible and intuitively appealing axioms which we will discuss later in this paper. Besides, they can also be intuitively motivated as reflecting certain internal attitudes of the decision maker vis-à-vis the choice problem, such as one’s risk aversion, the extent to which one is ready to iterate in the case of a tie, or the propensity to aim attention at particular characteristics of the sets (Arlegi, 2001).

Moreover, justification for these kinds of sequential processes in which only a limited number of attributes is taken into account can be found in the literature on “bounded rationality”, a concept introduced by Simon (Simon, 1955, 1956, 1990) which accounts for the observation that perfectly rational decisions are often not feasible in practice due to limitations in the computational resources available for making them. Simon argues that people rather “satisfice” than optimize; as soon as an alternative is found that meets one’s aspiration level, the decision process is cut short. The aforementioned Min and Max induced rankings, which imply that people concentrate on certain “focal features” of a decision situation, being the worst or the best outcome (or the second-worst or second best if the previous ones coincide), at the risk of ignoring potentially relevant information, perfectly fit the concept of bounded rationality. Furthermore, Gigerenzer and colleagues argue that simple step-by-step rules or heuristics can yield adaptive decisions in various situations (Gigerenzer & Selten, 2001). Research has also shown that simple heuristics are often more accurate in predicting actual decision behavior than models of optimal choice (Czerlinski et al., 1999; Dawes, 1979; Gigerenzer & Goldstein, 1999) and that they can even lead to better decisions (e.g., DeMiguel et al., 2009)

It is interesting to point out that almost all of the rankings described in the literature are “element-induced” (Arlegi, 2001), that is, preferences over sets are induced from the comparison of certain elements within the respective sets. Element induction, however, is not the only possible way of comparing and evaluating sets. For instance, a utility function can be defined over the universe of outcomes in addition to which additive operations could be performed. An example of such a decision rule is the Uniform Expected Utility criterion, which was characterized by Gravel et al. (2007). This model closely resembles the classical Expected Utility criterion: sets are ranked in accordance with the expected utility of their elements, but, since no information about the probabilities is available, the decision maker acts as if all outcomes within a set are equally probable.

So far, set rankings have only been adopted in some fields of economic analysis and models are mainly intended to be normatively appealing and useful in applied situations. The Min and Max induced rankings, for example, allow for the construction of a set ranking based exclusively on the observation of the preference relation over singletons. In our view, however, it would be interesting to investigate whether these rankings can prove themselves valid in the descriptive sense as well. So far we know of only one study that has adopted such a descriptive approach (i.e., Vrijdags, 2010). In the current article, we will examine the descriptive validity of the Min and Max induced rankings by two methods. First, we will attempt to test their overall validity by presenting a series of choices for which the Min induced rankings predict a different choice pattern than the Max induced rankings and by analyzing the number of participants showing each pattern. Secondly, since it will be shown that neither of these models fit the data, the individual axioms characterizing them will be tested in order to obtain a more detailed account of what goes wrong in particular. In the last section we will attempt to give some structure to the abundance of data gathered by testing every single axiom. The axioms will be grouped according to certain characteristics and the obtained clues as to how to model observed set rankings will be discussed.

Experimental method

The data for this paper were gathered in two experiments, Experiment I and Experiment II. For both experiments, participants were recruited by way of an e-mail in which they were asked to complete an Internet questionnaire regarding decision behavior. These questionnaires comprised a number of pairs of sets with numbers, representing monetary amounts in EUR, as the set elements. For each of these pairs the participants were asked to state their preference by clicking a "radio button" besides the preferred set. In the instructions it was explained that each set could 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 frequencies of each of the amounts in the container is unknown. For example, a set $\{28, 17\}$ can be thought of as a container holding an unknown number of tickets with "28" printed on them as well as an unknown number of tickets with "17" 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 euros. In this case, choosing a set comes down to choosing the lottery one would rather play.

Choices appeared as in Figure 1

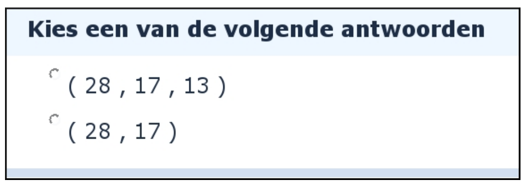

Figure 1. Choice representation for Experiment I and II. The sentence in Dutch literally translates to "Choose one of the following answers".

Each choice was presented twice with the position of the sets (first or second row) counterbalanced since, as explained in the next section, the proportion of preference reversals between replications of the same choice is needed to estimate the error rate for that choice.

In the first replicate, the choices were presented a first time. In the second replicate, the whole series was repeated. Within each replication, 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 101 people participated in Experiment I, which comprised 55 experimental choices, presented twice. Most of them were students in the Faculty of Psychology and Educational Sciences of Ghent University, 82% were female and 92% were between 18 and 25 years of age. Experiment II consisted of 33 choices, presented twice. A total of 131 students participated in Experiment II, 77% of which were female, and 91% of which were 23 years of age, or younger.

In order to test whether certain observed preference patterns are "real", rather than produced by random error, the "true and error" model proposed by Birnbaum (Birnbaum, 2004; Birnbaum & Bahra, 2007) was used. This model allows each person to have a different "true" pattern of preferences, and it allows each choice to have a different rate of error.

True and Error model

When choices are not too obvious to make, participants' responses to the same pair of sets cannot be expected to be fully consistent across multiple presentations. Since none of the representations in set ranking comes with a natural source of randomness, we need to impose some probabilistic structure. Different models of error arose in the literature on this subject (e.g., Sopher & Gigliotti, 1993; Harless & Camerer, 1994; Hey & Orme, 1994). In this paper the "true and error" model, proposed by Birnbaum (2004) is applied. This error model resembles that of Sopher & Gigliotti (1993), except that it uses repeated presentations of the same choices in order to unambiguously estimate the error rate for each distinct choice.

The true and error model assumes that preferences are subject specific, i.e., each subject is allowed to have a different "true" preference order. A subject's true preference order will not change when he or she is repeatedly presented with the same choice, but on any given trial there exists a possibility of making an "error" in evaluating his or her true preference. Presumably, the more "difficult" a choice is, the higher the error rate will be for that choice.

For example, consider a choice between two sets, F and S. If this choice is presented twice, there are four possible response patterns, FF if F is chosen both times, FS if one switches from F to S , SF if one switches from S to F, and SS if S is chosen both times. In this case, the observed pattern SS can come about in two mutually exclusive ways: the subject truly prefers S and made two correct reports, or the subject truly prefers F and made two errors. The theoretical probability that a person would choose the second gamble on both replicates, $P(SS)$, is then given by the following expression:

$$P(SS) = p(1-e)^2 + (1-p)e^2. \tag{1}$$

where p represents the probability of truly preferring S to F, and e represents the error rate for this choice.

Similarly, the probability of switching from F in the first replicate to S in the second replicate is given as follows:

$$P(FS) = pe(1-e) + (1-p)(1-e)e = e(1-e). \tag{2}$$

This model implies that the probability of switching from F to S equals the probability of switching from S to F. Therefore, the probability of either type of preference reversal equals $2e(1-e)$.

The probability of choosing the first gamble twice is given by:

$$P(FF) = pe^2 + (1-p)(1-e)^2. \tag{3}$$

The extension of this model to evaluate properties that comprise more than one choice is explained in the next section.

Comparison of Min and Max induced rankings and UEU

We do not assume that all people use the same decision rule for stating a preference between sets of outcomes. As pointed out, most rules that are standard in the field postulate that the decision maker starts by focussing on the elements she considers as representative or focal in each set, and then compares the sets by contrasting their respective focal elements.

People who are cautious or risk-averse, for example, might focus on the least attractive outcomes, the minima. If these people, in deciding between two sets, give absolute priority to the maximization of the minimum at the risk of ignoring potentially relevant information, they might be using the Maximin, Leximin, Minmax or Lexicographic Minmax rules or yet another rule that starts with comparing minimal outcomes. We will assign to this group of element-induced rules the collective name "Min induced rankings". Other people might be more inclined to focus on the most attractive outcomes, the maxima. If, for each choice, they start comparing sets by their maximal outcomes before anything else, their behavior accords with one of the "Max induced rankings", a term which we use to lump together all element-induced rules that incorporate comparison of the maxima as a first step, such as the Maximax, Leximax, Maxmin and Lexicographic Maxmin rules. Still others might not be using an element-induced rule at all, but judge or validate a set by some averaging operation, like the Uniform Expected Utility criterion (UEU), instead of limiting the scope to certain focal elements. In order to get a grasp of the distribution of these kinds of decision behaviors we set up an experiment in which we administered, amongst other choices, the series of three choices listed in Table 1.

Predictions

Each row in Table 1 represents a different pair of sets for which the participants were asked to state a preference. For example, predicted preference for the first set in choice 13 is indicated by "F_1" in Table 1, predicted preference for the second set is indicated by "S_1", and so on. As can be seen in Table 1, the minima and maxima of the sets to be compared never coincide. Under these circumstances, decision behavior according to the element induced-rules described above will fall into two categories: decisions based on the minimum if people follow one of the Min induced rankings, and decisions based on the maximum if their decision strategy matches one of the Max induced rules.

Table 1: Choice stimuli used for the examination of the predictive capacities of the Min and Max induced rankings and UEU.

Choice	Sets		Rankings		UEU
No.	First (F)	Second(S)	Min	Max	
13	$F_1 = \{6, 5\}$	$S_1 = \{7, 1\}$	F_1	S_1	F_1
25	$F_2 = \{25, 20\}$	$S_2 = \{21\}$	S_2	F_2	F_2
3	$F_3 = \{29, 3\}$	$S_3 = \{28, 27, 2\}$	F_3	F_3	S_3

The Min and Max induced rankings make different predictions for Choices 13, 25, and 3, administered in Experiment I. In the last three columns, F_i denotes a predicted preference for the first set; S_i indicates a predicted preference for the second set.

The data pattern implied by the Min induced rankings will be indicated by "$F_1S_2F_3$", which denotes the rankings $F_1 \succ S_1, S_2 \succ F_2$ and $F_3 \succ S_3$ in choices 13, 25 and 3, respectively. According to these decision rules, participants should choose F_1, S_2 and F_3 because all three of them have more favorable minima than their counterparts. Similarly, the Max induced rankings imply pattern $S_1F_2F_3$.

When the expected utilities are calculated for each set in the way prescribed by UEU (where we assume for simplicity that $u(x) = x$ for these rather small amounts[3]), these are the highest for sets F_1, F_2 and S_3 in choices 13, 25 and 3, respectively. Thereby, the predictions of UEU agree with the Min induced rankings in choice 13 and with the Max induced rankings in choice 25. For choice 3, both Min and Max induced rankings predict a preference for set F_3. Yet, UEU prescribes that people will choose set S_3; despite the fact that its minimal as well as its maximal outcomes are lower than those in set F_3, it still has a higher expected utility.

Evidently, other decision strategies that imply pattern $F_1F_2S_3$ can be thought of besides UEU. Consider for example a rule according to which people compare the ratio of 'high' and 'low' outcomes in each set in order to decide. In choice 13, for example, 1 EUR can be considered a low outcome, while 5 EUR, 6 EUR and 7 EUR can be regarded as higher ones. The first set thus contains nothing but high outcomes, whereas 50% of the elements in the second set is low, yielding a preference for the first set. The same rule would predict a preference for set F_2 in choice 25 since it contains one high and one low outcome as opposed to only one low outcome in set S_2, and set S_3 would be preferred in choice 3, where two out of three outcomes are high as opposed to one out of two in the first set.

