ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 45, 253–275 (1990)

An Empirical Evaluation of the Transitivity, Monotonicity, Accounting, and Conjoint Axioms for Perceived Risk

ELKE U. WEBER

Center for Decision Research, University of Chicago

AND

WILLIAM P. BOTTOM

John M. Olin School of Business, Washington University

This study tests the adequacy of the axioms underlying Luce and Weber's (1986) conjoint expected risk model. Risk judgments are found to be transitive. Monotonicity or the substitution principle per se seems to hold, but the related probability accounting assumption is violated. The conjoint structure assumptions about the effect of change of scale transformations on risk hold for negative-outcome lotteries but encounter some difficulty for positive-outcome lotteries. Possible explanations for violations are suggested, and implications of these results for the modeling of perceived risk are discussed. © 1990 Academic Press, Inc.

As confidence in expected utility models as descriptions of risky choice eroded (see e.g., Schoemaker, 1982; M. Weber & Camerer, 1987), researchers in the decision sciences have looked for additional variables responsible for preference. Perceived risk has been one such variable. Its role in judgment and choice has been investigated in a variety of areas—psychology (Coombs, 1969), marketing (Bettman, 1973), and economics (Markowitz, 1953; Libby & Fishburn, 1977). Consensus is growing that people may base their decisions on qualitative aspects of choice alternatives separate and/or orthogonal to expected utility, some of which may be captured by the label "risk." Weber, Anderson, and Birnbaum (1990) suggest that people encode and combine probability and outcome information in qualitatively different ways when judging the attractiveness of a choice alternative (with greater emphasis on the alternative's positive aspects) than when judging the same alternative's riskiness (with greater emphasis on negative components). Decision or choice might then be

This work was supported by National Science Foundation Grant SES 86-06750 to the first author. We thank R. Duncan Luce and two anonymous reviewers for helpful comments. Send requests for reprints to Elke U. Weber, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637.

0749-5978/90 $3.00

modeled as a function of two such constructs. A similar idea underlies Coombs' (1975) characterization of choice as a conflict between greed and fear.

This article addresses the measurement of the perceived riskiness of choice alternatives that can be conceptualized as "lotteries" or "gambles" (i.e., alternatives that have a known probability distribution of possible outcomes), where the outcomes can be described on a single dimension (e.g., dollar amounts or return on investment for financial "gambles," or mortality rates for medical treatment "gambles").

A historical account of developments in this area over the last 20 years can be found in Weber (1988). The purpose of the present article is to test whether people's risk judgments satisfy the axioms of the conjoint expected risk (CER) model of Luce and Weber (1986). The CER model is a revision (i.e., modification and extension) of the risk assumptions or axioms suggested by Luce (1980, 1981). This revision incorporates the results of several empirical investigations of perceived risk (Keller, Sarin, & M. Weber, 1986; Weber, 1984a, 1988). A short summary of the model, emphasizing its empirically interesting and testable assumptions, is provided in the next section. This is followed by an account of the empirical investigations of these assumptions.

Conjoint Expected Risk (CER) Model

The CER model makes a set of assumptions about people's judgments of perceived risk which, if satisfied, imply that people's risk judgments can be described by the following function:

$$\begin{aligned} R(X) = {} & A_0 Pr(X = 0) + A_+ Pr(X > 0) + A_- Pr(X < 0) \\ & + B_+ E[X^{k+} | X > 0] Pr(X > 0) \\ & + B_- E[|X|^{k-} | X < 0] Pr(X < 0), \end{aligned}$$

where $R(X)$ denotes the perceived riskiness of lottery X and is described as a linear combination of the probability of breaking even (zero outcomes or status quo), the probability of positive outcomes, the probability of negative outcomes, the conditional expectation of positive outcomes raised to some power k_+ weighted by the probability of winning, and the conditional expectation of negative outcomes raised to some power k_- weighted by the probability of losing, with k_+ and $k_- > 0$. Parameters A_0, A_+, A_-, B_+, and B_- are weights on the respective components.

Weber (1988) showed that the proportion of variance in risk ratings of lotteries accounted for by the CER function is significantly larger than that of other risk functions and usually approaches the squared reliability of the ratings. The CER function's parameters allow comparisons between subjects with respect to differences in risk perception. For a sample of college students, the values of k_+ and k_-, for example, were both small, indicating that the size of wins or losses was less salient for their risk judgments than the probability of winning or losing. Highschool teachers, in contrast, were more sensitive to the magnitude of wins (higher k_+) and especially of losses (higher k_-).

A more critical as well as more sensitive test of any axiomatic measurement model than the goodness-of-fit indices reported by Weber (1988) is an assessment of the empirical validity of the behavioral assumptions from which the model derives. Weber and Bottom (1989) reported empirical evidence supporting one central feature of the CER axiomatization, namely an additive (rather than multiplicative) combination of gains and losses. Just as Coombs and Lehner's (1984) bilinear model of risk, the CER model postulates that gains reduce risk to an extent that does not depend on the magnitude of the losses. In contrast, Fishburn's (1982, 1984) risk axiomatizations which are multiplicative over gains and losses (i.e., gains affect risk only by modifying the effects of losses) seemed to be ruled out by Weber and Bottom's data.

The assumptions about people's risk judgments made by Luce and Weber (1986) can be divided into two groups. First, the mixture space assumptions allow the CER function to retain the benefits of expectation models, namely a constant number of components and parameters regardless of the number of outcomes of the decision alternatives. This property is not shared by risk dimension models of the kind suggested by Payne (1973). The following mixture space axioms will be tested below.

(i) *Transitivity:* If $X \succcurlyeq Y$ and $Y \succcurlyeq Z$, then $X \succcurlyeq Z$. If risky prospect (lottery) X is judged to be as or more risky than Y, and prospect Y is judged to be as or more risky than prospect Z, then prospect X should also be judged to be as or more risky than Z. Thus, the symbol $\succcurlyeq$ stands for the comparison "at least as risky as," and the symbols X, Y, and Z symbolize random variable representations of risky choice alternatives.

(ii) *Monotonicity* (also known as the substitution principle): If $X \succcurlyeq Y$, then $X\text{o}_p Z \succcurlyeq Y\text{o}_p Z$, for every Z or p. If lottery X is judged as at least as risky as Y, then the compound lottery (probability mixture) of obtaining lottery X with probability p and Z otherwise should also be judged as at least as risky as the compound lottery of obtaining lottery Y with probability p and Z otherwise, regardless of the particular choice of lottery Z or of the level of p.

