The Geneva Papers on Risk and Insurance Theory, 20: 111-133 (1995)

A Comparison of the Estimates of Expected Utility and Non-Expected-Utility Preference Functionals

ENRICA CARBONE AND JOHN D. HEY
University of York, Department of Economics and Related Studies, Heslington, York, YO1 5DD, United Kingdom

Abstract

This paper extends the literature on the estimation of expected utility and non-expected-utility preference functionals (and the consequent exploration of the superiority of non-expected-utility over expected utility preference functionals) to a comparison of two different ways (pairwise choice and complete ranking) of experimentally obtaining data on such preferences. What is revealed is that the magnitude of the subject error is clearly conditional on the elicitation method used and, rather alarmingly, that the preference functional apparently employed by the subject may also be conditional on the elicitation method.

Key words: expected utility, non-expected-utility, experiments, pairwise choice, complete ranking

This paper concerns the estimation of expected utility and non-expected-utility preference functionals and the nature of the error term in this estimation. Previous work on the estimation of expected utility and non-expected-utility preference functionals has been carried out, inter alia, by Hey and Di Cagno [1990], Hey and Orme [1994], and Carbone and Hey [1994]. The first two of these papers estimated preference functionals using pairwise choice data, while the third used complete ranking data.

As the preference functionals implied by expected utility and the various alternative theories are all deterministic, in the estimation of them it was necessary to make some assumption about the stochastic structure underlying the observations. In all these papers it was assumed that subjects stated their preferences with some error, postulated there to be white noise with homoscedastic error. Other authors have made other assumptions.

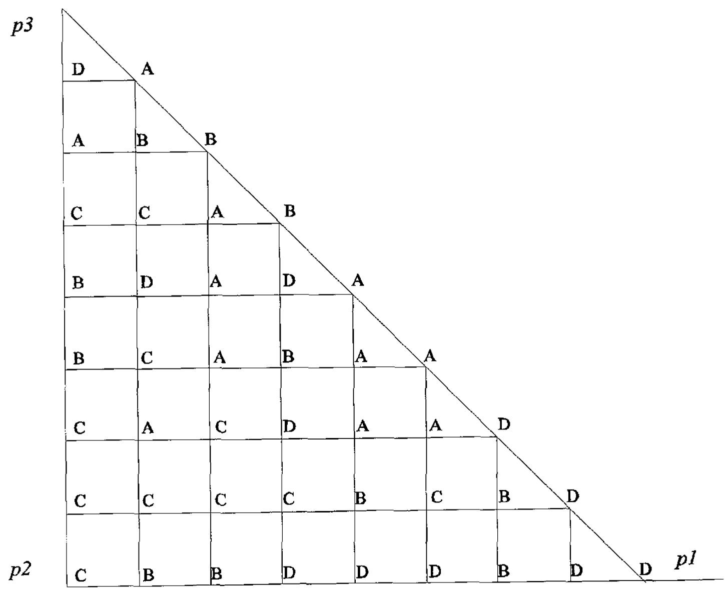
It is clear from the above papers that the error term is often sizable, and the question naturally arises as to whether the design of the experiment itself (to be specific, pairwise choice or complete ranking or whatever) contributes to the importance and structure of the error term. Unfortunately, we cannot answer this question using the data from the experiments in the papers cited above because the different types of experiment were run with different participants. The aim of the present paper is to try to answer this question using an experimental design in which participants complete both a pairwise choice experiment and a complete ranking experiment.

The plan of the paper is as follows: Section 2 describes the experimental design, Section 3 reports on the estimated preference functionals, and Section 4 discusses the estimation procedure. The results are presented in Section 5, and we conclude in Section 6. The main conclusion at this stage of the work is that the same individual in different experimental contexts apparently behaves according to different preference functionals. Moreover, the importance of the error term seems to vary from context to context.

2. The experimental design

The experiment results reported in this paper are the findings of two interlinked experiments, called *complete ranking* (CR) and *pairwise choice* (PC), conducted with forty subjects over a period of a week. Each participant was required to participate in both experiments, the two experiments had to be completed on two separate days, and we monitored the booking for the experiments in a way that half of the participants should[1] have done complete ranking first and the other half should have done pairwise choice first. The idea behind this was to try to detect whether the order in which the subjects completed the experiments affected their behavior. This did not appear to be the case.

The complete ranking experiment involved the participant in placing forty four risky prospects in the order of the participant's preference—from the one he or she preferred the most through to the one he or she preferred the least. The forty-four lotteries could be pictured as points in four Marschack-Machina triangles (see Figure 1). We divided each of the four triangle sides into eighths,[2] obtaining a grid with forty five intersection points, excluding the top vertex where the probability of the highest outcome was 1. We assigned



Key: A — triangle 1
B — triangle 2
C — triangle 3
D — triangle 4

p_1 — probability of worst outcome
p_2 — probability of middle outcome
p_3 — probability of best outcome

Figure 1. The chosen lotteries and the Marschak-Machina triangle.

to each point in the grid a position, then with a random number table we chose eleven positions for each triangle excluding those that were already chosen. There were just four final outcomes and they were chosen to be −£5, £5, £15, and £25.[3] From each triangle we took eleven lotteries; each lottery was represented in the form of colored pie chart. The size of each segment of the pie chart represented the probability of having a certain outcome, the outcome was labeled on each segment and each outcome corresponded to a color: −£5 red, £5 light blue, £15 medium blue, and £25 dark blue.[4]

The forty four prospects presented to the participants were initially divided (by us) into four separate piles each of eleven prospects (each pile corresponded to one of the four triangles; in each pile there were only three possible outcomes, so the first pile involved the three amounts −£5, £5, £15; the second pile −£5, £5, £25; the third pile −£5, £15, £25; and the fourth pile £5, £15, £25). Participants were advised to work pile by pile, ordering the eleven prospects in each pile, from the one that he or she preferred the most through to the one that he or she preferred the least. The participant was then asked to merge the four piles so as to obtain a complete ranking of all forty four prospects, from the one he or she preferred the most through to the one he or she preferred the least. We followed the procedure described above because we thought that to rank first the four piles (each with only three outcomes) would be easier for the participants and would enable them to develop their own ranking technique. We are aware that the final ranking so obtained might be different from the ranking we would have obtained if we had asked the participants to rank all forty four lotteries together.

To motivate the participants to carry out their ranking with appropriate care, we used the following incentive mechanism: when the participant had completed his or her ranking, two of the forty four risky prospects were chosen at random, and the one (of the two) highest in his or her ranking was played out and he or she was paid according to the outcome. (For obvious reasons, the actual playing out took place after the completion of both experiments.)

