*Journal of Behavioral Decision Making, Vol. 2, 113–131 (1989)*

# Axiomatic Measures of Perceived Risk: Some Tests and Extensions*

ELKE U. WEBER
*Center for Decision Research, University of Chicago, USA*
WILLIAM P. BOTTOM
*Washington University, USA*

ABSTRACT

The history of axiomatic measurement of perceived risk of uni-dimensional risky choice alternatives is briefly reviewed. Experiments 1 and 2 present data that distinguish between two general classes of risk functions (those that assume that gain and loss components of an alternative combine additively versus multiplicatively) on empirical grounds. The most viable risk model on the basis of these and other results is described. Experiment 3 presents data that call into question the descriptive adequacy of some of this risk model's assumptions, in particular the expectation principle. Suggestions for possible modifications are made.

KEY WORDS Perceived risk Allais paradox Reflection effect Axiomatic measures of risk

The role of perceived risk on judgment and choice has been investigated in a variety of areas such as psychology (e.g., Coombs, 1969), marketing (Bettman, 1973), and economics (Markowitz, 1959; Libby and Fishburn, 1977). Confidence in expected utility models (the longstanding normative theory of risky choice) as a descriptive model of choice has slowly eroded in the face of evidence of its shortcomings (e.g., Schoemaker, 1982; M. Weber and Camerer, 1987). This has led researchers in the decision sciences to a search for additional variables responsible for preference.

One promising candidate has been the familiar, but ill defined, concept of risk. Theories of choice that incorporate risk as a central concept have not achieved wide acceptance because of the absence of a descriptively adequate *measure* of risk. Consensus is nevertheless growing that people may base their decisions on qualitative aspects of choice alternatives separate and/or orthogonal to expected utility. These aspects may be captured by the label 'risk'. Weber, Birnbaum, and Anderson (1988), for example, suggest that people encode and combine probability and outcome information in qualitatively different ways when judging the attractiveness of a choice attractiveness (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 modeled as a function of two such constructs. A similar idea underlies Coombs' (1972) characterization of choice as a conflict between greed

*This work was supported by National Science Foundation Grant SES 86-06750 to the first author.

0894-3257/89/020113-19 $09.50


*Received 23 December 1987*
*Revised 15 June 1988*

and fear. Dyer and Sarin (1986) illustrate how the explicit inclusion of risk into decision models can explain many of the paradoxes of traditional utility theory.

In addition to affecting choice directly in the evaluation-of-alternative stage, perceived risk of decision alternatives may also affect further information acquisition (e.g., by acquiring more information about alternatives perceived as riskier). Thus, perceived risk may affect choice indirectly via qualitatively and/or quantitatively different information search. Schwartz and Griffin (1986) summarize the evidence of such effects of risk perceptions on preference in medical decision making.

There is, however, little agreement on a *definition* of risk. Decision making under risk (or risky choice) refers to choices among alternatives that can be described by probability distributions over possible outcomes. An additional assumption, often made implicitly, is that at least one of the possible outcomes ought to be undesirable, or at least less desirable than the others for risk to exist. In the case of social or technological risks, undesirable usually means some threat to life, health, or welfare (Fischhoff, Lichtenstein, Slovic, Derby and Keeney, 1981, p. 2). For financial risks, undesirable can mean a loss or a return below aspiration level (Fishburn, 1977; Kahneman and Tversky, 1979).

Researchers do not agree on the relative importance of the uncertainty of outcomes versus the undesirability of outcomes in determining subjective risk. Fishburn (1982) emphasizes the latter by assuming that any lottery that has no possibility of a loss has zero risk. Similar emphasis on the undesirable outcome aspect of risk comes from the psychological research on risk perception (Fischhoff, Slovic and Lichtenstein, 1982; Johnson and Tversky, 1984; Vlek and Stallen, 1981) which seeks to identify underlying dimensions of 'badness' (e.g., degree of voluntariness, disaster potential) perceived by people in social or technological risks.

Other researchers, particularly economists, emphasize the uncertainty of outcomes aspect of risk. Rothschild and Stiglitz (1970) use the terms 'riskier', 'more variable', and 'more uncertain' interchangeably. Kaplan and Garrick (1981) combine both aspects in their symbolic equation 'Risk = Uncertainty + Damage.'

This paper 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'). Psychological risk dimensions of the type discussed in the last paragraph are not considered in this research area, mainly because the alternatives that are being compared would all rate quite similar on such dimensions as voluntariness of exposure or 'dread'. The basic question in this area of research is how people combine the objective probability and outcome information to arrive at judgments of 'riskiness' for different alternatives. To the extent that people's perceptions of the riskiness of choice alternatives differ, a measurement model of perceived risk should be able to capture individual differences while attempting to account for between-subject similarities in judgments at the same time.

Measurement in this area has tended to be 'axiomatic' or foundational. That is, risk measures are mathematically deduced from a set of plausible qualitative assumptions (axioms) regarding properties of the construct under study (i.e., perceived risk). The advantage of an axiomatic measurement model is that evaluation of its empirical adequacy is not limited to goodness-of-fit measures which have well-documented inadequacies (Birnbaum, 1973). Instead, the model can also be evaluated by testing whether the *qualitative properties* or axioms are empirically satisfied. Tests of model axioms are not only more sensitive than goodness-of-fit measures, they also identify the locus of inadequate model fits and thus provide the basis for model revisions (i.e., by replacing an assumption that is empirically falsified with a more appropriate alternative). Examples of this approach are discussed in more detail below.

The next section provides a brief review of developments in this area over the last 20 years. This is followed by an account of three experiments. Experiments 1 and 2 successfully discriminate between two general classes of risk models. Experiment 3 calls into question the descriptive adequacy of some of the assumptions of the currently most viable risk measure in this domain.

## MEASUREMENT OF SUBJECTIVE RISK

Portfolio theory (Coombs, 1969, 1975) was the first psychological theory of choice that incorporated risk as an explanatory variable. When Coombs developed the theory, he left the nature of risk itself undefined. Prompted by portfolio theory, several researchers attempted to develop an independent measure of risk.

Pollatsek and Tversky (1970) derived a risk function that was a linear combination of an alternative's mean outcome and its variance. They incorporated some of the assumptions made by Coombs (1969, 1975) about subjective risk as axioms in a risk system closely related to an extensive structure. Coombs and Bowen (1971a), however, showed Pollatsek and Tversky's risk measure to be empirically inadequate. By using transformations on gambles that left expectation and variance unchanged, they found that risk varied systematically with the skewness of a gamble.

