# RANKING VERSUS CHOICE IN THE ELICITATION OF PREFERENCES


## Abstract

This paper shows that preferences elicited via a ranking procedure differ systematically from those generated by pairwise choices, even though nothing in standard decision theory entails any such differences. The data strongly suggest that one of the best-known violations of expected utility (EU) theory – the 'common ratio effect' – may to a substantial extent be an artefact of the pairwise choice procedure. However, EU theory is not rehabilitated by ranking: other non-EU behaviour persists. Our findings raise serious doubts about the generality of conclusions drawn from laboratory and field studies that are based primarily on pairwise choice.

96 words



Keywords: Preference elicitation, choice, ranking, experiment

JEL Classification: C91, D81

Many experimental tests of individual decision theories involve what appear to be the most elementary decision tasks possible: namely, pairwise choices between lotteries involving no more than three payoffs with straightforward probabilities.

At first sight, there seem to be compelling arguments for running experiments in this way. It is relatively easy to explain the tasks to respondents and to link decisions to incentives; and it sounds plausible to argue that if a supposedly general theory fails in even the simplest cases, then we have good reason to doubt whether it can succeed in more complex situations.

However, on closer examination these arguments may not be so compelling after all. To begin with, such stripped-down problems may be overly susceptible to respondents focusing on one aspect of the problem in some cases and then switching attention to a different aspect in other cases. Moreover, for many areas of economic analysis, such simple tasks are not necessarily the most appropriate template. Often, decision-makers face larger choice sets between more complex alternatives. They do not necessarily arrive at their decisions by considering those alternatives two at a time, but rather by weighing up the relative merits of options which may be better or worse to varying degrees on a number of different dimensions. Of course, standard theory assumes that what holds for simple problems also holds for more complex ones; but the objective of the series of experiments reported in this paper was to test that assumption and examine whether a different task might result in significantly different patterns of behaviour.

The task in question was a ranking procedure. Respondents were presented with a larger set of options and were asked to order them from the one they most preferred down to the one that they would least like to have. Compared with pairwise choice, such a procedure is arguably closer to what decision-makers do in many real world contexts. For many important decisions – such as buying a house, purchasing consumer durable goods or choosing holiday packages – people are liable to consider a number of alternatives and balance various competing characteristics against one another. Other decisions, such as selecting portfolios for investment, or identifying projects to be prioritised in some programme of corporate or public expenditure, are rather more like ranking than they are like pairwise choice.

The main objective of this study was to examine whether certain non-standard ‘effects’ observed so frequently in pairwise choices are largely specific to that particular format, or whether they are manifested just as strongly in ranking responses. In a series of three experiments we found robust and highly significant differences between choice and ranking. Expected utility (EU) theory and most non-EU theories of choice under uncertainty postulate for each person a single preference ordering over lotteries which governs both choice and ranking, and so cannot explain systematic differences between choice and ranking. Our findings call for explanations in which responses to tasks are influenced by context-dependent judgments. We suggest that similarity theory (Ariel Rubinstein, 1988; Jonathan Leland, 1994) organises some, although not all, of our results.

However, whatever the reasons, our results raise serious concerns about the use of simple pairwise choice experiments to test the performance of decision theories with respect to more complex environments. More generally, these results may give grounds for reconsidering the robustness of discrete choice formats of the kind used in market research and in some of the studies that aim to elicit public preferences in areas such as health, safety and the environment (see, for example, Ian Bateman *et al.*, 2002). In short, the implications of these results for substantial tranches of experimental and applied research, as well as for the use of such research in guiding real-world decision making, may be profound.

**I. Experiment 1**

This experiment set out to investigate the 'common ratio' form of the Allais paradox (see Maurice Allais, 1953; Daniel Kahneman and Amos Tversky, 1979), a violation of EU's independence axiom which has proved to be extremely robust in pairwise choice experiments. The design investigated whether this effect was replicated when preferences were elicited through ranking.

Table 1 shows the two sets of binary lotteries used in this experiment. In all cases, the other consequence of each lottery was a zero payoff: i.e. lottery A offered a 0.5 chance of £25 and a 0.5 chance of zero; and so on. The expected values are shown in the EV column. The particular lotteries relevant to *this* paper[1] are shown in bold and given labels for ease of subsequent identification.

[1] Besides examining the relationship between ranking and choice, we were also interested in exploring the relationship between directly elicited certainty equivalent values and those inferred from ranking responses. To that end, a number of lotteries were included in the overall design which are not pertinent to the focus of the present paper: details of those other data can be found in **anonymised**.

TABLE 1 HERE

Each lottery was represented as shown in Figure 1 below (X does not appear in Table 1, but was the example used in the instructions to respondents):

FIGURE 1 HERE

It was explained that if the respondent ended up with lottery X, it would be played out as follows. The respondent would be presented with a bag containing 100 discs, each bearing a different number from 1 to 100 inclusive, and would be asked to dip their hand into the bag and pick one disc at random. If that disc bore a number between 1 and 65 inclusive, X would pay £12.50 in cash; if the disc showed a number between 66 and 100 inclusive, the respondent would go away with nothing. In all cases, the widths of the payoff columns were drawn so that they were proportional to the probabilities of receiving their respective payoffs.

Each respondent was asked to undertake four tasks, on the understanding that there was an equal chance that any one of the four could turn out to provide the basis of their payment for participating in the experiment. The order in which these tasks were presented was varied to control for any 'order effects'.

One of these tasks, not relevant to this paper (see footnote 1), was to state a 'certainty equivalent' valuation for eight of the other lotteries. Another task involved making twelve pairwise choices. Each choice was displayed as in Figure 2 below:

FIGURE 2 HERE

In the event that a respondent's payment was to depend on this task, each of the twelve choices was equally likely to be selected at random as the basis for payment. Of the twelve, eight are pertinent to this paper: four involving lotteries from Set 1 – {£9 for sure versus A – illustrated above}, {£9, B}, {£6, C}, {£6, D} – and four involving lotteries from Set 2 – {G, E}, {G, F}, {H, F} and {H, G}. Notice that each of the pairs in the second group of four is a 'scaled-down' version of its counterpart in the first group: so, for example, the probabilities of receiving £9 or £25 offered by G and E are, respectively, one-fifth of the probabilities offered by the options of £9 for sure or A; the 'scale factors' in the other three cases are one-fifth, one-quarter and one-quarter.

These questions were included to test for the 'common ratio effect' (CRE). The essence of the CRE is as follows. While EU does not predict whether a respondent will prefer lottery A or £9 for sure, the independence axiom *does* entail that scaling down the probabilities of their respective positive payoffs by the same factor should leave the direction of preference unaffected: that is, someone who prefers A over £9 should also prefer E (= 0.2(A) + 0.8(0)) over G (= 0.2(£9) + 0.8(0)), and vice-versa.

There are two ways of violating this implication of independence: either to choose the safer option in the first pair (i.e. £9 for sure) and then the riskier option (E) from the scaled-down pair: or else to choose the riskier of the two scaled-up options (in this case, A) together with the safer one (G) in the second pair. While both forms of violation might occur occasionally as the result of 'noise' or 'error', the classic CRE finding is that the combination of £9 and E is generally observed very much more frequently than the combination of A and G, with the asymmetry between the frequencies of the two forms of violation generally being deemed as (highly) statistically significant. An important element in the experimental design was to see whether the usual asymmetry was replicated by our respondents in the pairwise choice task, and then examine how far the same patterns were observed in the ranking task.

