

Can ranking techniques elicit robust values?

Ian Bateman · Brett Day · Graham Loomes ·
Robert Sugden

Published online: 30 December 2006
© Springer Science + Business Media, LLC 2006

Abstract This paper reports two experiments which examine the use of ranking methods to elicit ‘certainty equivalent’ values. It investigates whether such methods are able to eliminate the disparities between choice and value which constitute the ‘preference reversal phenomenon’ and which thereby pose serious problems for both theory and policy application. The results show that ranking methods are vulnerable to distorting effects of their own, but that when such effects are controlled for, the preference reversal phenomenon, previously so strong and striking, is very considerably attenuated.

Keywords Preference reversal · Valuation · Ranking · Experiment

JEL Classification C91 · D81

There is now a substantial literature reporting evidence of apparently systematic violations of key axioms of expected utility (EU) theory—see, for example, Camerer (1995). One ‘anomaly’ which is particularly troublesome, not only for standard theory but also for applied work in various important areas of public policy, is the preference reversal (PR) phenomenon, first reported by Lichtenstein and Slovic (1971) and Lindman (1971). Although the PR anomaly can take a variety of forms (Tversky and Thaler, 1990; Seidl, 2002), the best known and most frequently replicated occurs when the preference ordering inferred from the *values* an individual separately attaches to two different items is contradicted by the *choice* that individual makes when considering them together.

Such reversals have been replicated within gains (Reilly, 1982) and losses (MacDonald et al., 1992); have been found in both individual and group responses (Mowen and Gentry, 1980); have been observed both in real-world lotteries (Bohm and Lind, 1993) and in those constructed in the laboratory (Knez and Smith, 1987); occur across lotteries which differ in

I. Bateman · B. Day
School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ

G. Loomes (✉) · R. Sugden
School of Economics, University of East Anglia, Norwich, NR4 7TJ
e-mail: g.loomes@uea.ac.uk

expected value (Cox and Epstein, 1989) and are priced using differing formats (Berg et al., 1985). Perhaps most tellingly PR phenomena have been shown to be robust against explicit attempts to design them out of responses (Grether and Plott, 1979). Such is the frequency, persistence and robustness of the anomaly that, in his review of the literature, Seidl (2002, p. 621) describes PR as “one of the most spectacular violations of procedure invariance.”

As noted, the most frequently replicated form of PR is that which elicits certainty equivalent (usually, selling price) values for two lotteries and also asks respondents to make a straight choice between the two. Normally in these experiments, one lottery gives a relatively small chance of a high payoff (and has come to be referred to as the \$-bet) while the other offers a much more modest payoff, but with a high probability of receiving it (and is often referred to as the P-bet). Transitivity requires that an individual who (strictly) prefers lottery X to lottery Y will both place a higher certainty equivalent value on X than on Y and also select X in a straight choice between the two. But the PR phenomenon shows that many people behave otherwise: a substantial proportion place a *higher value on the \$-bet*, but *choose the P-bet* in a straight choice between the two. In some experiments, this is the modal pattern and we shall refer to it as the *regular* form of reversal. The opposite violation—placing a higher value on the P-bet but choosing the \$-bet—is relatively rarely observed, and we shall call it the *counter* reversal.

One interpretation of anomalies such as PR is that people’s preferences are essentially well-behaved, but operate according to different principles than underpin EU (for a review, see Starmer, 2000). However, a radically different interpretation of such evidence is that, when presented with different types of tasks, many people are liable to use different cognitive processes or heuristics with which to construct their preferences (see Tversky and Kahneman, 1986). There is more than one variant of the heuristic explanation for the PR phenomena, but the essence of most of them is that the valuation and choice tasks bring somewhat different cognitive processes into play. The valuation task asks for a monetary response and therefore tends to focus respondents’ attention on the money payoffs, and in particular, evokes a tendency to anchor on the positive payoff and adjust downwards, but insufficiently, so that the valuation response for the higher-anchor \$-bet is liable to end up greater than the valuation of the lower-anchor P-bet. By contrast, the choice task encourages relatively more weight to be attached to the probability of getting some positive payoff, which favours the P-bet. The claim is that this difference between the ways the two tasks are processed is liable to lead to the disparity so frequently observed.

Might there be some elicitation procedure which is less susceptible (perhaps invulnerable) to such influences? And if so, will the standard PR pattern be attenuated (even eliminated)? To investigate this issue, our basic strategy, described in more detail in the next section, is to move away from the standard presentation of isolated valuation and choice tasks and adopt a format which encourages respondents to consider a variety of lotteries—including some that have the characteristics of \$-bets and others with the characteristics of P-bets—as well as a number of different sure amounts, and ask them to produce a single ranking of the full set of alternatives. This allows valuations (within some range) to be inferred for each lottery from its location between two sure amounts.

The intuition is that engaging respondents in a task where there is a wide spectrum of probabilities as well as numerous payoffs spread across a broad range, and where there is implicit encouragement to strike balances between the two dimensions, might make them less susceptible to anchoring on any particular component and may thereby greatly attenuate the disparity between choice and valuations.

However, even if that turned out to be the case, it would not, by itself, be sufficient to establish that the ranking procedure is immune from distortions or anomalies of its own. The

most obvious candidate for a ranking anomaly is that the ordering of some particular subset of lotteries may be systematically affected by changing some of the *other* lotteries being ranked.

An indication of the sort of thing that has been observed to happen in other contexts is reported in Robinson et al. (2001). In that study, respondents were asked to rank a number of descriptions of road accident injuries in order of how bad they were, and then score these on a scale where an index of 100 was assigned to ‘Normal Health’ while ‘Worst Outcome/Death’ was assigned a score of 0. Two sets of nine injury descriptions were compiled, of which Normal Health, Death and three injuries (labeled R, S and X) were common to both sets, while the other four descriptions differed between sets. In Set A, three of these other four were all injuries of a less serious nature, involving no permanent after-effects, whereas in Set B three of the four were permanently disfiguring and/or disabling. These differences in the ‘other’ injuries did not affect the ranking over R, S and X. But there *were* significant differences in the scores assigned to those injuries: the inclusion of the milder ‘other’ injuries in Set A pushed the scores for R, S and X down relative to those in Set B where the inclusion of several very unpleasant injuries made R—and even more so, S and X—seem less severe.

Of course, those scores were not certainly equivalent values of the kind being considered in the present study. But what those results illustrate is the potential for a ranking procedure to induce a form of ‘range-frequency (R-F) effect’¹ (Parducci and Weddell, 1986), and it is important to check whether something analogous may come into play in the ranking of lotteries.