Subsequent approaches to the problem of risk measurement considered how certain transformations of a risky choice alternative or gamble (e.g., changes of scale — multiplying all outcomes by a constant, or changes of origin — adding or subtracting a constant amount from all outcomes) affected people's perceptions of its riskiness (Coombs and Bowen (1971a, b), Coombs and Huang (1970a, b), Coombs and Lehner (1981)). Luce (1980, 1981) took this approach one step further by deriving some risk measures equivalent to certain functional equations relating the perceived risk to transformations on the gambles. He considered, first, the effect of a change of scale on risk and, in particular, studied the two simplest possibilities, an additive and a multiplicative effect. Second, he considered two ways in which outcomes and probabilities might be aggregated into a single risk value. The two simplest possibilities seemed to be, first, a form analogous to expected utility integration which resulted in an expected risk function already suggeted by Huang (1971a). The second possibility considered was that of the density undergoing a transformation before integration. From the combination of options considered at these two choice points, four distinct possible measures were derived. Luce left examination of their descriptive validity to empirical investigation. Several investigators undertook that task.

Weber (1984a) found Luce's first choice point between an additive or multiplicative effect of a change of scale on risk to be indeterminate. For Luce's second choice point regarding integration of probabilities and outcomes into a single risk value, the assumption of the density undergoing a transformation before integration leads to risk functions that are insensitive to change of origin of the random variable representing the gamble. This simple fact ruled out all measures of this type because it was shown empirically that the subjective risk of a gamble is significantly affected by a change of origin (Coombs and Bowen, 1971a; Keller, Sarin and M. Weber, 1986; Weber, 1984a).

Further empirical work by Weber (1984b, in press) eventually led to a revision and expansion of Luce's original (1980, 1981) set of risk axioms or assumptions. This new axiomatic risk function, called conjoint expected risk (CER), was developed in Luce and Weber (1986). The results of empirical tests of the theoretically interesting and testable assumptions of the CER model can be found in Weber and Bottom (1988).

The resulting risk function, R, is called CER for Conjoint Expected Risk:

$$R(X) = A_o \Pr(X = 0) + A_+ \Pr(X > 0) + A_- \Pr(X < 0) \\ + B_+ E[X^{k+} | X > 0] \Pr(X > 0) + B_- E[|X|^{k-} | X < 0] \Pr(X < 0).$$

Thus, the perceived risk of 'gamble' X is a linear combination of the probability of breaking even (zero outcomes), the probability of positive outcomes, the probability of negative outcomes, the conditional expectation of positive outcomes raised to some power $k_+$, and the conditional expectation of negative outcomes raised to some power $k_-$ with $k_+$ and $k_- > 0$. Parameters $A_0$, $A_+$, $A_-$, $B_+$, and $B_-$ are weights on the respective components.

A nice feature of the CER function is that it retains the benefits of expectation models, namely a constant number of parameters regardless of the number of outcomes, a property that is not shared by risk dimension models of the kind suggested by Payne (1973). Fits of the CER function to perceived riskiness judgments of monetary gambles are reported in Weber (1988). The CER function also does well in an evaluation against other empirical evidence regarding subjective riskiness. Coombs and Bowen (1971b) reported that when two gambles are convolved, the risk of the resulting gamble is not an additive function of the risk of the two component gambles. This fact was an additional strike against Pollatsek and Tversky's (1970) risk function and eliminated Coombs and Huang's (1970a) polynomial model of perceived risk because both models predicted additivity. It is easy to see that the CER function does not predict such additivity.

Another instance where CER seems to provide a superior prediction of empirical phenomena is the effect of a change in expected value on risk. One of Coombs' (1972) assumptions about risk was that relative risk order remains unaffected by changes in expected value. Payne, Laughhunn, and Crum (1980, 1981) on the other hand, brought about reversals in relative risk by increasing the expected value of two gambles by the same amount. With the right choice of parameters, CER predicts such reversals.

## ADDITIVE vs MULTIPLICATIVE COMBINATION OF GAINS AND LOSSES

One basic difference between classes of risk functions is the way positive and negative components of a choice alternative are assumed to combine. A central characteristic of the CER model is its *additivity* in the contributions of a gamble's positive components (gains) and its negative components (losses). In this respect the CER function is a special case of the bilinear risk model suggested by Coombs and Lehner (1984) and Huang's expected risk function (1971a), all of which postulate an additive combination of gains and losses (i.e., gains reduce risk to an extent that does not depend on the magnitude of the losses). This assumption is in direct contrast to the approach taken by Fishburn (1982) who derives —from a large set of elementary axioms — a set of risk functions that are all *multiplicatively* separable in gains and losses (i.e., gains affect risk by modifying the effect of losses, and every gamble with no chance of a loss is regarded as nonrisky).

Coombs and Lehner (1984) developed a test for the additivity of positive and negative components. However, their results were somewhat equivocal and perhaps open to alternative explanations. (See the discussion in the Appendix of Weber (1984a)). Birnbaum (1974, 1982) suggests to discriminate between an additive and a non-additive representation using difference judgments. Assume that a gamble is described by a positive component or gain, P, and a negative component or loss, N. If P and N combine additively to determine risk, i.e., $R(P_iN_i) = P_i + N_i$, the difference-in-risk judgment for pairs of lotteries that differ on one component (e.g., N) should not depend on the value of the other component if that component is the same for both members of the pair:

$$\begin{aligned} \mathrm{Diff}\{R(P_1N_1), R(P_1N_2)\} &= J(P_1 + N_1 - P_1 - N_2) = J(N_1 - N_2) \\ &= J(P_2 + N_1 - P_2 - N_2) = \\ &\quad \mathrm{Diff}\{R(P_2N_1), R(P_2N_2)\} \end{aligned}$$

where J is an monotonic response judgment function. This equality will not hold if P and N combine non-additively.

## EXPERIMENT 1

Experiment 1 was designed to discriminate between additive and non-additive combinations of positive and negative outcome contributions for the judgment of perceived riskiness.

### Method

One of three positive components for a gamble was factorially combined with one of three negative components to form nine gambles as shown in Exhibit 1. The positive and negative components were selected to provide a variety of payoff values ranging from low to high. One of the negative and one of the positive components consisted of a single outcome. The others consisted of two outcomes. Their combination thus resulted in two-, three-, or four-outcome lotteries. Lotteries will be referenced by the positive and negative components from which they were constructed, for example P1N3.

| | |
|---|---|
| P1: (+$55, .40) | N1: (–$195, .60) |
| P2: (+$12, .25; +$5, .15) | N2: (–$99, .20; –$45, .40) |
| P3: (+$160, .10; +$90, .30) | N3: (–S17, .30; –$8, .30) |
| P4: (+$210, .40) | N4: (–$63, .60) |
| P5: (+$110, .15; +$35, .25) | N5: (–$240, .2; –$135, .40) |

Exhibit 1. Positive (P1-P5) and negative (N1-N5) component of lotteries. The factorial combination of P1-P3 and N1-N3 formed the nine stimulus lotteries of Experiment 1. The factorial combination of P1-P5 and N1-N5 formed the 25 stimulus lotteries of Experiment 2

All 36 possible pairwise comparisons of the 9 stimulus gambles were presented to each of 14 participants who were undergraduates at the University of Illinois fulfilling a credit requirement for an introductory psychology course. Order of lotteries within each pair and order of pairs were random. Lottery pairs were presented in a questionnaire booklet, one pair per page, and were represented as follows. Outcomes were listed in increasing order from left to right. Monetary outcomes (losses indicated by a minus sign) appeared above the respective probabilities. The probabilities of outcomes were depicted by a proportionate number of X's as well as by their numerical values.