The third and fourth tasks were both ranking exercises. In addition to the ten lotteries from Table 1, ten sure amounts – £2, £3, £4, £5, £6, £7, £8, £9, £10 and £12 – were added to each Set, making twenty options to be ranked on each occasion[2]. Each lottery or sure amount was represented on a separate strip of card, and in each exercise respondents were asked to arrange the twenty strips from most preferred to least (with no ties allowed) and then record their ranking. In the event that a respondent's payout was to be based on one of the ranking exercises, the incentive system was that two of the twenty strips would be selected at random and the respondent would play out whichever of the two she had ranked higher[3].

[2] The purpose of including these sure amounts was to be able to compare the 'certainty equivalents' stated in the valuation task mentioned earlier with values for the same lotteries inferred from their positions in the rankings – again, see footnote 1.

[3] Thus the incentive mechanism was analogous to the one used for the pairwise choice task, thereby providing as level a playing field as possible for comparing the data generated by the two types of task.

I.A. Results from Experiment 1

Table 2 reports the relevant data[4]. The first two columns show the lotteries involved. For each case, the upper row reports the choice data while the lower row shows the preferences inferred from ranking. Columns 3-6 show the various combinations of choosing the safer (S) and riskier (R) alternatives. So those in the SS and RR columns were consistent with EU, while those in the SR column violated EU in the manner typical of CRE by choosing the safer alternative from the scaled-up pair but the riskier alternative from the scaled-down pair. The RS column reports the numbers who violated EU in the opposite direction. The last column shows whether the asymmetry between SR and RS was statistically significant. The test being used is an exact binomial test, with the null hypothesis being that the two forms of violation are equally likely to occur. This is the null hypothesis most commonly used in previous studies of the CRE.

TABLE 2 HERE

What Table 2 shows is that when the standard pairwise choice design was used, the CRE appeared strongly in all four cases – the asymmetry between SR and RS being statistically significant at the 1% level in three cases and at the 5% level in the other one. By contrast, when preferences were inferred from the ranking exercises, the asymmetries disappeared: there were still a number of violations, but in *these* responses the asymmetry was not statistically significant. Given the prevalence of CRE anomalies reported in the literature, this appears to constitute a substantial impact.

An alternative perspective on the data is provided in Table 3, which shows how far respondents' preferences over any pair of alternatives were consistent across choice and ranking. The fourth column headed $S_C R_R$ shows the numbers of individuals who opted for the safer alternative in the choice (hence S subscripted C) while preferring the riskier alternative in the ranking exercise (hence R subscripted R). The frequency of the opposite discrepancy, $R_C S_R$, is shown in the fifth column. The third and sixth columns show the cases where stated preferences were the same in both tasks. The null hypothesis is that there was no significant asymmetry between

[4] In this table, we report the data only from those respondents who answered both the choice *and* the ranking questions: 154 for the first case, 153 for the other three cases.

$S_C R_R$ and $R_C S_R$. Taking an agnostic position about the direction of an asymmetry, the measure of statistical significance is based on a two-sided exact binomial test.

TABLE 3 HERE

Table 3 shows that for all four scaled-down pairs, ranking reduced the propensity to favour the riskier options to a highly significant extent. For the scaled-up lotteries, the picture was more mixed. In two cases, there were no systematic differences, while in a third there was a difference in the same direction as with the scaled-down lotteries, albeit to a rather weaker extent. However, in the case of {£6, D} there was a significant difference in the opposite direction, with the ranking responses markedly *less* risk averse than the pairwise choices. Overall, then, ranking produced orderings which were rather less extreme than those generated by choices, with the net result that the CRE was removed in all four cases.

Given the robustness of the CRE in many previous experiments conducted during the previous three decades, this was a striking result. Indeed, it contrasted so greatly with the bulk of previous evidence that it was important to see whether it could be reproduced with different parameters. That was the purpose of the second experiment.

## II. Experiment 2

In this experiment there were four sets of 11 lotteries, as shown in Table 4 below. For each ranking task, the same 14 sure amounts were added (these being every whole pound value from £2 to £15 inclusive). Thus for each ranking exercise, the options were presented on 25 separate strips which respondents were asked to arrange in order of preference. Each respondent undertook three ranking exercises: which three of the four sets were presented to a particular individual, and the order in which they were presented, was determined at random.

TABLE 4 HERE

Two sets of pairwise choices were sandwiched between the ranking exercises. These two choice tasks each involved making 10 pairwise choices. Once a respondent had completed all ranking and choice tasks, one task was picked at random and then the same incentive mechanism was used as in the first experiment. As before, the lotteries which are relevant to this paper are shown in Table 4 in bold type and are given labels.

II.A. Results from Experiment 2

Within each Set, we can compare the patterns of preference inferred from respondents' ranking responses with their direct pairwise choices. There are, in total, 22 such comparisons, presented in Table 5 below using the same column notation as in Table 3. Since most pairs appeared in more than one Set, the Set is shown in the third column.

TABLE 5 HERE

The overall pattern was very much like that found in Experiment 1. The top ten rows show the pairings of scaled-up lotteries. There was little in the way of systematic differences between choice and ranking among those pairs, the only exception being {N, M}, where ranking gave a significantly *less* risk averse distribution. By contrast, among the scaled-down lotteries shown in the lower twelve rows, ranking elicited orderings which were very much *more* risk averse than those produced by pairwise choices. The impact of this on the CRE is shown in Table 6, which reports the data in the same format as in Table 2 above.

TABLE 6 HERE

The CRE comparison involves three scaled-up pairs {£15, J}, {£12, K} and {M, K} and their scaled-down counterparts, {R, Q}, {S, P} and {P, M}. A slight complication when reporting the results in Table 6 is that although each individual only made one choice per pair, they typically saw some pairs in two separate ranking exercises. For example, an individual randomly assigned to rank Sets 1, 2 and 3 (or 1, 3 and 4) saw £15, £12, J and K in the same set on two occasions, but their scaled-down counterparts P, Q, N and M only once; whereas those asked to rank Sets 1, 2, and 4 (or 2, 3 and 4) saw those scaled-up lotteries all together only once, but encountered their scaled-down counterparts together on two occasions. And in the case of L and K, a quarter of the sample – those assigned to rank Sets 2, 3 and 4 – did not see them in the same ranking exercise at all. So we have totals of 148 for the first two comparisons, and 112 for the third[5].

As in the first experiment, when the standard pairwise choice design was used, the CRE appeared very strongly ($p < 0.001$) in all three cases. And although in this second experiment it did not disappear *entirely* from the ranking data – the asymmetry between SR and RS is still significant at the 5% level in the case of {£15, J} & {P, N} – it became insignificant in the other two cases, with the main reason being the different preferences elicited by ranking for the scaled-down pairs.

## III. Reflections on the Results from Experiments 1 and 2

For EU and many non-EU theories which assume transitivity and procedural invariance, there is no reason to expect the kinds of differences between choice and ranking responses observed in the data reported above. However, it might be that the data from Experiments 1 and 2 can, to some extent at least, be explained in terms of similarity theory (ST), as proposed by Rubinstein (1988) and elaborated by Leland (1994).

Rubinstein conjectured that when individuals are choosing between pairs of lotteries that are perceived to be dissimilar on both dimensions – i.e. payoffs and probabilities – they are most likely to behave as risk averse EU maximisers[6]. However, if the lotteries are perceived as dissimilar on one dimension but similar on the other, the dissimilar dimension receives greater weight and/or becomes decisive; and this can result in CRE departures from EU.