(iii) *Probability accounting:* The following two lotteries should be judged as equally risky: $(X\text{o}_p Y)\text{o}_q Y = X\text{o}_{pq} Y$. That is, the two-stage lottery of obtaining with probability q the probability mixture of lotteries X (with probability p) and Y (with probability $1-p$) or with probability $1-q$ the lottery Y should be judged equally risky to the single-stage lottery of obtaining lottery X with probability pq and lottery Y with probability $1-pq$. The assumption has normative appeal because the final outcomes in the two situations are identical. However, equality in perceived risk will only obtain if people equate the uncertainties in the two situations following the rules of the probability calculus, i.e., by perceiving the subjective probability of pq as the product of the subjective probabilities of p and q. Demonstrations like the Allais (1953, 1979) paradox for preference which are often interpreted as implicating the validity of the monotonicity assumption may in fact be the result of violations of the probability accounting assumption (see Luce & Narens, 1985).

Weber and Bottom (1989) report data that put into question some of the mixture space (expectation) assumptions. Following a study by Keller *et al.* (1986), Weber and Bottom confirmed the existence of a risk analogue to the Allais paradox.

The main feature of the second group of axioms employed by Luce and Weber (1986) is the assumption that gambles are split into positive, negative, and zero outcome components for the determination of risk, and that the relative impact of these components on risk need not be symmetric. (An example is Coombs and Lehner's [1984] thought experiment that adding an extra $10 to the loss side of a 50:50 lottery of winning or losing $10 increases the perceived risk by more than the addition of the same amount to the gain side would decrease it.) The conjoint structure axioms of the CER model make assumptions about the effects of change of scale transformations (i.e., multiplication of all outcomes of a lottery by a constant a) for positive-outcome-only and negative-outcome-only lotteries separately. The following conjoint structure axioms will be tested below:

(iv) *Independence:* For any two positive (negative)-outcome-only lotteries X and Y and for any change of scale multiplication factors a and b which are members of the positive reals, $X \succcurlyeq Y$ iff $aX \succcurlyeq aY$, and $aX \succcurlyeq bX$ iff $aY \succcurlyeq bY$. If lottery X is judged to be at least as risky as lottery Y, then aX (the lottery created from X by multiplying all of its outcomes by the constant a) should also be judged to be as least as risky as aY, regardless of the particular value of a. Secondly, if a particular change of scale transformation of X, i.e., aX, is judged to be as least as risky as another change of scale transformation, bX, then the same should be true when the two transformations are applied to any other lottery Y.

(v) *Archimedean assumption:* If $X \succcurlyeq Y$, then there exists a positive real a such that $aY \succcurlyeq X$. This condition assumes that no lottery is infinitely more risky than any other. Should X be judged as at least as risky as Y, then there is some change of scale transformation a that, when applied to lottery Y, will make aY to be at least as risky as X.

Axiomatic Measurement and Error Theory

The next section reports the results of investigations of the axioms of the CER function. These tests like others in this area suffer from lack of an error theory to assess the fit of empirical (and thus noisy) data to the requirements of deterministic axioms.

EXPECTED RISK AXIOMS

Experiment 1: Transitivity

Method

Items for the transitivity experiment were constructed using the transformations of Coombs and Bowen (1971). The symmetric two-outcome gamble ($50, .50 −$50), i.e., a 50:50 chance of winning or losing $50, was subject in turn to the following transformations: skewness, origin, scale, and number of plays. Three different skewness levels (p = .25, .50, .75) were used to transform the lottery (a, .50, $-a$) into ($a\sqrt{q/p}$, p, $-a\sqrt{p/q}$), leaving the expectation and variance of the lottery unchanged. The change of origin transformation changed the gamble (y, p, $-z$) into ($y + b$, p, $-z + b$), using two level of b (b = $20, −$20). This changes the expectation by b but leaves the variance and skewness unchanged. The change of scale transformation changed the outcomes of gamble (y, p, $-z$) into (cy, p, $-cz$), using c = 1 and 3. This changes the expectation by c and the variance by c^2 but does not affect the skewness. The 12 two-outcome gambles thus produced were transformed into the stimulus set of three-outcome gambles shown in Table 1 by simulating the effect of playing each gamble twice and dividing the outcomes by two.

TABLE 1
TWELVE LOTTERIES USED IN EXPERIMENT 1

Lottery	
1	(+$104, .06; +$49, .38; −$9, .56)
2	(+$67, .06; +$9, .38; −$49, .56)
3	(+$70, .25; +$20, .50; −$30, .25)
4	(+$30, .25; −$20, .50; −$70, .25)
5	(+$49, .56; −$9, .38; −$67, .06)
6	(+$9, .56; −$49, .38; −$104, .06)
7	(+$320, .06; +$146, .38; −$27, .56)
8	(+$200, .06; +$26, .38; −$147, .56)
9	(+$210, .25; +$60, .50; −$90, .25)
10	(+$90, .25; −$60, .50; −$210, .25)
11	(+$147, .56; −$27, .38; −$200, .06)
12	(+$27, .56; −$146, .38; −$320, .06)

Note. Lotteries are of the form (outcome-1, probability-1; outcome-2, probability-2; outcome-3, probability-3).

Sixteen undergraduates from the University of Illinois served as subjects in the experiment. For this and all subsequent experiments, subjects participated as a credit requirement for an introductory psychology course. All tasks were carefully explained both verbally and in writing, and several practice items were provided to familiarize subjects with the tasks, the range of stimulus lotteries, and the response scales. Subjects worked at their own pace with sufficient time to carry out the requisite judgments or ratings for the different tasks without any time pressure.

Testing the transitivity of risk judgments assumption requires paired comparisons with respect to risk. Subjects' task was to first judge which member of a pair of gambles was riskier and then to rate the difference in risk on a scale from 0 (no difference at all, i.e., identical in risk) to 100 (extremely different in risk). Risk was left intentionally undefined, except for emphasizing to subjects that they were to judge differences in *risk* and *not* in *preference* or desirability. No subject found this distinction confusing.

All 66 possible pairwise combinations of the 12 gambles were presented. The order of pairs and of the gambles within each pair was random. Lotteries were represented as vertical bar graphs (i.e., the probabilities of outcomes were depicted graphically by a proportionate number of X's as well as by their numerical values). Monetary outcomes (losses indicated by a minus sign) appeared to the left of the respective probabilities.