Results

We used the true and error model (Birnbaum, 2004) presented in the introduction to estimate the error rates (e) and the true probability of preferring the second set (p) for each choice. Table 2 displays the number of people who showed each combination of stated preferences for the two presentations of each choice. The parameters, p and e, were estimated by minimizing the $\chi^2(1)$ between the observed frequencies and those predicted by the true and error model. The estimated parameters and corresponding $\chi^2(1)$'s are presented in the last three columns of Table 2. The true and error model appears to give a reasonable approximation to the data; indeed, none of the $\chi^2(1)$'s are significant ($\alpha = .05$).

[3] $u(x) = x$ is just one possible utility function for which these predictions hold. If the utility function were of the form $u(x) = x^\beta$, the predictions are correct for all β's within the interval $]0, 2.95]$.

Table 2: Replication data used to estimate $p =$ probability of true preference for the second set, and $e =$ error rate for each choice ($n = 101$).

Choice	Sets		Replication Patterns				$\hat{p}$	$\hat{e}$	$\chi^2(1)$
No.	First (F)	Second (S)	FF	FS	SF	SS			
13	$F_1 = \{6, 5\}$	$S_1 = \{7, 1\}$	90	4	4	3	0.031	0.041	0.00
25	$F_2 = \{25, 20\}$	$S_2 = \{21\}$	78	13	5	5	0.047	0.107	3.42
3	$F_3 = \{29, 3\}$	$S_3 = \{28, 27, 2\}$	12	12	17	60	0.865	0.176	0.86

Entries under FF, FS, SF, and SS are the observed numbers of people who showed each combination of choices for the two replicates of each choice. For example, 90 participants out of 101 chose the first set (F_1) in both presentations of Choice 13 (first row of the table).

Contrary to the predictions of the Max induced rankings, merely three percent of all participants is estimated to have a true preference for set S_1 in choice 13. For this choice, the majority thus seems to comply with the predictions of the Min induced rankings and UEU. A different phenomenon occurs in choice 25 where the results contradict the Min induced rankings. In choice 3, the modal response only agrees with UEU. In spite of its lower minimum and maximum, the estimated rate of true preference for the second set, S_3, is 86.5%.

Unlike the Min and Max induced rankings, which fail to predict the modal response in two out of three choices, UEU correctly predicts all three modal choices.

The frequencies of each response pattern for the complete sequence of three choices have been tabulated in Table 3. Since the three choices were presented twice, there are $2^6 = 64$ possible response patterns in total. Many of these 64 patterns have null frequencies, therefore, the data were pooled into 16 cells as follows. For each of the eight preference patterns in Table 3, 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 3 it is shown that 16 out of 101 participants demonstrated pattern $F_1F_2F_3$ on the first replicate, and 18 did so on the second replicate. Out of these, six people showed pattern $F_1F_2F_3$ on both replicates. The number of times $F_1F_2F_3$ was shown on either the first or the second replication, but *not* both is then given by $(16 - 6) + (18 - 6) = 22$. In order to avoid that responses with a different pattern 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 11. 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 3. From these 16 frequencies, which sum to the total number of participants, there are three error probabilities and eight true probabilities to be estimated.

Table 3: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns for Choices 13, 25, and 3 in Experiment I.

Response	Observed frequencies				Estimated true
pattern	Rep 1	Rep 2	Both	One not both	probability
$F_1F_2F_3$	16	18	6	11	0.079
$F_1F_2S_3$	68	58	44	19	0.821
$F_1S_2F_3$	4	7	2	3.5	0.038
$F_1S_2S_3$	6	11	1	7.5	0.005
$S_1F_2F_3$	4	4	1	3	0.023
$S_1F_2S_3$	3	3	2	1	0.035
$S_1S_2F_3$	0	0	0	0	0.000
$S_1S_2S_3$	0	0	0	0	0.000
Total	101	101	56	45	1

Entries under " Rep 1", "Rep 2" and "Both" represent the number of participants who showed each pattern for the first presentation of these three choices (the first replicate), the second presentation (the second replicate), and both presentations, respectively. The column labeled "One not both" contains the number of people who showed that pattern on one of the two replicates but not both divided by two. Estimated error rates are 0.041, 0.107, and 0.176 for Choices 13, 25, and 3, respectively. Evaluation of the true and error model yields $\chi^2(5) = 4.52$, an acceptable fit ($\alpha = .05$).

The true and error model can straightforwardly be extended to estimate the proportion of individuals with a given true preference pattern for the sequence of three choices (Birnbaum & Gutierrez, 2007; Birnbaum & Bahra, 2007; Birnbaum & Schmidt, 2008, 2010).

For example, suppose that a person shows the pattern predicted by UEU, i.e., $F_1F_2S_3$, on both replicates. There are eight different ways in which this pattern can occur, each of them corresponding to one of the eight possible true preference patterns. For instance, the conditional probability that someone demonstrates $F_1F_2S_3$ on both replicates, given that the person's true pattern is $F_1F_2F_3$ is as follows:

$$P(\text{“}F_1F_2S_3, F_1F_2S_3\text{”}|F_1F_2F_3) = (1-e_1)(1-e_2)e_3(1-e_1)(1-e_2)e_3.$$

The observed preference pattern is indicated with quotations marks, as opposed to a true pattern without quotation marks. Furthermore, e_1, e_2 and e_3 denote the error probabilities in the three respective choices. In this particular case, the decision maker stated her preferences correctly in each replicate of the first two choices, choices 13 and 25, but made an error on both replicates of the third choice.

The joint probability of displaying response pattern "$F_1F_2S_3, F_1F_2S_3$" and at the same time having the true pattern $F_1F_2F_3$ (making the same mistake in both replicates) is given by

$$P(\text{``}F_1F_2S_3, F_1F_2S_3\text{''} \cap F_1F_2F_3) = p(F_1F_2F_3)(1-e_1)(1-e_2)e_3(1-e_1)(1-e_2)e_3.$$

where $p(F_1F_2F_3)$ is the theoretical probability that someone's true pattern is $F_1F_2F_3$. The overall probability of a person demonstrating response pattern $F_1F_2S_3$ in both replicates, $P(\text{``}F_1F_2S_3, F_1F_2S_3\text{''})$, is the sum of eight terms like the above, one for each of the eight true patterns.

$$\begin{aligned}
P(\text{``}F_1F_2S_3, F_1F_2S_3\text{''}) \;=\; & p(F_1F_2F_3)(1-e_1)^2(1-e_2)^2(1-e_3)^2 + \\
& p(F_1F_2S_3)(1-e_1)^2(1-e_2)^2e_3^2 + p(F_1S_2F_3)(1-e_1)^2e_2^2(1-e_3)^2 + \\
& p(F_1S_2S_3)(1-e_1)^2e_2^2e_3^2 + p(S_1F_2F_3)e_1^2(1-e_2)^2(1-e_3)^2 + \\
& p(S_1F_2S_3)e_1^2(1-e_2)^2e_3^2 + p(S_1S_2F_3)e_1^2e_2^2(1-e_3)^2 + \\
& p(S_1S_2S_3)e_1^2e_2^2e_3^2.
\end{aligned}$$

Similarly, the overall probability of showing $F_1F_2S_3$ in one replicate (the first or the second), irrespective of the response on the other replicate is given by:

$$\begin{aligned}
P(\text{``}F_1F_2S_3\text{''}) \;=\; & p(F_1F_2F_3)(1-e_1)(1-e_2)(1-e_3) + p(F_1F_2S_3)(1-e_1)(1-e_2)e_3 + \\
& p(F_1S_2F_3)(1-e_1)e_2(1-e_3) + p(F_1S_2S_3)(1-e_1)e_2e_3 + \\
& p(S_1F_2F_3)e_1(1-e_2)(1-e_3) + p(S_1F_2S_3)e_1(1-e_2)e_3 + \\
& p(S_1S_2F_3)e_1e_2(1-e_3) + p(S_1S_2S_3)e_1e_2e_3.
\end{aligned}$$

Using elementary algebra and probability theory, it can be shown that the average probability of showing pattern $F_1F_2S_3$ on one replicate or the other, but not both is obtained with:

$$P(\text{“}F_1F_2S_3\text{”}) - P(\text{“}F_1F_2S_3, F_1F_2S_3\text{”}).$$

The extended true and error model was fit to the 16 frequencies in Table 3. Error estimates obtained from the replication data in the previous analysis were used. 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 true and error model. The column labeled "estimated true probability" in Table 3 shows the estimated true probabilities of each pattern, which were estimated to minimize the χ^2 between observed and predicted frequencies. For this analysis, $\chi^2(5)$ equals 4.52, which is not significant (with $\alpha = 0.05$), suggesting that the general true and error model can be retained. With this model, 82.1% of all participants are estimated to have $F_1F_2S_3$ ($p(F_1F_2S_3) = 0.821$), the pattern predicted by UEU, as their true pattern. The patterns predicted by the Min induced rankings and the Max induced rankings were both very rare: $p(F_1S_2F_3)$ =3.8% and $p(S_1F_2F_3)$ =2.3%, respectively.

Discussion

With the aim of providing an indication as to what extent real life decision makers actually make decisions in accordance with the element-induced decision rules described in the introduction, we devised a sequence of three choices for which Min induced rankings, Max induced rankings and UEU each predict a different choice pattern.

The results provide a strong refutation of the Min and Max induced rules since only a very small percentage of people seems to base their decisions on minimal or maximal outcomes only, ignoring other relevant information. In this study, UEU performs much better, the pattern predicted by UEU was followed by the great majority of participants. Of course, UEU is not the only possible decision rule that can account for these results.

The results do not reveal which specific decision rule(s) is (are) used by the vast proportion of participants who showed pattern $F_1F_2S_3$, but they do demonstrate, however, that it is not a Min or Max induced ranking.

As pointed out in the introduction, each of these rankings has been characterized by means of axioms imposed on the preference relation between sets. It has been clearly demonstrated that the Min and Max induced rankings fail to fit the data, which indicates that at least some of these axioms are not satisfied by the participants. In the next sections we will explore exactly which axioms are systematically violated and which appear to be valid[4]. Knowledge thereof might be useful if one aims to design a decision model which is able to accommodate actual choice behavior. In order to construct empirical tests for the axioms, we attempted, where possible, to devise our choice stimuli in such a way that we believe they are most likely to yield violations. In most cases this was done by assuming that real-life decision makers perform some averaging operation over the outcomes in each set.

Axiom tests

Prior to continuing, some notation should be clarified. An *ordering* is a reflexive, transitive and complete binary relation, and a *linear ordering* is an antisymmetric ordering. A *linear preference ordering* is a linear ordering with the interpretation "at least as good as". Let X be a nonempty and finite universal set of outcomes. The set of all nonempty finite subsets of X is denoted by $\mathcal{X}$. Elements of $\mathcal{X}$ are interpreted as sets of possible outcomes under uncertainty, where the decision maker does not know any probability distribution, nor any likelihood ranking of the possible outcomes. Let R denote a linear preference ordering over X. The asymmetric factor of R is denoted by P. Let $\succsim$ be an ordering over $\mathcal{X}$. The interpretation of $\succsim$ is such that $A \succsim B$ if and only if the set of possible outcomes $A \in \mathcal{X}$ is considered at least as good as the set of possible outcomes $B \in \mathcal{X}$ by the decision maker. Furthermore, $\succ$ denotes the asymmetric factor of $\succsim$.