Students first judged which lottery of a given pair was riskier, then rated the difference in risk on a numerical 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 students that they were to judge differences in *risk* and *not* in preference or desirability. Nobody found this distinction difficult.

A set of six practice trials familiarized subjects with the task as well as with the range of items they would encounter in the experiment. The experiment was replicated with the same subjects once more on the same day and twice more, two days later.

### Analysis and results

In order to differentiate judgments that gamble A is riskier than gamble B from judgments that B is riskier than A, the risk difference ratings were recoded as negative values for all those cases where a subject judged B to be riskier than A. A repeated measures one-way analysis of variance was performed on subject's difference ratings (a separate analysis for every subject) with the 36 lottery pairs as levels of analysis.

For Birnbaum's difference test, the hypothesis of additivity of positive and negative components predicts equality of risk difference ratings for certain lottery pairs. Specifically, these were three subsets of three pairs each that differed on the negative component but had the same positive component and three subsets of three pairs each where the opposite was true. To maintain a reasonable family-wise Type I error rate, two different procedures (Least Significant Difference, LSD, and the Scheffé test), were used to conduct the pairwise comparisons.

Exhibit 2 summarizes the percentage of pairwise comparisons which were significantly different from each other (and thus constitute violations of additivity). The results convincingly favour additivity. Using the least-significant difference procedure which (for multiple comparisons) strongly biases the results towards rejecting the null hypothesis (i.e., towards rejecting additivity), there are only a small number of violations of additivity (8.6% and 1.6% across subjects for the 'negative' and 'positive' comparisons, respectively). Most subjects had a violation rate of 0% to 5%. Using the Scheffé procedure, which is less biased towards rejecting the null hypothesis, no subjects showed *any* violations of additivity, thus providing evidence that the violations detected with the LSD procedure were not very severe. In combination, these results confirm Coombs and Lehner's (1984) conclusion of additivity of positive and negative components. The data seem to rule out Fishburn's (1982) multiplicatively separable risk functions on empirical grounds.

| | | LSD procedure[a] | | Scheffé procedure[b] | |
|---|---|---|---|---|---|
| Experiment | N | NEGATIVE[c] | POSITIVE[d] | NEGATIVE[c] | POSITIVE[d] |
| 1 | 14 | M = 8.6%<br>(S = 6.3%) | M = 1.6%<br>(S = 3.9%) | M = 0%<br>(S = 0%) | M = 0%<br>(S = 0%) |
| 2 | 39 | M = 14.3%<br>(S = 6.9%) | M = 13.8%<br>(S = 6.5%) | M = 2.0%<br>(S = 4.5%) | M = 1.9%<br>(S =4.2%) |
| 2[e] | 36 | M = 11.8%<br>(S = 5.1%) | M = 10.4%<br>(S = 4.7%) | M = 0.4%<br>(S = 1.3%) | M = 0.3%<br>(S = 1.2%) |

[a] Least significant difference procedure, i.e., $t_{.975;108}\sqrt{MSE(1/4 + 1/4)}$ for Experiment 1 and $t_{.975;1500}\sqrt{MSE(1/6 + 1/6)}$ for Experiment 2, least conservative procedure for multiple comparisons in the sense that it will most likely result in rejection of the null hypothesis of equality (additivity).
[b] Scheffé procedure, i.e., $3\, F_{.95;35,108}\sqrt{MSE(1/4 + 1/4)}$ for Experiment 1 and $5\, F_{.95;299,1500}\sqrt{MSE(1/6 + 1/6)}$ for Experiment 2, multiple comparison procedure less biased towards rejection of the null hypothesis of equality (additivity).
[c] Compared pairs of risk difference judgments came from lotteries that differed on the negative component but had the same positive component.
[d] Compared pairs of risk difference judgments came from lotteries that differed on the positive component but had the same negative component.
[e] Subset of 36 subjects, excluding three subjects with a larger number of violations.

Exhibit 2. Results of tests of additivity of a lottery's negative and positive components in their effect on subjective risk using the least significant difference (LSD) and Scheffé multiple comparison procedures. Table lists the mean (M) percentage of violations across the individual subjects' test and corresponding standard deviations (S)

## EXPERIMENT 2

Experiment 2 is a replication of Experiment 1 which uses a different procedure to determine the difference-in-risk data necessary for the additivity analysis. In Experiment 1, subjects judged the difference in risk between two lotteries directly. Pairs that entered the additivity analysis were those that had one common component (either positive or negative). It is possible that subjects would 'edit out' (i.e., ignore) those common components and base their difference in risk judgment entirely on the values of the other (differing) component. Such behavior would give the appearance of additivity of positive and negative components, but in fact does *not* speak to the issue of how the two components are combined when they are *both* being considered.[1] Experiment 2 was designed to test this hypothesis.

### Method

In this experiment, 39 participants (from the same pool as in Experiment 1) were asked to rate the riskiness of *individual* lotteries. Risk differences between pairs of lotteries were subsequently computed as the difference in risk judgments for the individual members of the pair. Ratings of risk for individual lotteries (rather than difference ratings between all possible pairs) allowed us to use a larger number of items. The five positive components of Exhibit 1 were factorially combined with the five negative components to form 25 lotteries.

Subjects rated the riskiness of these lotteries which were presented individually and in a random order. Instructions and presentation format of lotteries were similar to those of Experiment 1. The experiment was replicated with the same subjects twice more on the same day and three additional times, two days later.

### Analysis and results

For each of the 39 participants and for each replication of the experiment, risk differences for all possible $\binom{25}{2} = 300$ pairs of lotteries were computed as the difference between the risk judgments given for the members of a pair. This was done separately for all 6 replications. A repeated measures one-way analysis of variance was performed on each subject's difference ratings with the 300 pairs as levels of analysis.

The additivity hypothesis predicts equality of risk difference ratings for ten subsets of ten pairs each that differ on the negative component but have the same positive component and for ten subsets of ten pairs each where the opposite is true. Exhibit 2 summarizes the percentage of pairwise comparisons which were significantly different from each other (and thus constitute violations of additivity). As in Experiment 1, the evidence argues *in favor of addivitity*, thus ruling out the possibility that 'editing' of common components was responsible for the result.