From this perspective, the scaled-up lotteries can be regarded as being dissimilar on both dimensions: the positive payoffs are clearly different, and the probability of a positive payoff offered by the safer alternative is, in the cases reported in Tables 2 and 6, between 0.2 and 0.5 greater than the corresponding probability in the riskier alternative. However, as the probabilities are scaled down, the differences between them reduce: these are either 0.05 or 0.1 in all the scaled-down pairs in Tables 2 and 6. ST proposes that it is not (only) the ratios of the probabilities that respondents are concerned with, but rather (or also) the simple differences, which are

[5] Whenever the same pair was ranked twice by an individual, the inferred choice on each occasion is given a weight of 0.5. Individuals who ranked the two alternatives in the same order on both occasions therefore register as a single observation in one of the four columns, whereas individuals who ranked the alternatives differently on each occasion register as 0.5 in each of two different columns.

[6] As will become apparent later, we do not necessarily accept that EU is the 'default' position.

easier to process cognitively. And although there is no absolute and general threshold at which respondents switch from perceiving a pair as dissimilar to perceiving a pair as similar, it is clearly possible that some proportion of respondents will consider a difference of 0.1 to be below that threshold, and that an even larger proportion will regard 0.05 to be too small a difference to warrant much weight.

In this respect, ST appears to organise our CRE choice data very successfully. In Experiment 1, when the pairs of lotteries became more similar, the majority of those who chose S in the scaled-up pairs switched to R. Moreover, in the two cases where the difference between probabilities was just 0.05 – that is, {G, F} and {H, G} – the proportion who switched was markedly bigger than in the other two cases, where the difference was 0.1 – an average of 79% as compared with 64% for {G, E} and {H, F}. Much the same pattern was repeated in Experiment 2, with the majority of those who chose S in the scaled-up pairs switching to R when the probabilities became more similar: on average, 57% switched in the two cases where the difference was 0.1, and 84% in the case where the difference was 0.05.

But if ST is at work, why should the ranking task change behaviour in the way observed? One possible reason is that similarity judgments in pairwise choices are liable to lead to violations of transitivity in the scaled-down pairs, whereas ranking does not allow such violations and therefore produces a different ordering over at least one pair. To see how this works, consider three lotteries such as F, G and H from Experiment 1. When the three pairwise choices {H, G}, {G, F} and {H, F} are presented separately, similarity judgments can operate to different degrees on each choice in turn. In particular, such judgments are liable to produce a preference for G over H in the first pair and a preference for F over G in the second pair, since the probabilities differ by only 0.05 in these cases. By transitivity, this would entail a preference for F over H. However, when a respondent is presented separately with that pair, where the probabilities are less similar, she may give greater weight to risk aversion and choose the safer option H, thereby violating transitivity.

Such cycles – which we shall designate RRS because they involve choosing the riskier option in the two more similar pairs and the safer option in the less similar pair – are an implication of ST when applied to pairwise choices. But of course they cannot occur in a ranking task: ranking imposes transitivity over the set of lotteries under consideration, so that at least one, and possibly more, of the pairwise choice orderings must be reversed.

For this 'conflict resolution' account of the difference between choice and ranking to carry conviction, we should expect to see some evidence that the pairwise choices did indeed exhibit the intransitive patterns implied by ST. And in fact, such patterns are clearly evident. For the {H, G, F} triple, there were two ways in which intransitive cycles could occur: either the RRS cycle – F $\succ$ G, G $\succ$ H and H $\succ$ F – consistent with ST; or else the opposite cycle H $\succ$ G, G $\succ$ F and F $\succ$ H, which we designate SSR and which is inconsistent with ST. Out of a total of 159 respondents who answered all three pairwise choices, 22 violated transitivity in one direction or the other. However, the null hypothesis that the two cycles occurred by chance with equal frequency can be confidently rejected: there were 5 cycles of the SSR type, but 17 of the RRS type – an asymmetry which, if the null hypothesis were true, would occur with probability less than 0.01.

Such cycles are reminiscent of those reported by Amos Tversky (1969), who also accounted for them in terms of similarity judgments[7]. However, that triple was the only one in Experiment 1 which gave any scope for such cycles. By contrast, in Experiment 2 respondents were asked to make choices between all six pairwise combinations of P, Q, R and S. The numbers of cycles in each direction for each triple are shown in Table 7 below.

TABLE 7 HERE

Of course, the four rows are not independent; but even the least asymmetric of the four rejects the null hypothesis at the 0.1% significance level. The preponderance of RRS cycles over SSR cycles reinforces the result from Experiment 1 and is consistent with the idea that significant similarity effects were at work in the pairwise choice tasks, and that those judgments – and hence the CRE patterns associated with them – were greatly attenuated by the ranking task. Indeed, what actually appeared to happen, by and large, was that the degree of risk aversion involved in choices between

[7] Both Rubinstein (1988) and Leland (especially 2002) acknowledge Tversky's early (1969) identification of the potential role of similarity judgments as contributing to intransitive choice patterns. However, Tversky's lottery-related data were generated by just eight respondents, who had themselves been selected from an initial sample of eighteen on the basis of a 'screening' procedure which established a predisposition to violate transitivity in that particular direction. So there was a question of how robust such cycles would be in samples typical of those used in many experiments; and we can find no subsequent body of evidence of such cycles: whether this means that researchers tried to find such evidence but failed, or whether no such attempts were made, we cannot say.

scaled-up lotteries was also applied to those lotteries in the ranking exercise; and that when scaled-down lotteries were being ranked, much the same balance between payoffs and probabilities was struck for them too.

It is possible that the ranking task encourages respondents to take a more holistic view of each lottery. In pairwise choices – especially ones with only two non-zero payoffs and their respective probabilities to attend to – a similarity heuristic is readily available. By contrast, the ranking task presents respondents with a variety of payoffs and with probabilities drawn from all parts of the 0-1 range. Perhaps this, together with the fact that ranking imposes transitivity, overrides – or at least, greatly attenuates – any influence of similarity judgments.

To explore the issues further, we organised a third experiment, with three main objectives. First, we needed to check the robustness of the intransitive choice patterns among scaled-down pairs: the finding of such strong patterns in an unscreened sample was a striking result, and it was important to know whether it would replicate. Second, we wished to see how far similarity effects applied to other pairs of lotteries not considered so far – and if they did, whether they too were attenuated by ranking. Third, we wanted to know whether a departure from EU that did ***not*** appear to be due to ST would also be affected by ranking.

## IV. Experiment 3

Table 8 shows the two Sets of fourteen lotteries used in this experiment.

TABLE 8 HERE

Each respondent undertook two ranking exercises and made two series of twenty pairwise choices (with choice and ranking tasks alternating, and with the order varied from one session to the next to achieve an overall balance). For each ranking exercise, six sure amounts – £12, £10, £8, £6, £4 and £2 – were added. The incentive mechanisms were the same as in the previous two experiments.

The twenty pairwise choices relating to Set 1 were as follows[8]. There were six choices between every pairwise combination of £10, A, B and D, and another six between their scaled-down counterparts J, K, L and Z, thereby providing six CRE

[8] The twenty choices relating to Set 2 involved £10 for sure and the corresponding lotteries as shown in Table 8.

cases and opportunities to look for cycles between those pairs. The other eight choices were concerned with testing for betweenness.

To see what was involved here, consider the Marschak-Machina (M-M) triangle diagram in Figure 3. Since Mark Machina (1982), this diagram has frequently been used to depict lotteries involving different probability distributions over the same three payoffs. The vertical edge of the triangle shows the probability of the highest payoffs (in this case, £20) and the horizontal edge shows the probability of the lowest payoff (here, 0). Any residual is the probability of the intermediate payoff (£10). Thus, for example, E = (£20, 0.45; £10, 0.25; 0, 0.3) is located at 0.45 on the vertical axis and 0.3 on the horizontal. Exactly the same diagram can be drawn for the other set of cases where the high payoff was £25, except that N, E, F, G, T and U would be replaced, respectively, by B, Q, R, S, H and I.