Leximin & Leximax

Axioms and Stimuli. In this section, we will explain in detail the axioms characterizing Leximin and Leximax as formulated by Pattanaik & Peleg (1984) as well as the stimuli used to test them. It should be noted that every axiom in this paper was tested by means of more than one (combination of) choice(s). In order to avoid

[4] Although Maximin and Maximax are widely mentioned as basic decision rules for ranking sets of outcomes, we did, however, not find an explicit axiomatization in the context of set ranking. As a consequence, we will not provide any detailed empirical analysis of these decision rules. Nonetheless, it is suspected that the axioms needed for the characterization of Maximin and Maximax will not differ substantially from the axioms tested in the next sections.

needlessly complicating this paper, only one test for each axiom will be reported, namely the test for which the most violations were observed[5].

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

1. $[xPy \text{ for all } y \in A] \Rightarrow A \cup \{x\} \succ A$;
2. $[yPx \text{ for all } y \in A] \Rightarrow A \succ A \cup \{x\}$;

Dominance requires that adding an element which is better (worse) than all elements in a given set A produces a set that is better (worse) than the original set. Given the interpretation of $\succsim$, Dominance is often considered a basic requirement for the ranking of sets of possible outcomes.

As shown in Table 4, this axiom was tested in Experiment I, with choices 31 and 32. In choice 31, the second set contains the same outcomes as the first set plus one additional outcome, 13 EUR, which is worse than every outcome in F_1, resulting in set $S_1 = \{28, 17, 13\}$ which, in the case where Dominance holds, should be considered less attractive than set $F_1 = \{28, 17\}$. Similarly, people should also prefer $F_2 = \{37, 28, 17\}$ to $S_1 = \{28, 17\}$ according to Dominance.

Neutrality: For all $A, B \in \mathcal{X}$, for all one-to-one mappings $\varphi : A \cup B \to X$,

$$([xRy \Leftrightarrow \varphi(x)R\varphi(y) \text{ and } yRx \Leftrightarrow \varphi(y)R\varphi(x)] \text{ for all } x \in A, \text{ for all } y \in B)$$
$$\Rightarrow (A \succsim B \Leftrightarrow \varphi(A) \succsim \varphi(B) \text{ and } B \succsim A \Leftrightarrow \varphi(B) \succsim \varphi(A)).$$

Neutrality implies that the relation R is not affected by changes in two sets A and B that preserve the relative rankings of all elements in $A \cup B$. This rules out any utilization of the relative differences in position of the outcomes. Consider choices 1 and 15 in Table 4 and suppose that a decision maker prefers $F_1 = \{15, 10\}$ to $S_1 = \{11\}$. In this case, Neutrality requires that she prefers $F_2 = \{25, 15\}$ to $S_2 = \{24\}$ as well, since the mapping from F_1 and S_1 to F_2 and S_2 does not change the relative rankings of the elements in both sets.

Bottom Independence: For all $A, B \in \mathcal{X}$, for all $x \in X$ such that yPx for all $y \in A \cup B$,

[5] An exhaustive listing of all the data gathered in Experiment I and II can be found in Table 15.

Table 4: Stimuli used to empirically validate the axioms characterizing Leximin and Leximax (Pattanaik & Peleg, 1984).

Axiom tested	Choice	Experiment	Set First (F)	Set Second(S)
Dominance	31*	II	$F_1 = \{28, 17\}$	$S_1 = \{28, 17, 13\}$
	32	II	$F_2 = \{37, 28, 17\}$	$S_2 = \{28, 17\}$
Neutrality	1*	II	$F_1 = \{15, 10\}$	$S_1 = \{11\}$
	15	II	$F_2 = \{25, 15\}$	$S_2 = \{24\}$
Bottom Ind	19	I	$F_1 = \{20\}$	$S_1 = \{20, 18\}$
	20*	I	$F_2 = \{20, 1\}$	$S_2 = \{20, 18, 1\}$
Top Ind	21	I	$F_1 = \{4, 3\}$	$S_1 = \{3\}$
	22	I	$F_2 = \{35, 4, 3\}$	$S_2 = \{35, 3\}$
Disjoint Ind	29*	I	$F_1 = \{100\}$	$S_1 = \{95, 90\}$
	30	I	$F_2 = \{100, 10\}$	$S_2 = \{90, 95, 10\}$

$*$ Choices marked with an asterisk were presented with set S in the first position, and set F in the second position when presented for the first time. These positions were reversed in the second replicate. Unmarked choices were arranged in the opposite fashion.

$$A \succ B \Rightarrow A \cup \{x\} \succ B \cup \{x\}.$$

Top Independence: For all $A, B \in \mathcal{X}$, for all $x \in X$ such that xPy for all $y \in A \cup B$,

$$A \succ B \Rightarrow A \cup \{x\} \succ B \cup \{x\}.$$

Disjoint Independence: For all $A, B \in \mathcal{X}$ such that $A \cap B = \emptyset$,
for all $x \in X \backslash (A \cup B)$,

$$A \succ B \Rightarrow A \cup \{x\} \succ B \cup \{x\}.$$

Bottom Independence requires that if there exists a strict preference between to sets A and B, adding the same element to both sets, an element which is worse than any of the outcomes in either of the two original sets, should lead to a relative ranking of the resulting sets which is the same as that of the originals. Top Independence is its dual, where the added element is better than all of those present in the original sets. Disjoint Independence applies to situations where A and B are disjoint. Here, the element to be added simply needs to be outside both original sets.

Choices 19 and 20 in Table 4 were used to test Bottom Independence. In choice 19, one is sure of receiving 20 EUR if set $F_1 = \{20\}$ is chosen. In set $S_1 = \{20, 18\}$, on the other hand, there is a certain, unknown probability of ending up with only 18 EUR. In this case, we would expect most people to choose set F_1, since there is no apparent reason to do otherwise, except if the decision is based on maximal outcomes only, in which case one would be indifferent between both sets. In choice 20, the original sets are enlarged with the same outcome, 1 EUR, which is considerably lower than all other outcomes. If most people prefer the first set in choice 19, then, according to Bottom Independence, they should also do so in choice 20. However, in choice 20, the average over the three outcomes in set S_2 is higher than the average over the two outcomes in set F_2. People who evaluate sets by performing some averaging operation over the outcomes might prefer the first set in choice 19, and the second set in choice 20, and would thereby be violating Bottom Independence. The next two choices in Table 4, were used to test Top Independence. These choices were constructed following the reverse recipe as for Bottom Independence. First, participants are asked to choose between a binary set, $F_1 = \{4, 3\}$, and a singleton, $S_1 = \{3\}$. Here, both sets have the same minimum, while in the test for Bottom Independence, the sets in the first pair share their maximum. In choice 22, a considerably higher outcome, 35 EUR, is added to the original sets, which again results in an evident inequality between the first and the second set in terms of the averages over the outcomes. Consequently, people who consider the average over the elements in each set might demonstrate pattern F_1S_2, i.e., they state preferences $F_1 \succ S_1$ and $S_2 \succ F_2$, a violation of Top Independence. Our test for Disjoint Independence consists of Choices 29 and 30[6]. Sets in Choice 29 do not have any outcomes in common. Here, everyone should prefer $F_1 = \{100\}$, because its certain outcome is higher than any of the outcomes $S_1 = \{95, 90\}$. In Choice 30, the same lower consequence was added to both original sets. Even though in this last choice, the maximum is still higher in set $F_2 = \{100, 10\}$ than in set $S_2 = \{95, 90, 10\}$, the average over the outcomes is higher in the latter. People who choose F_1 in choice 29 and S_2 in choice 30 would be violating Disjoint Independence.

[6] The same set of stimuli could also be used to test Bottom Independence, but as rate of violations was higher for choices 19 and 20, both tests were reported.

Results. The number of people demonstrating each preference pattern for the two presentations of the eight experimental choices is reported in Table 5. From these data, the rates of error and the true probabilities of preferring the second set were estimated for each choice.

Table 5: Replication data used to estimate the true probability and error rate for each choice.

Choice	Set		Replication Patterns				$\hat{p}$	$\hat{e}$	$\chi^2(1)$
(Exp.)	First (F)	Second(S)	FF	FS	SF	SS			
Dominance									
31(II)	$F_1 = \{28, 17\}$	$S_1 = \{28, 17, 13\}$	111	11	6	3	0.020	0.073	1.44
32(II)	$F_2 = \{37, 28, 17\}$	$S_2 = \{28, 17\}$	125	1	4	1	0.007	0.023	1.67
Neutrality									
1(II)	$F_1 = \{15, 10\}$	$S_1 = \{11\}$	111	12	6	2	0.011	0.078	1.95
15(II)	$F_2 = \{25, 15\}$	$S_2 = \{24\}$	7	7	11	106	0.944	0.076	0.88
Bottom Independence									
19(I)	$F_1 = \{20\}$	$S_1 = \{20, 18\}$	85	8	6	2	0.017	0.076	0.28
20(I)	$F_2 = \{20, 1\}$	$S_2 = \{20, 18, 1\}$	1	2	4	94	0.991	0.032	0.65
Top Independence									
21(I)	$F_1 = \{4, 3\}$	$S_1 = \{3\}$	99	1	1	0	0.000	0.010	0.01
22(I)	$F_2 = \{35, 4, 3\}$	$S_2 = \{35, 3\}$	31	14	21	35	0.536	0.227	1.39
Disjoint Independence									
29(I)	$F_1 = \{100\}$	$S_1 = \{95, 90\}$	94	3	4	0	0.000	0.036	0.27
30(I)	$F_2 = \{100, 10\}$	$S_2 = \{90, 95, 10\}$	6	4	8	83	0.937	0.067	1.30

Entries under FF, FS, SF, and SS are the observed frequencies for each combination of choices on the two replicates. The $\chi^2(1)$'s in the right-most column evaluate the fit of the true and error model. All are acceptable ($\alpha = .05$).

The first row in the table shows that 111 out of the 131 participants in Experiment II chose set F_1 on both presentations of choice 31, and only 3 chose S_1 both times. For this choice, merely 2% of the participants are estimated to have a true preference for the second set, and the estimated error rate on this choice is 7.3%. For the majority of people, i.e., 98%, choices thus comply with the Dominance axiom. The same holds for choice 32 where 99.3% is estimated to have a true preference for set F_2, thereby satisfying Dominance. According to Neutrality, people should choose either the first or the second set on both choices 1 and 15. Instead, the results show that 98.9% truly preferred F_1 in choice 1, while only 5% chose F_2 in choice 15. The modal choices of the participants in Experiment I also violate Top Independence, Bottom Independence and Disjoint Independence, since most people chose set F_1 on the first choice of each respective test, and set S_2 on the second. The last column of Table 5 contains the chi-square tests of the true and error model; none is significant ($\alpha = 0.05$), indicating that this error model can be accepted.

Table 6: Estimated true probabilities of each response pattern in the tests of Dominance, Neutrality, and the three independence axioms.