Results

Table 2 lists the percentage of X, Y, Z triples that satisfied the conditions $X \succcurlyeq Y$ and $Y \succcurlyeq Z$ but violated the condition $X \succcurlyeq Z$. Keeping in mind the earlier caveat regarding the difficulty of interpreting violation rates for deterministic axioms, it nevertheless seems that transitivity holds reasonably well for risk comparisons. The mean and median violation rate across subjects was 6% and only three of the 16 subjects had violation rates above 10%.

TABLE 2

PERCENTAGE OF RISK JUDGMENTS VIOLATING TRANSITIVITY

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Percentage of violations[a]	4	1	6	6	11	5	14	12	7	4	4	2	5	7	1	8

[a] The number of tests (i.e., triplets X, Y, and Z for which $X \succcurlyeq Y$ and $Y \succcurlyeq Z$) varied somewhat from subject to subject (between 102 and 147). The percentages of violations (i.e., cases where $Z \succcurlyeq X$) were computed using the appropriate individual's number of tests.

Experiment 2: Monotonicity

Method

Some of the lotteries used in the transitivity test were employed in the monotonicity test. Of the lotteries shown in Table 1, items 7 and 10 formed X–Y pair I, items 1 and 12 formed pair II, and items 4 and 6 formed pair III. Both members of the three X–Y pairs were combined with an additional gamble, Z, to generate complex gambles of the form (Xo_pZ) and (Yo_pZ). The probability of obtaining the original gamble was varied (p = .05, .30, .60, .80, .95). Five different gambles were used for Z. These included three two-outcome gambles, as well as a sure win of $45 and a sure loss of $45. All possible 25 pairings of Z and p with an original pair of X–Y gambles were formed. The particular Z gamble and level of p associated with the X gamble in a pair was always identical to the ones associated with the Y gamble. This produced 75 pairs of gambles. Lotteries were presented in the format of Experiment 1 (i.e., as single-stage lotteries), which means that the "monotonicity structure" of a pair (i.e., the fact that it was constructed using identical p and Z) was not obvious to subjects.

Twelve undergraduates received a booklet containing the 75 pairs in random order. The order of gambles within a pair was also random. Similar to the instructions for the transitivity test, subjects were asked to indicate which member of a given pair was riskier. They were then to rate by how much the indicated lottery was riskier on a scale from 0 to 100.

Results

Table 3a shows the number of times subjects selected the X member or the Y member as the riskier lottery for each of the three X–Y pairs. Each subject made 25 different risk comparisons for each X–Y pair, counting over all p and Z combinations. Because the riskier member of a pair is probably best defined by a subject's modal judgment across the 25 comparisons, the number of violations of the monotonicity assumptions can be defined as the number of times a subject selected the nonmodal member of a pair.

Allowing thus for individual differences in perceptions of relative risk was, however, hardly necessary. Across subjects there was considerable consistency in the modal risk comparisons. For pairs I and II, all 16 subjects agreed that Y was the riskier member of the pair. For pair III, 13 of the 16 subjects gave Y as their modal choice. The response consistency for pairs I and II is perhaps not surprising in light of the fact that (in

TABLE 3
RESULTS OF MONOTONICITY TEST

(a) Lottery pair[a]	I		II		III		
Subject	X	Y	X	Y	X	Y	% Violations
1	1	24	1	24	10	15	16%
2	5	20	2	23	2	23	12%
3	1	24	0	25	4	21	7%
4	0	25	2	23	9	16	15%
5	1	24	2	23	8	17	15%
6	3	22	1	24	8	17	16%
7	1	24	1	24	6	19	11%
8	1	24	1	24	10	15	16%
9	3	22	1	24	7	18	15%
10	1	25	1	24	11	14	17%
11	1	24	3	22	12	13	21%
12	2	25	1	24	18	7	13%
13	1	24	1	24	12	13	19%
14	0	25	1	24	19	6	9%
15	0	25	2	23	12	13	19%
16	9	16	2	23	17	8	25%
% Violations	7%		6%		33%		
(b)							
	X	Y	X	Y	X	Y	% Violations
1	0	25	2	23	20	5	9%
2	1	24	0	25	7	18	11%
3	0	25	0	25	0	25	0%
4	0	25	0	25	2	23	3%
5	1	24	0	25	6	19	9%
6	0	25	0	25	0	25	0%
7	1	24	2	23	3	22	8%
8	0	25	0	25	4	21	5%
9	0	25	0	25	16	9	12%
10	0	25	19	6	14	11	5%
11	0	25	2	23	23	2	23%
12	1	24	4	21	18	7	16%
% Violations	1%		5%		19%		

Note. Number of choices (out of 25 combinations of p and Z) for three pairs X–Y for which lottery Xo_pZ or lottery Yo_pZ, respectively, was judged the riskier of the pair. (a) Lotteries Xo_pZ and Yo_pZ are presented as one-stage lotteries. (b) Lotteries Xo_pZ and Yo_pZ are presented as two-stage lotteries.

[a] Lottery pair I: X = lottery 7; Y = lottery 10 (see Table 1); lottery pair II: X = lottery 1; Y = lottery 12; lottery pair III: X = lottery 4; Y = lottery 6.

addition to differences in skewness and variance) the two members of pairs I and II both differed in expected value (Pair I: $E(X) = +\$60$, $E(Y) = -\$60$; Pair II: $E(X) = +\$20$, $E(Y) = -\$20$). For pair III, on the other hand, lotteries X and Y differed only in skewness and variance ($E(X) = E(Y) = -\$20$). While it is true that these two variables affect perceived risk, their contributions to risk are generally smaller and more variable than the contribution of expected value (Weber, 1984a).

Table 3a summarizes the violation rates of the monotonicity assumption by both summing across the three pairs for every subject and across subjects for every item pair. Violation rates per subject are substantial (mean = 15.3% and SD = 4.5%). However, the vast majority of violations occurs for lottery pair III (33%) for which the average difference in risk between X and Y (as judged in Experiment 1) was quite small and for which between-subject differences in relative risk assessment exist. This suggests that the violations observed for this pair may not be an indictment of the monotonicity assumption in general, but mainly an indication of an instability in the risk comparison for lotteries that are very similar in risk.

However, violation rates even for pairs I and II (7% and 6%, respectively) may leave reason for concern in light of the risk analogue of the Allais paradox reported by Keller *et al.* (1986) and replicated by Weber and Bottom (1989). The Allais paradox has traditionally been interpreted as implicating the monotonicity assumption. However, Luce and Narens (1985) argue that the probability accounting assumption could be inaccurate and monotonicity per se hold. This possibility was investigated in two ways. We repeated the monotonicity test, using a new sample of 12 undergraduates, in the same way as described above except for presenting the lotteries of the stimulus pairs as compound two-stage lotteries (see Fig. 1b). Violations of the monotonicity postulate in this version of the test would not be attributable to probability accounting mechanisms. In addition, Experiment 3 is a direct test of the probability accounting assumption.