FIGURE 3 HERE

David Buschena and David Zilberman (1999) suggested that when all pairs of lotteries are transformations of some base pair such as {£10, N}, then the distances between alternatives in the M-M triangle would be primary indicators of similarity. They qualified this suggestion with the conjecture that if one alternative but not the other involved certainty or quasi-certainty, this would cause the pair to be perceived as less similar, while if both alternatives had the same support, they would be regarded as more similar.

On this basis, the eight pairs drawn from Figure 3 were, in (weak) order from least to most similar on each line, and with the safer alternative listed first in each pair: {£10, E}, £10, F}, {£10, G}, {F, E}, {G, F}; and {Y, T}, {U, T}, {Y, U}. The corresponding sequence for the other lotteries involving a high payoff of £25 rather than £20 is: {£10, Q}, £10, R}, {£10, S}, {R, Q}, {S, R}; and {Y, H}, {I, H}, {Y, I}. Thus in addition to testing for the existence of cycles between the scaled-down pairs identified above, it is also possible to see whether there is a tendency for other variants of the two 'base pairs' {£10, N} and {£10, B} to exhibit similarity effects.

In particular, there is the potential for 'betweenness cycles'. Betweenness is a property of EU that imposes regularity on all linear combinations of any pair of lotteries. For example, consider an EU maximiser who prefers N over £10 for sure, and then identify any mixture of those two lotteries lying on the straight line joining them – F, for example. This mixture of N and £10 should not only be preferred to £10 but to any other linear combination to the south-west of F on that line; but it should be less preferred than any mixture to the north-east of it on that line. Thus if N is strictly preferred to £10, we should expect N $\succ$ E $\succ$ F $\succ$ G $\succ$ £10. Alternatively, if an EU maximiser prefers £10 to N, he should exhibit the opposite set of preferences for the lotteries lying on that line: that is, £10 $\succ$ G $\succ$ F $\succ$ E $\succ$ N.

However, if similarity effects lead to a propensity to choose the riskier alternative more often in more similar pairs, these orderings could be disrupted. For example, suppose that similarity favours F in {G, F} and G in {£10, G}, but that because £10 and F are less similar, the safer option £10 is more often favoured in *that* pairwise comparison. This could result in betweenness cycles in the direction F $\succ$ G, G $\succ$ £10, £10 $\succ$ F, but not in the opposite direction: that is, RRS cycles but not SSR cycles.

At this point, a distinction needs to be made between such cycles, which may be an implication of the similarity heuristic, and another violation of betweenness which does not involve intransitivity: namely, a preference for a linear combination of two lotteries over *both* of the two constituent lotteries: for example, a preference for G in {£10, G} *and* a preference for G in {G, F}. *This* form of violation of betweenness is not implied by similarity as set out above, although other non-EU models such as the 'rank dependent expected utility' (RDEU) class of models[9] do entail convex rather than linear indifference loci in the triangle, with the implication that some mixtures may be preferred to both of their constituent lotteries. Such models are transitive and the convex indifference curves they entail are not implied by ST, so that any attenuation of similarity judgments in the ranking task may not affect such violations of betweenness.

IV.A. Results from Experiment 3: Intransitivity

A total of 100 respondents took part in Experiment 3 (although one respondent missed a page in the Set 2 choice booklet, so there were only 99 responses for {£10,

[9] An early form of this type of model was John Quiggin's (1982) 'anticipated utility' theory. Subsequently, Chris Starmer and Robert Sugden (1989) proposed a form which incorporated a reference point and allowed for losses being treated differently from gains. Essentially the same basic idea is at the heart of Amos Tversky and Daniel Kahneman's (1992) 'cumulative prospect theory'. Other variants have also been proposed. What is common to all of them is the notion that probabilities are transformed into decision weights in a manner which can accommodate violations of independence and betweenness while respecting transitivity and stochastic dominance.

P} and {U, T}). We begin by considering the extent to which cyclical choices were made, as shown in Table 9. There is a contrast between the triples involving scaled-down lotteries and those involving betweenness lotteries.

TABLE 9 HERE

For the scaled-down lotteries, the patterns are much like those found in Experiments 1 and 2, with the null hypothesis that both kinds of cycles are equally likely being rejected in favour of the alternative hypothesis consistent with ST in all eight cases (seven times at the 1% level and once at the 5% level).

For the betweenness lotteries, there is *some* asymmetry in the direction consistent with ST, but it appears to be rather weaker and is largely concentrated in triples {Y, U, T} and {Y, I, H}. This patchy picture is somewhat reminiscent of earlier findings. For example, Buschena and Zilberman (1999) looked for such cycles, and found some evidence suggesting them in one area of the triangle, but no significant evidence in another area. It may be that the various mixtures were not similar enough to elicit strong effects: adjacent points on the lines in Figure 3 were still separated by differences of 0.15 in the probabilities of the high payoff, whereas the effects reported in Tables 2 and 6 involved probability differences of 0.05 or 0.10 at most. But for whatever reason, the evidence of betweenness cycles in this experiment was mixed.

IV.B. Results from Experiment 3: The Common Ratio Effect

The intransitive patterns among the scaled-down lotteries are consistent with the existence of similarity judgments. The extent to which ranking attenuated such judgments and reduced the strength of the CRE can be seen in Tables 10 and 11, which show the comparisons for Sets 1 and 2, starting with the three cases where the difference between the scaled-down probabilities was 0.05 (and where similarity effects might therefore be expected to be strongest), followed by the two cases where the difference was 0.10, with the last case being the one where the difference was 0.15.

TABLES 10 AND 11 HERE

In the six cases (three in each table) where ST suggests that the CRE is likely to be strongest, the asymmetries in choices were all significant at the 1% level. Ranking did not fully eliminate all of the asymmetries, but once again it did substantially attenuate them: one remained significant at the 1% level and two at the 5% level, but two were no longer significant and one actually became significant in the opposite direction.

Of the four cases involving scaled-down probability differences of 0.1, there were two cases where ranking reduced the total number of violations by between 30% and 40% but left the degree of asymmetry much the same. In the other two cases, there was actually some tendency in the choice data towards asymmetries in the *un*usual direction, and in the Set 2 case this registered as statistically significant. Here, ranking toned down these asymmetries to the point where neither was significant. Finally, in the two cases where the scaled-down probability differences were 0.15, there was no significant asymmetry in the choice cases, and the distributions for choice and ranking were quite alike.

Overall, then, there was strong evidence of the CRE in the choices where the probabilities were most similar, and it was among these cases that ranking produced the greatest reductions in the asymmetries. Even so, some CRE patterns persisted in the ranking data – these arguably being of more consistent strength across the various pairs. So although similarity effects may be a major source of CRE patterns, they might not be the only contributory factor.

IV.C. Results from Experiment 3: Betweenness

Table 12 reports the relevant data. Because ranking imposes transitivity, we have in each case omitted those individuals who violated transitivity in the relevant choice questions, as reported in Table 9. Thus Table 12 is organised as follows. For each triple there are six possible orderings. Those listed first and second in each section respect betweenness. The third and fourth violate betweenness by having the mixture *less* preferred than its two constituent lotteries, while the fifth and sixth orderings have the mixture *more* preferred.

TABLE 12 HERE

With one exception, slightly more than half of the transitive responses to each triple respected betweenness. However, of the large minority who did not, there was a clear asymmetry between those whose orderings placed the mixture *above* the two constituent lotteries and those whose orderings placed it *below* them both, with the former significantly outweighing the latter in four cases out of six in the choice task. However, unlike our findings for the CRE, ranking did not attenuate these asymmetries. On the contrary, they were at least as pronounced under ranking as under choice in the four cases above, while in the other two cases – {£10, G, F} and {£10, S, R} – insignificant asymmetries in the choice data became highly significant in the ranking data, so that all six cases manifested rank orderings consistent with convex rather than linear indifference loci in the Marschak-Machina triangle.