Choice	Set		Parameter Estimates					$\chi^2(2)$
(Exp)	First (F)	Second (S)	$\hat{e}$	$\hat{p}(F_1F_2)$	$\hat{p}(F_1S_2)$	$\hat{p}(S_1F_2)$	$\hat{p}(S_1S_2)$	
Dominance								
31(II)	$F_1 = \{28, 17\}$	$S_1 = \{28, 17, 13\}$	0.073	**0.970**	0.008	0.021	0.000	0.51
32(II)	$F_2 = \{37, 28, 17\}$	$S_2 = \{28, 17\}$	0.023					
Neutrality								
1(II)	$F_1 = \{15, 10\}$	$S_1 = \{11\}$	0.078	**0.064**	0.922	0.000	**0.014**	0.83
15(II)	$F_2 = \{25, 15\}$	$S_2 = \{24\}$	0.076					
Bottom Independence								
19(I)	$F_1 = \{20\}$	$S_1 = \{20, 18\}$	0.076	**0.012**	0.982	0.000	**0.006**	0.23
20(I)	$F_2 = \{20, 1\}$	$S_2 = \{20, 18, 1\}$	0.032					
Top Independence								
21(I)	$F_1 = \{4, 3\}$	$S_1 = \{3\}$	0.010	**0.476**	0.524	0.000	**0.000**	0.99
22(I)	$F_2 = \{35, 4, 3\}$	$S_2 = \{35, 3\}$	0.227					
Disjoint Independence								
29(I)	$F_1 = \{100\}$	$S_1 = \{95, 90\}$	0.036	**0.057**	0.943	0.000	**0.000**	0.24
30(I)	$F_2 = \{100, 10\}$	$S_2 = \{90, 95, 10\}$	0.067					

Tests of the true and error 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 the axiom on that line.

Table 6 shows the results of the true and error model, extended to combinations of two choices[7]. For each axiom test, this model assumes that a participant's true preferences correspond to one of the four possible patterns: F_1F_2, F_1S_2, S_1F_2, and S_1S_2.

According to Dominance, people should demonstrate pattern F_1F_2. The first row in Table 6 shows that the estimated probability of having F_1F_2 as true preference pattern in the test for Dominance equals 97%. In this test, almost everybody's choices agree with Dominance.

If Neutrality holds, people should show either F_1F_2 or S_1S_2. Instead, it is estimated that 92.2% of participants truly switch from F_1 in Choice 1 to S_2 in Choice

[7] After extending the true and error model to a model capable of analyzing patterns consisting of three choices, adaptation for the analysis of patterns consisting of two choices is self-explanatory (e.g., Birnbaum, 2008).

15, hereby violating Neutrality.

The bottom part of Table 6 displays the tests for Bottom Independence, Top Independence and Disjoint Independence. Patterns where preferences are reversed, i.e., S_1F_2 and F_1S_2, are incompatible with the independence axioms, patterns S_1S_2 and F_1F_2 agree with the axioms. In all three tests, most people are estimated to violate the axioms. In none of the rows did the sum of $\hat{p}(F_1F_2)$ and $\hat{p}(S_1S_2)$ reach a majority. In the test for Bottom Independence (choices 19 and 20), 98.2% are estimated to have F_1S_2 as their true preference pattern, while true preferences are estimated to comply with Bottom Independence for only 1.8% ($\hat{p}(F_1F_2) + \hat{p}(S_1S_2)$). The rates of violations of Top Independence and Disjoint Independence are 52.4% and 94.3%, respectively. Surprisingly, even though the choice stimuli were constructed using the same—albeit reversed—recipe, the violation rate is much lower for Top Independence than for Bottom Independence. In fact, although almost everybody chose $S_2 = \{20, 18, 1\}$ instead of $F_2 = \{20, 1\}$ in the test of Bottom Independence, also 46.4% truly preferred $F_2 = \{35, 4, 3\}$ over $S_2 = \{35, 3\}$ in the test of Top Independence (see Table 5). The latter preference statement does not match the assumption implied by models like UEU, according to which these stimuli were designed, i.e., that people rank sets according to the average over the values of the outcomes.

Discussion. The only axiom of Pattanaik and Peleg's (1984) axiomatization of the Leximin and Leximax rules that could not be refuted empirically is Dominance, which is considered an extremely plausible requirement, given the interpretation of $\succsim$.

Our test of Neutrality demonstrated that most participants changed their preferences as the relative values of the outcomes in the sets to be compared changed, even though the comparative order of those outcomes over both sets remained the same. This finding entails that relative differences between outcomes appear to play a role in choosing between sets of monetary consequences.

We were able to bring about substantial violation rates for all three independence axioms. Presumably, these violations occurred because the choice stimuli were constructed in such a way that the addition of a common element resulted in a reversal of the ranking of the set averages. Particularly for those people who actually evaluate sets by their average, such stimuli should give rise to preference statements that disagree with the independence axioms. Surprisingly, we found the violations to be substantially lower in number for Top Independence as compared to Bottom Independence and Disjoint Independence. Apparently, people are more likely to choose the set with the higher average if this is also the set with the highest cardinality, than when the set with the highest average has the least number of elements. This observation contradicts decision models like UEU, and is reminiscent of the Richness Appeal axiom[8], proposed by Arlegi (2007) as an attitude towards uncertainty in the context of sequentially consistent rankings. Preferences complying with Richness Appeal might be indicative of a positive attitude towards a diversification uncertainty within the range of the minimum and maximum of a set; the decision maker prefers one more possible outcome instead of being constrained to the two extremes, even when the value of this middle outcome is very close to the minimum, resulting in a lower set average.

Minmax & Maxmin

Axioms and Stimuli. Arlegi (2003) postulated six axioms for the characterization of Minmax and Maxmin, each of which is described in more detail below.

Simple Uncertainty Aversion: For all $x, y, z \in X$,

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

Simple Uncertainty Appeal: For all $x, y, z \in X$

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

Simple Uncertainty Aversion and Simple Uncertainty Appeal are contrasting empirical assumptions regarding the decision maker's attitude towards uncertainty when comparing singletons and binary sets. Simple Uncertainty Aversion establishes that receiving an outcome y with certainty is always preferred to an uncertain prospect where either outcome x which is better than y or z which is worse than y may materialize.

Simple Uncertainty Appeal reflects the dual attitude. This axiom prescribes that the possibility of receiving an outcome which is better than y, despite the fact

[8] ***Richness Appeal***: For all $x, y, z \in X$, $xPyPz \Rightarrow \{x, y, z\} \succ \{x, z\}$.
Richness Appeal stipulates that any set of three outcomes is always considered better than the binary set consisting of its minimum and maximum.

that it is linked with the possibility of receiving an outcome worse than y, is always considered more attractive than receiving y with certainty.

Choices 15 and 16 in Table 7 were devised to test Simple Uncertainty Aversion and Simple Uncertainty Appeal. Like Neutrality, these axioms assume that only ordinal information about the outcomes is taken into account by the decision maker. Consequently, if $\succsim$ represents uncertainty aversion, the certain option should be preferred in both choices, regardless of the fact that the certain outcome is much closer to the worst outcome of the uncertain prospect $\{25, 15\}$ in Choice 16, i.e., $S_2 = \{16\}$, than in Choice 15, i.e., $S_1 = \{24\}$. Similarly, if $\succsim$ represents uncertainty appeal, the uncertain option, $\{25, 15\}$, should be preferred in both choices. If the relative differences in the positions of the objects do appear to matter (as was demonstrated in the test for Neutrality in the preceding section), it is not unthinkable that the certain outcome is chosen in the first pair of sets, and the uncertain prospect in the second pair.

Table 7: Stimuli used to empirically validate the axioms characterizing Minmax and Maxmin (Arlegi, 2003).

Axiom tested	Choice	Experiment	Set	
			First (F)	Second(S)
Simple Uncertainty Aversion/Appeal	15	II	$F_1 = \{25, 15\}$	$S_1 = \{24\}$
	16	II	$F_2 = \{25, 15\}$	$S_2 = \{16\}$
Simple Top Monotonicity	45∗	I	$F_1 = \{29, 3\}$	$S_1 = \{27, 3\}$
Simple Bottom Monotonicity	38	I	$F_1 = \{20, 2\}$	$S_1 = \{20, 1\}$
Monotone Consistency	38	I	$F_1 = \{20, 2\}$	$S_1 = \{20, 1\}$
	39	I	$F_2 = \{20, 2, 1\}$	$S_2 = \{20, 1\}$
Robustness	47∗	I	$F_1 = \{59, 5, 1\}$	$S_1 = \{56, 5, 1\}$
	48	I	$F_2 = \{59, 5, 1\}$	$S_2 = \{55, 5, 1\}$
	49∗	I	$F_3 = \{59, 5, 1\}$	$S_3 = \{56, 55, 5, 1\}$

∗ Choices marked with an asterisk were presented with set S in the first position, and set F in the second position in the first replicate. In the second replicate, these positions were reversed. Unmarked choices were arranged in the opposite fashion.

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

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

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

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

Given a singleton $\{z\}$ and two outcomes x and y, both of which are strictly better than z, Simple Top Monotonicity implies that $\{z\}$ enlarged with the better of x and y will be strictly preferred to $\{z\}$ enlarged with the worse of those two outcomes. Simple Bottom Monotonicity is its dual, and applies to enlarging a singleton with worse outcomes. These axioms can be regarded as "rationality conditions" for choice under complete uncertainty (Bossert et al., 2000), and were assessed with Choices 45 and 38 in Table 7.

Monotone Consistency : For all $A, B \in \mathcal{X}$,

$$A \succsim B \Rightarrow A \cup B \succsim B.$$

Monotone Consistency requires that if a set A is at least as good as another set B, then B cannot be strictly better than the union of both A and B. Indeed, the union of A and B contains the same outcomes as B, plus those in A, which made the decision maker evaluate A as better than B. However, if one allows A and B to have at least one 'high' outcome in common, it is possible to establish a reversal of the ranking of the set averages from the first to the second choice, which might cause the decision maker to violate Monotone Consistency. This can be illustrated with Choices 38 and 39 in Table 7. In Choice 38, sets $F_1 = \{20, 2\}$ and $S_1 = \{20, 1\}$ differ only in their minima. Consequently, as is also predicted by Simple Bottom Monotonicity, we expect people to choose F_1, the set with the most favorable minimum. If we compare the union of F_1 and S_1, i.e., $F_2 = \{20, 2, 1\}$, with $S_2 = \{20, 1\}$, the minima and maxima are equal in both sets, but F_2 has one more low outcome. The average over the outcomes is now more favorable in the second set. People who choose set F_1 in Choice 38 and set S_2 in Choice 39 would be violating Monotone Consistency.

Robustness: For all $A, B, C \in \mathcal{X}$,

$$A \succsim B \text{ and } A \succsim C \Rightarrow A \succsim B \cup C.$$

Robustness ensures that if A is weakly better than both B and C, then A is weakly better than the union of B and C.

Choices 47, 48 and 49 in Table 7 constitute a test for Robustness. By allowing the first and second set to share their lowest outcomes, we were again able to produce a reversal in the ranking of the set averages from the first two choices to Choice 49, which might lead some subjects to exhibit response pattern $F_1F_2S_3$ for the three choices, a violation of Robustness.

Table 8: Replication data used to estimate true probability and error rate for each choice.

Choice	Set		Replication Patterns				$\hat{p}$	$\hat{e}$	$\chi^2(1)$
(Exp.)	First (F)	Second(S)	FF	FS	SF	SS			
Simple Uncertainty Aversion/Appeal									
15(II)	$F_1 = \{25, 15\}$	$S_1 = \{24\}$	106	11	7	7	0.944	0.076	0.88
16(II)	$F_2 = \{25, 15\}$	$S_2 = \{16\}$	117	7	2	5	0.039	0.041	2.60
Simple Top Monotonicity									
45(I)	$F_1 = \{29, 3\}$	$S_1 = \{27, 3\}$	100	1	0	0	0.000	0.007	0.84
Simple Bottom Monotonicity									
38(I)	$F_1 = \{20, 2\}$	$S_1 = \{20, 1\}$	101	0	0	0	0.000	0.000	0.00
Monotone Consistency									
38(I)	$F_1 = \{20, 2\}$	$S_1 = \{20, 1\}$	101	0	0	0	0.000	0.000	0.00
39(I)	$F_2 = \{20, 2, 1\}$	$S_2 = \{20, 1\}$	31	7	13	50	0.621	0.116	1.77
Robustness									
47(I)	$F_1 = \{59, 5, 1\}$	$S_1 = \{56, 5, 1\}$	101	0	0	0	0.000	0.000	0.00
48(I)	$F_2 = \{59, 5, 1\}$	$S_2 = \{55, 5, 1\}$	100	0	1	0	0.000	0.007	0.84
49(I)	$F_1 = \{59, 5, 1\}$	$S_3 = \{56, 55, 5, 1\}$	4	5	9	83	0.960	0.079	1.12

Entries under FF, FS, SF, and SS show the observed frequencies of each combination of choices on the two replications. The $\chi^2(1)$'s in the last column assess the fit of the true and error model. All are acceptable fits ($\alpha = .05$).