Hence the actual outcome of the experiment (that is, the payment received by the participant) depended on the participant's ranking, on which two of the forty four prospects were randomly chosen, and on the outcome of the playing out of the preferred prospect of these two.[5]

The pairwise choice experiment consisted of ninety four pairwise choice questions, each asking which of two risky prospects the participant preferred. In each of the ninety four questions the two prospects were represented in the form of a pie chart with particular segments representing particular outcomes as in the complete ranking experiment. The pairs of lotteries were taken from the same Marschak-Machina triangles that we used for the complete ranking experiment. In order to build up the question set, we took from each triangle all the nondominated pairs; after that we reversed those questions (we moved the lottery that originally was on the right to the left and the lottery that originally was on the left to the right). We also included two dominated pairs from each triangle. Finally we randomized the order of these questions. The participant was required to write on an answer sheet which of the two prospects in each pair he or she preferred. The participants were not allowed to express indifference between the two prospects.[6]

Once the participant had answered all ninety four questions (indicating which of the two prospects in each question was preferred), one question was chosen at random and the prospect the participant said he or she preferred on that question was played out. (As before, this playing out took place after the participant had completed both experiments.)

In both experiments the chosen lottery was played out by placing the copy of the chosen circle on the top of a spinning device (a continuous roulette wheel) with a freely spinning pointer, and the experimenter set the pointer spinning.

In both the complete ranking and the pairwise choice experiments the participant ended up with −£5, £5, £15, or £25. The total would have been thus one of −£10, £0, £10, £20, £30, £40, or £50. In addition, each participant was given a participation fee of £5, so the final payment was one of −£5, £5, £15, £25, £35, £45, or £55. Because of the possibility of a loss of £5, we asked the participant to bring a deposit of £5 when he or she came for his or her first experiment.

3. The preference functionals estimated

First, some notation. Let $\mathbf{x} = (x_1, x_2, x_3, x_4)$ denote the four outcomes used in the experiment; let $\mathbf{r} = (r_1, r_2, r_3, r_4)$ denote the respective probabilities in a gamble. Let $V(\mathbf{r})$ denote a preference functional. We describe below each estimated preference functional's form, the normalization adopted, and the parameters to be estimated. Note that an additional parameter to be estimated is the parameter defining the *spread* or noise in the error ter; we discuss this in Section 4.

3.1. *Subjective expected utility theory* (EU)

$$V(\mathbf{r}) = r_1 u(x_1) + r_2 u(x_2) + r_3 u(x_3) + r_4 u(x_4)$$

For expected utility theory, as is well known, the utility function is unique only up to a linear transformation, which means that one can set *two* utility values arbitrarily. We choose to normalize so that $u(x_1) = 0$ and $u(x_4) = 30$. So the parameters to be estimated are $u(x_2)$ and $u(x_3)$.

3.2. *Risk neutrality* (RN)

This is a special case of expected utility theory, preference being determined simply by the expected value of the prospect:

$$V(\mathbf{r}) = r_1 x_1 + r_2 x_2 + r_3 x_3 + r_4 x_4.$$

3.3. *Prospective reference theory* (PR)

Here the preference functional is

$$V(\mathbf{r}) = \lambda[r_1 u(x_1) + r_2 u(x_2) + r_3 u(x_3) + r_4 u(x_4)] + (1 - \lambda)[c_1 u(x_1) + c_2 u(x_2) + c_3 u(x_3) + c_4 u(x_4)],$$

where $c_i = 1/n(\mathbf{r})$ if $r_i > 0$ and 0 otherwise, and where $n(\mathbf{r})$ is the number of nonzero elements in the vector $\mathbf{r}$. Prospective reference's preference functional can be thought as a weighted average of the expected utility functional using the correct probability weights and the expected utility functional using equal probability weights for the nonnull outcomes. Clearly, the normalization here is the same as the expected utility normalization: $u(x_1) = 0$ and $u(x_4) = 30$. Thus, the parameters to be estimated are $u(x_2)$ and $u(x_3)$, as in expected utility theory, plus the additional parameter λ. Viscusi [1989] refers to λ as the weight of the "relative information content. . .associated with the stated lottery" and $(1 - \lambda)$ as the weight of the "relative information content. . .associated with the reference lottery." Note that if $\lambda = 1$, then prospective-reference theory reduces to expected utility theory.

3.4. Disappointment aversion theory (DA)

Here our characterization appears different from that in Gul [1991], but it can be shown that they are identical (see Hey and Orme [1994]), ours is more useful for our purposes. Here the preference functional is

$$V(\mathbf{r}) = \min\{V_1, V_2, V_3\} \text{ where,}$$

$$V_1 = [(1+\beta)r_1u(x_1) + (1+\beta)r_2u(x_2) + (1+\beta)r_3u(x_3) + r_4u(x_4)]/(1+\beta r_1+\beta r_2+\beta r_3)$$

$$V_2 = [(1+\beta)r_1u(x_1) + (1+\beta)r_2u(x_2) + r_3u(x_3) + r_4u(x_4)]/(1+\beta r_1+\beta r_2)$$

$$V_3 = [(1+\beta)r_1u(x_1) + r_2u(x_2) + r_3u(x_3) + r_4u(x_4)]/(1+\beta r_1).$$

Again, we normalize so that $u(x_1) = 0$ and $u(x_4) = 30$. The parameters to be estimated are $u(x_2)$ and $u(x_3)$, as in expected utility theory, plus the additional parameter β. If $\beta = 0$, disappointment aversion theory reduces to expected utility theory.

3.5. Weighted utility theory (WU)

Here the preference functional is

$$V(\mathbf{r}) = [w_1r_1u(x_1) + w_2r_2u(x_2) + w_3r_3u(x_3) + w_4r_4u(x_4)]/[r_1w_1 + r_2w_2 + r_3w_3 + r_4w_4].$$

Here w_1, w_2, w_3, and w_4 are the *weights* attached to x_1, x_2, x_3, and x_4. Once again, the normalization is that $u(x_1) = 0$ and $u(x_4) = 30$; in addition, we set the weights attached to x_1 and x_4 equal to unity. The parameters to be estimated are $u(x_2)$ and $u(_3)$, as in expected utility theory, plus the additional parameters w_2 and w_3. Note that weighted utility theory reduces to expected utility theory if w_2 and w_3 are both equal to unity.