IV.D. Results of Experiment 3: Choice vs Ranking

Table 13 groups the twenty pairwise choices according to whether they involved scaled-up lotteries, their scaled-down counterparts, or whether they were included to examine for violations of betweenness and betweenness cycles.

TABLE 13 HERE

As in Experiments 1 and 2, there was relatively little systematic difference between choice and ranking in the scaled-up pairs: in nine of those twelve pairs there was no significant asymmetry between $S_C R_R$ and $R_C S_R$. In two of the three cases where the asymmetry *was* significant – {£10, D and {£10, P} – we once again see the effect operating to produce significantly *less* risk aversion in ranking than in choice[10]. And although we cannot conduct the standard test on the totals, since these were not all generated by different individuals, the overall picture is that choice and ranking produce broadly similar patterns in the scaled-up cases.

The picture for the scaled-down lotteries was markedly different: while there was no particular overall trend for the pairs in the first three rows of those sections, where the probability differences were 0.1 or more, five of the six cases where the

[10] Notice that, just as in Experiment 1, these differences occur in comparisons between a certainty and a lottery involving a 0.8 chance of a better payoff but a 0.2 chance of zero. One possible interpretation is that choices in these cases manifested what Kahneman and Tversky (1979) referred to as a 'certainty effect' – that is, a strong attraction towards certainty when the EVs are not very different, so that the zero represents a relatively large loss compared with the possible gain offered by the lottery. Although the evidence is limited, we speculate that this effect may also be moderated by ranking.

probability difference was 0.05 showed a significant asymmetry favouring $R_CS_R$ over $S_CR_R$: that is, in the cases most susceptible to similarity effects in choice, the ranking task shifted the balance significantly towards the safer options. So while the totals show much the same numbers of $S_CR_R$ cases as found among the six scaled-up pairs, the numbers of $R_CS_R$ cases were about 70% higher for both Sets.

There were also a number of significant effects among the betweenness pairs, all of them in the direction of greater risk aversion in ranking than in choice. Table 13 shows that the effects tended to be strongest among the pairs with the shortest distances between them in Figure 3 – that is, in the bottom five rows of the table – and weakest among the more dissimilar pairs (in the top three rows of those sections). Since there was rather less evidence of intransitive choice among these lotteries, the 'conflict resolution' explanation has less force here. Perhaps it is that ranking encourages more holistic judgments. But note that since ST cannot account for the violations of betweenness, and since ranking tends, if anything, to produce even stronger violations, such holistic judgments do not conform with EU.

## V. Discussion

What the three experiments reported in this paper demonstrate beyond any doubt is that there are certain patterns of pairwise choice among standard experimental stimuli which do not generalise to larger choice sets. The preferences elicited via ranking diverged significantly from those expressed through pairwise choices, with ranking producing substantially more risk averse orderings over scaled-down lotteries. It appeared that the kind of similarity judgment effects posited by Rubinstein (1988) were prevalent, especially in choices where the differences in probabilities of positive payoffs were 0.05, but that these effects were greatly attenuated when respondents were asked to consider such lotteries as part of larger choice sets.

Even so, CRE patterns did not disappear *altogether* under ranking. Moreover, when violations of betweenness were examined, the patterns of violation were, if anything, *more* pronounced in the ranking data than in the pairwise choice responses.

These results, taken together, appear to have two implications. First, even though similarity judgments may be implicated in the best-known and arguably most influential violation of EU in the form of the CRE, as well as in striking patterns of intransitive choice, trying to control for such effects by using a ranking procedure will not necessarily result in EU-like behaviour.

Second, the results caution against the assumption that data from simple choice tasks can be reliably used to draw more general inferences about the structure of preferences and/or estimate preference functions intended for more general application. Besides its significance for theory testing, this finding also has potentially important implications for a substantial body of applied work which uses discrete choice methods in market research or in order to inform public policy about people's values and preferences in areas such as health, safety and environmental protection.

It has been known for some time that the preferences inferred from rankings over larger sets of options are liable to differ considerably from those inferred from smaller choices sets – see, for example, Jerry Hausman and Paul Ruud (1987) and Moshe Ben-Akiva *et al*. (1991). However, there has been an inclination to suppose that it is the ranking data that are defective, due to the greater complexity and heavier cognitive loads supposedly involved in ranking. Thus studies involving discrete choices from two or three options at a time continue to be the norm. By contrast, our results suggest – for lotteries, at least – that it may be choices that are more vulnerable to 'effects' and biases, and that ranking may give a more consistent picture (albeit one which diverges systematically from EU). Whether these conclusions carry over to studies concerned with predicting demand for new products or eliciting values for health, safety and the environment benefits is something that remains to be thoroughly investigated; but in view of the importance of those areas of economic activity, our results make a strong case for such investigations to be given a high priority.

## References


Allais, Maurice, (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axioms de l'ecole Americaine. *Econometrica*, **21**, 503-46.

Bateman, Ian, Carson, Richard, Day, Brett, Hanemann, Michael, Hanley, Nick, Hett, Tannis, Jones-Lee, Michael, Loomes, G., Mourato, Susana, Ozdemiroglu, Ece, Pearce, David, Sugden, Robert and Swanson, John, (2002). *Economic Valuation with Stated Preference Techniques: A Manual*. Edward Elgar.

Ben-Akiva, Moshe, Morikawa, Takayuki, and Shiroishi, Fumiaki, (1991). Analysis of the reliability of preference ranking data. *Journal of Business Research*, **23**, 253-68.

Buschena, David and Zilberman, David, (1999). Testing the Effects of Similarity on Risky Choice: Implications for Violations of Expected Utility. *Theory and Decision*, **46**, 251-76.

Hausman, Jerry, and Ruud, Paul, (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, **34**, 83-104.

Kahneman, Daniel, and Tversky, Amos, (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, **47**, 263-91.

Leland, Jonathan, (1994). Generalized similarity judgments: an alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, **9**, 151-72.

Machina, Mark, (1982). 'Expected Utility' Theory Without the Independence Axiom. *Econometrica*, **50**, 277-323.

Quiggin, John, (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, **3**, 323-43.

Rubinstein, Ariel, (1988). Similarity and decision making under risk: is there a utility theory resolution to the Allais paradox? *Journal of Economic Theory*, **46**, 145-53.

Starmer, Chris and Sugden, Robert, (1989). Violations of the independence axiom in common ratio problems: an experimental test of some competing hypotheses. *Annals of Operations Research*, **19**, 79-102.