Results. The true and error model was applied to estimate the proportion of participants who truly preferred the second set in each choice. The results are shown in Table 8. The estimates indicate that 94.4% of all participants truly preferred singleton S_2 in choice 15, whereas only 3.9% chose the singleton in choice 16. These results entail a systematic violation of both Simple Uncertainty Appeal and Simple Uncertainty Aversion.

The next two rows in Table 8 show that for Simple Top Monotonicity as well as Simple Bottom Monotonicity, everyone is estimated to choose accordingly. In fact, even the observed frequencies comply almost perfectly with the axioms; only one out of 101 participants showed pattern F_1S_1 on the two presentations of Choice 45, all others choose F_1F_1 in both axiom tests.

Considering the estimated choice probabilities for choices 38 and 39, we can see that everybody chose set F_1 in the first choice and 62% were estimated to truly prefer set S_2 in the second choice. This means that 62% of all participants truly violated Monotone Consistency, since, according to this axiom, people who prefer F_1 to S_1, should also prefer the union of F_2 to S_2.

In the last three rows in Table 8 it is shown that all of the participants truly preferred the first set in the first two choices of the test for Robustness. Finally, 96% are estimated to prefer the S_3 in Choice 49, thereby violating Robustness.

Discussion. It has been shown that the Minmax and Maxmin rules, both members of the family of Min and Max induced rankings, are poor predictors of actual choice behavior. An empirical examination of the axiomatic foundations of these rules yielded that the only axioms withstanding the test are Simple Top Monotonicity and Simple Bottom Monotonicity.

Both Simple Uncertainty Aversion and Simple Uncertainty Appeal were violated on a large scale. People do not seem to be drawn to or repelled by uncertainty in the way prescribed by these axioms, which is rather restrictive. Since the certain outcome was the only difference between both tests, we can assume that it was the change in relative position of this certain outcome in relation to the outcomes in the binary set that made people change their preference from one test to the other. These findings dovetail with the results of our test of Neutrality in the previous section and imply that people's preferences over outcomes are not merely ordinal.

The choice patterns demonstrated by the majority of participants also disagreed with Monotone Consistency and Robustness. By allowing sets to have either their highest (cf. Monotone Consistency) or their lowest (cf. Robustness) outcomes in common, we were able to establish a reversal of the rankings of the average set values which might have caused the participants to violate these axioms' requirements. Still, in the test of Monotone Consistency, a rather large portion of individuals preferred $\{20, 2, 1\}$ over $\{20, 1\}$. This pair of sets and its estimated choice proportion closely resemble the second choice in the test of Top Independence[9]. In both cases, a significant number of participants opted for the three-element set, despite the fact that its average over the outcomes is much lower than in the binary set. As mentioned, these unexpected results match the implications of Richness Appeal (Arlegi, 2007)

[9]This is the choice between $\{35, 4, 3\}$ and $\{35, 3\}$, see section "Leximin & Leximax".

and might be reflective of a positive attitude towards a diversification of the possible outcomes within the range bounded by the minimum and maximum of the set.

Lexicographic Minmax & Maxmin

Axioms and Stimuli. The axioms for the characterization of the Lexicographic Minmax and the Lexicographic Maxmin rules as proposed by Bossert et al. (2000), as well as the stimuli used to verify them empirically are presented below.

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

$$xPy \Rightarrow \{x\} \succ \{x, y\} \succ \{y\}$$

Simple Monotonicity ensures that, if an outcome x is strictly preferred to another outcome y, then having x with certainty will be preferred to the uncertain prospect with the two possible outcomes x and y. The latter, in turn, will be preferred to having y with certainty. Obviously, Simple Monotonicity is fairly unquestionable in the context of choice under uncertainty. The first two rows in Table 9 contain the pairs of sets that were used to test Simple Monotonicity.

Simple Uncertainty Aversion: For all $x, y, z \in X$,

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

Simple Uncertainty Appeal: For all $x, y, z \in X$,

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

These axioms were already explained and tested in the previous section. The main results for the tests of Simple Uncertainty Aversion and Simple Uncertainty Appeal will be recaptured in the discussion

Type 1 Dominance For all $A \in \mathcal{X}$ and all $x, y \in X$,

$$[xPaPy \text{ for all } a \in A] \Rightarrow \{x, y\} \succ A \cup \{y\}.$$

Type 2 Dominance For all $A \in \mathcal{X}$ and all $x, y \in X$,

Table 9: Stimuli used to empirically validate the axioms characterizing Lexicographic Minmax and Lexicographic Maxmin (Bossert, Pattanaik & Xu, 2000).

Axiom tested	Choice	Set	
		First (F)	Second(S)
Simple Monotonicity	29∗	$F_1 = \{11\}$	$S_1 = \{11, 10\}$
	13	$F_2 = \{11, 10\}$	$S_2 = \{10\}$
Type 1 Dominance	14	$F_1 = \{29, 3\}$	$S_1 = \{28, 27, 3\}$
Type 2 Dominance	17	$F_1 = \{20, 3, 2\}$	$S_1 = \{20, 1\}$
Type 1 Extension Principle	9	$F_1 = \{30\}$	$S_1 = \{50, 17\}$
	10∗	$F_2 = \{50, 30, 17\}$	$S_2 = \{50, 17\}$
Type 2 Extension Principle	5∗	$F_1 = \{40, 15\}$	$S_1 = \{20\}$
	6	$F_2 = \{40, 15\}$	$S_2 = \{40, 20, 15\}$
Type 1 Monotonicity	21∗	$F_1 = \{27\}$	$S_1 = \{35, 19\}$
	22	$F_2 = \{27\}$	$S_2 = \{40, 19\}$
	23∗	$F_3 = \{27\}$	$S_3 = \{40, 35, 19\}$
Type 2 Monotonicity	9∗	$F_1 = \{50, 17\}$	$S_1 = \{30\}$
	27∗	$F_2 = \{50, 16\}$	$S_2 = \{30\}$
	28	$F_3 = \{50, 17, 16\}$	$S_3 = \{30\}$
Extension Independence	29∗	$F_1 = \{11\}$	$S_1 = \{11, 10\}$
	30	$F_2 = \{13, 11, 3\}$	$S_2 = \{13, 11, 10, 3\}$

All of these tests were administered in Experiment II

$*$ In the first replicate, choices marked with an asterisk were presented with set S in the first position, and set F in the second position. In the second replicate, these positions were reversed. Unmarked choices were arranged in the opposite fashion.

$$[xPaPy \text{ for all } a \in A] \Rightarrow A \cup \{x\} \succ \{x, y\}.$$

Type 1 Dominance ensures that if an outcome x is better than every outcome in a given set A, and y is worse than every outcome in A, then $\{x, y\}$ will be preferred to $A \cup \{y\}$. Analogously, if x and y are strictly better and strictly worse than every outcome in A, Type 2 Dominance ensures that $A \cup \{x\}$ is preferred to $\{x, y\}$. These conditions are stronger versions of Simple Top Monotonicity and Simple Bottom Monotonicity, which could not be refuted in the previous section. Yet, in contrast to the latter, Type 1 and Type 2 Dominance must also hold in the case where $A \cup \{y\}$ and $A \cup \{x\}$ contain more than two elements. Under those circumstances, however, it is possible to choose the "middle" elements in such a way that the average over the outcomes in the larger set is either higher or lower than in the binary set. Choices 14 and 17 in Table 9 were used to test Type 1 Dominance and Type 2 Dominance, respectively. In choice 14, the set average of $S_1 = \{28, 27, 3\}$ is higher than the average of the two outcomes in $F_1 = \{29, 3\}$, which might lead people to violate Type 1 Dominance and choose S_1. Similarly, choice 17 assesses Type 2 Dominance, but here $F_1 = \{20, 3, 2\}$ has the lower average. This might cause people to choose $S_1 = \{20, 1\}$, which would be a violation of Type 2 Dominance.

Type 1 Extension Principle For all $A \in \mathcal{X}$ and all $x, y \in X \backslash A$,

$$([\{a\} \succ \{x, y\} \text{ for all } a \in A] \text{ and } [A \succ \{x, y\}]) \Rightarrow A \cup \{x, y\} \succ \{x, y\}.$$

Type 2 Extension Principle For all $A \in \mathcal{X}$ and all $x, y \in X \backslash A$,

$$([\{x, y\} \succ \{a\} \text{ for all } a \in A] \text{ and } [\{x, y\} \succ A]) \Rightarrow \{x, y\} \succ A \cup \{x, y\}.$$

In a decision situation where outcomes x and y are outside of a given set A, and every distinct element a of A is, when received with certainty, preferred to the uncertain prospect $\{x, y\}$, and A itself is preferred to $\{x, y\}$ as well, Type 1 Extension Principle implies that $A \cup \{x, y\}$ will be preferred to $\{x, y\}$. Type 2 Extension Principle is the dual which applies to situations where A, as well as all distinct elements a of A received with certainty, are ranked as less attractive than $\{x, y\}$. In order to reduce the number of choices needed to test these axioms, we constrained the decision problem to the simple case where A is a singleton, $\{a\}$. In an attempt to produce violations of Type 1 Extension principle, we used choice stimuli for which the value of a is situated somewhere in between the values of x and y. The value of a needs to be sufficiently high so that not too many subjects prefer $\{x, y\}$ over a, because such choice patterns cannot be used to evaluate Type 1 Extension Principle. On the other hand, the value of a should not be too high, since, in order to obtain violations, we need $\{x, y\}$ to be preferred over $\{x, a, y\}$. The reverse holds for Type 2 Extension Principle. Several combinations of values for a, x and y were tested. Those which yielded the most violations for either axiom are presented as choices 9, 10, 5, and 6 Table 9.

Type 1 Monotonicity For all $x \in X$ and all $A, B \in \mathcal{X}$,

$$(\{x\} \succ A \text{ and } \{x\} \succ B) \Rightarrow \{x\} \succ A \cup B.$$

Type 2 Monotonicity For all $x \in X$ and all $A, B \in \mathcal{X}$,

$$(A \succ \{x\} \text{ and } B \succ \{x\}) \Rightarrow A \cup B \succ \{x\}.$$

Type 1 Monotonicity requires that if the certainty of having x is preferred to each of the sets A and B, then receiving x with certainty will also be preferred to the union of A and B. Type 2 Monotonicity is the dual of Type 1 Monotonicity.

In order to test Type 1 Monotonicity, A and B are both operationalized as binary sets which have their minimal element in common. That way, taking the union of both sets will result in a situation where the higher outcomes outweigh the lower one, as is shown in choices 21, 22, and 23 in Table 9. Suppose that a decision maker prefers the certain outcome, i.e., F_1 and F_2, in the first two choices. In this case, Type 1 Monotonicity requires that she will also prefer the certain outcome in the third choice. However, as $S_3 = \{40, 35, 19\}$ contains two outcomes which are higher than 27 EUR instead of just one, we might expect some people who safely opted for the certainty of receiving 27 EUR in choices 21 and 22 to switch to the uncertain prospect S_3 in the third choice.