The present study reports the results of two experiments designed to explore whether there might be some ranking procedure that could reduce or even eliminate PR without generating some other subversive anomaly of its own.² The next two sections report the two experiments we conducted. The first showed that there was indeed a strong procedural effect associated with the ranking task, and that this was liable to confound the preference reversal data. The second experiment attempted to both explore and control for that procedural effect. This showed that in cases where the effect was controlled for, the preference reversal phenomenon was greatly attenuated. The final section discusses possible interpretations and implications.

1 Experiment 1

1.1 Design

The experiment was designed to achieve two objectives relevant to this paper: first, to replicate the usual PR phenomenon using a standard two-valuations-and-a-choice design of the kind that has generated it previously; and second, to implement a ranking procedure that would allow the necessary comparisons with the choices and values elicited by the standard methods, while also providing checks on the stability of valuations across different sets.

¹ In essence, the R-F effect states that the value assigned to an item—in the example from Robinson et al. (2001), that is the score on the scale from 0 to 100—reflects not only the ‘intrinsic’ value of that item but also its ranking in the set of items within which it is embedded.

² The experiments in question also addressed certain other anomalies, such as the ‘common ratio effect.’ However, because the total volume of evidence was too great to fit into a single paper, companion papers focus on the other anomalies and compare the patterns generated purely within pairwise choice tasks with those inferred from ranking. A more detailed account of those results can be found at http://www.uea.ac.uk/eco/people/add_files/loomes/

Table 1 The lotteries in Experiment 1

Set 1	Label	Lottery	EV ^a	Set 2	Label	Lottery	EV ^a
High EV	A	$0.5 \times \pounds 25$	12.50	Low EV	P	$0.1 \times \pounds 25$	2.50
	B	$0.75 \times \pounds 15$	11.25		Q	$0.15 \times \pounds 15$	2.25
	C	$0.6 \times \pounds 15$	9.00		R	$0.2 \times \pounds 9$	1.80
	D	$0.8 \times \pounds 9$	7.20		S	$0.25 \times \pounds 6$	1.50
P-Bets	E	$0.7 \times \pounds 12$	8.40	\$-Bets	N	$0.1 \times \pounds 60$	6.00
	F	$0.85 \times \pounds 10$	8.50		M	$0.15 \times \pounds 45$	6.75
	G	$0.9 \times \pounds 9$	8.10		L	$0.2 \times \pounds 35$	7.00
	H	$0.95 \times \pounds 8$	7.60		K	$0.3 \times \pounds 25$	7.50
Common	I	$0.4 \times \pounds 18$	7.20	Common	I	$0.4 \times \pounds 18$	7.20
	J	$0.55 \times \pounds 13$	7.15		J	$0.55 \times \pounds 13$	7.15

^a Expected value of lottery.

Table 1 shows the two sets of lotteries used, with each lottery expressed as some probability of a positive payoff (with the other payoff always being zero). Expected values are shown in the EV columns. The first four rows show two groups of lotteries, A-D and P-S, that were primarily intended to test for violations of independence (the details of which can be found in the companion papers referred to in footnote 2 above). Notice that A-D are relatively high EV lotteries whilst P-S are relatively low EV lotteries.

Lotteries E-H in Set 1 constituted four P-bets, while lotteries K-N in Set 2 were \$-bets. By asking respondents to rank these lotteries together with a set of sure amounts, we could infer their values for each of the P-bets from the ranking of Set 1 and for each of the \$-bets from the separate ranking of Set 2. These values could then be compared a) with values for each bet elicited by the standard method, and b) with the direct choices between various {P, \$} pairs.

Lotteries I and J were common to both sets. The objective was to test whether including them in a ranking exercise where the majority of the other lotteries had higher EVs and offered better probabilities of a positive payoff, as in Set 1, would cause them to be valued differently than when they were included in Set 2, where most of the other lotteries offered a lower EV and a smaller probability of winning. If something analogous to the effect observed by Robinson et al. (2001) were to occur, the values inferred for both I and J from the Set 1 ranking exercise would be markedly lower than the values inferred from ranking them in Set 2. On the other hand, if it turned out that there were no significant differences between the values inferred for I and J even though any R-F effect was being given every chance of manifesting itself in these cases, confidence in the robustness of certainty equivalent values inferred from ranking would be increased. The alternative possible findings and their interpretations might be summarized as follows.

- (i) No R-F effect is observed and the PR phenomenon is absent from the ranking data, suggesting that ranking has the potential to elicit values that correspond with choices and indicating that it might be worthwhile exploring further the use of ranking-based methods for eliciting values in a broader range of applications.
- (ii) No R-F effect, but the persistence of the PR phenomenon in the ranking data. This might suggest that even though we may disable the particular response mode effects that psychologists regard as likely to be responsible for PR in the classic two-valuations-and-a-choice design, other heuristics may be at work. Alternatively, it may be that non-EU

	1-65	66-100
X	£12.50	0
	65%	35%

Fig. 1 A typical lottery display

preferences are responsible for the phenomenon and that these are robust to different elicitation procedures.

- (iii) An R-F effect is observed, indicating that ranking is not a ‘neutral’ procedure for eliciting values. In this case, the possible implications for PR would need to be considered, and further experimental work would be required to examine the feasibility of controlling for any such effect.

1.2 Implementation

Respondents were recruited via a general e-mail invitation to undergraduate and graduate students throughout the University of East Anglia. The experiment was conducted in 12 sessions with up to 16 respondents in any one session. On arrival, respondents were seated at large desks with partitions separating them. They were given some introductory notes which the moderator read through with them at the start of the session, inviting them to ask for clarification if there was any aspect they did not understand. Those notes are given in the Appendix to this paper and included an example of a typical lottery display, reproduced in Fig. 1.

It was explained that if a respondent ended up with a lottery such as X, it would be played out by drawing a disc at random from a bag containing 100 discs numbered from 1 to 100. In the case of X, if the disc bore a number from 1 to 65, the experiment would pay the respondent £12.50 in cash; but if the disc bore a number from 66 to 100, the respondent would receive nothing. Respondents were told that they would be asked various different types of question that would be presented in different booklets, each with its own brief instructions and labeled with a playing card suit—i.e., ♠, ♥, ♣, or ♦. Copies of the introductory notes and of each booklet are available from the corresponding author on request.