Using the least-significant difference procedure, subjects show an average violation-of-additivity rate of 14.3% and 13.8% for 'negative' and 'positive' comparisons, respectively. A large part of this average violation rate is attributable to three subjects (S4, S17, and S22). Exhibit 2 lists violation rates both with and without those subjects. Using the Scheffé procedure (less biased towards rejecting the null hypothesis of additivity), violation rates again dropped dramatically, indicating that most of the violations detected with the LSD procedure were not very severe.

Ratings of risk (averaged across all 39 subjects) are plotted in Exhibit 3 for the 25 stimulus lotteries as a function of their positive and negative component. Each line connects lotteries that have the same negative component but different positive components. Positive components are ordered along the abscissa by the size of their marginal risk ratings. For the across-subjects ratings of Exhibit 3, the

[1] We thank Jerry Busemeyer (personal communication, August 7, 1987) for pointing out this possibility.

relative (increasing) order of positive components with respect to risk was P4, P3, P5, P1, and P2. The relative (increasing) risk order for negative components was N3, N4, N2, N1, and N5. Additive separability implies parallel lines. With the exception of the lowest and the highest data point (possibly floor and ceiling effects), the data clearly support additivity.

Plotting the risk ratings of individual subjects provides a similar picture. The plot for Subject 33 in Exhibit 4 is representative for most of the subjects. The risk order of negative components was identical for all subjects. There was some variability in their ordering of positive components with respect to risk. (Mainly in the relative order of P3, P4, and P5.) Exhibit 4 also shows the risk ratings for the three subjects (S17, S4, and S22) who contributed most of the violations to the additivity analysis. For intermediate risk values, even their plots show a lot of parallelism. There certainly is no evidence of a multiplicative combination of components as postulated by Fishburn (1982).

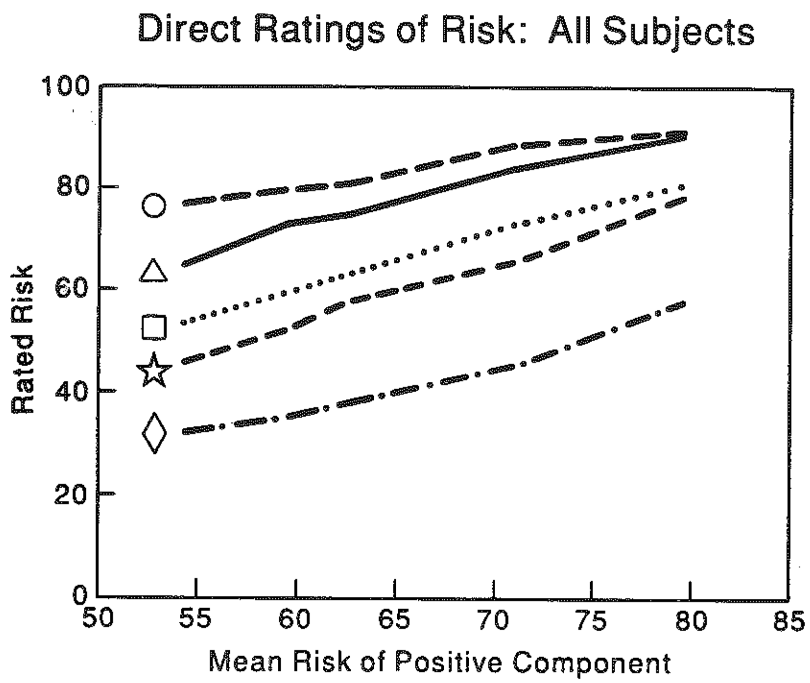


Exhibit 3. Mean risk ratings averaged across subjects for 25 stimulus lotteries plotted as a function of positive and negative components
Circle = N5; Triangle = N1; Square = N2; Star = N4; Diamond = N3

## EXPECTATION PRINCIPLE FOR RISK

In light of Experiments 1 and 2, the conjoint expected risk (CER) model and Coombs and Lehner's (1984) bilinear risk model appear to be the most viable models of perceived risk. The CER model can be considered a generalization of the bilinear risk model as the latter applies only to two-outcome lotteries.

One of the axioms on which the CER model is based (see Luce and Weber (1986)) is the assumption that risk judgments satisfy the regularities of a mixture space (see, Fishburn, 1970). Therefore, the risk of a choice alternative that is a (probability) mixture of components X and Y is equal to $R(X \circ_p Y)=$