3.6. Rank-dependent expected utility theory (RP *and* RQ)

The rank-dependent expected utility model is the outcome of a number of contributions, including Quiggin (1982) and Yaari (1987). It models behavior as ranking the outcomes in order of preference and then distorting the decumulative probabilities (of getting at least a given outcome) through some probability weighting function. We estimate two versions, the first (RP) assuming that the probability weighting function takes the specific functional form of the power function:

$$V(\mathbf{r}) = u(x_1)W_1(r) + u(x_2)W_2(r) + u(x_3)W_3(r) + u(x_4)W_4(r),$$

where

$$W_1(\mathbf{r}) = \pi(r_1 + r_2 + r_3 + r_4) - \pi(r_2 + r_3 + r_4)$$
$$W_2(\mathbf{r}) = \pi(r_2 + r_3 + r_4) - \pi(r_3 + r_4)$$
$$W_3(\mathbf{r}) = \pi(r_3 + r_4) - \pi(r_4)$$
$$W_4(\mathbf{r}) = \pi(r_4),$$

where the weighting function π takes the specific form as follows:

$$\pi(r) = r^{\gamma}.$$

The second version (RQ) of rank-dependent expected utility assumes that the weighting function $\pi(.)$ takes the specific functional form recommended by Quiggin [1982]. It allows the probability function to be S-shaped:

$$\pi(r) = r^{\gamma}/[r^{\gamma} + (1 - r)^{\gamma}]^{1/\gamma}.$$

The normalization again is that $u(x_1) = 0$ and $u(x_4) = 30$; thus the parameters to be estimated are $u(x_2)$ and $u(x_3)$, as in expected utility theory, plus the additional parameter γ. In both the rank dependent formulations, if $\gamma = 1$, then $\pi(r) \equiv r$, and this preference functional reduces to expected utility.

4. The estimation procedure

In order to estimate the various functionals, we need to make some assumption about the stochastic process generating our data. For both data sets (complete ranking and pairwise choice) we assume that the participant evaluates each risky prospect by some deterministic preference functional but that there is a random component added to this deterministic evaluation—with the final decisions being made on the basis of the deterministic evaluation *plus* the stochastic term. We naturally term this stochastic component the *error* term.

More precisely, we proceed as follows. Suppose $\mathbf{r}$ is a given risk prospect (a vector of probabilities) and suppose that $V(.)$ is the subject's true underlying deterministic preference function. Then the subject's *actual* evaluation of $\mathbf{r}$ is given by $V(\mathbf{r})$. However, we assume that the subject bases his or her decisions on the value of

$$V(\mathbf{r}) + \epsilon$$

where ϵ is an error term.

So far this is innocuous; to give it empirical content, we have to specify an assumption about ϵ. We simply assume that the ϵ (across all risky prospects) are independently and identically distributed with an extreme value distribution with parameters 0 and σ. Hence, for all ϵ, we have

$$\text{Prob}(\epsilon \leq t) = \exp[-\exp(-t/\sigma)].$$

The mean of this distribution is zero—implying an assumption that there is no *bias* in the individual's decision-making. The parameter σ determines the spread of the distribution: the larger σ is, the more spread out the distribution. Or, as far as the individual's decision-making is concerned, the greater is σ, the greater is the error or noise in the individual's decisions.

The above, of course, is the basis for the logit specification—relating to binary choice, which is precisely the problem in (each of) the pairwise choice decision tasks in the pairwise choice experiment. More precisely, as far as the pairwise choice data are concerned, the stochastic assumptions made above lead naturally to the logit model: left is preferred to right on any one pairwise choice question if and only if

$$V(\mathbf{r}_l) + \epsilon_l > V(\mathbf{r}_r) + \epsilon_r,$$

where $\mathbf{r}_l$ and $\mathbf{r}_r$ are, respectively, the probability vectors in the left-hand and right-hand risky prospects. With ϵ_l and ϵ_r being independently and identically distributed extreme value variables, we immediately get the conventional logit specification.

Moreover, it follows (see Beggs, Cardell, and Hausman [1981], that for the data coming from the complete ranking experiment, the appropriate stochastic specification is an ordered logit specification on the ranked data obtained from the experiment. The stochastic specification described above implies that the *relative* position of two lotteries in the ranking is independent of which other lotteries are to be ranked. Note further that one very attractive property of the extreme value distribution is that the expression $P(Y_1 > Y_2 > \ldots > Y_n)$ can be explicitly evaluted, whereas with the normal distribution (which was assumed by Hey and Di Cagno [1990] and Hey and Orme [1994]) the evaluation of this expression requires numerical approximation of an $(n-1)$-fold integral.

Then, for an observed ordinal ranking of the choices put in descending order, the probability of the subject's observed ranking is

$$P(Y_1 > Y_2 > \ldots > Y_n) = \prod_{i=1}^{n-1} \left[\frac{\exp[V(\mathbf{r}_i)]}{\sum_{j=i}^{n} \exp[V(\mathbf{r}_j)} \right] = \prod_{i=1}^{n} \left[\frac{\exp[V(\mathbf{r}_i)]}{\sum_{j=i}^{n} \exp[V(\mathbf{r}_j)} \right].$$

Hence the log-likelihood function LL is given by

$$\text{LL} = \sum_{i=1}^{n} V(\mathbf{r}_i) - \sum_{i=1}^{n} \log \left[\sum_{j=1}^{n} \exp[V(\mathbf{r}_j) \right].$$

For both specifications, estimation was carried out using the MAXLIK maximum likelihood procedure within the GAUSS suite of programs. This gives estimated values for all the parameters, their (asymptotic) standard errors, and various goodness-of-fit measures and diagnostic tests.

5. Estimation results

We estimated the seven preference functionals described in Section 3 for each of the forty participants and for three data sets: data set PC from the pairwise choice experiment, data set CR from the complete ranking experiment, and data set PCR from the combined data from both experiments. For each estimated preference functional for each subject on each data set, we can measure the goodness of fit of the functional using the maximized value of the log-likelihood. However, since different preference functionals have differing numbers of parameters, some correction to the log-likelihoods is necessary before meaningful comparisons are possible. We used the correction based on the Aikake information criterion (AIC) (see Amemiya [1980]). Table 1 lists the corrected log-likelihoods using the Aikake information criterion for the fitted preference functionals for each model and each subject. The Akaike information criterion is given by

$$\text{AIC} = -2 \log L(\hat{\alpha})/T + 2k/T,$$

where $L(\hat{\alpha})$ is the maximized log-likelihood for a particular estimated preference functional, k is the number of the estimated parameters in that functional, and T is the number of observations. Akaike suggests the ranking of different models on the basis of this: the smaller (in absolute value) is AIC, the better the model. Since T is constant across all models, this implies ranking the models according to the magnitude of CLL $= \log L(\hat{\alpha}) - k$ (here we are using CLL to denote the corrected log-likelihood).

The GAUSS maximum likelihood routine failed to converge on the PC data set in two cases (model WU for subjects 25 and 34); and on the CR data set in a number of cases (model PR for subject 15; models EU, DA, and PR for subject 18; while for subjects 22 and 24 none of the models apart from RN converged). There are a number of reasons why that might be so, including the fact that the likelihood function is not well behaved for the DA model: in certain cases the likelihood function has a cusp at a β value of zero (the EU special case), in which case either the maximum occurs at $\beta = 0$ (the likelihood function has its maximum at $\beta = 0$) or there are two local maxima either side of β (so the likelihood function is bottom-shaped) that have to be checked numerically. Additionally, for the other models, there are occasionally identification problems—perhaps not surprising in view of the nature of the data—which means that the maximizing routine drifts all over the parameter space without converging. This, for example, was the case with the CR data set and subjects 22 and 24, who were quite clearly risk neutral. In such cases, there are infinitely many sets of parameters for the EU and the top-level functionals that will fit the data perfectly.