The booklet labeled ♠ contained 8 questions, each asking the respondent to state the minimum sure amount she would accept to sell the right to play out a particular lottery. The standard incentive system designed to encourage truthful and accurate answers³ was explained. Four of the questions asked for values of the P-bets (E, F, G and H) and four asked for values of the \$-bets (K, L, M and N), with one \$-bet and one P-bet presented on each page—in each case being the pair that the respondent was asked to choose between in the pairwise choice exercise described below.

The booklet labeled ♥ contained 12 questions, each being a pairwise choice. Four were the {P, \$} pairs—{E, N}, {F, M}, {G, L}, {H, K}—while the other eight were pairs designed to replicate the common ratio effect (see footnote 4). In the event that a respondent’s payout was based on the pairwise choice exercise, the incentive system was the standard one whereby one of the 12 questions was picked at random and the respondent played out whichever lottery she had chosen in that question.

³ Based on the mechanism originally proposed in Becker et al. (1963); see the booklet instructions for details.

	1-100
	£9
	100%

Fig. 2 The display for a sure amount

The third booklet had two answer pages, one labeled ♣ and the other labeled ♦, respectively relating to Set 1 and Set 2, together with instructions about how to undertake the ranking exercise and record their decisions. The essence was as follows. For a particular Set, respondents were given two envelopes. One envelope contained ten strips of card, each depicting a particular lottery displayed as in Fig. 1. The other envelope contained ten more strips, each offering a sure amount—£2, £3, £4, £5, £6, £7, £8, £9, £10, or £12: an example is shown in Fig. 2.

Respondents were asked to take the ten lotteries and set them out on table, arranging them from most preferred to least (with no ties allowed); and then to take the ten sure amounts and integrate them into the ranking, until all twenty strips of card were arranged in order of preference. At this point, they were asked to record their ordering on the appropriate answer sheet.⁴ The twenty strips were then put back in their envelopes, which were collected before any further task was undertaken.

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 were selected at random (by rolling a 20-sided die) and the respondent played out whichever of the two she had ranked higher.

So by the end of the experiment, a respondent had answered four sets of questions, each set of answers labeled by a different suit. It was explained at the outset that once they had finished all of the tasks, each respondent would pick one card at random (and with replacement) from a standard pack of playing cards. His/her set of answers labeled with that suit would then be recovered, and one of those questions would be picked at random. Each respondent's payment would then depend entirely on how his/her decision in that question played out.

1.3 Results

Before considering the results in detail, a brief note about data that were excluded. Because this was a 'pen-and-paper' exercise conducted with up to 16 respondents at a time, it was not possible to build in automated consistency checks in the way that would be possible for computerised exercises. And indeed, we were interested to see how well respondents handled the tasks in the absence of such checks and prompts. But as a result, some respondents' answers exhibited what would be widely regarded as basic mistakes.

In particular, in the ranking exercise, respondents occasionally ranked a certainty of £X higher than a certainty of £Y even though $£X < £Y$. In such cases, it is impossible to infer a respondent's valuation of lotteries ranked between £X and £Y, since these lotteries are

⁴ In the exposition, we shall refer to each lottery by the label given to it in Table 1. However, in order to minimise problems of unclear handwriting, the strips of card depicting lotteries actually had two-letter labels. The sure amount strips did not have labels: when recording where these came in their ranking, respondents were asked to write down the amount itself, prefaced with a £ sign.

apparently valued less than the lower amount, £X, while simultaneously being valued greater than the higher amount, £Y. Of course, mistakes in the ordering of the certainties might be accepted as one-off errors, attributable perhaps to a lapse in concentration. Alternatively, they might signify some fundamental misunderstanding or misinterpretation of the task. We adopted a simple rule designed to distinguish between these two possible sources of error. If one certainty could be removed from the ranking such that the remaining certainties were correctly ordered, then this was treated as evidence of a one-off error. The offending certainty was removed from the ranking and the rest of the data were used in the analysis. On the other hand, if more than one certainty had to be removed in order to establish a correct order, this was taken as signifying a fundamental error. All data generated by the respondent for that ranking task were excluded from the analysis.⁵

After those adjustments, values for each lottery were inferred as follows. In all cases where a lottery was ranked below one certainty and above another, the inferred value was set halfway between the two. In cases where the lottery was ranked above £12, we assigned it a value of £12.50, and in cases where it was ranked below £2, we assigned it a value of £1.50. These last two assignments are, of course, somewhat arbitrary. However, applying them consistently across all cases and, where necessary, implementing corresponding truncation of the values elicited directly in the exercise, allows us to make the comparisons we are interested in.

When presenting the data from this experiment, we start with the results relating to the R-F effects, since it turned out that these had implications for the preference reversal patterns.

To see whether the value inferred for any one lottery is liable to be influenced by the nature of the other lotteries in the set being ranked, the most direct test is to compare on a within-subject basis the values inferred for I and J in Set 1 with those inferred in the context of Set 2. Both I and J were ranked very much higher in Set 2, which contained more relatively low EV lotteries—an average ranking of 2.30 and 2.49 respectively—than they were in Set 1, which contained more relatively high EV lotteries, where their average ranks were 8.30 and 8.56.⁶ If there were no particular effect of ranking upon values, any differences between the values inferred from each set would be randomly distributed around zero means.

However, the evidence shows that there was a very powerful effect indeed. When the sure amounts were incorporated into the rankings, the value for I inferred from Set 2 was strictly higher than the value inferred from the Set 1 for 121 of the 154 respondents. That is to say, those respondents ranked I above some sure amount X in Set 2 but ranked I below that same sure amount in Set 1. Another 16 gave the same value in both sets—i.e. ranked I between the same two sure amounts on both occasions; and 17 gave I a lower value in Set 2. For lottery J, the corresponding breakdown was 120, 28 and 6.

If differences in the inferred valuations across ranking tasks were occurring simply as a result of randomness on the part of respondents, we would expect as many respondents to err in their valuations in one direction as in the other. Using this null, we calculate an exact binomial test of proportions in matched pairs to compare the numbers erring in each

⁵ On the basis of this criterion, we found that 5 of the 162 people who took part made a one-off error in the ♣ ranking exercise, 3 made a one-off error in the ♦ ranking exercise, with one individual making one-off errors in both exercises. Four respondents' answers to both ranking exercises were excluded from the analysis because they were deemed to have fundamentally misunderstood the task in both cases. Another 4 committed fundamental errors only in the ♦ exercise. Since the results presented in this paper involve comparisons across the ♦ and ♣ ranking exercises, the analysis is based on the 154 people not excluded from either exercise. (However, note that occasional failures by respondents to answer every question may cause the number of observations in some instances to fall below 154.)