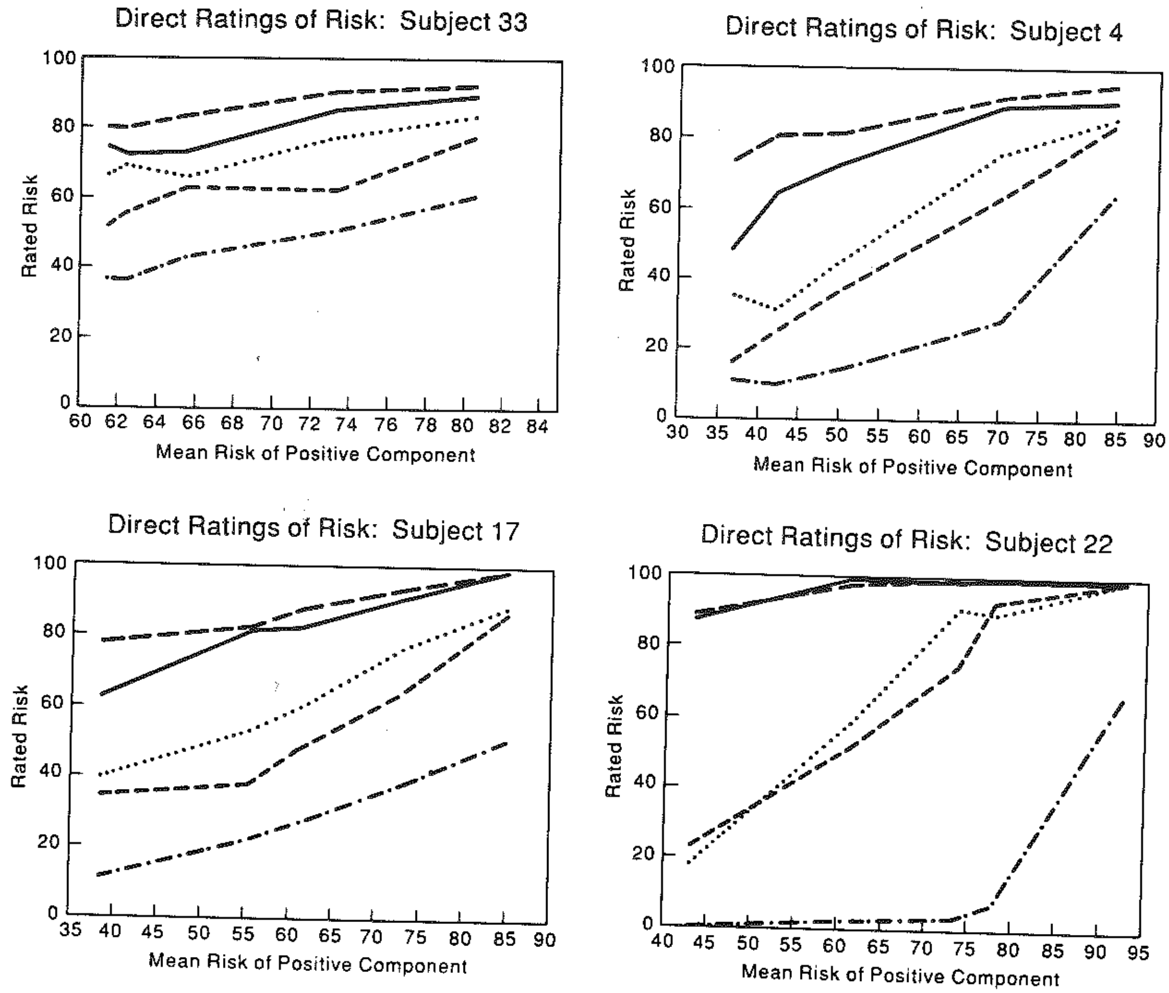


Exhibit 4. Individual subject risk ratings averaged across six replications for 25 stimulus lotteries plotted as a function of positive and negative components. Legend as in Exhibit 3

pR(X) + (1−p)R(Y), where p is the probability with which X can be expected to occur in the compound gamble X $\circ_p$ Y. This axiom (henceforth referred to as the expectation principle) was first proposed for risk by Huang (1971a).

When applied to *preference*, the assumptions of a mixture space lead to the well-known expected utility representation. In the preference of choice domain, some of the mixture space assumptions have been found to be empirically inadequate. One early demonstration of violations of the expectation principle for preference was the common ratio test described by Allais (1953), generally referred to as the Allais paradox. Subjects in the common ratio test often violate the substitution axiom which asserts that if gamble X is preferred to gamble Y, then any (probability) mixture X $\circ_p$ Z must be preferred to Y $\circ_p$ Z, regardless of the particular value of *p* or component Z.

In the domain of *risk* judgments, a number of empirical studies have reported data in which the expectation principle was satisfied (e.g., Aschenbrenner, 1978; Coombs and Bowen, 1971b; Huang, 1971b). However, Keller, Sarin, and M. Weber (1986) reported significant violations of the expectation principle. In close analogy to the stimuli used by Allais (1953), they designed six pairs of lotteries (an example is shown in Exhibit 5) consisting of three 'original pairs' and three 'transformed pairs.' In the original pairs, the 'WP' lottery has the *worse probability* of loss whereas the 'WL' lottery has the *worse* possible *loss*. The transformed pair of lotteries is constructed by taking a $p$ probability of getting the original lottery and a $(1-p)$ probability of getting $0. To be consistent with the expectation principle, WP' should be riskier than WL' in a transformed pair of lotteries if and only if WP was judged riskier than WL in the corresponding pair of original lotteries.

Keller *et al.* (1986) found that only about one-third of the risk comparisons provided by U.S. and German business graduate students (total N = 70) were consistent with the expectation principle. Furthermore, the pattern of violation was systematic, such that for pairs with high probability of loss (the 'original' pairs), the gamble with the worse probability of loss was considered riskier, whereas for pairs with low probability of loss (the 'transformed' pairs), the gamble with the higher loss amount was judged to be riskier.

| Original Pair (High probability of loss) | Transformed Pair (Low probability of loss) |
|---|---|
| WP: [ 5%, $0; 95%, -$300] | WP': [90.5%, $0; 9.5%, -$300] |
| WL: [25%, $0; 75%, -$400] | WL': [92.5%, $0; 7.5%, -$400] |

Exhibit 5. Sample of stimulus pairs used by Keller, Sarin and Weber (1986) to test the expectation principle

As discussed by Luce and Weber (1986), the Allais paradox is only an indirect indictment of the substitution principle (or monotonicity assumption, as it is called in the axiomatic measurement literature). (See also Weber and Bottom (1988) for further discussion.) It may well be that the substitution principle or monotonicity per se holds, but the people employ subjective probabilities that are nonlinear functions of the stated probabilities when estimating subjective risk. Experiment 3 was designed to replicate and expand Keller *et al.*'s (1986) investigation of the applicability of the expectation principle for risk judgments.

## EXPERIMENT 3

### Method

Seventeen undergraduates from the University of Illinois participated for credit in an introductory psychology class. To compare violations of the expectation principle for risk to violations for preference (i.e., the Allais paradox, which has been investigated using mainly positive outcome lotteries), we created parallel positive and negative outcome sets of lottery pairs (i.e., in one set of lottery pairs all outcomes were positive, while the other set contained the lotteries where every gain was replaced by a loss of equivalent size). The positive and negative outcome sets consisted of the six lottery pairs employed by Keller *et al.* (1986), the four lottery pairs used by Kahneman and Tversky (1979, [Problems 3 and 4, 7 and 8]) as well as 18 lottery pairs used by MacCrimmon and Larsson (1979) to investigate the substitution principle for preference.

For the MacCrimmon and Larsson (1979) pairs, one member consisted of outcome $0_1$ with probability $P$, and zero outcome otherwise; the other member was constructed as outcome $0_1$ multiplied by factor F with probability $4P/5$, and zero outcome otherwise. Three levels of $P$ (1.0, 0.75, 0.05), three levels of $0_1$ (\$10, \$100, and \$1000), and two levels of F (1.5 and 3) were orthogonally combined to create the 18 pairs. Pairs equal in $0_1$ and F but differing in $P$ (e.g., $P'$ and $P''$) are equivalent to the 'original' and 'transformed' pairs of Keller *et al.* (1986), in the sense that one can be considered as arising from a probability mixture of the other, with zero outcome otherwise, where the $p$ of the probability mixture is equal to the ratio $P'/P''$, where $P' < P''$.

For the six $0_1$ by F combinations, two Allais-type comparisons were made, those from $P' = 1.0$ to $P'' = 0.05$ and those from $P' = 0.75$ to $P'' = 0.05$. This makes for 12 Allais-type comparisons for both the positive and the negative outcome set. There were two comparisons for the four Kahneman and Tversky (1979) item pairs, and three for the six Keller *et al.* (1986) item pairs.