The results of the two-stage representation version of the monotonicity test are shown in Table 3b. The violation rates (both at the individual subject and at the item pair level) are considerably lower than for the single-stage representation version (Table 3a). The mean violation rate at the subject level was reduced from 15.3% to 8.4%. (Without subjects 10 and 12 whose judgments appeared erratic on other grounds, the rate was 6.2%.) The average violation rate for pair I was reduced from 7% to 1% and that for pair II from 6% to 5% (or 2% without subjects 10 and 12). Thus, it seems that a significant proportion of the violations observed in the single-stage representation version can be attributed to failure of the probability accounting assumption. The vast majority of the remaining

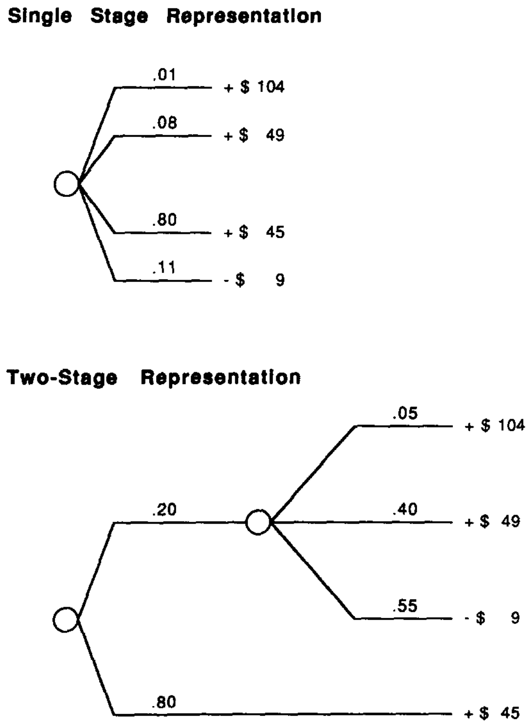

FIG. 1. Sample of single-stage and two-stage representation of a lottery of type Xo_pZ, where X = (+\$104, .05; +\$49, .40; −\$9, .55), Z = (+\$45), and p = .20.

violations occurred for stimulus pair III (a reduction from 33% to still 19% (or 15% without subjects 10 and 12), suggesting that nonstability of risk comparisons because of similarity in risk is an additional factor contributing to apparent violations of monotonicity.

Experiment 3: Probability Accounting

Method

Eleven of the complex gambles (Xo_pZ or Yo_pZ) used in Experiment 2 were chosen at random to examine the probability accounting assumption. Eight of these were five-outcome gambles; the remainder were four-outcome gambles. Each gamble was represented in two ways as shown in Fig. 1, first as a single-stage gamble with the probabilities for each outcome multiplied out by the probabilities for obtaining the component gamble. The second representation was as a two-stage lottery, showing separately the probability of obtaining the component gambles and the probabilities of the various outcomes for each component.

Thirteen undergraduates were given a booklet which contained the 11-item pairs in random order, interspersed with 11 other items. In addition, one item pair was presented twice to act as a reliability check. Subjects were asked to first indicate which of the two gambles was riskier (without being told that the two were different representations of the same gamble in terms of final outcome position). As before, subjects then rated the difference in risk between the two lotteries on a scale from 0 to 100.

Results

If a subject's probability accounting for risk obeys the expected risk axiom, then the subject should indicate that the two representations are equal in risk and/or give a difference rating close to zero. Table 4a lists for every subject the number of items for which either the single-stage (SS) or the two-stage (TS) representation was judged to be riskier as well as the number of indifference (I) judgments. As the table indicates, indifference with respect to risk was quite rare (15%, averaged across subjects). In general, subjects were more likely to judge the two-stage representation of an item as riskier (51% across subjects) than the single-stage version (34%). (A Wilcoxon signed-rank test across subjects of the difference in the number of items for which the TS version was judged as riskier than the SS version was marginally significant at the .10 level.) Table 4b lists for each of the 11 items the number of subjects who judged the single-stage form to be riskier, the two-stage form to be riskier, and the two forms to be equal in risk. For items 8 and 9, the percentage of risk indifference or equality judgments was higher (54%) than for the rest of the items (15% on average). These items had a "simpler" two-stage structure than the other items, and thus might have appeared more "similar" to their single-stage version than the other items. There was no indication that subjects noticed the correspondence between the two representa-

TABLE 4
RESULTS OF PROBABILITY ACCOUNTING TEST

(a) Subject	1	2	3	4	5	6	7	8	9	10	11	12	13
SS	4	1	3	4	3	3	5	6	6	1	4	3	5
TS	7	10	6	6	6	8	6	4	2	8	4	3	3
I	0	0	2	1	2	0	0	1	3	2	3	5	3

(b) Item	1	2	3	4	5	6	7	8	9	10	11
SS	8	3	7	6	5	4	4	3	2	2	4
TS	5	10	5	6	7	8	9	3	4	8	8
I	0	0	1	1	1	1	0	7	7	3	1

Note. Part (a) lists for every subject the number of items (out of 11) for which either the single-stage (SS) or the two-stage (TS) representation was judged to be riskier and the number of indifference (I) judgments. Part (b) lists for every item the number of subjects (out of 13) who judged the single-stage (SS) form to be riskier, the two-stage (TS) form to be riskier, and the two forms to be equal in risk (I).

tions and/or perceived the two versions as "the same." The result that the TS version was generally perceived as riskier than the SS version also holds at the item level (Wilcoxon signed-rank test was significant at the .05 level).

Given the results reported in Table 4, the probability accounting axiom appears strongly violated. However, one could argue that subjects who are asked to judge the difference in risk between two members of a pair may simply be hesitant to give an *in*difference rating. They may feel compelled to judge one of the two members of the pair to be the riskier one, but do so essentially at random and without conviction. There is, however, evidence that discounts this argument at least partially. The correlation between the risk difference judgments provided by subjects for the replicated item pair was $r = .96$ ($p < .001$, 11 *df*). If subjects judged risk difference arbitrarily, one would hardly expect such high reliability. Along the same line, the fact that there was some systematicity in the violations (TS generally judged riskier than SS) seems to speak against risk indifference. On the other hand, judged differences in risk between the two versions were often small (the grand mean of the absolute value of risk difference judgments across subjects and items was 21 on a scale from 0 to 100).