⁶ The highest ranked lottery was assigned a rank of 1, down to 10 for the lowest ranked lottery.

Table 2 Pairwise choice and direct valuation in Experiment 1

{P, \$} pair	n ^a	P, P ^b	P, \$ ^c	\$, P ^d	\$, \$ ^e	Statistical significance ^f
{E, N}	147	83	42	6	16	***
{F, M}	137	75	35	4	23	***
{G, L}	127	66	26	10	25	**
{H, K}	116	45	36	6	29	***

^aNumber of respondents.

^bNumber preferring P-bet in choice and P-bet in valuation (consistent with standard theory).

^cNumber preferring P-bet in choice and \$-bet in valuation ('regular' preference reversal).

^dNumber preferring \$-bet in choice and P-bet in valuation ('counter' preference reversal).

^eNumber preferring \$-bet in choice and \$-bet in valuation (consistent with standard theory).

^fTests null hypothesis that both types of reversal are equally likely to occur by chance, where ***, ** and * denote significant differences in the proportions of regular and counter reversal at 1%, 5% and 10% levels respectively.

direction. For each of these two lotteries, the probabilities of the observed asymmetries occurring by chance are less than 10^{-6} .

Analysed another way, the mean value for lottery I was £8.64 when inferred from the Set 2 ranking, as compared with £5.80 when inferred in the context of Set 1. For lottery J the corresponding figures were £8.40 and £5.82. These differences are highly significant; a paired *t*-test of equality of means returned *t*-statistics greater than 12 in both cases. In short, however the data are processed, it is clear that the ranking procedure as implemented in this experiment had a dramatic effect on the inferred valuation of a given lottery—in these cases, increasing the mean values by between 40% and 50%—depending on the nature of the *other* lotteries in the set.

If that was true for I and J, it was also liable to be true for the other lotteries in each set. Given the results for lotteries I and J, the rankings of each of the P-bets within Set 1, and of each of the \$-bets within Set 2, were liable to be a factor in their relative valuations. That should be borne in mind when considering the PR results. The evidence was as follows.

Table 2 shows the patterns generated by the standard two-valuations-and-a-choice design (booklets ♠ and ♥), with the column headings showing the lottery that was chosen followed by the lottery that was valued higher. So, for example, the column headed P,\$ shows the numbers of respondents committing the 'regular' reversal—choosing the P-bet but valuing the \$-bet higher—while the column headed \$,P shows the numbers exhibiting the 'counter' reversal. Those who gave the same value to both lotteries are included either in the P,P column (if they chose P) or in the \$,\$ column (if they chose \$): that is, since no *strict* reversal has been observed, they are counted as cases which are consistent with conventional theory. Cases where respondents gave a value that was equal to or greater than the payoff offered by the lottery were excluded from the analysis of that pair. This happened more often as the probability offered by the P-bet approached 1, so the number of observations, *n*, declines as we move from {E, N} to {H, K}.

In line with the previous literature, the null hypothesis is that there is no significant difference between the numbers of regular as compared with counter reversals. Since the number of observations is not excessive we are able to calculate an exact binomial test of proportions in matched pairs, the results of which are reported in the final column of Table 2. The usual asymmetry between P,\$ and \$,P occurred to an extent that is significant at the 1%

Table 3 Pairwise choice and inferred valuation in Experiment 1

{P, \$} pair	n	P, P	P, \$	\$, P	\$, \$	Statistical significance
{E, N}	151	100	28	8	15	***
{F, M}	144	99	16	12	17	–
{G, L}	141	76	22	12	31	*
{H, K}	130	54	31	2	43	***

Notes: As per Table 2.

level in three cases out of four; in the fourth case involving {G, L}, the asymmetry was in the expected direction and significant at the 5% level. In short, when using the standard design, we replicated the usual preference reversal phenomenon.

Table 3 shows the results when values were inferred from the ranking exercise. If the P-bet and the \$-bet in any pair were ranked between the same two certainties and therefore were assigned the same inferred values, they were counted as cases which conform with standard theory. Analogous to the exclusion criterion applied to direct valuations, if a respondent ranked a lottery with payoff X above the certainty of X, that observation was excluded from the analysis of that pair.

Clearly, significant PR patterns were also observed in the ranking data, although the strength of the effect seemed rather more variable. For {H, K} and {G, L} the numbers of reversals and their degree of asymmetry was not very different from those observed in the standard two-valuations-and-a-choice procedure (Table 2). For {E, N} the number of regular reversals was reduced by a third (although the asymmetry still remained highly significant). By contrast, for {F, M} the asymmetry was virtually eliminated, with counter reversals somewhat higher under ranking while regular reversals were less than half as frequent as in Table 2.

Of course, bearing in mind the results concerning lotteries I and J, it seems likely that the inferred values of the different P and \$ bets may have been confounded by their rankings in their respective sets. Table 4 shows the mean ranking of each lottery prior to the incorporation of the sure amounts.

If the R-F effect were operating on inferred values, we should expect it to favour the inferred value of \$ over the inferred value of P most in the case of {H, K}, where K is the highest ranked of the \$ bets while H is ranked third among the P bets; conversely, \$ should be least favoured over P in the case of {F, M}, where F is the highest ranked P bet while M is only ranked third among the \$ bets. A comparison of the data in Tables 2 and 3 confirms this expectation. K was valued at least as highly as H by 65 respondents in the standard choice and valuation task, but this figure was increased to 74 in the ranking task. By contrast,

Table 4 Mean ranking of P and \$ bets in Experiment 1

P-Bet	Mean rank ^a	\$-Bet	Mean rank ^a
E	5.32	N	5.53
F	3.66	M	5.08
G	4.53	L	4.08
H	5.16	K	2.69

^aMean rank calculated by assigning a value of 1 if lottery placed first in a respondent’s ranking, 2 if placed second and so on down to 10 if placed last.

whereas M was valued at least as highly as F by 58 respondents in the standard task, the number fell to 33 in the ranking task. The comparisons for the other two pairs were also consistent with the operation of an R-F effect.⁷

Since patterns of PR inferred from the ranking procedure seemed to have been affected by range-frequency considerations, we conducted a second experiment which aimed to test for PR while trying to control for—or at least, monitor—R-F effects.

2 Experiment 2

2.1 Design

It could be argued that Experiment 1 had encouraged the most extreme form of R-F effect in two respects: (i) lotteries I and J were at one end of the range of EVs in one set and at the opposite end in the other set; and (ii) the certainties were only inserted after the initial ranking of lotteries had been completed. So Experiment 2 was designed to examine the R-F effects for lotteries spread more widely across the EV and probability spectra, and to drop the initial separation between lotteries and certainties.