Participants were asked to compare all 56 lottery pairs (28 each for the positive and negative outcome set) with respect to risk ('Which lottery is riskier?') and with respect to preference ('Which lottery would you prefer to play if you had to play one?'). The same individuals gave pairwise comparisons of the relative riskiness and preferability of all positive and all negative lottery pairs (i.e., a completely within-subject design). For every rating (e.g., 'Lottery A is riskier'), subjects also provided an index of the degree to which they judged the chosen lottery riskier or more preferable than the other (on a numerical rating scale ranging from 0 [equally preferable/risky] to 100 [extremely different in preferability/risk]).

The order of pairs within the list and the order of lotteries within a pair were randomized. Lotteries were represented as described in Experiment 1, with probabilities shown both graphically and numerically. Keller *et al.*'s graphic representation of probabilities employed chance wheels, whereas we used a proportionate number of X's. The two representations appear similar in the extent to which they reveal (or rather fail to reveal) the common-ratio aspect of the stimulus pairs.

**Results**

Exhibit 6 shows the percentages of judgments from a total of 17 subjects that were *consistent* with the expectation principle. This is further broken down separately for risk and preference comparisons and for the three 'classes' of stimuli. It is obvious that violations occur in both risk and the preference domain.

With the exception of risk judgments for positive outcome lotteries, only slightly more than half of all risk and preference judgments are consistent with the expectation principle. We do not have a ready explanation for the apparent reduction of violations in the risk judgments for positive outcome lotteries. One possible explanation could lie in the greater between-subject agreement in risk ratings for the positive than for the negative domain (the standard deviations in risk difference ratings tend to be smaller) which may lead one to infer that within-subject agreement may also be higher. The reason for this may be that for mixed or negative outcome lotteries, the two definitions of risk discussed in the introduction (risk as loss, and risk as uncertainty or variability of outcomes) are both present (and perhaps competing). For positive outcome lotteries, however, only the latter definition is relevant and thus greater consistency might be expected. At present time, this is pure speculation. Furthermore, an elaboration of the CER model incorporating the concept of aspiration level as an outcome evaluation reference point (see Weber and Bottom, 1988) would introduce 'losses' as deviations from aspiration level even for positive outcome lotteries.

Keller *et al.* (1986), using a sample of business graduate students for risk judgments in the negative outcome domain, found somewhat more violations of the expectation principle (69%) than this study with its sample of undergraduates (48%). One possible explanation for this could be differences in the visual representation of the lottery stimuli, as found by Keller (1985) for violations of the substitution principle for preference judgments.

Kahneman and Tversky (1979) and MacCrimmon and Larsson (1979) unfortunately do not report the percentage of pairs of judgments consistent with the expectation principle in their studies. (They

% Ratings *Consistent* with Expectation Principle for RISK

| Outcomes | MacCrimmon | Kahneman and Tversky | Keller *et al*[a] | Combined |
|---|---|---|---|---|
| | Items | | | |
| NEGATIVE | 55% | 48% | 53% (31%) | 54% |
| POSITIVE | 70% | 72% | 67% | 70% |

% Ratings *Consistent* with Expectation Principle for PREFERENCE

| Outcomes | MacCrimmon | Kahneman and Tversky | Keller *et al* | Combined |
|---|---|---|---|---|
| | Items | | | |
| NEGATIVE | 63% | 37% | 36% | 55% |
| POSITIVE | 59% | 53% | 33% | 53% |

[a] Figure in brackets is corresponding result reported by Keller, Sarin, and Weber (1986).

Exhibit 6. Results of test of expectation principle for risk and preference for positive-outcome and negative-outcome gambles.

only tabulate the percentages of choices (i.e., preferences) of either alternative in the two comparison pairs, from which violation of the expectation principle cannot be calculated). However, in their discussion, Kahneman and Tversky (1979) put the number of violations (for preference judgments in the positive and in the negative outcome domain) at 'less than 50%' which agrees well with the violation rates obtained in this experiment.

Thirdly, the data suggest that the expectation principle may be less often violated for risk than for preference. A sign test of the number of violations (across subjects) for risk compared to the number of violations for preference for the 34 comparisons is marginally significant at the .08 level. When broken down into negative and positive outcome set comparisons, there is no significant difference in the number of violations for risk over those for preference for the negative outcome set comparisons. There is, however, a highly significant effect ($p < .005$) for the positive outcome set, where violations for preference judgments exceed those for risk judgments.

Other questions to which the data speak are the following: (*a*) Are there patterns of violations, in the sense that certain item pair comparisons elicit more violations than others? (*b*) Are there individual differences in the frequency or the pattern of violations? (*c*) Do violations of the expectation principle co-occur for risk and for preference judgments (i.e., when a comparison shows a violation for, e.g., preference, will a violation for risk be more, less, or equally likely)?

The answer to the first two questions is essentially negative for the range of variables in our data. No systematic individual differences in violations of the expectation principle were found for either preference or risk comparisons. The MacCrimmon and Larsson items allow us to examine the effects of outcome size ($0_1$), transformation from certain outcome to uncertain outcome (1.0 to 0.05) versus from uncertain outcome to smaller uncertain outcome (0.75 to 0.05), and gain or loss domain, on violation frequency for risk or for preference.

Aside from a main effect of domain (more violations in the loss that in the gain domain (see Exhibit 6), none of the factors affected violation frequency for risk comparisons. For preference comparisons, there was no main effect for gain versus loss domain, and no effect of outcome size. There was, however, a marginally significant effect of the probability manipulation ($p < 0.06$).

Violations occurred more frequently when the original pair contained one member that was a certain outcome ($p = 1.0$) which was subsequently reduced to probability 0.05 than when that outcome in the original pair had only occurred with probability $p = 0.75$ and was then reduced to 0.05. Such a result might be predicted on the basis of Kahneman and Tversky's (1979) certainty effect and points towards nonlinearity between objective and subjective probabilities as the cause of the violations.