Conclusions

The following pattern seems to emerge with respect to the mixture space axioms. Transitivity appears to hold reasonably well, a result that agrees with Aschenbrenner (1978). Monotonicity in its purest form (i.e., presenting compound lotteries in their two-stage version) seems to be reasonably satisfied as well. Violations of the expectation principle usually attributed to violation of monotonicity (e.g., Allais-paradox equivalents for risk) may instead be due to two other factors. One is an inadequate and thus unstable difference in the riskiness between the original lotteries X and Y. When the original difference is sufficiently large, violations of monotonicity virtually disappear. The other factor is the apparent failure of the probability accounting axiom. Failure of this axiom indicates that the perceived risk of compoundly uncertain situations may not be reducible to a distribution of possible final outcomes. Implications of this result for the CER model and for the modeling of perceived risk in general can be found in the discussion section below.

CONJOINT STRUCTURE AXIOMS

Experiment 4: Independence for Negative Outcome and for Positive Outcome Lotteries

Method

Skewness and multiple play transformations as described in Experiment 1 were applied to the lottery (+\$1, .5, −\$1) to generate three two-outcome and three three-outcome lotteries. By subtracting a constant \$2 from all outcomes of the gambles, the negative-outcome set shown in Table 5 was created. For the equivalent positive-outcome set, all negative outcomes of the lotteries of Table 5 were replaced by positive outcomes of the same absolute value.

The independence property makes assumptions about risk judgments of positive or negative-outcome lotteries after change of scale transformations. Therefore five levels of a change of scale (a = 1, 5, 9, 13, 17) were factorially combined with the six lotteries of the positive and negative-outcome set to generate two sets of 30 gambles. For each set, the requisite pairwise combinations of lotteries were generated to test both aspects of the independence property. To test the first aspect of independence (that the relative riskiness of gambles should be independent of the scale factor on the outcomes), all 75 possible pairs of the form aX–aY were generated. To test the second aspect of independence (that the effects of changes of scale should be independent of the gambles they affect), all 60 possible pairs of the form aX–bX were generated.

Twenty-two undergraduates received booklets that contained the 135 negative-outcome lotteries paired comparisons in random order. A different set of 13 undergraduates received the 135 positive-outcome lotteries paired comparisons. The order of lotteries within a pair was random. Subjects first indicated which member of the pair was riskier, then rated the difference in risk on a scale from 0 to 100.

Results

Negative outcome set. Violation rates of the independence assumptions for each subject are shown in Table 6. Part I pertains to the risk order of pairs of gambles across changes of scale. An index of the prevalence of violations of the type: $aX \geqslant aY$ but $bX \geqslant bY$, was computed for each subject as the proportion of risk judgments for each X–Y pair whose

TABLE 5
SIX NEGATIVE-OUTCOME LOTTERIES USED IN INDEPENDENCE AXIOM TEST OF EXPERIMENT 6

Lottery	
1	(−\$3.00, .50; −\$1.00, .50)
2	(−\$3.75, .25; −\$1.50, .75)
3	(−\$2.50, .75; −\$0.25, .25)
4	(−\$6.00, .25; −\$4.00, .50; −\$2.00, .25)
5	(−\$7.50, .06; −\$5.25, .38; −\$3.00, .56)
6	(−\$5.00, .56; −\$2.75, .38; −\$0.50, .06)

TABLE 6
RESULTS OF INDEPENDENCE TEST FOR THE 22 SUBJECTS OF THE NEGATIVE-OUTCOME SET TEST AND FOR THE 13 SUBJECTS OF THE POSITIVE-OUTCOME SET TEST

Negative outcome set			Positive outcome set		
Subject	Part I	Part II	Subject	Part I	Part II
1	.04	.03	1	.13	.13
2	.12	.03	2	.08	.05
3	.22	.00	3	.17	.02
4	.20	.13	4	.09	.05
5	.10	.00	5	.12	.18
6	.20	.30	6	.08	.00
7	.13	.03	7	.17	.05
8	.15	.05	8	.12	.00
9	.04	.00	9	.07	.03
10	.10	.00	10	.12	.03
11	.10	.00	11	.23	.05
12	.08	.00	12	.03	.00
13	.10	.02	13	.05	.00
14	.09	.07			
15	.09	.07			
16	.15	.03			
17	.09	.05			
18	.25	.03			
19	.13	.00			
20	.05	.00			
21	.20	.10			
22	.21	.15			

Note. Part I: Violation rates for assumption $aX \geqslant aY$ but $bX \ngeqslant bY$ (aggregated over the 15 different X–Y pairs and five levels of scale factor). Part II: Violation rates for assumption $aX \geqslant bX$ but $aY \ngeqslant bY$ (aggregated over the 10 different a–b combinations and six levels of lottery).

directionality did not agree with the modal choice of a given subject for a given pair. The modal choice (across five levels of scale factors) between a pair of gambles was used as the best indication of a subject's "true" risk order for that pair. Accounting for violations in this way avoids the problem of dependencies between violations that occurs when pairs of pairwise comparisons are used as the unit of analysis.

The violation rates of individual subjects for part I ranged from .04 to .25, with a median of .11. To see whether certain item pairs were involved in more independence violations than others, violation rates were also broken down by item pair, as shown in Table 7. The range of violation rates (from .01 to .30) is considerable. The hypothesis that violation rates are equal for the different item pairs was significantly violated ($\chi^2(13\ df) = 115.25$, $p < .001$). Given the result of the monotonicity test that ap-

TABLE 7
RESULTS OF INDEPENDENCE TEST ASSUMPTION $aX \geq aY$ BUT $bX \geq bY$

	Violation rate	
Item pair[a]	Negative outcome set[b]	Positive outcome set[c]
1–2	.07	.16
1–3	.24	.09
1–4	.07	.11
1–5	.24	.05
1–6	.06	.11
2–3	.04	.09
2–4	.20	.03
2–5	.07	.16
2–6	.24	.18
3–4	.04	.15
3–5	.23	.09
3–6	.06	.14
4–5	.01	.08
4–6	.30	.15
5–6	.06	.09

Note. Violation rates (aggregated over the respective number of subjects and five levels of scale factor) for the 15 different X–Y pair comparisons of the negative and of the positive outcome set.

[a] Lottery numbers correspond to those in Table 5.

[b] 22 subjects.