In order to make sense of Table 2 and subsequent analyses, we must make clear the nested structure of the various models; this is illustrated in Figure 2. So, for example, two parameter

Table 1. Aikake corrected log-likelihoods.

a. Data Set PC

S	RN	EU	DA	PR	RQ	RP	WU
1	-39.58	-19.08	-19.36	-10.89	-11.90	-17.67	-14.89
2	-63.24	-23.22	-24.07	-16.95	-15.42	-18.92	-5.00
3	-56.85	-40.44	-41.36	-31.89	-31.35	-36.33	-25.18
4	-38.74	-25.64	-26.37	-24.40	-24.22	-26.62	-21.66
5	-48.75	-19.81	-20.34	-20.49	-20.40	-20.63	-21.33
6	-60.94	-15.96	-16.44	-14.78	-14.74	-14.90	-14.98
7	-65.17	-30.29	-30.90	-31.27	-31.29	-30.50	-31.42
8	-64.51	-14.89	-14.48	-11.00	-13.82	-13.44	-15.73
9	-48.22	-33.47	-32.12	-33.33	-29.30	-28.15	-29.37
10	-30.53	-16.56	-17.06	-16.92	-17.10	-16.99	-15.45
11	-39.58	-32.63	-32.95	-31.46	-27.76	-29.76	-29.04
12	-51.19	-47.13	-48.13	-47.39	-47.24	-48.08	-47.05
13	-52.92	-12.58	-13.58	-5.39	-5.39	-5.39	-6.39
14	-36.01	-28.96	-29.90	-29.88	-29.96	-29.90	-30.89
15	-62.62	-13.81	-14.51	-14.77	-14.65	-14.78	-15.28
16	-57.15	-19.50	-20.50	-19.87	-20.28	-20.17	-21.46
17	-64.62	-7.99	-8.92	-8.96	-8.99	-8.91	-8.85
18	-33.98	-17.95	-18.95	-18.89	-18.95	-18.93	-19.94
19	-38.74	-26.39	-26.88	-26.47	-26.23	-25.80	-27.07
20	-46.53	-22.50	-23.50	-23.23	-23.18	-23.50	-21.72
21	-44.68	-30.72	-31.64	-31.65	-31.65	-31.66	-32.65
22	-63.80	-5.25	-5.39	-5.39	-5.39	-5.39	-6.39
23	-36.01	-12.52	-13.52	-13.51	-13.31	-12.40	-13.57
24	-29.24	-17.18	-18.14	-15.82	-12.97	-17.37	-15.40
25	-65.81	-14.97	-15.97	-14.23	-9.92	-15.28	
26	-52.08	-29.88	-30.43	-21.62	-22.64	-28.74	-18.50
27	-44.68	-37.50	-37.69	-26.73	-30.29	-38.30	-28.17
28	-53.32	-36.81	-37.81	-35.70	-36.16	-36.62	-37.84
29	-56.85	-21.47	-19.60	-22.45	-22.47	-22.45	-18.70
30	-49.77	-30.62	-31.48	-31.33	-31.14	-31.61	-32.16
31	-41.17	-35.58	-36.58	-33.43	-34.46	-36.46	-26.45
32	-41.17	-26.68	-27.68	-27.52	-27.57	-25.05	-23.19
33	-56.21	-29.99	-30.99	-30.62	-30.75	-30.84	-28.79
34	-47.67	-43.83	-44.83	-39.97	-40.71	-36.24	
35	-33.98	-15.72	-16.72	-15.72	-15.37	-16.61	-15.71
36	-33.98	-23.86	-24.86	-23.04	-23.39	-23.20	-23.07
37	-60.07	-26.18	-26.74	-24.46	-25.02	-24.80	-25.76
38	-55.20	-47.61	-48.61	-48.33	-48.27	-48.40	-47.56
39	-29.24	-13.54	-14.54	-13.70	-14.15	-14.48	-15.05
40	-33.98	-25.48	-26.48	-26.34	-26.47	-26.11	-23.21

Table 1. Continued.

b. Data Set CR

S	RN	EU	DA	PR	RQ	RP	WU
1	-58.32	-46.67	-47.67	-47.51	-47.54	-46.15	-45.58
2	-83.65	-59.66	-60.66	-58.08	-57.10	-60.47	-59.08
3	-98.99	-48.06	-49.06	-42.02	-43.42	-49.02	-46.15
4	-70.41	-58.62	-59.60	-39.95	-44.57	-53.41	-49.39
5	-59.13	-54.92	-55.91	-55.88	-55.42	-55.83	-55.50
6	-97.57	-33.85	-34.85	-34.85	-34.84	-31.40	-35.54
7	-99.44	-50.42	-50.94	-50.50	-50.83	-51.42	-51.95
8	-97.64	-84.59	-85.59	-74.59	-73.99	-82.09	-74.97
9	-92.62	-85.03	-86.03	-85.62	-86.02	-85.90	-86.91
10	-70.90	-66.71	-67.47	-65.67	-60.63	-63.80	-62.63
11	-31.48	-31.52	-32.51	-30.69	-31.36	-31.42	-31.82
12	-51.67	-43.09	-44.09	-43.92	-43.94	-43.87	-43.40
13	-84.89	-53.11	-54.11	-53.63	-53.98	-47.58	-53.95
14	-74.39	-66.80	-66.72	-65.92	-65.24	-67.62	-66.40
15	-125.60	-127.34	-128.34		-124.03	-124.03	-126.72
16	-38.01	-32.01	-33.01	-32.98	-33.00	-32.74	-33.57
17	-103.77	-38.33	-35.17	-37.23	-32.78	-36.99	-36.31
18	-23.65				-21.97	-22.06	-21.92
19	-66.82	-65.09	-66.09	-58.20	-58.03	-57.45	51.01
20	-93.65	-69.96	-70.96	-69.00	-68.05	-70.29	-69.29
21	-42.91	-43.25	-43.80	-44.16	-44.25	-44.23	-45.09
22	-23.65						
23	-84.25	-51.34	-52.34	-52.34	-52.32	-49.66	-45.96
24	-23.65						
25	-98.80	-31.27	-32.27	-31.20	-31.53	-30.22	-31.27
26	-90.22	-47.54	-48.49	-48.10	-48.51	-48.01	-48.82
27	-69.08	-61.39	-62.39	-61.82	-62.15	-60.37	-59.17
28	-102.35	-49.29	-50.29	-49.77	-47.32	-49.29	-47.84
29	-89.23	-61.79	-62.79	-58.32	-58.60	-56.22	-59.26
30	-84.24	-46.79	-47.79	-47.77	-47.79	-44.74	-48.77
31	-108.38	-104.32	-105.31	-104.94	-105.30	-105.23	-104.52
32	-68.29	-52.56	-53.56	-53.01	-53.55	-53.53	-52.73
33	-85.04	-81.52	-82.52	-74.10	-67.62	-70.70	
34	-71.25	-70.72	-71.62	-71.35	-71.41	-69.72	-69.19
35	-102.36	-103.09	-97.26	-104.07	-104.08	-103.23	-101.18
36	-125.11	-126.45	-127.29	-126.52	-126.31	-126.84	-124.21
37	-84.71	-61.35	-62.35	-62.07	-61.92	-60.19	-62.57
38	-90.82	-51.93	-52.93	-48.36	-49.66	-47.98	-50.34
39	-64.14	-64.39	-65.21	-60.51	-60.67	-65.03	-64.84
40	-73.65	-73.18	-74.18	-74.08	-74.10	-73.75	-73.07