Exhibit 7 addresses the question of whether violations of the expectation principle for risk choices co-occur with those for preference choices. Out of a possible total of 289 violations of the expectation principle in the negative stimulus set (i.e., 17 comparisons by 17 subjects), only 69 (or 24%) were co-occurrances (i.e., *both* risk and preference judgments showed violations. Given a violation rate for 46% for risk and 45% for preference (see Exhibit 6), one would expect 21% co-occurrences by chance alone. The corresponding figures in the positive stimulus set are 41 co-occurrences (or 14%). Given a violation rate of 30% for risk and 47% for preference, 14% co-occurrences is exactly what one would expect by chance alone. Thus, the evidence seems to indicate that violations of the expectation principle for risk are independent of those for preference. This result also holds when one analyzes the data at the individual item or individual subject level.

| | Stimulus Domain | |
|---|---|---|
| *Type of Violation* | Negative outcomes | Positive outcomes |
| same direction | 20 | 1 |
| opposite direction | 49 | 40 |
| Total: | 69 | 41 |

Exhibit 7. Contingency table of frequency of comparisons for which expectation principle was violated simultaneously for risk and preference judgments as a function of stimulus set (negative *vs.* positive outcome domain) and of type of violation (in same direction for risk and preference, or into opposite directions)

Studying the relationship between risk and preference decisions, provides one with a novel and perhaps more direct way of assessing an individual's 'risk attitudes'. Historically, risk attitudes have been derived indirectly from the shape of utility functions. The Arrow-Pratt (Arrow, 1971; Pratt, 1964) measure of risk attitude, for example, is defined as $-u''/u'$ where $u'$ and $u''$ are the first and second derivative of the utility function, respectively. A concave utility function is seen as evidence for 'risk aversion' (because individuals with such utility will accept certainty equivalents below the expected value of a lottery), whereas a convex utility function is evidence for 'risk seeking.' Thus 'risk attitudes' are descriptive but redundant labels for the shape of utility functions. Dyer and Sarin (1982, 1986) describe a measure of 'relative risk attitude' which separates strength of preference for outcomes from attitudes towards uncertainty (which are confounded in the Arrow-Pratt measure). This distinction is important, but the assessment of risk attitude is still indirect (i.e., via utility or value functions).

A novel, different, and perhaps more useful and appropriate definition and classification of 'risk attitude' involves the question whether an individual will or will not prefer an alternative that he or she perceives as more risky to an alternative perceived as less risky (*ceteris paribus*). Given the data of

Experiment 3, we can assess risk attitudes in this way. Given a set of pairs of lotteries (e.g., the positive outcome or negative outcome set with 28 pairs each), individuals who consistently select/prefer the member of the pair that they judge to be riskier (less risky) will be labelled 'risk seeking' ('risk averse'). Those individuals whose preference selection does not follow from their risk selection will be labelled 'risk neutral.'

The relationship between risk and preference selection was assessed with a binomial test ($H_o$: probability of risk/preference reversal = 0.5) separately for the gain and loss domain and for each of the 17 subjects. Given 28 lottery pairs and a Type I error level of 0.05, the critical values for $Y$, the number of pairs for which a different member is selected when judging preference than when judging riskiness, are $Y <= 8$ for risk seeking behavior and $Y >= 20$ for risk averse behavior. Intermediate values of $Y$ are, of course, evidence in favor of the null hypothesis, i.e., of risk neutrality. The results of this analysis are summarized in Exhibit 8.

Consistent with prospect theory assumptions, there are more individuals with risk averse or risk neutral behavior in the gain domain than in the loss domain. Somewhat contrary to prospect theory and more in line with data presented by Cohen, Jaffray, and Said (1987) however, not every individual with risk averse attitudes in the gain domain becomes risk seeking in the loss domain. Risk seekers, while more frequent in the loss than in the gain domain, are in the minority even in the loss domain.

| | Loss domain attitude | | | |
|---|---|---|---|---|
| | Risk averse | Risk neutral | Risk seeking | |
| Gain domain atitude: | | | | |
| Risk averse | 4 | 3 | 1 | 8 |
| Risk neutral | 1 | 5 | 2 | 8 |
| Risk seeking | 1 | 0 | 0 | 1 |
| Total | 6 | 8 | 3 | |

Exhibit 8. Number of subjects with particular risk attitudes in gains and loss domain inferred from relationship between preference and risk judgments

Further evidence of the prevalence of risk averse attitudes (and the differential in prevalence in the loss and gain domain) comes from an analysis of the types of violations of the expectation principle for risk and for preference decisions. Violations can be of two types: Member A is selected in the 'original' pair and member B in the 'transformed' pair (type A/B) or vice versa (type B/A). These two types of violations could occur with different frequencies. However, these frequencies should not systematically differ for risk and preference judgments unless there is a relationship between the two judgments. The frequencies (summed across subjects and items) of violation types for the two types of judment are shown in Exhibit 9, broken down by gain and loss domain as well as combined.

Chi-square tests of association between type of judgment and type of violation are significant in all three cases ($\chi^2 = 8.15, p < .005$, for the combined data; $\chi^2 = 35.05, p < .001$, for the gains domain; and $\chi^2 = 4.02, p < .05$, for the loss domain). Again the effect is more significant for the gains domain than for the loss domain.

| | Stimulus domain | | | | | |
|---|---|---|---|---|---|---|
| | Gains | | Losses | | Combination | |
| Direction of Violation: | B/A | A/B | B/A | A/B | B/A | A/B |
| *Judgment* | | | | | | |
| Risk | 47 | 21 | 62 | 59 | 109 | 80 |
| Preference | 36 | 102 | 81 | 46 | 117 | 148 |

Exhibit 9. Contingency table of frequency of two types of violations for risk judgments and for preference judgments (summed across item pairs and subjects)

Unrelated to concern about the expectation principle, the data also speak to the current controversy regarding the existence of prevalence of Kahneman and Tversky's (1979) reflection effect between positive and negative outcome sets (i.e., the choice of one member of a pair if the outcomes are in the positive domain, but of the other member of the pair if the outcomes are translated into the negative domain). This type of reflection—reflections of preference between gain and loss domain—is different from the reversals discussed in the last section, namely reversals within one domain between different types of judgments (risk *vs.* preference).

Kahneman and Tversky's claim of a general reflection effect was based on a between-subject analysis of choice data. Within-subject tests subsequently conducted by Hershey and Schoemaker (1980), Schneider and Lopes (1986), and Cohen, Jaffray, and Said (1987) generally have failed to find the effect or found it to occur with much smaller frequency. The within-subject design of experiment 3, employing the same pairs of lotteries in the positive and in the negative outcome domain using risk as well as preference, judgments, allows several 'reflection' analyses.

For any lottery pair, judgments (of risk or preference) either are reversed between the positive and the negative outcome domain or they are not. If there is no relationship between the judgments in the positive and in the negative domain (i.e., independence), the probability of a reflection is 0.5. Under the 'reflection' hypothesis, this probability is greater than 0.5, and under the 'identical pattern' hypothesis it is less than 0.5.

The frequency of reflections can be tabulated both across subjects for the 28 item pairs and across items for the 17 subjects. In both cases, reflection occurs reliably less frequently for risk comparisons than for preference comparisons. (By sign test, for 23 out of the 28 items, $p < .001$, and for 14 out of the 17 subjects, $p < .02$). More informatively, the frequency of reversal between the positive and negative domain can be tested for evidence of 'consistency', 'independence', and 'reflection' as defined above by means of a binomial test ($H_o$: probability of reversal = 0.5). This was done separately for risk and for preference comparisons.