Tversky, Amos and Kahneman, Daniel, (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**, 297-323.

| **Table 1: The Lotteries in Experiment 1** | | | | | | |
|---|---|---|---|---|---|---|
| Set 1 | | | | Set 2 | | |
| | | EV | | | | EV |
| | | | | | | |
| **A** | **0.5 x £25** | **£12.50** | | **E** | **0.1 x £25** | **£2.50** |
| **B** | **0.75 x £15** | **£11.25** | | **F** | **0.15 x £15** | **£2.25** |
| **C** | **0.6 x £15** | **£9.00** | | **G** | **0.2 x £9** | **£1.80** |
| **D** | **0.8 x £9** | **£7.20** | | **H** | **0.25 x £6** | **£1.50** |
| | | | | | | |
| | 0.7 x £12 | £8.40 | | | 0.1 x £60 | £6.00 |
| | 0.85 x £10 | £8.50 | | | 0.15 x £45 | £6.75 |
| | 0.9 x £9 | £8.10 | | | 0.2 x £35 | £7.00 |
| | 0.95 x £8 | £7.60 | | | 0.3 x £25 | £7.50 |
| | | | | | | |
| | 0.4 x £18 | £7.20 | | | 0.4 x £18 | £7.20 |
| | 0.55 x £13 | £7.15 | | | 0.55 x £13 | £7.15 |

| Table 2: The CRE in Choice and Ranking in Experiment 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | SS | SR | RS | RR | Stat sig |
| | | | | | | | |
| {£9, A} | 1 x 9 vs 0.50 x 25 | Choice | 17 | 39 | 9 | 89 | *** |
| & {G, E} | 0.20 x 9 vs 0.10 x 25 | Ranking | 40 | 30 | 23 | 61 | |
| | | | | | | | |
| {£9, B} | 1 x 9 vs 0.75 x 15 | Choice | 13 | 55 | 10 | 75 | *** |
| & {G, F} | 0.20 x 9 vs 0.15 x 15 | Ranking | 32 | 29 | 27 | 65 | |
| | | | | | | | |
| {£6, C} | 1 x 6 vs 0.60 x 15 | Choice | 21 | 30 | 17 | 85 | ** |
| & {H, F} | 0.25 x 6 vs 0.15 x 15 | Ranking | 31 | 22 | 31 | 69 | |
| | | | | | | | |
| {£6, D} | 1 x 6 vs 0.80 x 9 | Choice | 18 | 62 | 13 | 60 | *** |
| & {H, G} | 0.25 x 6 vs 0.20 x 9 | Ranking | 15 | 27 | 37 | 74 | |

*** and ** denote significance at 1% and 5% levels respectively

| Table 3: Within-Subject Choice vs Ranking Patterns in Experiment 1 | | | | | | |
|---|---|---|---|---|---|---|
| | | $S_CS_R$ | $S_CR_R$ | $R_CS_R$ | $R_CR_R$ | Stat sig |
| **Scaled-Up Pairs** | | | | | | |
| {£9, A} | 1 x 9 vs 0.50 x 25 | 42 | 14 | 28 | 70 | ** |
| {£9, B} | 1 x 9 vs 0.75 x 15 | 43 | 25 | 18 | 67 | |
| {£6, C} | 1 x 6 vs 0.60 x 15 | 34 | 17 | 19 | 83 | |
| {£6, D} | 1 x 6 vs 0.80 x 9 | 26 | 54 | 16 | 57 | ‡‡‡ |
| | | | | | | |
| **Scaled-Down Pairs** | | $S_CS_R$ | $S_CR_R$ | $R_CS_R$ | $R_CR_R$ | Stat sig |
| {G, E} | 0.20 x 9 vs 0.10 x 25 | 16 | 10 | 47 | 81 | *** |
| {G, F} | 0.20 x 9 vs 0.15 x 15 | 15 | 8 | 44 | 86 | *** |
| {H, F} | 0.25 x 6 vs 0.15 x 15 | 26 | 12 | 36 | 79 | *** |
| {H, G} | 0.25 x 6 vs 0.20 x 9 | 17 | 14 | 35 | 87 | *** |

***, ** and * denote significance at 1% and 5% and 10% levels respectively;
‡‡‡ indicates significance at 1% in opposite direction to the asterisked cases

| Table 4: The Lotteries in Experiment 2 | | | | | |
|---|---|---|---|---|---|
| Label | Set 1 | Set 2 | Set 3 | Set 4 | EV |
| | | | | | |
| **J** | **0.5 x £35** | | **0.5 x £35** | | **£17.50** |
| | 0.9 x £18 | | | 0.9 x £18 | £16.20 |
| | | 0.25 x £60 | 0.25 x £60 | | £15 |
| **K** | **0.6 x £25** | | **0.6 x £25** | | **£15** |
| **L** | **0.75 x £20** | 0.75 x £20 | **0.75 x £20** | 0.75 x £20 | **£15** |
| | 0.85 x £16 | | 0.85 x £16 | | £13.60 |
| | | 0.35 x £35 | 0.35 x £35 | | £12.25 |
| **M** | **0.8 x £15** | | | **0.8 x £15** | **£12** |
| **N** | **0.9 x £12** | | **0.9 x £12** | | **£10.80** |
| | 0.5 x £20 | 0.5 x £20 | 0.5 x £20 | 0.5 x £20 | £10 |
| | | 0.2 x £40 | 0.2 x £40 | | £8 |
| | | 0.15 x £50 | 0.15 x £50 | | £7.50 |
| | 0.7 x £10 | | | 0.7 x £10 | £7 |
| | 0.75 x £8 | | | 0.75 x £8 | £6 |
| | 0.25 x £20 | 0.25 x £20 | 0.25 x £20 | 0.25 x £20 | £5 |
| **P** | | **0.15 x £25** | | **0.15 x £25** | **£3.75** |
| **Q** | | **0.1 x £35** | | **0.1 x £35** | **£3.50** |
| **R** | | **0.2 x £15** | | **0.2 x £15** | **£3** |
| **S** | | **0.25 x £12** | | **0.25 x £12** | **£3** |

| Table 5: Within Subject Choice vs Ranking Patterns in Experiment 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | $S_CS_R$ | $S_CR_R$ | $R_CS_R$ | $R_CR_R$ | Stat sig |
| **Pairs** | | | | | | | |
| Pair | Parameters | Set | | | | | |
| {£15, J} | 1 x 15 vs 0.50 x 35 | 1 | 57 | 16 | 12 | 28 | |
| {£15, J} | 1 x 15 vs 0.50 x 35 | 3 | 56 | 10 | 19 | 25 | |
| {£12, K} | 1 x 12 vs 0.60 x 25 | 1 | 49 | 18 | 9 | 37 | |
| {£12, K} | 1 x 12 vs 0.60 x 25 | 3 | 49 | 12 | 12 | 37 | |
| {M, K} | 0.80 x 15 vs 0.60 x 25 | 1 | 37 | 19 | 13 | 43 | |
| {N, L} | 0.90 x 12 vs 0.75 x 20 | 1 | 23 | 12 | 8 | 70 | |
| {N, L} | 0.90 x 12 vs 0.75 x 20 | 3 | 22 | 9 | 13 | 66 | |
| {L, K} | 0.75 x 20 vs 0.60 x 25 | 1 | 74 | 15 | 13 | 11 | |
| {L, K} | 0.75 x 20 vs 0.60 x 25 | 3 | 73 | 16 | 15 | 6 | |
| {N, M} | 0.90 x 12 vs 0.80 x 15 | 1 | 29 | 25 | 12 | 47 | ‡‡ |
| {S, Q} | 0.25 x 12 vs 0.10 x 35 | 2 | 47 | 20 | 19 | 24 | |
| {S, Q} | 0.25 x 12 vs 0.10 x 35 | 4 | 41 | 21 | 14 | 34 | |
| {S, P} | 0.25 x 12 vs 0.15 x 25 | 2 | 26 | 20 | 29 | 37 | |
| {S, P} | 0.25 x 12 vs 0.15 x 25 | 4 | 26 | 16 | 28 | 42 | * |
| {R, Q} | 0.20 x 15 vs 0.10 x 35 | 2 | 41 | 4 | 23 | 43 | *** |
| {R, Q} | 0.20 x 15 vs 0.10 x 35 | 4 | 29 | 11 | 23 | 47 | * |
| {R, P} | 0.20 x 15 vs 0.15 x 25 | 2 | 11 | 2 | 41 | 57 | *** |
| {R, P} | 0.20 x 15 vs 0.15 x 25 | 4 | 6 | 4 | 34 | 66 | *** |
| {P, Q} | 0.15 x 25 vs 0.10 x 35 | 2 | 44 | 9 | 28 | 30 | *** |
| {P, Q} | 0.15 x 25 vs 0.10 x 35 | 4 | 30 | 21 | 26 | 33 | |
| {S, R} | 0.25 x 12 vs 0.20 x 15 | 2 | 33 | 21 | 21 | 36 | |
| {S, R} | 0.25 x 12 vs 0.20 x 15 | 4 | 35 | 12 | 31 | 32 | *** |