However, while we were interested in gathering more information about the nature and strength of the R-F effect, our principal objective was to test whether PR would persist, or reduce, or even disappear altogether under ranking if the R-F effect could be controlled for. To that end, the design was as follows.

Table 5 organises the lotteries into groups. The first three rows show the lotteries—E, K and Q—which were common to all four sets and which were intended to gauge the extent of any R-F effects across the sets. Then there were four lotteries—A, D, F and J—that we shall call ‘High EV’ lotteries, and that were included only in Sets 1 and 3; while in Sets 2 and 4 there were four ‘Low EV’ lotteries—R, S, T and V—shown at the bottom of the table. In between, there were four \$-bets—C, G, L and M—and four P-bets—B, H, N and P. Thus the design may be seen as follows:

Set 1: {E, K, Q}, {P-bets}, {High EV}

Set 2: {E, K, Q}, {\$-bets}, {Low EV}

Set 3: {E, K, Q}, {\$-bets}, {High EV}

Set 4: {E, K, Q}, {P-bets}, {Low EV}

So the four P-bets appear in both Sets 1 and 4 while the four \$-bets appear in both Sets 2 and 3. To try to control for R-F effects, the seven lotteries *other than* the P-bets in Set 1 were exactly the same as the seven lotteries *other than* the \$-bets in Set 3. Likewise, the seven ‘other’ lotteries in Set 2 were the same as the seven ‘others’ in Set 4, although these seven are clearly inferior to the seven in Sets 1 and 3, because they include the four ‘Low EV’ bets, each of which is strictly dominated by at least one of the ‘High EV’ bets which appear in Sets 1 and 3.

This may not be a perfect control for R-F effects, because the inferred value of any one P-bet (say, N) may be influenced by how it is ranked in relation to the other three P-bets, while a \$-bet with which it is compared (say, G) may be influenced differently by the other

⁷ L was slightly better ranked in its set than G was in its, and 53 respondents’ inferred values for L were at least as high as their values for G, compared with 51 in the standard task; whereas for E and N, the number fell from 58 in the standard task to 43 in the ranking task, where N’s relative ranking was slightly lower than E’s.

Table 5 The lotteries in Experiment 2

	Label	Set 1	Set 2	Set 3	Set 4	EV ^a
Shared	E	0.75 × £20	0.75 × £20	0.75 × £20	0.75 × £20	15.00
	K	0.5 × £20	0.5 × £20	0.5 × £20	0.5 × £20	10.00
	Q	0.25 × £20	0.25 × £20	0.25 × £20	0.25 × £20	5.00
High EV	A	0.5 × £35		0.5 × £35		17.50
	D	0.6 × £25		0.6 × £25		15.00
	F	0.85 × £16		0.85 × £16		13.60
	J	0.9 × £12		0.9 × £12		10.80
-\$-Bets	C		0.25 × £60	0.25 × £60		15.00
	G		0.35 × £35	0.35 × £35		12.25
	L		0.2 × £40	0.2 × £40		8.00
	M		0.15 × £50	0.15 × £50		7.50
P-Bets	B	0.9 × £18			0.9 × £18	16.20
	H	0.8 × £15			0.8 × £15	12.00
	N	0.7 × £10			0.7 × £10	7.00
	P	0.75 × £8			0.75 × £8	6.00
Low EV	R		0.15 × £25		0.15 × £25	3.75
	S		0.1 × £35		0.1 × £35	3.50
	T		0.2 × £15		0.2 × £15	3.00
	V		0.25 × £12		0.25 × £12	3.00

^a Expected Value of lottery.

-\$-bets against which *it* is ranked. The ideal control would be to have everything common to two sets except a single P-bet in one set and a single -\$-bet in the other. However, in the interests of gathering more data from a limited research budget, we chose to look at varying four bets across the sets. If this was sufficient to *largely* control for R-F effects, and if inferring the values of the P-bets and -\$-bets has the potential to reduce or eliminate PR *once R-F effects are controlled for to this extent*, we should see this when the values of the P-bets inferred from Set 1 are compared with the -\$-bet values inferred from Set 3, as well as when the -\$-bet values inferred from Set 2 are compared with the P-bet values from Set 4.

To see how any R-F effects might be detected, consider first the three lotteries E, K and Q. It turned out that, overall, P-bets were chosen more often than -\$-bets. On that basis, Set 2 has the least attractive combination of ‘other’ (i.e. not E, K or Q) lotteries—namely, {-\$-bets} and {Low EV}, so that it is in that set that we should find the highest rankings of E, K and Q and the greatest upward influence on their inferred values. By contrast, Set 1 contains the two *most* desirable groups of ‘other’ lotteries, namely {P-bets} and {High EV}, which should combine to exert the greatest downward pressure on the rankings and values of E, K and Q. While the effects in Sets 3 and 4 can be expected to lie between those in the ‘extreme’ Sets 1 and 2, it is not obvious *ex ante* whether the effect of the {Low EV} versus {High EV} contrast will be greater than the opposite influence of the {P-bet} versus {-\$-bet} comparison.

Likewise, we should expect to see the rankings and therefore values of the P-bets elicited from Set 4 (which contains the Low EV group) being higher than their counterparts elicited from Set 1 (where they will be compared against the High EV group). By the same token, the rankings and values of the -\$-bets elicited from Set 2 should be higher than those from Set 3. Similarly, any differences between the rankings and values of the High EV bets should

favour those in Set 3 over those in Set 1, while if there is any difference among the Low EV bets, they should be higher ranked and valued in Set 2 than in Set 4.

2.2 Implementation

As before, respondents were recruited via a general e-mail invitation and participated under similar conditions as in Experiment 1: that is, in one of 12 sessions, seated at separated desks and with instructions read through so as to give an opportunity to ask for clarification. Copies of the introductory notes and of each booklet are available from the corresponding author on request.

Each respondent undertook three ranking exercises, involving three of the four sets of 11 lotteries shown in Table 5 plus, in each case, the same 14 certainties—these being every whole pound from £2 to £15 inclusive: thus each ranking exercise involved a total of 25 strips of card. Sandwiched between the ranking exercises were two sets of 10 pairwise choices, including eight {P, \$} pairs, with each P-bet and each \$-bet appearing twice, as follows: {B, C}, {B, M}, {H, C}, {H, G}, {N, G}, {N, L}, {P, L} and {P, M}. The aim was to have a reasonable diversity of pairs, with some favouring the \$-bet in choice, and some others favouring the P-bet to an even greater extent than is usual in PR experiments.⁸

For each ranking exercise, respondents were asked to empty all 25 strips out of a single envelope, arrange them in order of preference and record their order on the appropriate answer sheet. So in this experiment, by contrast with Experiment 1, no distinction was made between lotteries and sure amounts. The order in which ranking tasks were presented, and the combinations of three of the four sets, were all systematically varied from one session to the next so as to control for order effects. Likewise, the order of the two sets of 10 choices spliced in between the ranking tasks was alternated. Finally, once a respondent had completed all five tasks, one task was picked at random (independently for each respondent) and then the same incentive mechanisms were used as in the first experiment.