Table 1. Continued.

c. Data Set PCR

S	RN	EU	DA	PR	RQ	RP	WU
1	-98.32	-70.90	-71.90	-71.42	-71.30	-71.41	-70.47
2	-154.66	-90.17	-91.17	-81.38	-78.14	-88.52	-77.62
3	-154.91	-100.36	-101.36	-90.01	-88.27	-98.89	-89.31
4	-108.18	-84.02	-84.65	-69.16	-71.71	-82.58	-71.92
5	-112.83	-84.32	-85.32	-85.32	-85.24	-84.35	-85.87
6	-158.79	-48.52	-49.52	-48.68	-48.17	-49.29	-48.55
7	-168.83	-82.54	-83.38	-82.63	-83.03	-83.05	-82.97
8	-166.05	-107.84	-108.84	-100.43	-99.71	-105.10	-107.30
9	-140.17	-117.64	-118.54	-117.20	-117.96	-117.90	-118.53
10	-101.35	-85.92	-86.42	-85.03	-79.53	-80.94	-81.15
11	-83.56	-70.38	-70.93	-71.37	-71.22	-71.26	-71.51
12	-114.23	-102.24	-103.24	-103.24	-103.22	-102.94	-103.96
13	-137.23	-67.59	-68.59	-65.18	-66.55	-56.72	-66.02
14	-109.76	-102.16	-101.42	-101.92	-101.04	-101.09	-102.12
15	-188.62	-178.18	-178.49	-177.42	-179.18	-176.09	-174.58
16	-124.42	-91.11	-92.11	-91.17	-91.45	-89.31	-90.72
17	-170.35	-46.29	-42.23	-44.11	-39.17	-45.28	-42.01
18	-70.69	-50.89	-51.73	-51.85	-51.69	-51.43	-52.61
19	-104.78	-90.12	-91.12	-87.99	-88.22	-88.30	-85.99
20	-139.86	-93.38	-94.38	-93.93	-93.85	-94.15	-93.19
21	-97.09	-79.42	-80.42	-80.17	-79.96	-80.16	-80.66
22	-144.10	-100.96	-101.96	-101.43	-101.55	-100.73	-102.12
23	-121.61	-71.40	-72.40	-71.93	-71.29	-71.85	-64.46
24	-61.94	-30.05	-31.05	-30.77	-30.21	-30.98	-31.25
25	-170.59	-54.73	-55.73	-55.35	-55.50	-54.88	-55.65
26	-141.36	-78.21	-79.21	-72.95	-74.26	-79.11	-77.02
27	-113.88	-96.28	-97.28	-90.52	-92.23	-96.62	-88.63
28	-155.06	-103.17	-104.17	-102.39	-103.77	-101.42	-104.45
29	-146.23	-82.25	-83.25	-80.85	-80.80	-78.67	-82.77
30	-133.10	-79.58	-80.58	-80.15	-79.74	-78.73	-80.93
31	-156.75	-154.09	-155.09	-153.80	-154.28	-154.34	-151.97
32	-108.73	-100.31	-101.25	-101.30	-101.21	-101.14	-99.94
33	-142.16	-114.41	-115.41	-106.80	-103.63	-109.16	-102.48
34	-119.32	-112.30	-113.30	-111.80	-111.93	-104.30	-101.84
35	-143.78	-141.04	-136.71	-140.73	-140.40	-139.37	-138.39
36	-181.77	-182.63	-183.61	-180.52	-180.79	-182.34	-176.56
37	-148.00	-91.29	-92.29	-90.38	-90.49	-88.56	-91.75
38	-145.38	-109.93	-110.93	-110.53	-110.88	-109.21	-109.90
39	-92.66	-83.17	-84.08	-80.45	-78.74	-83.24	-82.81
40	-107.09	-98.64	-99.64	-99.40	-99.28	-99.45	-96.12

S = Subjects.

Table 2. Results of likelihood ratio test.

a. Data Set PC

S	EU vRN	DA vEU	PR vEU	RQ vEU	RP vEU	WU vEU
1	**		**	**	*	**
2	**		**	**	**	**
3	**		**	**	**	**
4	**		*	*		**
5	**					
6	**		*	*	*	
7	**					
8	**		**	*	*	
9	**	*		**	**	**
10	**					*
11	**		*	**	**	**
12	**					
13	**		**	**	**	**
14	**					
15	**					
16	**					
17	**					
18	**					
19	**					
20	**					
21	**					
22	**					
23	**					
24	**		*	**		*
25	**			**		—
26	**		**	**	*	**
27	**		**	**		**
28	**		*			
29	**	*				**
30	**					
31	**		*	*		**
32	**				*	**
33	**					*
34	**		**	**	**	—
35	**					
36	**					
37	**		*	*	*	
38	**					
39	**					
40	**					*
	40**	0	8**	11**	6**	12**
	40*	2*	7*	5*	6*	4*

Table 2. Continued.

b. Data Set CR

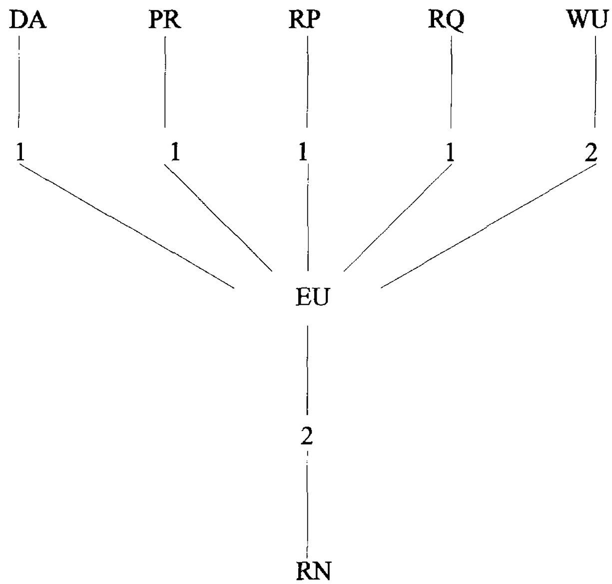
S	EU vRN	DA vEU	PR vEU	RQ vEU	RP vEU	WU vEU
1	**					*
2	**		*	**		
3	**		**	**		*
4	**		**	**	**	**
5	**					
6	**				**	
7	**					
8	**		**	**	**	**
9	**					
10	**		*	**	**	**
11						
12	**					
13	**				**	
14	**			*		
15			—	**	**	
16	**					
17	**	**	*	**	*	*
18	—	—	—	—	—	—
19	*		**	**	**	**
20	**		*	*		
21						
22	—	—	—	—	—	—
23	**				*	**
24	—	—	—	—	—	—
25	**				*	
26	**					
27	**				*	*
28	**			*		*
29	**		**	**	**	*
30	**				*	
31	**					
32	**					
33	**		**	**	**	—
34					*	*
35		**				*
36						*
37	**				*	
38	**		**	*	**	*
39			**	**		
40						
	28**	2**	8**	11**	10**	5**
	1*	0	4*	4*	7*	10*
	3nc	3nc	4nc	3nc	3nc	4nc