***, ** and * denote significance at 1% and 5% and 10% levels respectively;
‡‡ denotes significance at 5% in the opposite direction to all the asterisked cases.

| Table 6: CRE Patterns from Choice and Ranking in Experiment 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | SS | SR | RS | RR | Stat sig |
| | | | | | | | |
| {£15, J} | 1 x 15 vs 0.50 x 35 | Choice | 40 | 50 | 20 | 38 | *** |
| & {R, Q} | 0.20 x 15 vs 0.10 x 35 | Ranking | 66.5 | 26.5 | 12 | 43 | ** |
| | | | | | | | |
| {£12, K} | 1 x 12 vs 0.60 x 25 | Choice | 34 | 49 | 20 | 45 | *** |
| & {S, P} | 0.25 x 12 vs 0.15 x 25 | Ranking | 48.5 | 28.5 | 23 | 48 | |
| | | | | | | | |
| {M, K} | 0.80 x 15 vs 0.60 x 25 | Choice | 9 | 47 | 3 | 53 | *** |
| & {R, P} | 0.20 x 15 vs 0.15 x 25 | Ranking | 30.5 | 19.5 | 21 | 41 | |

*** and ** denote significance at 1% and 5% levels respectively

| Table 7: Intransitive Patterns in Experiment 2 | | | |
|---|---|---|---|
| Triple | SSR | RRS | Stat Sig |
| | | | |
| {P, Q, R} | 3 | 18 | *** |
| {P, Q, S} | 3 | 20 | *** |
| {P, R, S} | 1 | 20 | *** |
| {Q, R, S} | 4 | 20 | *** |

*** denotes significance at 1% level

| Table 8: The Lotteries in Experiment 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| Set 1 | | | | Set 2 | | | |
| | | EV | | | | | EV |
| A | 0.4 x £40 | £16.00 | | M | 0.4 x £30 | | £12.00 |
| B | 0.6 x £25 | £15.00 | | N | 0.6 x £20 | | £12.00 |
| C | 0.5 x £25 | £12.50 | | O | 0.5 x £20 | | £10.00 |
| D | 0.8 x £15 | £12.00 | | P | 0.8 x £12.50 | | £10.00 |
| E | 0.45 x £20; 0.25 x £10 | £11.50 | | Q | 0.45 x £25; 0.25 x £10 | | £13.75 |
| F | 0.3 x £20; 0.5 x £10 | £11.00 | | R | 0.3 x £25; 0.5 x £10 | | £12.50 |
| G | 0.15 x £20; 0.75 x £10 | £10.50 | | S | 0.15 x £25; 0.75 x £10 | | £11.25 |
| H | 0.3 x £25 | £7.50 | | T | 0.3 x £20 | | £6.00 |
| I | 0.15 x £25; 0.25 x £10 | £6.25 | | U | 0.15 x £20; 0.25 x £10 | | £5.50 |
| Y | 0.5 x £10 | £5.00 | | Y | 0.5 x £10 | | £5.00 |
| J | 0.1 x £40 | £4.00 | | V | 0.1 x £30 | | £3.00 |
| K | 0.15 x £25 | £3.75 | | W | 0.15 x £20 | | £3.00 |
| L | 0.2 x £15 | £3.00 | | X | 0.2 x £12.50 | | £2.50 |
| Z | 0.25 x £10 | £2.50 | | Z | 0.25 x £10 | | £2.50 |

| Table 9: Intransitivity in Scaled-down and Betweenness Lotteries | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Set 1 Scaled-down Lotteries | | | | | Set 2 Scaled-down Lotteries | | | |
| Triples | SSR | RRS | Stat Sig | | Triples | SSR | RRS | Stat Sig |
| | | | | | | | | |
| {Z, L, K} | 8 | 22 | *** | | {Z, X, W} | 1 | 16 | *** |
| {Z, L, J } | 1 | 18 | *** | | {Z, X, V} | 1 | 10 | *** |
| {Z, K, J} | 3 | 18 | *** | | {Z, W, V} | 3 | 13 | ** |
| {L, K, J} | 3 | 14 | *** | | {X, W, V} | 2 | 17 | *** |
| | | | | | | | | |
| {£20, £10, 0} Betweenness Lotteries | | | | | {£25, £10, 0} Betweenness Lotteries | | | |
| Triples | SSR | RRS | Stat Sig | | Triples | SSR | RRS | Stat Sig |
| | | | | | | | | |
| {£10, F, E} | 5 | 4 | | | {£10, R, Q} | 1 | 5 | * |
| {£10, G, F} | 5 | 4 | | | {£10, S, R} | 4 | 6 | |
| {Y, U, T} | 0 | 13 | *** | | {Y, I, H} | 4 | 9 | * |

***, ** and * denote significance at 1%, 5% and 10% levels respectively

| Table 10: CRE Patterns in Experiment 3, Set 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | SS | SR | RS | RR | Stat sig |
| | | | | | | | |
| {£10, D} | 1 x 10 vs 0.80 x 15 | Choice | 23 | 37 | 16 | 24 | *** |
| & {Z, L} | 0.25 x 10 vs 0.20 x 15 | Ranking | 33 | 8 | 25 | 34 | ‡‡‡ |
| | | | | | | | |
| {D, B} | 0.80 x 15 vs 0.60 x 25 | Choice | 36 | 36 | 4 | 24 | *** |
| & {L, K} | 0.20 x 15 vs 0.15 x 25 | Ranking | 44 | 23 | 9 | 24 | ** |
| | | | | | | | |
| {B, A} | 0.60 x 25 vs 0.40 x 40 | Choice | 27 | 38 | 4 | 31 | *** |
| & {K, J} | 0.15 x 25 vs 0.10 x 40 | Ranking | 42 | 28 | 11 | 19 | *** |
| | | | | | | | |
| {£10, B} | 1 x 10 vs 0.60 x 25 | Choice | 29 | 21 | 27 | 23 | |
| & {Z, K} | 0.25 x 10 vs 0.15 x 25 | Ranking | 43 | 14 | 16 | 27 | |
| | | | | | | | |
| {D, A} | 0.80 x 15 vs 0.40 x 40 | Choice | 34 | 29 | 11 | 26 | *** |
| & {L, J} | 0.20 x 15 vs 0.10 x 40 | Ranking | 49 | 23 | 5 | 23 | *** |
| | | | | | | | |
| {£10, A} | 1 x 10 vs 0.40 x 40 | Choice | 43 | 13 | 21 | 23 | |
| & {Z, J} | 0.25 x 10 vs 0.10 x 40 | Ranking | 47 | 14 | 11 | 28 | |