[c] 13 subjects.

parent violations of an axiom may be the result of an unstable risk order between pairs of lotteries, we computed the correlation between the mean risk difference ratings between item pairs (averaged across scale factors and subjects) and the number of violations involving a pair. The observed correlation of $-.92$ ($p < .001$, 13 *df*) indicates more violations for those pairs that subjects judge as being closer together in risk and thus provides an explanation for apparent independence violations similar to that invoked for the monotonicity test.

There were few violations for part II of the independence assumption which holds that the relative effects of two different scale factors on risk should be independent of the gamble to which they are applied. The violation rates of individual subjects ranged from .00 to .30, but with a median value of .03. Eight subjects showed no violations at all. To see whether certain pairs of scale factors generated more violations than others, violation rates were broken down not by subject but by scale factor pair as shown in Table 8. There was little variability in the violation rates (mean = .05, SD = .02) which did not differ significantly for the different scale factor pairs ($\chi^2(9\ df) = 10.03$, n.s.). Thus it seems that the (infrequent) instances of violation can probably be attributed to random error.

TABLE 8
RESULTS OF INDEPENDENCE TEST ASSUMPTION $aX \geqslant bX$ BUT $aY \geqslant bY$

Scale factor pair	Violation rate: Negative outcome set[a]	Violation rate: Positive outcome set[b]
1–5	.02	.03
1–9	.03	.03
1–13	.04	.04
1–17	.07	.03
5–9	.06	.05
5–13	.04	.06
5–17	.04	.08
9–13	.05	.05
9–17	.06	.08
13–17	.09	.03

Note. Violation rates (aggregated over the respective number of subjects and six levels of lottery) for the 10 different *a–b* scale factor combinations separately for the negative and of the positive outcome set.
[a] 22 subjects.
[b] 13 subjects.

Positive outcome set. Part I violation rates (pertaining to the risk order of lottery pairs across changes of scale) for the positive-outcome set were larger than for the negative-outcome set. Individual subject rates (Table 6) ranged from .03 to .23 with a median of .12. When broken down by item pair (Table 7), there was no significant difference in violation rates for the different items pairs ($\chi^2(14\ df) = 15.34$, n.s.) and only a marginally significant correlation between risk difference ratings and the number of violations involving each pair ($r = -.54$, $p < .10$, 13 *df*), another deviation in results from the negative-outcome set.

Part II violation rates (pertaining to the effect of changes of scale across lotteries) on the other hand were similar to the pattern observed for the negative-outcome set. They were lower than for part I (ranging from .00 to .18 with a median of .03, with four out of 13 violation-free subjects), and showed no significant differences when broken down by scale factor pair (Table 8) ($\chi^2(9\ df) = 5.6$, n.s.).

Experiment 5: Archimedean Assumption for Negative and for Positive Outcome Lotteries

Method

Five pairs of three-outcome lotteries from the negative-outcome set of Experiment 4 were used in this study. Additional items were created by multiplying all of the outcomes of the first member of a pair, $aX–Y$, as well as all of the outcomes of the second member of a pair, $X–aY$ by a positive constant a (a = 10, 50, or 100). This generated a total of 35 items. A corresponding set of 35 positive-outcome items was generated by selecting the five starting pairs from the positive-outcome set of Experiment 4.

Eighteen undergraduates received a booklet containing the 35 negative-outcome item pairs in random order. Another group of 12 undergraduates received the 35 positive-outcome item pairs. The position of lotteries within a pair was also random. Subjects first indicated which lottery in a pair was riskier and then rated the difference in risk on a scale from 0 to 100.

Results

Negative outcome set. A subject's relative risk judgment for the five original X–Y pairs determined which additional pairs were used to test the Archimedean assumption. If a subject originally judged X to be riskier than Y, then the subject's responses to the pairs X–aY were examined. According to the Archimedean axiom, there should be a value of a such that the subject would judge lottery aY to be riskier than X. The subject should also judge aY to be riskier than X for all a greater than this sufficiently large value. (For subjects who judged Y riskier than X in the original X–Y pair, responses to the pairs aX–Y provided the necessary information.)

The Archimedean assumption was strongly supported for the negative-outcome set. While subjects' judgments about the relative risk of the original lottery pairs varied, all subjects reversed their original judgments for every one of the five pairs when the previously less risky lottery underwent a change of scale transformation. Only one subject (and only for one of the five pairs) showed inconsistency in the pattern of reversal by reversing his original judgment for a = 10, reverting to the original relative risk judgment for a = 50, but then switching again for a = 100. This violation rate of one instance in 90 (18 subjects × 5 items) is clear evidence for the adequacy of the Archimedean assumption for negative-outcome sets.

The Archimedean condition requires merely that there is *some* value of a that will satisfy the requirement, so that the condition may be verified but cannot be falsified with a finite set of test items. However, by providing the Archimedean assumption with some "process" interpretation, one can make it more amenable to experimental test. In particular, it seems plausible to assume that the eventual switch in relative risk judgment required by the condition comes about by a decrease in perceived risk difference between the two lotteries as the originally less risky option undergoes change of scale transformations with increasing values of a, to a point where the two options will be perceived as equally risky and beyond which the required switch in relative risk will occur.

We used subjects' risk difference ratings for the item pairs to test this conjecture. Figure 2 shows the risk difference ratings (positive values indicating that X was judged as riskier than aY, negative values the opposite) averaged over subjects as a function of the multiplier a ($a = 1, 10, 50$, and 100), with a separate curve for each of the five pairs X–aY. The picture clearly confirms our hypothesis. Not surprisingly, in a repeated measures ANOVA of the difference ratings with Pair and Multiplier as within factors, the main effect for Multiplier was highly significant ($F_{3,51} = 93.42, p < .001$).

Positive outcome set. The results for the positive-outcome set showed far greater variability than for the negative-outcome set. Only two of the 12 subjects reversed their original risk judgments on all five pairs. Another subject reversed her judgment for three of the five pairs. No other subject reversed judgments for more than one pair. Four subjects did not reverse any judgments. One subject reported no difference in risk for all of the 35 pairs.