Table 2. Continued.

c. Data Set PCR

S	EU vRN	DA vEU	PR vEU	RQ vEU	RP vEU	WU vEU
1	**					
2	**		**	**	*	**
3	**		**	**	*	**
4	**		**	**	*	**
5	**					
6	**					
7	**					
8	**		**	**	**	
9	**					
10	**			**	**	**
11	**					
12	**					
13	**		**	*	**	*
14	**			*	*	
15	**				*	**
16	**				*	
17	**	**	*	**	*	**
18	**					
19	**		*	*	*	**
20	**					
21	**					
22	**					
23	**					**
24	**					
25	**					
26	**		**	**		*
27	**		**	**		**
28	**				*	
29	**		*	*	**	
30	**					
31	**					*
32	**					
33	**		**	**	**	**
34	**				**	**
35	**	**			*	**
36			*	*		**
37	**				**	
38	**					
39	**		**	**		
40	**					*
	39**	2**	9**	10**	7**	13**
	39*	4*	4*	5*	10*	4*

S = Subjects.
**Additional parameters significant at 1 percent.
*Additional parameters significant at 5 percent.



Key: DA: Disappointment Aversion theory
EU: Expected Utility theory
PR: Prospective Reference theory
RN: Risk Neutrality
RP: Rank Dependent with Power weighting function
RQ: Rank Dependent with Quiggin weighting function
WU: Weighted Utility theory

Figure 2. The nested structure of the estimated functionals.

restrictions take us from WU down to EU, and two more restrictions take us down to RN. Of course, this is all specific to our particular experimental setup and, in particular, to the fact that our experiment involved just four final outcomes. More generally, PR and DA each have always just one parameter more than EU, while EU has two more parameters than RN (when there are just four final outcomes); WU, in contrast, has $(n - 2)$ more parameters than EU (where n is the number of final outcomes); in general, the number of additional parameters associated with the RQ and RP models depend on the specific form of the weighting function; in our two specifications they have only one additional parameter, since the two weighting functions involve just one extra parameter.

We can make use of this nested structure to test whether the more general models fit better than the less general models. So, for example, since going from WU to EU requires two parameter restrictions, we can test the null hypothesis that the restricted model (less general) is accepted and hence that the added parameters are not significant, using the usual likelihood ratio test. More specifically, if LL_a and LL_b are the estimated log-likelihoods for models a and b, respectively, and if model a is obtained from model b by imposing k parameter restrictions, then under the null hypothesis that these restrictions are satisfied, the test statistic $2(LL_b - LL_a)$ has a chi-squared distribution with k degrees of freedom. We have applied this test looking for significance at 1 percent (5 percent). Table 2 gives the results obtained from these tests.

For data set PC, EU always fits the data better than RN. DA's additional parameter β is never significant at 1 percent (only twice at 5 percent level). The rank dependent models, in both the Quiggin and the Power version (RQ and RP), have significant additional parameters, respectively for 11 and 6 (16 and 12) subjects, while PR has a significant additional parameter for eight (fifteen) subjects and WU for twelve (sixteen) subjects.

For data set CR, the EU model fits the data better than RN for twenty-eight (twenty-nine) subjects while the GAUSS routine does not converge for three subjects. Only twice is DA's additional parameter β significant; the rank dependent models still perform well: RQ's additional parameter is significant for eleven (fifteen) subjects, RP for ten (seventeen) subjects, and PR for eight (twelve) subjects. WU performs worse on the data set CR because its parameters are significant only for five (fifteen) subjects.

For data set PCR, EU's additional parameters (relative to RN) are significant for all except one subject; and DA fits the data better than EU for two subjects. WU's parameters are significant for thirteen (seventeen) subjects. The additional parameters of the top level models RQ, RP, and PR are significant, respectively, for ten (fifteen), seven (seventeen), and nine (thirteen) subjects.

In Table 3 we summarize the results coming from Table 1 and Table 2 by defining for each subject the best model or the "winner" (at 5 percent). We define the winner as RN if neither EU nor one of the top-level models have significant additional parameters; or EU if it had significant additional parameters compared to RN and if none of the top-level functionals fitted significantly better than EU. If RN and EU were rejected in favor of one or more of the top-level preference functionals, we considered the winner that model that had the largest corrected log-likelihood. Let us look first at data sets PC and CR: DA is always the worse model, being never the "winner" for these two data sets. For data set PC, EU is the winner eighteen times, WU ten times, PR six times, RQ four times, and RP two times. For data set CR, RN performs better than the other models, being the winner for ten subjects, EU is the winner eight times, the RQ eight times, RP seven times, and PR two times. In data set PCR, EU is again the best model, being the winner sixteen times; after that WU, which is the winner nine times; while DA model is definitely the one that performs the worst.

In Table 4 we present a matrix showing whether models that were winners in the PC data set, kept this position in the CR data set. The results show that the only model that kept the position was EU, for four subjects. This has interesting implications for debate concerning which model is the true model.

Table 5 reports on the Chow test on the stability of the parameters across the two data sets: if we denote by LL_{PC}, LL_{CR}, and LL_{PCR} the respective maximized log-likelihoods of a particular estimated model under the data sets PC, CR, and PCR, then under the null hypotheses that the coefficients are the same in both data sets, the statistic

$$2[(LL_{PC} + LL_{CR}) - LL_{PCR}]$$

Table 3. Summary of the "winners" at 5 percent for the three data sets.