*** and ** denote significance at 1% and 5% levels respectively;
‡‡‡ indicates significance at 1% in opposite direction to the one usually observed

| Table 11: CRE Patterns in Experiment 3, Set 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | SS | SR | RS | RR | Stat sig |
| | | | | | | | |
| {£10, P} | 1 x 10 vs 0.80 x 12.50 | Choice | 61 | 29 | 5 | 4 | *** |
| & {Z, X} | 0.25 x 10 vs 0.20 x 12.50 | Ranking | 54 | 15 | 22 | 9 | |
| | | | | | | | |
| {P, N} | 0.80 x 12.50 vs 0.60 x 20 | Choice | 13 | 44 | 8 | 35 | *** |
| & {X, W} | 0.20 x 12.50 vs 0.15 x 20 | Ranking | 42 | 14 | 7 | 37 | |
| | | | | | | | |
| {N, M} | 0.60 x 20 vs 0.40 x 30 | Choice | 47 | 33 | 8 | 12 | *** |
| & {W, V} | 0.15 x 20 vs 0.10 x 30 | Ranking | 59 | 23 | 10 | 8 | ** |
| | | | | | | | |
| {£10, N} | 1 x 10 vs 0.60 x 20 | Choice | 46 | 12 | 27 | 15 | ‡‡ |
| & {Z, W} | 0.25 x 10 vs 0.15 x 20 | Ranking | 50 | 14 | 16 | 20 | |
| | | | | | | | |
| {P, M} | 0.80 x 12.50 vs 0.40 x 30 | Choice | 43 | 26 | 14 | 17 | ** |
| & {X, V} | 0.20 x 12.50 vs 0.10 x 30 | Ranking | 56 | 17 | 7 | 20 | ** |
| | | | | | | | |
| {£10, M} | 1 x 10 vs 0.40 x 30 | Choice | 62 | 10 | 16 | 12 | |
| & {Z, V} | 0.25 x 10 vs 0.10 x 30 | Ranking | 58 | 13 | 12 | 17 | |

*** and ** denote significance at 1% and 5% levels respectively

‡‡ denotes significance at 5% in opposite direction to the one usually observed

| Table 12: Orderings Over Betweenness Triples | | | | | | |
|---|---|---|---|---|---|---|
| {£20, £10, 0} Set | | | | {£25, £10, 0} Set | | |
| Ordering | Choice | Ranking | | Ordering | Choice | Ranking |
| | | | | | | |
| £10 > F > E | 29 | 28 | | £10 > R > Q | 18 | 33 |
| E > F > £10 | 22 | 18 | | Q > R > £10 | 32 | 18 |
| £10 > E > F | 4 | 7 | | £10 > Q > R | 12 | 4 |
| E > £10 > F | 4 | 3 | | Q > £10 > R | 1 | 3 |
| F > £10 > E | 9 | 15 | | R > £10 > Q | 9 | 11 |
| F > E > £10 | 23 | 20 | | R > Q > £10 | 22 | 25 |
| | | | | | | |
| £10 > G > F | 19 | 27 | | £10 > S > R | 18 | 24 |
| F > G > £10 | 34 | 31 | | R > S > £10 | 44 | 33 |
| £10 > F > G | 12 | 6 | | £10 > R > S | 5 | 4 |
| F > £10 > G | 6 | 2 | | R > £10 > S | 8 | 4 |
| G > £10 > F | 7 | 6 | | S > £10 > R | 3 | 7 |
| G > F > £10 | 13 | 19 | | S > R > £10 | 12 | 18 |
| | | | | | | |
| Y > U > T | 34 | 47 | | Y > I > H | 30 | 40 |
| T > U > Y | 4 | 6 | | H > I > Y | 16 | 5 |
| Y > T > U | 8 | 4 | | Y > H > I | 4 | 9 |
| T > Y > U | 6 | 4 | | H > Y > I | 2 | 3 |
| U > Y > T | 24 | 19 | | I > Y > H | 23 | 13 |
| U > T > Y | 10 | 6 | | I > H > Y | 12 | 17 |

**Table 13: Within-Subject Choice vs Ranking Patterns in Experiment 3**

| Pairs | $S_CS_R$ | $S_CR_R$ | $R_CS_R$ | $R_CR_R$ | Stat sig | Pairs | $S_CS_R$ | $S_CR_R$ | $R_CS_R$ | $R_CR_R$ | Stat sig |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Scaled-Up Lotteries, Set 1** | | | | | | **Scaled-Up Lotteries, Set 2** | | | | | |
| {£10, A} | 46 | 10 | 15 | 29 | | {£10, M} | 61 | 11 | 10 | 18 | |
| {£10, B} | 39 | 11 | 18 | 32 | | {£10, N} | 49 | 9 | 15 | 27 | |
| {D, A} | 56 | 7 | 16 | 21 | * | {P, M} | 60 | 9 | 13 | 18 | |
| {D, B} | 57 | 15 | 10 | 18 | | {P, N} | 40 | 17 | 16 | 27 | |
| {£10, D} | 33 | 27 | 8 | 32 | ‡‡‡ | {£10, P} | 64 | 26 | 4 | 5 | ‡‡‡ |
| {B, A} | 51 | 14 | 19 | 16 | | {N, M} | 72 | 8 | 10 | 10 | |
| Total reversals | | 84 | 86 | | | Total reversals | | 80 | 68 | | |
| **Scaled-Down Lotteries, Set 1** | | | | | | **Scaled-Down Lotteries, Set 2** | | | | | |
| {Z, J} | 43 | 21 | 15 | 21 | | {Z, V} | 60 | 18 | 10 | 12 | |
| {Z, K} | 39 | 17 | 20 | 24 | | {Z, W} | 57 | 16 | 9 | 18 | |
| {L, J} | 30 | 15 | 24 | 31 | | {X, V} | 44 | 13 | 19 | 24 | |
| {L, K} | 26 | 14 | 27 | 33 | * | {X, W} | 16 | 5 | 33 | 46 | *** |
| {Z, L} | 30 | 9 | 28 | 33 | *** | {Z, X} | 56 | 11 | 20 | 13 | |
| {K, J} | 23 | 8 | 30 | 39 | *** | {W, V} | 44 | 11 | 25 | 20 | ** |
| Total reversals | | 84 | 144 | | | Total reversals | | 74 | 116 | | |
| **Betweenness Lotteries in {20, 10, 0} Triangle** | | | | | | **Betweenness Lotteries in {25, 10, 0} Triangle** | | | | | |
| {£10, E} | 37 | 9 | 19 | 35 | * | {£10, Q} | 35 | 9 | 16 | 40 | |
| {£10, F} | 26 | 16 | 16 | 42 | | {£10, R} | 26 | 6 | 16 | 52 | * |
| {Y, T} | 72 | 8 | 10 | 10 | | {Y, H} | 59 | 7 | 16 | 18 | * |
| {£10, G} | 25 | 17 | 12 | 46 | | {£10, S} | 22 | 13 | 15 | 50 | |
| {G, F} | 32 | 12 | 26 | 30 | ** | {S, R} | 28 | 9 | 31 | 32 | *** |
| {F, E} | 54 | 12 | 17 | 17 | | {R, Q} | 41 | 9 | 32 | 18 | *** |
| {U, T} | 60 | 8 | 23 | 8 | ** | {I, H} | 58 | 11 | 22 | 9 | * |
| {Y, U} | 40 | 9 | 22 | 29 | ** | {Y, I} | 30 | 10 | 31 | 29 | *** |
| Total reversals | | 91 | 145 | | | Total reversals | | 74 | 179 | | |

***, ** and * denote significance at 1% and 5% and 10% levels respectively;
‡‡‡ denotes significance at 1% in the opposite direction to all the asterisked cases.

**Figure 1: How Lotteries Were Displayed**

| | 1-65 | 66-100 |
|---|---|---|
| X | £12.50 | 0 |
| | 65% | 35% |

**Figure 2: Choice Display**

**Which of A and B would you prefer to play out?**

| | 1-50 | 51-100 |
|---|---|---|
| **A** | £25 | 0 |
| | 50% | 50% |

| | 1-100 |
|---|---|
| **B** | £9 |
| | 100% |

**Chosen option for Q1:** ☐

**Figure 3: Betweenness Lotteries in the {£20, £10, 0} Marschak-Machina Triangle**