2.3 Results

Altogether, 151 people took part. We applied the same exclusion criteria as in the first experiment: if a respondent made a single ‘basic mistake’ in ranking the sure amounts, we treated that particular response as a missing value; but if a respondent’s answers contained two or more such mistakes, we excluded all data generated by that respondent from the analysis. Two respondents were excluded on the grounds of multiple errors in the ordering of the certainties, so the results presented here are for a sample of 149. A further two respondents made one-off errors that were dealt with by removing the incorrectly ordered certainty.

The data relating to R-F effects are shown in Tables 6–8. Table 6 reports the mean rankings of each lottery within each set, while Tables 7 and 8 report the mean within-respondent differences between the values of each lottery inferred from different sets.

Table 6 shows that the effects on rankings were generally much as anticipated. For the common lotteries E, K and Q, Set 2 always generated the highest rankings while Set 1 always produced the lowest. If respondents’ values were simply reflecting their ‘true’ preferences, they would not have been influenced by these ranking differences. However, as

⁸ Ideally, we would also have had a direct valuation exercise of the kind undertaken in the ♠ section of Experiment 1. However, we were asking respondents to undertake three ranking exercises involving 25 strips rather than two exercises involving 20 strips, and 20 pairwise choices rather than 12, and we were concerned not to overload respondents.

Table 6 Mean rankings of each lottery within each set in Experiment 2

Label	Lottery	EV ^a	Mean rankings ^b			
			Set 1	Set 2	Set 3	Set 4
A	0.5 × £35	17.50	4.37	–	3.45	–
B	0.9 × £18	16.20	1.83	–	–	1.37
C	0.25 × £60	15.00	–	2.77	5.38	–
D	0.6 × £25	15.00	4.92	–	3.88	–
E	0.75 × £20	15.00	3.45	1.50	2.75	2.18
F	0.85 × £16	13.60	4.15	–	3.71	–
G	0.35 × £35	12.25	–	3.53	7.14	–
H	0.8 × £15	12.00	5.42	–	–	3.15
J	0.9 × £12	10.80	5.76	–	5.23	–
K	0.5 × £20	10.00	7.73	3.16	6.59	4.36
L	0.2 × £40	8.00	–	5.43	8.68	–
M	0.15 × £50	7.50	–	6.35	9.04	–
N	0.7 × £10	7.00	8.52	–	–	5.74
P	0.75 × £8	6.00	9.53	–	–	6.80
Q	0.25 × £20	5.00	10.31	6.46	10.16	6.75
R	0.15 × £25	3.75	–	9.04	–	8.54
S	0.1 × £35	3.50	–	9.61	–	8.52
T	0.2 × £15	3.00	–	9.21	–	9.42
V	0.25 × £12	3.00	–	8.96	–	9.16

^aExpected Value of lottery.

^bMean rank calculated by assigning a value of 1 if lottery placed first in a respondent’s ranking, 2 if placed second and so on.

Table 7 Mean differences of within-respondent values for lotteries common to all sets (E, K and Q)

Label	Mean difference of within-respondent implied valuations in different sets (<i>p</i> -values < .1 from 2-tailed paired <i>t</i> -test for equivalence of valuations ^a)					
	Sets 1 vs. 2	Sets 1 vs. 3	Sets 1 vs. 4	Sets 2 vs. 3	Sets 2 vs. 4	Sets 3 vs. 4
E	– 0.76*** (.003)	– 0.15	–0.13	0.66** (.024)	0.41	–0.31
K	– 1.67*** (.000)	–0.47	0.13	1.06*** (.001)	1.05*** (.002)	0.13
Q	– 0.91*** (.002)	–0.39	–0.04	0.61** (.015)	0.57** (.036)	0.25

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

Table 7 shows, there were clear effects on the values of these lotteries, with Set 2 producing significantly higher values than any of the other three sets, with the size of the difference and its significance being greatest for K, followed by Q, in every comparison. K involved a win probability of 0.5, so was most similar to I and J from Experiment 1. However, even the largest difference in this experiment—£1.67 in the Set 1 vs Set 2 comparison—constituted less than 18% of the average value of K, whereas the differences for I and J in Experiment 1 amounted to between 36% and 40% of their average values.

But even if it was weaker than in Experiment 1, the R-F effect had certainly not disappeared, and was also in evidence among the \$-bets. All four \$-bets were ranked considerably higher in Set 2 than in Set 3, and, as shown in Table 8, the two where the mean rankings were

Table 8 Mean differences of within-respondent values for \$-bets, P-bets, high-EV and low-EV lotteries

Mean difference of within-respondent implied valuations in different sets (<i>p</i> -values < .1 from 2-tailed paired <i>t</i> -test for equivalence of valuations ^a)							
\$-bets		P-bets		High-EV		Low-EV	
Label	Sets 2 vs. 3	Label	Sets 1 vs. 4	Label	Sets 1 vs. 3	Label	Sets 2 vs. 4
C	0.43	B	−0.04	A	−0.41	R	0.10
G	1.00*** (.003)	H	−0.21	D	0.08	S	0.05
L	0.54** (.033)	N	0	F	−0.15	T	0.28
M	0.20	P	−0.01	J	−0.25	V	0.34

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

most different—G and L—also registered significantly higher values. By contrast, although there were differences in the rankings of the P-bets—all four being ranked higher in Set 4 than in Set 1, with H, N and P more than two places higher on average—this did *not* translate into significant value differences. One possible interpretation is that the greater dissimilarity between \$-bets and sure amounts makes them more susceptible to influences, whether these be ‘anchoring-and-adjustment’ biases in the standard valuation task or R-F effects as in this experiment; by contrast, with P-bets offering much higher probabilities of winning and thus being closer to certainties, there is less scope for their values to be influenced by procedural effects.