S	PC	CR	PCR
1(CR)	PR	WU	EU
2(CR)	WU	RQ	WU
3(CR)	WU	PR	RQ
4(CR)	WU	PR	PR
5(CR)	EU	EU	EU
6(PC)	RQ	RP	EU
7(PC)	EU	EU	EU
8(CR)	PR	RQ	RQ
9(CR)	RP	EU	EU
10(CR)	WU	RQ	RQ
11(CR)	RQ	RN	EU
12(CR)	EU	EU	EU
13(PC)	PR	RP	RP
14(PC)	EU	RQ	RQ
15(PC)	EU	RN	WU
16(CR)	EU	EU	RP
17(CR)	EU	RQ	RQ
18(CR)	EU	—	EU
19(PC)	EU	WU	WU
20(PC)	EU	RQ	EU
21(PC)	EU	RN	EU
22(CR)	EU	RN	EU
23(CR)	EU	WU	WU
24(PC)	RQ	RN	EU
25(PC)	RQ	RP	EU
26(CR)	WU	EU	PR
27(CR)	PR	WU	WU
28(PC)	PR	RQ	RP
29(PC)	WU	RP	RP
30(CR)	EU	RP	EU
31(CR)	WU	EU	WU
32(PC)	WU	EU	EU
33(PC)	WU	RQ	WU
34(PC)	RP	RN	WU
35(PC)	EU	RN	DA
36(PC)	EU	RN	RN
37(PC)	PR	RP	RP
38(PC)	EU	RP	EU
39(PC)	EU	RN	RQ
40(PC)	WU	RN	WU
Number of times that each model was the winner at the 5 percent level.			
DA	0	0	1
PR	6	2	2
RQ	4	8	6
RP	2	7	5
WU	10	4	9
EU	18	8	16
RN	0	10	1

S = Subjects.

Table 4. Transition matrix between data set PC and data set CR.

		Data set CR						
		RN	EU	DA	PR	WU	RP	RQ
	RN	0	0	0	0	0	0	0
	EU	5	4	0	0	2	2	3
	DA	0	0	0	0	0	0	0
Data set PC	PR	1	0	0	0	2	2	2
	WU	1	3	0	2	0	1	3
	RP	1	1	0	0	0	0	0
	RQ	2	0	0	0	0	2	0

Table 5. Tests of parameter stability.

Model	Significant at 5%	Significant at 1%
RN	18	16
EU	37	37
DA	35	35
PR	36	36
RQ	38	38
RP	38	38
WU	35	35

Note: If the test is significant, it indicates that the estimated coefficients on the two data sets or experiments are significantly different from each other.

will have a chi-square distribution with k degreees of freedom where k is the number of the estimated parameters in that particular model. Table 5 summarizes the results of carrying out this test for the seven preference functionals at both 5 percent and 1 percent significance level. Apart from RN, which has stable parameters for 60 percent of the subjects, the tests for the other models show that the parameters for each model and each subject are not stable across the data sets for about 90 percent of the subjects for each model (more precisely, it varies from 88 to 95 percent across the models). This is not encouraging news for those who look for stability across experimental environments. Indeed, it suggests strongly that different experimental designs lead to fundamentally different perceptions of the underlying true preference functional.

Table 6 summarizes some information concerning the relative magnitude of the error term in the PC experiment and that in the CR experiment. Recall that we have, for each subject, each preference functional, and each data set (experiment), an estimate of the magnitude of σ—the spread parameter in the extreme value distribution representing the error process. A higher (estimated) σ indicates a noisier error structure—that is, a less accurate decision process. Table 6 summarizes, for each preference functional, how many (of the forty subjects) had a higher estimated σ on the PC experiment than on the CR experiment. So, for example, with the RN functional, the estimated σ on the PC experiment was larger than on the CR experiment for twelve subjects and smaller for the remaining

Table 6. Comparison of the estimated σ in the two experiments.

Model	$\sigma_{PC} > \sigma_{CR}$	$\sigma_{PC} < \sigma_{CR}$	Total
RN	12	28	40
EU	16	21	37
DA	15	20	35
PR	15	21	36
RQ	16	22	38
RP	18	20	38
WU	11	24	35

twenty-eight subjects. So, for the RN formulation, the error magnitude is generally lower on the PC experiment than on the CR experiment. This finding contradicts the intuition of one of the authors (Hey). However, the situation is not so pronounced with the other preference functionals—where the percentage of the subjects for whom (the estimate) σ on the CR experiment was larger than (the estimated) σ on the PC experiment varied between fifty-two and fifty-eight.

Generally, error is important—as has already been well documented. Some insight into its magnitude (and possible cause) can be obtained by looking at particular sets of questions. First, one can look at the PC experiment and, in particular, at the questions, which appeared twice (though with left and right interchanged). Table 7 gives a summary. For each of the forty-three (or rather forty-two, see below) repeated pairs, and for each of the forty subjects, the table indicates whether the answers the subject gave were consistent (that is, left-right or right-left) or inconsistent (left-left or right-right). The rows indicate the questions pairs, and the columns the subjects. An asterisk (*) denotes consistency. It will be seen that consistency varies considerably—from a high of 100 percent consistent (question pairs 8 and 37) to a low of 60 percent (question pair 39). It also varies considerably across individuals. (Note that question pair 40, which has an apparently very high inconsistency rate, indicates a typographical error in inputting the data to the pie-drawing program: what we had thought was a reversed pair of questions was, in fact, not so because of this typographical error.)

A consistency rate as low as 50 percent would be obtained if answers were random; our consistency rates are well above this, and, for some questions, reach 100 percent. Of course, the consistency rate depends on two things: how close together the subjects perceive the two risky choices—the magnitude of $V(\mathbf{r}_l) - V(\mathbf{r}_r)$ and the magnitude of σ, the spread of the error term. A high consistency rate might result from a high $|V(\mathbf{r}_l) - V(\mathbf{r}_r)|$ or a low σ (or both). Whether Table 7 indicates that σ may vary from risky choice to risky choice, or simply that $V(\mathbf{r}_l) - V(\mathbf{r}_r)$ varies from question to question, is not immediate; further exploration is required.

What is startling, however, are the results of tests for the satisfaction or violation of *(first-order) dominance*: Table 8 reports on this for the PC experiment. Here there were eight questions (two in each triangle) for which one of the two risky choices dominated the other. Table 8 shows that over all eight questions and over all forty subjects there was just one case where dominance was violated—that is, where the subject chose the dominated choice. This is a mean violation rate of just 0.3 percent. In contrast, the average inconsistency

Table 7. Consistency in the answers of the repeated questions.