In order to test Type 2 Monotonicity, A and B are again binary sets, now sharing their maximum. This results in a less favorable distribution of higher and lower outcomes if we take the union of both. Considering choices 24, 25, and 26 in Table 9, someone showing pattern $F_1F_2S_3$, would be contravening the implications of Type 2 Monotonicity.

Extension Independence For all $A, B \in \mathcal{X}$ and all $x, y \in X \backslash (A \cup B)$,

$$[xPzPy \text{ for all } z \in A \cup B] \Rightarrow [A \succsim B \Leftrightarrow A \cup \{x, y\} \succsim B \cup \{x, y\}].$$

Extension Independence requires that if every alternative in $A \cup B$ is worse than x and better than y, then the relative ranking of $A \cup \{x, y\}$ and $B \cup \{x, y\}$ is the same as the relative ranking of A and B.

Extension Independence was tested following a similar recipe as for Bottom Independence (see section "Leximin & Leximax"). First, as shown in choice 29, participants are asked to choose between a singleton, $F_1 = \{11\}$, and a binary set, $S_1 = \{11, 10\}$ of which the maximum equals the amount to be won if one chooses the singleton and the minimum is only slightly lower. Next, both sets are enlarged with one outcome which is considerably lower than the original ones and one which is just a little higher. In choice 30, we end up with two sets with the same minimum of which $S_3 = \{13, 11, 10, 3\}$ has one more high outcome and thus a higher set average than $F_3 = \{13, 11, 3\}$. With these stimuli we attempted to elicit pattern $F_1 S_2$, a violation of Extension Independence.

Results. Again, the true and error model (Birnbaum, 2004) was applied. For each choice, the error rates (e) and the true probability of preferring the second set (p) were estimated from the data in Table 10.

Table 10: Replication data used to estimate true probability and error rate for each choice.

Choice	Set		Replication Patterns				$\hat{p}$	$\hat{e}$	$\chi^2(1)$
	First (F)	Second(S)	FF	FS	SF	SS			
Simple Monotonicity									
29	$F_1 = \{11\}$	$S_1 = \{11, 10\}$	110	8	11	2	0.011	0.080	0.47
13	$F_2 = \{11, 10\}$	$S_2 = \{10\}$	128	1	1	1	0.008	0.008	0.00
Type 1 Dominance									
14	$F_1 = \{29, 3\}$	$S_1 = \{28, 27, 3\}$	11	11	18	91	0.910	0.130	1.67
Type 2 Dominance									
17	$F_1 = \{20, 3, 2\}$	$S_1 = \{20, 1\}$	83	8	15	25	0.225	0.101	2.09
Type 1 Extension Principle									
9	$F_1 = \{30\}$	$S_1 = \{50, 17\}$	75	16	15	25	0.237	0.137	0.03
10	$F_2 = \{50, 30, 17\}$	$S_2 = \{50, 17\}$	124	2	4	1	0.007	0.025	0.65
Type 2 Extension Principle									
5	$F_1 = \{40, 15\}$	$S_1 = \{20\}$	102	7	11	11	0.092	0.076	0.88
6	$F_2 = \{40, 15\}$	$S_2 = \{40, 20, 15\}$	3	9	6	113	0.978	0.062	0.60
Type 1 Monotonicity									
21	$F_1 = \{27\}$	$S_1 = \{35, 19\}$	60	18	16	37	0.373	0.153	0.12
22	$F_2 = \{27\}$	$S_2 = \{40, 19\}$	49	16	15	51	0.511	0.137	0.03
23	$F_3 = \{27\}$	$S_3 = \{40, 35, 19\}$	15	9	4	103	0.876	0.056	1.86
Type 2 Monotonicity									
24	$F_1 = \{15, 8\}$	$S_1 = \{10\}$	96	10	13	12	0.102	0.098	0.39
25	$F_2 = \{15, 7\}$	$S_2 = \{10\}$	90	9	9	23	0.200	0.074	0.00
26	$F_3 = \{15, 8, 7\}$	$S_3 = \{10\}$	70	18	14	29	0.281	0.143	0.50
Extension Independence									
29	$F_1 = \{11\}$	$S_1 = \{11, 10\}$	110	8	11	2	0.011	0.080	0.47
30	$F_2 = \{13, 11, 3\}$	$S_2 = \{13, 11, 10, 3\}$	5	14	7	105	0.964	0.092	2.28

Entries under FF, FS, SF, and SS show the observed frequencies of each combination of choices on the two replicates. The $\chi^2(1)$'s in the right-most column evaluate the fit of the true and error model. All are acceptable fits ($\alpha = .05$).

Matching the implications of Simple Monotonicity, we found that the great majority of participants chose the first option in both choices 29 and 13.

As regards choice 14, $p = .91$ according to the true and error model, indicating that the preferences of only 9% of the participants conform to Type 1 Dominance. A different result was found for Type 2 Dominance, where as much as 77.5% appears to comply with the axiom. We did not expect that one of the two counterparts would be systematically violated while the other is not, since the choice stimuli for both axioms were constructed following the same reasoning.

Both Type 1 and Type 2 Extension Principle require that the people who choose set F_1 in the first choice should also be choosing F_2 in the second choice of the respective tests. This seems to be the case for Type 1 Extension Principle, as almost everybody is estimated to prefer F_2 in choice 10. For Type 2 Extension Principle, however, the modal choice proportions imply pattern F_1S_2 for most people. A similar result is observed for Extension Independence where 98.9% of the participants is estimated to truly prefer the first set in choice 29, while only 3.6% prefers the first set in choice 30.

Type I Monotonicity implies that people who prefer F_1 and F_2 in choices 21 and 22, should prefer the first set in choice 23 as well. However, the estimated choice proportions suggest otherwise. Indeed, the majority chose F_1 in the first choice, but only 48.9% did so in the second choice and merely 12.4% chose the first set in the third choice. For Type 2 Monotonicity, modal choice proportions appear to be in line with the axiom, but whether the 71.9% who chose the first set in choice 26 also chose the first set in the two previous choices is a question that can only be addressed by analyzing the choice patterns for the sequence of three choices.

Simple Uncertainty Aversion and Simple Uncertainty Appeal, both tested in the section "Minmax & Maxmin", were violated by almost everyone which led us to conclude that besides ordinal information, also the relative position of elements in and between sets is taken into account. The effect of the relative position of outcomes is also nicely illustrated by the results for the first two choices of the tests for Type 1 and Type 2 Monotonicity, i.e., choice pairs 21-22 and 24-25. Only one outcome changes from the first to the second choice in both pairs. In both cases, the percentages of participants choosing the second set change in the corresponding direction. For example, in choice 25, the minimum is 1 EUR higher than in the preceding choice which should elicit a lower preference rate for the second set in choice 25 as compared to choice 24. Indeed, Table 10 shows that $\hat{p}$ increases from 10.2% to 20%.

Further analyses of the demonstrated response patterns for the axiom tests comprising two choices are presented in Table 11.

Table 11: Estimated true probabilities of each response pattern in the tests of Simple Monotonicity, Type 1 Extension Principle, Type 2 Extension Principle, and Extension Independence.

Choice	Set		Parameter Estimates					$\chi^2(2)$
	First (F)	Second (S)	$\hat{e}$	$\hat{p}(F_1F_2)$	$\hat{p}(F_1S_2)$	$\hat{p}(S_1F_2)$	$\hat{p}(S_1S_2)$	
Simple Monotonicity								
29	$F_1 = \{11\}$	$S_1 = \{11, 10\}$	0.080	**0.981**	0.009	0.011	0.000	0.19
13	$F_2 = \{11, 10\}$	$S_2 = \{10\}$	0.008					
Type 1 Extension Principle								
9	$F_1 = \{30\}$	$S_1 = \{50, 17\}$	0.142	**0.765**	0.010	0.226	0.000	0.34
10	$F_2 = \{50, 30, 17\}$	$S_2 = \{50, 17\}$	0.025					
Type 2 Extension Principle								
5	$F_1 = \{40, 15\}$	$S_1 = \{20\}$	0.076	**0.026**	0.890	0.000	0.085	0.28
6	$F_2 = \{40, 15\}$	$S_2 = \{40, 20, 15\}$	0.062					
Extension Independence								
29	$F_1 = \{11\}$	$S_1 = \{11, 10\}$	0.080	**0.042**	0.950	0.000	**0.014**	1.59
30	$F_2 = \{13, 11, 3\}$	$S_2 = \{13, 11, 10, 3\}$	0.092					

Tests of the true and error 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 the axiom on that line.

The first row shows that the estimated probability of having F_1F_2 as the true preference pattern for choices 29 and 13, as prescribed by Simple Monotonicity, equals 98.1%.

Not all response patterns are eligible for the evaluation of both types of Extension Principle. Only those with a true preference for F_1 in the first choice of the respective tests—i.e., F_1F_2 and F_1S_2—can be considered, which adds up to a total of 77.5% ($\hat{p}(F_1F_2)+\hat{p}(F_1S_2)$) of all responses for Type 1 Extension Principle and 91.6% for Type 2 Extension Principle. Considering Type 1 Extension Principle, 98.7% (0.765/0.775) of all participants who had a true preference for F_1 in choice 9, also chose F_2 in choice 10, thus complying with the axiom. Conversely, 91.6% (0.89/0.971) of the respondents choosing F_1 in the test for Type 2 Extension Principle violated the axiom by choosing S_2 in choice 6.

Considering the last two choices in Table 11, patterns F_1F_2 and S_1S_2 are consistent with Extension Independence. However, it is estimated that 95% of all participants truly switched from F_1 in Choice 29 to S_2 in Choice 30, hereby violating

Extension Independence.

Table 12 shows the analysis of the response patterns for the sequence of three choices in the test of Type 1 Monotonicity. For this analysis, we used the extended true and error model described in the section "Comparison of Min and Max induced rankings and UEU". By taking into account the complete sequence, we were able to determine whether the participants choosing F_1 and F_2 in the first two choices also opted for F_3 in the third choice, as prescribed by Type 1 Monotonicity. Of the 131 participants, 50.5% is estimated to have a true preference for the first set in choices 21 and 22 ($\hat{p}(F_1F_2F_3) + \hat{p}(F_1F_2S_3)$)). The great majority thereof, i.e., 71.9% (0.363/0.505), appears to be violating the axiom by demonstrating pattern $F_1F_2S_3$.

Table 12: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns in the test of Type 1 Monotonicity. Choices 21, 22, and 23 in Experiment II. $\chi^2(5) = 8.81$

Response	Observed frequencies				Estimated true
pattern	Rep 1	Rep 2	Both	One not both	probability
$F_1F_2F_3$	18	16	11	6	0.142
$F_1F_2S_3$	40	37	21	17.5	0.363
$F_1S_2F_3$	1	2	1	0.5	0.008
$F_1S_2S_3$	19	21	5	15	0.061
$S_1F_2F_3$	2	0	0	1	0.000
$S_1F_2S_3$	5	11	1	7	0.000
$S_1S_2F_3$	3	1	0	2	0.000
$S_1S_2S_3$	43	43	30	13	0.425
Total	131	131	69	62	1.000

Entries under " Rep 1", "Rep 2" and "Both" represent the number of participants who showed each pattern for the first presentation of these three choices (the first replicate), the second presentation (the second replicate), and both presentations, respectively. The column labeled "One not both" contains the number of people who showed that pattern on one of the two replicates but not both. The sum of the frequencies in the "One not both" and "Both" columns adds to the total number of participants so none of the response patterns is counted twice. Estimated error rates are 0.041, 0.107, and 0.176 for Choices 21, 22, and 23, respectively. Evaluation of the true and error model yields $\chi^2(5) = 4.52$, an acceptable fit.