As discussed in the previous section, the failure to reverse for a limited range of a is not sufficient justification for rejecting the Archimedean property. It is conceivable that subjects might reverse their judgments for some larger value of a. Given the pattern of decreases in risk difference judgments leading to an eventual switch in relative risk judgment observed for the negative-outcome set, one would expect to see for the positive-outcome set a corresponding trend of decreases in risk difference judgments as a increases for the Archimedean condition to hold, even if

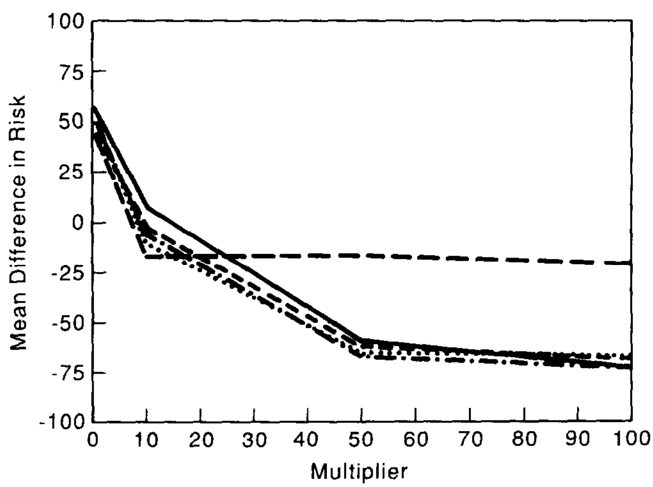

FIG. 2. Risk difference averaged across 18 subjects between members of lottery pairs aX and Y as a function of multiplier a for negative-outcome stimuli. Different lines indicate different pairs aX–Y.

the magnitude of a is not sufficient to lead to risk equality and/or reversal. To check on this possibility we performed a repeated measures ANOVA of the difference ratings for the nonreversing subjects with Pair (5 levels) and Multiplier (4 levels) as within factors. If insufficient range of a were responsible for the failure to reverse, the Multiplier factor should show some effect (i.e., show at least a trend to decrease risk differences). However, neither the main effect of Multiplier nor the interaction effect was significant. The absence of *any* decreasing trend in the risk difference judgments can also be seen visually in Fig. 3.

Conclusions

The following pattern seems to emerge with respect to the conjoint structure axioms. Both the independence assumption and the Archimedean assumption seem to hold reasonably well for negative-outcome lotteries. Violations that occur can be attributed either to insufficient difference in the riskiness of lotteries under test (and thus to an unstable risk order) or to other sources of random error. Manipulation of scale factor (part II of the independence condition) seems to be a more powerful and reliable method of affecting risk than our choices of lotteries (part I). For positive-outcome lotteries, violation rates for the independence assumptions are not as easily attributed to axiom-unrelated factors and the Archimedean assumption seems unlikely to hold.

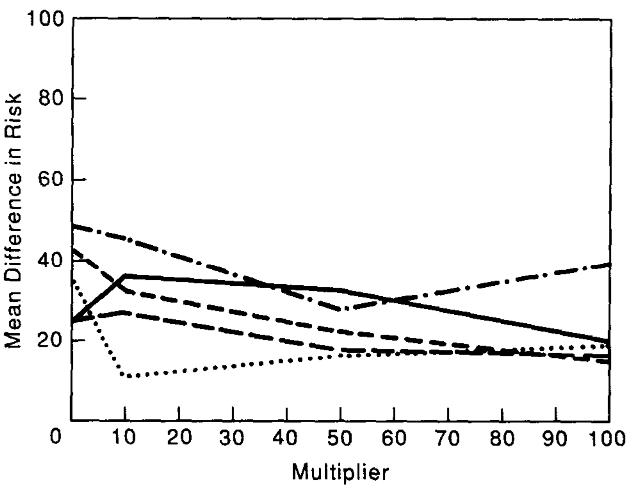

FIG. 3. Risk difference averaged across 12 subjects between members of lottery pairs aX and Y as a function of multiplier a for positive-outcome stimuli. Different lines indicate different pairs aX–Y.

GENERAL DISCUSSION

Our empirical investigation of the postulates of the conjoint expected risk model identified the following two deviations of people's risk judgments from the assumptions of the model. The first one is the failure of the probability accounting axiom in the mixture space group, the second is the failure of the independence and the Archimedean axioms for positive-outcome lotteries. Implications of these failures and possible modifications of the model to account for them are discussed in this section.

The first violation, that of the probability accounting assumption, and its consequences (e.g., risk equivalents of the Allais paradox) are similar to those observed for preference judgments of risky prospects (see Kahneman & Tversky, 1979). The exact nature of probability accounting violations for risk needs to be further explored. It may be possible to account for them using "risk weights" that are a monotonic but nonlinear function of the stated probabilities, similar to the weighting function in Kahneman and Tversky's (1979) prospect theory for preference. These "risk weights" would then replace the objective probabilities in the CER model. Such a nonlinear probability-to-risk-weight transformation, r, could for example predict that the risk of the compound lottery $(X\text{o}_p 0)\text{o}_q 0$ is perceived as greater than that of the equal-in-final-outcomes lottery $(X\text{o}_{pq} 0)$ where 0 is the prospect of getting zero, if $r(p)r(q) > r(pq)$. Satisfying this condition would require an r-function different from Kahneman and Tversky's decision weight function, π. (In particular, r would require some *over*weighting of probabilities.) However, the generality of the result that two-stage compound lotteries are perceived as riskier than single-stage lotteries needs to be further explored. A speculative psychological interpretation of the phenomenon might take the form of making "risk" some function of the emotional reaction to the resolution of uncertainty. A compound lottery which involves two such resolutions stages (uncertainty nodes) would then be perceived as riskier than the corresponding single-stage lottery with only one resolution stage.

A weaker version of probability accounting, requiring only the equivalence of compound lotteries $(X\text{o}_p Y)\text{o}_q Y$ and $(X\text{o}_q Y)\text{o}_p Y$, along with transitivity and monotonicity was shown by Luce and Narens (1985) to lead to the dual bilinear representation of preference. Luce (1988) generalized this model to lotteries with more than two consequences. A similar weakening of the expectation axioms may be necessary for the CER model as well.

The second violation was the failure of the conjoint structure axioms for positive-outcome-only lotteries. One subject in our sample rated all positive-outcome lotteries as having zero risk. Other subjects did discriminate between different lotteries of the positive-outcome set with respect to risk, but did so with visibly more difficulty and effort than subjects who judged the negative-outcome set. In Weber (1988), the relative weight parameters for the effect of positive outcomes on risk (A_+ and B_+) tended to be smaller than the corresponding parameters for negative outcomes. For some subjects, A_+ and B_+ were quite small, suggesting that risk comparisons between positive-outcome-only lotteries may be difficult. Such subjects may respond randomly or use idiosyncratic simplifying strategies to make such comparisons.