1	40
1	******************************1*********
2	0****0**1*****0**********0****0*****0***
3	**********1*0******1*****1***0*******0**
4	***********1*0***1**1*****1***11*1***0**
5	*************1************1************1
6	***0**1***0****0*0*0*******0************
7	*************************************1**
8	**
9	***0**0****1******10*******0***1********
10	*************1*0***********0*****1******
11	******0*************************11******
12	1*0********1*****************0***00****1
13	*******************1*******11***10***1**
14	**
15	********1****1************1*************
16	*****************01*1*********0*******00
17	****1*************0*0******0****1****0**
18	*******1***0**1**********************0**
19	1*******0*00*******0**********1**0***1***
20	*****1**0*1***********1**1**********0***
21	***********1*0*****************1*******0
22	******0***01*0*****0**********0****1*0**
23	*******************************0********
24	**
25	****0****0*0*****0000******0****0***01**
26	******1*1**1*****0*10*1**************1**
27	****0*1*0***********0*********0**1******
28	**********1******************01**0******
29	0**0*********1***********0*******1******
30	**0******0*******000**0**************00*
31	***********0****************************
32	***0******11*0****************1110**1**0
33	***********0*************************0**
34	***********************************1****
35	*************************************0**
36	**0*************************************
37	**
38	***0*0**00**************1**1*****1*00***
39	******10110**1**0*1*1******11*****11*1**
40	0*0*0000*0000000000000*0000*0000000*0*00
41	****************************1***********
42	*******0********************************
43	***************************0*1*****0****

Note: The columns go from subject number 1 to subject number 40; the rows go from question pair 1 to question pair 43.

* = the answer is consistent.

0 = the answer is left-left.

1 = the answer is right-right.

Table 8. Results of the test for dominance violations.

1	40
93	**
26	**
91	**
16	**
66	***********!****************************
76	**
54	**
81	**

Note: The columns go from subject number 1 to subject number 40; the rows indicate particular questions.

* = the answer respects dominance.

! = the answer violates dominance.

rate of the repeated pairs was 12 percent. On the CR experiment, there were 589 possible combinations involving a pair of risky choices in which one choice dominated the other, but violations of dominance (in the subjects' reported rankings) were observed just 0.7 percent of the time.

This suggests that subjects are generally aware of situations in which one choice dominates the other and rank accordingly but that when neither dominates the other, their stated preferences include some random (error) component.

6. Conclusions

For at least 55 percent of the subjects, on the three data sets considered, EU appears to fit no worse than any of the other models. For data set PC it would appear that, among the five top-level preference functionals, WU is the strongest contender to EU, followed by RQ, PR, and RP. DA performs always very badly. The model RN performs well only in data set CR, being not worse than EU for eight subjects; this result is possibly due to the fact that people find it easier to use the expected value to rank the circles rather than using a more complex criterion. The tables of the winners show that EU is the best model, even if in data set CR it loses a lot toward RN; its strongest contender is WU, followed by RQ, RP, and PR; RQ and RP perform well enough in CR data set, while PR performs well in PC data set; DA is always the worse.

The comparison of the two data sets estimated σ (the spread of the error term) shows that for RN and WU about 70 percent of the subjects have a standard deviation larger in data set CR than in PC; for the other models this percentage varies between 52 and 58. The higher standard deviation in CR could be explained by an argument concerning the difficulty of ranking the forty-four circles. The evidence concerning the inconsistency rate among repeated pairs, and the violations of dominance in dominating pairs, indicates that our maintained hypothesis about the nature of the error term may well be deficient. Given our assumption, it would appear that violations of dominance would be observed with a greater frequency than that actually observed. An alternative story (but one that should lead to *no* violations of dominance in the PC experiment) is that suggested by Loomes and Sugden [1995]: namely, that the preference functional is itself stochastic (a return to the random utility type of model). Perhaps the truth is somewhere in between (see Hey [1995]).

But perhaps the most startling findings of this experiment are those summarized in Tables 4 and 5: to the effect that the preference functional apparently used by a subject in a particular experiment is crucially dependent on the nature of the experiment itself. This is a particularly serious example of a framing effect.

Acknowledgments

We are grateful to the Leverhulme Trust for a grant (given to John Hey for collaborative experiments with young economists) that financed the running of these experiments. We are also grateful to Mark Machina and other participants at the twenty-first Seminar of the European Group of Risk and Insurance Economists at Toulouse, September 19 to 21, 1994, for helpful comments.

Notes

1. At the last minute, one participant changed bookings in such a way as to reverse the order in which that subject completed the two experiments.
2. We followed earlier practice in keeping all probabilities as multiples of one-eighth, for reasons discussed earlier [Hey and Orme 1994].
3. We arrived at this set of final outcomes after much discussion. Earlier experiments had used a wider range with a much larger loss, but it was felt that the set of participants so induced was possibly a rather biased set and was rather more risk-loving than the population at large.
4. The circles were printed on paper, attached to cardboard, and cut around the circumference, so people could easily handle them.
5. Some people have argued that this is not a particularly strong incentive mechanism and that subjects may just respond by putting the circles in a random order. Observation suggests that this was in fact not the case: the participants took time and care over the experiment. A factual calculation also reveals that a risk-neutral subject would expect to lose £3.52 by putting the lotteries in a random order rather than in the truly preferred (ranked by expected value) order. The (monetary equivalent of the) loss for a risk averter would be greater than this. Given the nature of our subjects, and the time they took to answer the CR experiment—about forty-five minutes—this should provide an appropriate incentive.
6. On earlier experiments we had allowed indifference to be expressed. However, there are difficulties involved with the interpretation of replies of indifference: if an individual truly is indifferent between the two lotteries, all three answers (prefer left, prefer right, indifferent) are equally attractive to the subject given the incentive mechanism.
7. See Hey and Orme [1994].
8. For further discussion, see Grant and Kajii [1994].

References

AMEMIYA, T. [1980]: "Selection of Regressors," *International Economic Review*, 21, 331–354.

BEGGS, S., CARDELL, S., and HAUSMAN, J. [1981]: "Assessing the Potential Demand for Electric Cars," *Journal of Econometrics*, 16, 1–19.

CARBONE, E., and HEY, J.D. [1994]: "Estimations of Expected and Non-Expected Utility Preference Functionals Using Complete Ranking Data," in *Models and Experiments on Risk and Uncertainty*, B. Munier and M.J. Machina (eds.), Kluwer, Boston.

GRANT, S., and KAJII, A. [1994]: "AUSI Expected Utility: An Anticipated Utility Theory of Relative Disappointment," Working Paper.

GUL, F. [1991]: "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

HEY, J.D. [1995]: "Experimental Investigations of Errors in Decision Making Under Risk," *European Economic Review*, 39, 633–640.

HEY, J.D., and DI CAGNO, D. [1990]: "Circles and Triangles: An Experimental Estimation of Indifference Lines in the Marschak-Machina Triangle," *Journal of Behavioral Decision Making*, 3, 279–306.

HEY, J.D., and ORME, C.D. [1994]: "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

HOLT, C.A. [1986]. "Preference Reversals and the Independence Axiom," *American Economic Review*, 76, 508–515.

LOOMES, G., and SUGDEN, R. [1995]: "Incorporating a Stochastic Element into Decision Theories," *European Economic Review*, 39, 641–648.

QUIGGIN, J. [1982]: "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

VISCUSI, W.K. [1989]: "Prospective Reference Theory: Toward an Explanation of the Paradoxes," *Journal of Risk and Uncertainty*, 2, 235–264.

YAARI, M.E. [1987]: "The Dual Theory of Choice Under Risk," *Econometrica*, 55, 95–115.