To see what implications this may have had for PR, the relevant data are reported in Table 9. Within each possible pairing of sets, the rows relate to the eight {P, \$} pairs, listed from the case where the P-bet was most frequently chosen down to the one where the \$-bet was most often preferred in the pairwise choice. Columns 3 to 6 show the combinations of choice and inferred value, with Columns 4 and 5 reporting the numbers of ‘regular’ and ‘counter’ reversals. Column 7 shows the statistical significance of the asymmetry between regular and counter reversals, based on an exact binomial test, making no prior assumption about the direction of any asymmetry.

Perhaps the first thing to notice from the table as a whole is that the PR phenomenon was generally much less in evidence than in most previous studies using the two-valuations-and-a-choice format. To the extent that reversals did occur, they occurred predominantly in combinations involving Set 2. The combination of Set 1 and Set 2, which had exhibited the strongest R-F effects for E, K and Q, produced the most significant asymmetries. In particular, the asymmetries for {N, L} and {N, G} (involving the two \$-bets, L and G, whose values were raised most by R-F effects) were significant at the 1% level.

In the comparisons between Set 2 and Set 4, where all lotteries other than the \$-bets and P-bets were the same, the asymmetries for {N, L} and {N, G} were less pronounced, but were still significant at the 5% and 10% level respectively. So the attempt to control for R-F effects by holding all other lotteries constant appears to have been only partially successful in this case.

However, the combination of Sets 1 and 3 appear to have provided a more successful control for R-F effects: although Table 7 showed that the values of E, K and Q (and of three of the other four lotteries) were all higher in Set 3 than Set 1, none of those differences registered as statistically significant. So although there may have been *some* upward influence on the values of the \$-bets in Set 3 relative to the P-bets in Set 1, the usual PR asymmetry was significant in only one of the eight pairs.

Table 9 Pairwise choice and inferred valuation in Experiment 2

{P, \$}	Lotteries	P, P	P, \$	\$, P	\$, \$	Statistical significance
Sets 1 & 2						
{B, M}	.9 × 18 vs .15 × 50	68	2	1	5	
{H, G}	.8 × 15 vs .35 × 35	47	8	3	15	
{B, C}	.9 × 18 vs .25 × 60	48	4	6	18	
{H, C}	.8 × 15 vs .25 × 60	36	11	3	23	*
{N, L}	.7 × 10 vs .2 × 40	32	12	1	29	***
{P, M}	.75 × 8 vs .15 × 50	29	10	4	30	
{N, G}	.7 × 10 vs .35 × 35	21	17	1	35	***
{P, L}	.75 × 8 vs .2 × 40	21	7	4	41	
Sets 2 & 4						
{B, M}	.9 × 18 vs .15 × 50	63	2	3	5	
{H, G}	.8 × 15 vs .35 × 35	41	9	5	17	
{B, C}	.9 × 18 vs .25 × 60	41	1	9	22	**†
{H, C}	.8 × 15 vs .25 × 60	32	5	5	30	
{N, L}	.7 × 10 vs .2 × 40	25	11	2	33	**
{P, M}	.75 × 8 vs .15 × 50	25	7	1	33	*
{N, G}	.7 × 10 vs .35 × 35	18	10	3	40	*
{P, L}	.75 × 8 vs .2 × 40	12	8	4	42	
Sets 1 & 3						
{B, M}	.9 × 18 vs .15 × 50	65	4	3	3	
{H, G}	.8 × 15 vs .35 × 35	52	7	4	10	
{B, C}	.9 × 18 vs .25 × 60	52	6	4	13	
{H, C}	.8 × 15 vs .25 × 60	39	14	3	17	**
{N, L}	.7 × 10 vs .2 × 40	40	4	5	25	
{P, M}	.75 × 8 vs .15 × 50	32	9	5	25	
{N, G}	.7 × 10 vs .35 × 35	31	12	7	24	
{P, L}	.75 × 8 vs .2 × 40	22	4	10	35	
Sets 3 & 4						
{B, M}	.9 × 18 vs .15 × 50	62	2	4	4	
{H, G}	.8 × 15 vs .35 × 35	48	7	4	12	
{B, C}	.9 × 18 vs .25 × 60	46	2	8	16	
{H, C}	.8 × 15 vs .25 × 60	37	5	4	25	
{N, L}	.7 × 10 vs .2 × 40	24	9	7	26	
{P, M}	.75 × 8 vs .15 × 50	26	8	4	28	
{N, G}	.7 × 10 vs .35 × 35	19	12	5	30	
{P, L}	.75 × 8 vs .2 × 40	19	2	8	37	

Notes: As per Table 2 though with † indicating asymmetry in opposite direction to the one usually observed.

Each individual in our sample provided responses pertinent to just one ‘control’ case (that is, they provided inferred valuations of P-bets in Set 1 and \$-bets in Set 3 or they provided valuations of P-bets in Set 2 and \$-bets in Set 4). To get an overall picture of how ‘controlling’ for R-F effects impacts on the incidence of PR in our sample, Table 10 details responses from these ‘control’ cases. In this case, only three asymmetries survive, and these are only significant at the 10% level. Overall, ‘regular’ reversals account for fewer than 10% of the observations. Considering the historical robustness of the PR phenomenon—where regular

Table 10 Pairwise choice and inferred valuation in Experiment 2 when combining Sets 2 & 4 With Sets 1 & 3

{P, \$}	n	P, P	P, \$	\$, P	\$\$	Statistical significance
{B, M}	148	128	6	6	8	
{H, G}	145	93	16	9	27	
{B, C}	148	93	7	13	35	
{H, C}	145	71	19	8	47	*
{N, L}	145	65	15	7	58	
{P, M}	137	57	16	6	58	*
{N, G}	145	49	22	10	64	*
{P, L}	137	34	12	14	77	

Notes: As per Table 2.

reversals often exceed 30% of the total number of observations and may in some cases be the modal response—this is a considerable attenuation of that phenomenon.⁹

Finally, when we consider the combination of Sets 3 and 4, which showed no systematic differences among E, K and Q, *none* of the eight pairs displayed an asymmetry that was significant at the 10% level, and overall the numbers of {P, \$} and {\$, P} observations were very evenly balanced, with a total of 47 regular reversals and 44 counter reversals. While it would be unwise to place too much weight on a particular combination of two sets of lotteries, this result is at least consistent with the idea that the preference reversal phenomenon may be largely or even completely eliminated if values can be inferred from a ranking procedure, so long as R-F effects are effectively annulled.