The extended true and error model was also fit to the frequencies in Table 13. Here, a different pattern prevails. The percentage of participants that can be taken into account for the examination of Type 2 Monotonicity equals 79.6% ($\hat{p}(F_1F_2F_3) + \hat{p}(F_1F_2S_3)$)). The true preferences of 86.8% (0.691/0.796) of the people who chose the first set in the first two choices of the test agree with the axiom as they preferred the first set in the third choice as well, hence demonstrating pattern $F_1F_2F_3$. Merely 13.2% (0.105/0.796) are estimated to be violating Type 2 Monotonicity by demonstrating pattern $F_1F_2S_3$.

Table 13: Observed frequencies and estimated true probabilities for each of the 8 possible response patterns in the test of Type 2 Monotonicity. choices 24, 25, and 26 in Experiment II. $\chi^2(5) = 9.66$

Response	Observed frequencies				Estimated true
pattern	Rep 1	Rep 2	Both	One not both	probability
$F_1F_2F_3$	70	71	55	15.5	0.691
$F_1F_2S_3$	22	22	9	13	0.105
$F_1S_2F_3$	12	8	3	7	0.036
$F_1S_2S_3$	2	8	1	4	0.013
$S_1F_2F_3$	3	3	0	3	0.000
$S_1F_2S_3$	4	3	0	3.5	0.000
$S_1S_2F_3$	3	2	1	1.5	0.009
$S_1S_2S_3$	15	14	7	7.5	0.145
Total	131	131	76	55	1.000

Discussion. It has been demonstrated that Lexicographic Minmax and Maxmin do not perform well at all in predicting choices between sets of possible monetary outcomes. In order to get a better understanding of why these models fail to accommodate the preferences of real-life decision makers, the axioms characterizing these models, as proposed by Bossert et al. (2000), were tested one by one.

Simple Monotonicity had the lowest violation rate, almost everybody's choices complied with this axiom.

As demonstrated in the section "Minmax & Maxmin", Simple Uncertainty Aversion and Simple Uncertainty Appeal were both systematically violated, indicating that the strength of preference between outcomes is taken into consideration when making a decision.

Choice patterns for the vast majority of participants also disagreed with Extension Independence and Type 1 Dominance. For constructing the stimuli in these tests, we started once again from the assumption that people rank sets according to the average over their outcomes. This seems to have worked fine for Extension Independence as well as for Type 1 Dominance. However, in the test of Type 2 Dominance a much smaller (but not insignificant) portion of the participants demonstrated a violating choice pattern. An inconsistency arises that bears some resemblance with the ones observed in the tests of Bottom Independence, Top Independence and Monotone Consistency, namely, that it appears to be easier to provoke violations when the set with the highest average also has the highest number of elements than when this is not the case.

In the construction of the stimuli for the tests of Type 1 and 2 Extension Principle and Type 1 and 2 Monotonicity, two important difficulties arose. First, it was not possible to use stimuli that would evoke violations if people consistently choose the set with the highest average. In fact, in the tests of Type 1 and Type 2 Extension Principle, no specific strategy was followed in order to provoke violations; we just used a binary set and a singleton with an outcome situated somewhere in the middle between the ones in the binary set. Consequently, it rather came as a surprise that an extremely large portion of participants violated Type 2 Extension Principle. However, for Type 1 Extension Principle, violations were virtually nonexistent. Again, it is shown that, when confronted with the choice between a binary set and a set with the same elements enlarged with some extra outcome in the middle, people are inclined to choose the larger set, which perfectly fits the implications of the Richness Appeal axiom. In order to challenge Type 1 and Type 2 Monotonicity, the proportion of outcomes higher/lower than the certain outcome was manipulated in the uncertain prospect. This yielded a high violation rate for Type 1 Monotonicity, while the violation rate for Type 2 Monotonicity was rather low. A second difficulty in testing these four axioms is that for the choice patterns to be eligible for the evaluation of the axiom, one or two "conditional" preferences which were not so obvious to elicit had to be stated. Conversely, the other axiom tests in this paper consisting of more than one choice lend themselves to evaluation with stimulus pairs where the desired response in the "conditional" choice(s) would occur if people choose according to Simple Top or Bottom Monotonicity or Dominance, which was as good as always the case.

Notwithstanding these impediments, we were able to bring about substantial violation rates for one of the two counterparts in both axiom pairs. Knowing that any couple of two dual axioms is supposed to represent the same property, albeit in different directions, we are inclined to believe that the actual properties represented by these axioms do not hold in general and that it is a rather unfortunate choice of stimuli which is responsible for the lack of violations of Type 1 Extension Principle and Type 2 Monotonicity.

Summary and General Discussion

This paper presents an extensive experimental investigation of some decision rules for the ranking of sets of uncertain outcomes: Maximax, Maximin, Minmax, Maxmin and their lexicographic extensions Leximin, Leximax, Lexicographic Minmax, and Lexicographic Maxmin. These rules, which were grouped under the umbrella term "Min and Max induced rankings", have in common that they require preferences over sets to be induced from the comparison of the best and/or worst elements in those sets. The lexicographic extensions also permit the induction of preferences from comparisons of the second, third, or fourth best and/or worst elements if the previous ones coincide. Furthermore, Min and Max induced rankings imply that only ordinal information about the outcomes is taken into account by the decision maker.

This family of models stands in sharp contrast with the Uniform Expected Utility criterion (UEU) according to which preferences are not induced from the comparison of certain elements, but sets are ranked on the basis of the arithmetic average of the utilities of all their elements. By assuming a utility function, one allows that the strength of preference over the elements plays a role in establishing a set ranking.

In the first experimental section of this paper, it was shown that the Min and Max induced rankings perform particularly worse than UEU in predicting real-life decision makers' preferences over sets of monetary outcomes. When presented with a binary set and a three-element set with a higher average over the elements but a lower minimum and maximum than the binary set, the grand majority chose the set with the better average, contrary to the implications of the Min and Max induced rankings.

Notwithstanding the fact that these rankings were rather easily rejected, we decided to further investigate their axiomatic underpinnings. Finding out which axioms hold, which do not, and whether we can we group them in some way according to certain characteristics might facilitate the development of models that are more accurate in the descriptive sense. In fact, analysis of the axioms characterizing

Min and Max induced rankings yielded four large categories: Dominance and Monotonicity conditions, axioms ensuring that only ordinal information is used, axioms that prevent rankings to be based on "average-goodness", and a rest category with four axioms.

Very few violations were observed for Dominance and its weaker version Simple Monotonicity. Both these axioms rule out rankings of sets that are based on"total-goodness" criteria (Pattanaik & Peleg, 1984). Dominance, for example, implies that enlarging a set with an element which is worse than all the original elements will decrease the attractiveness of that set, although enlarging a set would always increase its "total worth". Since as good as every participant's choices complied with Dominance as well as Simple Monotonicity, we can conclude that in the current context, where sets represent uncertain prospects for which only one outcome can materialize, people do not evaluate sets by their total goodness.

We were unable to provoke even the slightest "true" violation of Simple Top Monotonicity and Simple Bottom Monotonicity. Both these monotonicity conditions ensure that replacing one set element with a better one yields a better set. Like Dominance and Simple Monotonicity, these axioms seem fairly unquestionable, which was confirmed by the results.

We tested three axioms that ensure that only ordinal information about the outcomes is considered by the decision makers: Neutrality, Simple Uncertainty Appeal, and Simple Uncertainty Aversion. All three of them could be refuted on a large scale. Although the Min and Max induced rankings are convenient for constructing rankings over sets from nothing more but the ordinal data of people's rankings over outcomes, we clearly demonstrated that the relative differences in the positions of monetary outcomes have a substantial effect on stated preferences. It is suspected that this finding also applies to other types of outcomes. For example, in the case of election candidates the decision maker does not only prefer some candidate X over another candidate Y, but, in many cases she also prefers X "to some extent" over Y. Thus, if one's goal is to develop a descriptively plausible model for the ranking of sets of possible outcomes, the strength of preference over set elements should be incorporated in that model.

Several axioms in the characterizations of Min and Max induced rankings prevent preferences to be based on "average-goodness" considerations as implied by UEU, i.e., calculating the arithmetic average of the utilities of the outcomes[10]. In order to test these axioms, stimuli were constructed in such a way that the axiom in question is violated if the subjects consistently choose the set with the higher average over the outcomes. This method yielded major violations—violation rates ranging from 52% to 89%—for all but one of the axioms in this group. For unknown reasons, the test of Type 2 Dominance produced only 22.5% violations. Although non-negligible, this violation rate is substantially lower than those resulting from the tests of the other axioms in this group. Possibly, a stronger refutation of Type 2 Dominance would be established if choices were administered with an even larger difference in the averages of the sets to be compared.

More difficult to interpret are the results for the remaining four axioms: Type 1 and Type 2 Extension Principle, and Type 1 and Type 2 Monotonicity. In these cases, it was not possible to construct stimuli in such a way that consistently choosing the set with the higher average would result in violations. Instead, in the test of Type 1 and Type 2 Monotonicity, we attempted to elicit violations by creating a shift in the proportions of high versus low outcomes, which seems to have worked well for one of the two counterparts, but not for the other. With regard to the Extension Principle axioms, no specific strategy was followed besides not allowing the average over the outcomes in the sets to be compared to differ too much. Nonetheless, just about everybody violated the Type 1 variant, while almost no violations were observed for the Type 2 variant. What is more, is that the subjects needed to state some less evident preferences in the first choices of each test in order for their preference patterns to be eligible for the evaluation of these axioms. Nevertheless, under these circumstances, violation rates were in both cases very high for one of the two counterparts. Since any couple of dual axioms is intended to represent the same underlying property in opposing directions, it is suspected that the underlying properties do not hold and that the lower violation rates for Type 1 Extension Principle and Type 2 Monotonicity are mainly due to a poor choice of stimuli.

[10] Axioms that prevent rankings to be based on average-goodness are: Bottom Independence, Top Independence, Disjoint Independence, Monotone Consistency, Robustness, Type 1 Dominance, Type 2 Dominance, and Extension Independence.

Experimental research in the domain of set ranking has hitherto given us some hints as to what requirements should be met by rankings that are intended to model actual choice behavior for a significant portion of people. First of all, we suspect that such a ranking has to be transitive (Vrijdags, 2010). Furthermore, it should meet the requirements of Dominance, Simple Top and Simple Bottom Monotonicity. It must also take into account the strength of preference between outcomes, and it should not prevent sets to be evaluated according to their average-goodness.

The UEU criterion appears to satisfy every single one of these requirements, yet some observations were made that cast serious doubt on the validity of UEU as a descriptive model.

Table 14: Choices for which a significant portion of the estimated true preferences seem to conflict with UEU

Choice	Experiment	Set		$\hat{p}$	$\hat{e}$	$\chi^2(1)$
		First	Second			
22	I	$F = \{35, 4, 3\}$	$S = \{35, 3\}$	0.536	0.227	1.39
39	I	$F = \{20, 2, 1\}$	$S = \{20, 1\}$	0.621	0.116	1.77
17	II	$F = \{20, 3, 2\}$	$S = \{20, 1\}$	0.225	0.101	2.09

Entries under $\hat{p}$ show the true probability of preferring the second set (S) in each choice, estimated with the true and error model. The error rates for these choices are listed under $\hat{e}$. The $\chi^2(1)$'s in the right-most column evaluate the fit of the true and error model, which is acceptable in all three cases ($\alpha = .05$).