On the other hand, there are numerous examples of positive-outcome lotteries for which people find it easy to make relative risk discriminations. The experimental work of Lopes (e.g., 1984, 1986), for example, makes extensive use of positive-outcome-only monetary lotteries which her subjects discriminate easily and reliably with respect to risk. Outside the laboratory, a large percentage of investment decisions involve positive-outcome-only lotteries (i.e., alternatives whose worst outcomes provide a zero-rate of return on investment). Yet, few investors would claim that all such options have zero or even equal risk. In both of these situations, the concept of an aspiration level for outcomes may help to explain the ease of making relative risk judgments. An aspiration level or reference point that is used to encode outcomes as either gains (positive deviations) or losses (negative deviations) is a central feature of Kahneman and Tversky's (1979) prospect theory for preference. The utility of the aspiration level concept in explaining choice pattern was convincingly demonstrated by Payne, Laughhunn, and Crum (1980, 1981). Lopes (1984) gives the aspiration level concept a central role in risk judgments as well, by suggesting that "the perception of risk depends critically on one's aspiration level" and that individual differences in risk attitudes (i.e., "risk seeking" vs "risk avoiding") may in fact be caused and/or modeled by differences in aspiration level.

In Lopes' adoption of welfare economics' Lorenz curves to represent risky choice alternatives, aspiration level differences are modeled as differences in value of a parameter of an inequality index. (See Lopes [1984] for details.) The CER model can incorporate aspiration level effects perhaps more directly by taking as its domain not lotteries with the objectively stated outcomes but instead lotteries whose outcomes have been recoded as deviations from a subject's aspiration level. In that case, lotteries that objectively have only positive outcomes will be encoded as such only when a subject's aspiration level is zero or the status quo. (This, in the absence of expectations or instructions to the contrary, was probably the result in our experiments.) On the other hand, if a subject's aspiration level takes a positive value, objectively positive outcomes that fall short of the aspiration level will be encoded as "losses." In Lopes' experiments, subjects received implicit or explicit aspiration level instructions. In investment decisions, indices such as the Dow-Jones provide readily available positively valued aspiration levels.

When the notion of an aspiration level is incorporated into a model as in prospect theory or potentially into the CER model, it becomes necessary to provide an adjunct theory of the origin of and changes in aspiration levels. Only with such a theory would a choice or risk model that employs the aspiration level concept maintain its predictive power. A series of studies to that effect can be found in Bottom (1989).

REFERENCES

Allais, M. (1953). Le comportement de l'homme rationel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, **21**, 503–546.

Allais, M. (1979). The foundation of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school In: M. Allais, & O. Hagen (Eds.), *Expected utility hypotheses and the Allais paradox*. (pp. 27–145). Dordrecht: Reidel.

Bettman, J. R. (1973). Perceived risk and its components: A model and empirical test. *Journal of Marketing Research*, **10**, 184–190.

Bottom, W. P. (1989). *A theory of adaptive reference points in decision making and negotiation*. Unpublished doctoral dissertation, University of Illinois. Urbana-Champaign, IL.

Coombs, C. H. (1969). Portfolio theory: A theory of risky decision making. In *La decision*. Paris: Centre National de la Recherche Scientifique.

Coombs, C. H. (1975). Portfolio theory and the measurement of risk. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision* (pp. 63–68). New York: Academic Press.

Coombs C. H., & Bowen, J. N. (1971). A test of VE-theories of risk and the effect of the central limit theorem. *Acta Psychologica*, **35**, 15–28.

Coombs, C. H., & Lehner, P. E. (1984). Conjoint design and analysis of the bilinear model: An application to judgments of risk. *Journal of Mathematical Psychology*, **38**, 1–42.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns. *American Economic Review*, **67**, 116–126.

Fishburn, P. C. (1982). Foundations of risk measurement. II. Effects of gains on risk. *Journal of Mathematical Psychology*, **25**, 226–242.

Fishburn, P. C. (1984). Foundations of risk measurement. I. Risk as probable loss. *Management Science*, **30**, 396–406.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrika*, **47**, 263–291.

Keller, L. R., Sarin, R. K., & Weber, M. (1986). Empirical investigation of some properties of the perceived riskiness of gambles. *Organizational Behavior and Human Decision Processes*, **38**, 114–130.

Libby, R., & Fishburn, P. (1977). Behavioral models of risk taking in business decisions: A survey and evaluation. *Journal of Accounting Research*, **15**, 272–292.

Lopes, L. (1984). Risk and distributional inequality. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 465–485.

Lopes, L. (1986). Between hope and fear: The psychology of risk. *Advances in experimental social psychology*, **20**, 255–295.

Luce, R. D. (1980). Several possible measures of risk. *Theory and Decision*, **12**, 217–228.

Luce, R. D. (1981). Correction to 'Several possible measures of risk.' *Theory and Decision*, **13**, 381.

Luce, R. D. (1988) Rank-dependent subjective expected utility representations. *Journal of Risk and Uncertainty*, 1, 305–332.

Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology*, 29, 1–72.

Luce, R. D., & Weber, E. U. (1986). An axiomatic theory of conjoint, expected risk. *Journal of Mathematical Psychology*, 30, 188–205.

Payne, J. W. (1973). Approaches to decision making under risk: Moments versus risk dimensions. *Psychological Bulletin*, 80, 439–453.

Payne, J. W., Laughhunn, D. J., & Crum, R. (1980). Translation of gambles and aspiration level effects in risky choice behavior. *Management Science*, 26, 1039–1060.

Payne, J. W., Laughhunn, D. J., & Crum, R. (1981). Further tests of aspiration level effects in risky choice behavior. *Management Science*, 27, 953–959.

Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence, and limitations. *Journal of Economic Literature*, 20, 529–563.

Weber, E. U. (1984a). Combine and conquer: A joint application of conjoint and functional approaches to the problem of risk measurement. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 179–194.

Weber, E. U. (1984b). *Risk as an independent variable in risky choice*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Weber, E. U. 1988. A descriptive measure of risk. *Acta Psychologica*, 69, 185–203.

Weber, E. U., Anderson, C., & Birnbaum, M. H. (1990). *Perceived risk and attractiveness of lotteries*. Manuscript submitted for publication.

Weber, E. U., & Bottom, W. P. (1989). Axiomatic measures of perceived risk: Some tests and extensions. *Journal of Behavioral Decision Making*, 2, 113–131.

Weber, M., & Camerer, C. (1987). Recent developments in modelling preference under risk. *OR Spectrum*, 9, 129–151.

RECEIVED: February 18, 1988