3 Concluding remarks

It is clear that a ranking procedure *per se* is not an effect-free method of inferring the values of individual lotteries: the value of any given lottery can be affected by changing the other lotteries in the set being ranked. Such results add further to the evidence that values and preferences are not, as standard theory has generally supposed, invariant to the procedure used to elicit them. Even in the cases of well-defined lotteries whose payoffs provide transparent upper and lower bounds for values, it is arguable that the degree of imprecision in many respondents' preferences is such that inferred values are liable to be influenced by factors that are conventionally assumed to be irrelevant. For lotteries, it may be that those with higher variances, such as \$-bets, may be more susceptible than those with lower variances, such as P-bets. For goods where upper and lower bounds are not so transparent, and for which preferences may be even more imprecise—such as the kinds of health, safety and environmental goods which are often the subject of value elicitation surveys—there is the possibility that such influences may be far more potent.

Having said that, Experiment 2 also showed that, to the extent that some such influences can be controlled for, a relatively straightforward ranking procedure can be implemented which attenuates the preference reversal phenomenon to an extent which, as far as we are

⁹ At the time of writing, a meta-analysis of preference reversals is currently being undertaken by Nick Bardsley, Peter Moffatt, Chris Starmer and Robert Sugden: across all of the studies they have reviewed so far, regular reversals account for an average of 28% of all observations.

aware, has not been achieved in individual decision experiments to date. We believe that such a finding constitutes a useful addition to the literature in its own right. It suggests that methodological development of ranking procedures may provide a fruitful avenue for future research in the search for robust and consistent approaches to preference elicitation.

Appendix: General introductory instructions for Experiment 1

We are interested in people’s preferences for different chances of receiving different sums of money. A particular chance of receiving a particular sum of money will be called an option.

An option will look like this:

	1-65	66-100
X	£12.50	0
	65%	35%

If you ended up with option X (which is only an example), it would be played out as follows. We have a bag containing 100 discs, each bearing a different number from 1 to 100 inclusive. You will dip your hand into the bag, pick a single disc and pull it out. If you happened to choose a disc with a number between 1 and 65 inclusive, the experiment would pay you £12.50 in cash; if you happened to choose a disc with a number between 66 and 100 inclusive, you would go away with nothing.

We shall be asking you three types of question, in no particular order:

- To make choices between pairs of options
- To rank a number of options from most to least preferred by you
- To place values on particular options, in the form of the amounts you would sell them for

Different people with different tastes will answer the questions in different ways. We are interested in each person answering the questions according to their own tastes. To give you an incentive to answer according to your own tastes, at the end of the session, one of your decisions will be chosen at random and will be played out for real. What you get paid for taking part in this experiment will depend ENTIRELY on how your decision in that randomly-selected question turns out. So we suggest you answer each question in turn as if it is THE one on which everything depends—because that may in fact turn out to be the case.

We often run experiments using computers, but this time we are using pen and paper. To make this easier to organise, and to help with the random selection of the question which will determine your payment, we are going to divide the questions into four groups, which we shall label ♣ or ♦ or ♥ or ♠. Then at the end of the session each of you will pick a card at random from a standard pack of playing cards, and the suit will determine which group your payout question is picked from. Thereafter, any one of the decisions within that group is equally likely to be played out for real by you.

If you need clarification once the experiment has started, please do not disturb others by calling out. Please raise your hand and one of the organisers will come to you.

[INSTRUCTIONS FOR THE DIRECT VALUATION TASKS FOLLOWED. THESE AND INSTRUCTIONS FOR EXPERIMENT 2 ARE AVAILABLE ON REQUEST FROM THE CORRESPONDING AUTHOR]

Acknowledgments This research was undertaken as part of the U.K. Economic and Social Research Council's Award M535255117. We are grateful to Shepley Orr for assistance with the conduct of the experiments.

References

- Becker, Gordon, Morris DeGroot, and Jacob Marschak. (1963). "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science* 9, 226–232.
- Berg, Joyce, John Dickhaut, and John O'Brien. (1985). "Preference Reversal and Arbitrage," In V. Smith (ed.), *Research in Experimental Economics* Vol. 3 (pp. 31–72). Greenwich: JAI Press.
- Bohm, Peter, and Hans Lind. (1993). "Preference Reversal, Real-world Lotteries and Lottery-interested Subjects," *Journal of Economic Behavior and Organization* 22, 327–348.
- Camerer, Colin. (1995). "Individual Decision Making." In John Kagel and Al Roth (eds.), *The Handbook of Experimental Economics*. Princeton: Princeton University Press.
- Cox, James and Seth Epstein. (1989). "Preference Reversals Without the Independence Axiom," *American Economic Review* 79, 408–426.
- Grether, David and Charles Plott. (1979). "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69, 623–638.
- Knez, Marc and Vernon Smith. (1987). "Hypothetical Valuations and Preference Reversals in the Context of Asset Trading," In Al Roth (ed.), *Laboratory Experimentation in Economics: Six Points of View* (pp. 131–154). Cambridge: Cambridge University Press.
- Lichtenstein, Sarah and Paul Slovic. (1971). "Reversals of Preferences Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology* 89, 46–55.
- Lindman, Harold. (1971). "Inconsistent Preferences Among Gamblers," *Journal of Experimental Psychology* 89, 390–397.
- MacDonald, Don, William Huth, and Paul Taube. (1992). "Generalized Expected Utility Analysis and Preference Reversals: Some Initial Results in the Loss Domain," *Journal of Economic Behavior and Organization* 17, 115–130.
- Mowen, John and James Gentry. (1980). "Investigation of the Preference Reversal Phenomenon in a New Product Introduction Task," *Journal of Applied Psychology* 65, 715–722.
- Parducci, Allen and Douglas Weddell. (1986). "The Category Effect with Rating Scales: Number of Categories, Number of Stimuli and Method of Presentation," *Journal of Experimental Psychology: Human Perception and Performance* 12, 496–516.
- Reilly, Robert. (1982). "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design," *American Economic Review* 72, 576–584.
- Robinson, Angela, Michael Jones-Lee, and Graham Loomes. (2001). "Visual Analog Scales, Standard Gambles and Relative Risk Aversion," *Medical Decision Making* 21, 17–27.
- Seidl, Christian. (2002). "Preference Reversal: A Literature Survey," *Journal of Economic Surveys* 6, 621–655.
- Starmer, Chris. (2000). "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk," *Journal of Economic Literature* 38, 332–382.
- Tversky, Amos and Daniel Kahneman. (1986). "Rational Choice and the Framing of Decisions," *Journal of Business* 59, S251–S278.
- Tversky, Amos and Richard Thaler. (1990). "Anomalies: Preference Reversals," *Journal of Economic Perspectives* 4, 201–211.