In all three choices in Table 14, a substantial number of people truly preferred set F, the set with the lowest average. Such preferences are hard to explain with UEU, unless an extremely risk averse utility function is assumed for all subjects choosing the first set, which seems rather implausible. One way to avoid having to estimate utility functions is by verifying whether these people's preferences satisfy the behavioral axioms characterizing UEU. Therefore, an experimental study of these axioms will be the focus of our future research. This way, we hope to determine to what extent UEU applies.

A possible explanation for the results in Table 14 might be found along the lines of Arlegi's Richness Appeal axiom. (Arlegi, 2007) which states that any set of three outcomes is always considered better than the binary set containing only the minimum and maximum. Richness Appeal does not strictly apply to Choice 17, since these sets do not share their minimum. However, some stronger version of the axiom can be thought of that explains a preference for the three element set in this choice as well. The main underlying idea would be that some people appear to be drawn to a diversification of uncertainty in the form of larger sets, as long as this diversification does not go below the worst possible outcome. This unanticipated, yet interesting phenomenon appears to be fairly robust, similar choices in other tests reported in this paper endorse it, as well as the unreported data[11]. Considering the results for the test of Dominance, we can assume that this inclination towards larger sets is not the result of mistakenly evaluating sets by their total goodness.

In conclusion, the Min and Max induced rankings do a pretty bad job; they are clearly and easily rejected. Their main virtue is that their use does not require the elicitation of any parameter which is probably why they have been so popular in the literature, despite their low plausibility. In this paper, a descriptive attitude was adopted, a quite different perspective. For the modeling of observed set rankings, UEU seems to be a far more valuable candidate. However, for a considerable group of people we suspect that they do not use UEU, neither any of the Min and Max induced rankings investigated in this paper, but some other decision rule which might be characterized by a preference for a particular diversification of uncertainty.

References

Arlegi, R. (2001). *Rational evaluation of actions under complete uncertainty.* (mimeo - unpublished)

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.

Arlegi, R. (2007). Sequentially consistent rules of choice under complete uncertainty. *Journal of Economic Theory*, *135*, 131–143.

Arrow, K. J., & Hurwicz, L. (1972). An optimality criterion for decision-making under ignorance. In C. Carter & J. Ford (Eds.), *Uncertainty and expectations in economics: Essays in honour of G.L.S. Shackle* (pp. 1–11). Oxford: Basil Blackwell.

[11]Multiple tests were performed for each axiom, but only those with the highest number of violations are reported. An overview of all the data gathered in Experiment I and Experiment II can be found in Table 15.

Barberà, S., Bossert, W., & Pattanaik, P. K. (2004). Ranking sets of objects. In S. Barberà, P. Hammond, & C. Seidl (Eds.), *Handbook of utility theory. Volume 2 Extensions* (pp. 893–977). Dordrecht: Kluwer Academic Publishers.

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.

Birnbaum, M. H. (2008). New tests of cumulative prospect theory and the priority heuristic: Probability-outcome tradeoff with branch splitting. *Judgment and Decision Making*, *3*, 304–316.

Birnbaum, M. H., & Bahra, J. (2007). Gain-loss separability and coalescing in risky decision making. *Management Science*, *53*, 1016–1028.

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organizational Behavior and Human Decision Processes*, *104*, 96–112.

Birnbaum, M. H., & Schmidt, U. (2008). An experimental investigation of violations of transitivity in choice under uncertainty. *Journal of Risk and Uncertainty*, *37*, 77–91.

Birnbaum, M. H., & Schmidt, U. (2010). Testing transitivity in choice under risk. *Theory and Decision*, *69*, 599–617.

Bossert, W., Pattanaik, P. K., & Xu, Y. (2000). Choice under complete uncertainty: Axiomatic characterization of some decision rules. *Economic Theory*, *16*, 295–312.

Cohen, M., & Jaffray, J.-Y. (1980). Rational behavior under complete ignorance. *Econometrica*, *48*, 1281–1299.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Financial Studies*, *22*, 1915–1953.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–95). New York: Oxford University Press.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality : The adaptive toolbox.* Cambridge, Massachusetts: MIT Press.

Gravel, N., Marchant, T., & Sen, A. (2007, July). *Ranking completely uncertain decisions by the uniform expected utility criterion* (IDEP Working Papers No. 0705). Institut d'économie publique (IDEP), Marseille, France. Available from http://ideas.repec.org/p/iep/wpidep/0705.html

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*, 1251–1290.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*, 1291–1326.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions.* New York: Wiley.

Maskin, E. (1979). Decision-making under ignorance with implications for social choice. *Theory and Decision*, *11*, 319–337.

Milnor, J. (1954). Games against nature. In R. Thrall, C. Coombs, & R. Davis (Eds.), *Decision processes* (pp. 49–59). New York: Wiley.

Pattanaik, P. K., & Peleg, B. (1984). An axiomatic characterization of the lexicographic maximin extension of an ordering over a set to the power set. *Social Choice and Welfare*, *1*, 113–122.

Savage, J. L. (1956). *The foundations of statistics.* New York: Wiley.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19.

Sopher, B., & Gigliotti, G. (1993). Intransitive cycles: Rational choices or random error? An answer based on estimation of error rates with experimental data. *Theory and Decision*, *35*, 311–336.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Vrijdags, A. (2010). An experimental investigation of transitivity in set ranking. *Theory and Decision*, *68*, 213–232.

Table 15: Raw data for all choices in Experiment I and Experiment II.

No.	Set		Patterns		
	First	Second	FF	FS	SF
Experiment I					
1	{23, 20}	{20}	79	14	2
3	{28, 17}	{28, 17, 13}	83	6	8
4	{29, 3}	{28, 27, 2}	12	12	17
5	{37, 35, 27, 7, 2}	{35, 27, 7, 2}	91	4	3
7	{20}	{24, 7}	86	6	4
8	{20}	{24, 20, 7}	71	5	13
9	{24, 20, 7}	{24, 7}	90	3	4
10	{37, 9, 7}	{19}	19	10	13
11	{37, 9, 7}	{37, 19, 9, 7}	2	3	11
12	{37, 19, 9, 7}	{19}	25	12	16
13	{6, 5}	{7, 1}	90	4	4
14	{6, 5}	{7, 6, 5, 1}	71	15	9
15	{7, 6, 5, 1}	{7, 1}	90	4	3
16	{70, 20, 10}	{30, 25}	29	19	7
17	{70, 20, 10}	{70, 30, 25, 20, 10}	5	6	14
18	{70, 30, 25, 20, 10}	{30, 25}	42	27	5
16'	{94, 12}	{45, 43, 32}	18	15	9
17'	{94, 12}	{94, 45, 43, 32, 12}	4	11	8
18'	{94, 45, 43, 32, 12}	{45, 43, 32}	52	16	10
19	{20}	{20, 18}	85	8	6
20	{20, 1}	{20, 18, 1}	1	2	4
21	{4, 3}	{3}	99	1	1
22	{35, 4, 3}	{35, 3}	31	14	21
23	{20, 19, 18}	{20, 19, 18, 16}	83	8	4
24	{20, 19, 18, 2}	{20, 19, 18, 16, 2}	10	16	15
25	{25, 20}	{21}	78	13	5
26	{80, 75, 25, 20}	{80, 75, 21}	38	20	21
27	{62, 60}	{62, 59, 58}	83	8	6
28	{62, 60, 7, 2}	{62, 59, 58, 7, 2}	6	9	8
29	{100}	{95, 90}	94	3	4
30	{100, 10}	{95, 90, 10}	6	4	8
31	{5, 3}	{2}	96	3	2
32	{50, 45, 5, 3}	{50, 45, 2}	40	10	11
33	{40, 1}	{11, 10}	43	17	11
34	{40, 37, 25, 22, 1}	{37, 25, 22, 11, 10}	20	14	21
35	{67, 49, 45}	{70, 50, 2}	89	8	2
36	{67, 49, 45, 5, 1}	{70, 50, 5, 2, 1}	74	13	5
35'	{100, 9}	{29, 27}	45	15	11
36'	{100, 94, 93, 92, 9}	{94, 93, 92, 29, 27}	19	22	15
37	{15, 10, 9, 1}	{15, 9, 7, 1}	97	3	1
38	{20, 2}	{20, 1}	101	0	0
39	{20, 2, 1}	{20, 1}	31	7	13
40	{55, 10, 8}	{55, 9, 6}	98	2	1
41	{55, 10, 9, 8, 6}	{55, 9, 6}	42	12	16
42	{35, 10, 5}	{35, 10, 3}	96	4	1
43	{35, 10, 5, 3}	{35, 10, 3}	36	11	12
44	{29, 3}	{28, 3}	101	0	0
45	{29, 3}	{27, 3}	100	1	0
46	{29, 3}	{28, 27, 3}	8	7	16
47	{59, 5, 1}	{56, 5, 1}	101	0	0
48	{59, 5, 1}	{55, 5, 1}	100	0	1
49	{59, 5, 1}	{56, 55, 5, 1}	4	5	9
50	{99, 13, 10}	{95, 12, 10}	98	1	1
51	{99, 13, 10}	{97, 11, 10}	95	4	1
52	{99, 13, 10}	{97, 95, 12, 11, 10}	12	12	17
Experiment II					
1	{11}	{15, 10}	2	6	12
2	{15, 11, 10}	{15, 10}	72	9	31
3	{2}	{7, 0}	22	18	13
4	{7, 2, 0}	{7, 0}	78	20	15
5	{20}	{40, 15}	11	11	7
6	{40, 20, 15}	{40, 15}	113	9	6
7	{19, 9}	{14}	49	25	17
8	{19, 9}	{19, 14, 9}	3	4	2
9	{50, 17}	{30}	25	15	16
10	{50, 17}	{50, 30, 17}	1	4	2
11	{10, 4}	{7}	35	9	13
12	{10, 4}	{10, 7, 4}	3	9	4
13	{11, 10}	{10}	128	1	1
14	{29, 3}	{28, 27, 3}	11	11	18
15	{24}	{25, 15}	106	11	7
16	{25, 15}	{16}	117	7	2
17	{20, 1}	{20, 3, 2}	25	15	8
18	{10}	{13, 7}	51	25	16
19	{10}	{14, 7}	34	15	13
20	{10}	{14, 13, 7}	8	14	3
21	{27}	{35, 19}	60	18	16
22	{27}	{40, 19}	49	16	15
23	{27}	{40, 35, 19}	15	9	4
24	{15, 8}	{10}	96	10	13
25	{15, 7}	{10}	90	9	9
26	{15, 8, 7}	{10}	70	18	14
27	{50, 16}	{30}	25	15	6
28	{50, 17, 16}	{30}	15	10	8
29	{11}	{11, 10}	110	8	11
30	{13, 11, 3}	{13, 11, 10, 3}	5	14	7
31	{28, 17}	{28, 17, 13}	111	11	6
32	{37, 28, 17}	{28, 17}	125	1	4
33	{25}	{18, 2}	128	2	1

Entries under FF, FS, and SF show the observed frequencies of each combination of choices on the two replicates. Frequencies for SS are not given since they can be derived as follows: $SS = n - (FF + FS + SF)$ with $n = 101$ for Experiment I and $n = 131$ for Experiment II.