Given 28 lottery pairs and a Type I error rate of 0.05, the critical values for $Y$, the number of pairs for which a different member is selected in the gain than in the loss domain are $Y <= 8$ for the 'consistency' hypothesis and $Y >= 20$ for the 'reflection' hypothesis. Intermediate values of $Y$ are, of course, evidence in favor of the null hypothesis of 'independence.' The results of this analysis are shown in Exhibit 9. The same analysis can be conducted on the reversal frequencies summed across subjects. Given 17 subjects and a Type I error rate of 0.05, the critical values for $Y$ are $<= 4$ and $>= 13$. The results, cross classified for risk and preference are also shown in Exhibit 10.

| | | Item pattern | | | | Subject pattern | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Risk | | | | Risk | | | |
| | | R | I | C | | R | I | C | |
| Preference | R | 1 | 3 | 0 | 4 | 0 | 5 | 1 | 6 |
| | I | 0 | 23 | 1 | 24 | 0 | 5 | 4 | 9 |
| | C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| | | 1 | 26 | 1 | | 0 | 10 | 7 | |

Exhibit 10. Frequencies of 'reflection' (R), 'independence' (I), and 'consistency' (C) between gain and loss domain for risk and preference judgments. Frequencies of patterns are for subjects (analyzed across the items) and for items (analyzed across subjects)

With cell sizes too small to permit meaningful statistical analysis, the results in Exhibit 9 are still suggestive of qualitative differences between risk and preference comparisons when it comes to loss—gain reflection. While six subjects showed statistically significant 'reflection' for preference, none did for risk. Only two subjects showed comparison 'consistency,' i.e., 'nonreflection' for preference, but seven did so for risk.

## DISCUSSION

Experiments 1 and 2 established that loss and gain components of a risky prospect combine additively in the determination of its perceived risk. This result rules out (at least for money lotteries with outcome ranges studied here) all of Fishburn's (1982) risk functions which are multiplicatively separable in gain and loss components. To what extent additive separability holds for different outcome domains (e.g., health outcomes, or mortality estimates) or even for different ranges of financial outcomes is open to further empirical verification.

The Luce and Weber (1986) CER model thus appears to be one of the most viable quantitative risk models. Even though derived from different considerations, the CER model could be seen as an elaboration and extension of Coombs and Lehner's (1984) bilinear risk model which had as its domain only two-outcome gambles. Both are additively separable in gain and loss components. In both models, probabilities and outcomes combine in a multiplicative fashion. (In the CER model, probabilities also enter as separate additive components.) The CER model is more specific that the bilinear model by specifying the particular candidates for the four subjective transformation functions of the probability of gain and of loss, and for the gain and loss outcomes postulated by Coombs and Lehner. As already mentioned, it also specifies how multiple positive (negative) outcomes may be integrated into a single gain (loss) contribution.

Experiment 3 indicates that at least some of the assumptions underlying the CER model will need to be modified. It was demonstrated that expectation principle violations, while less frequent for risk than for preference, do occur. As discussed by Luce and Narens (1985), such violations may be the result of monotonicity violations *per se*, or they may be the result of 'probability accounting' violations. The (subjective) probability of $pq$ may not equal the product of the subjective probabilities of $p$ and $q$.

The CER model in its current form does not make use of a subjective probability transformation and thus cannot model such probability accounting patterns. These two explanations can be experimentally differentiated, and data suggesting that monotonicity *per se* is probably not violated are reported in Weber and Bottom (1988). The unsystematic violations in Experiment 3 also seems to suggest that something pervasive such as nonlinear subjective probabilities may be the source of the effect.

If true, the CER model would need to be modified to a subjective conjoint expected risk model. One puzzling result, in this context, is the failure of Experiment 3 to find co-occurrences of risk and preference expectation principle violations. If nonlinear subjective probability values are causing the Allais-type violations, why are violations for risk and preference seemingly unrelated? One possible explanation is that subjective probabilities are assessed differently when the judgment or comparison is one of risk rather than of preference. This interesting possibility would need further investigation.

While outside the scope of a current study, a comparison of different (direct and indirect) measures of risk attitude may provide useful insights into the relationship between risk perception and preference or choice.

## REFERENCES

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

Arrow, K. J. *Essays in the theory of risk bearing*. Chicago: Markham, 1971.

Aschenbrenner, K. M. 'Expected risk and single-peaked preferences,' *Acta Psychologica*, 42 (1978), 343–356.

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

Birnbaum, M. H. 'The devil rides again: Correlation as an index of fit', *Psychological Bulletin*, 79 (1973), 239–242.

Birnbaum, M. H. 'The nonadditivity of personality impressions', *Journal of Experimental Psychology*, 102 (1974), 543–561.

Birnbaum, M. H. 'Controversies in psychological measurement', in Wegener, B. (ed.), *Social attitudes and psychophysical measurements*. Hillsdale, N.J.: Lawrence Erlbaum, 1982.

Cohen, M., Jaffray, J.-Y. and Said, T. 'Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses,' *Organizational Behavior and Human Decision Processes*, 39 (1987), 1–22.

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

Coombs, C. H. 'A review of the mathematical psychology of risk and risk-taking', Michigan Mathematical Psychology Program Report 72-6, 1972.

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

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

Coombs, C. H. and Bowen, J. N. 'Additivity of risk in portfolios', *Perceptions and Psychophysics*, 10 (1971b), 43–46.

Coombs, C. H. and Huang, L. 'Polynomial psychophysics of risk', *Journal of Mathematical Psychology*, 7 (1970a), 317–338.

Coombs, C. H. and Huang, L. 'Tests of a portfolio theory of risk preference', *Journal of Experimental Psychology*, 85 (1970b), 23–29.

Coombs, C. H. and Lehner, P. E. 'Evaluation of two alternative models for a theory of risk: I. Are moments of distributions useful in assessing risk?' *Journal of Experimental Psychology: Human Perception and Performance*, 7 (1981), 1110–1123.

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

Dyer, J. S. and Sarin, R. K. 'Relative risk aversion', *Management Science*, 28 (1982), 8.

Dyer, J. S. and Sarin, R. K. 'Measuring risk attitudes in risk analysis', in Covello, V. T., Mankos, J. and J. Mumpower (eds.), *Risk evaluation and management*. New York: Plenum Publishing, 1986.

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L. and Keeney, R. L. *Acceptable risk*. Cambridge University Press, 1981.
Fischhoff, B., Slovic, P. and Lichtenstein, S. 'Lay foibles and expert fables in judgments about risk', *American Statistician, 36 (1982), 240–255.*
Fishburn, P. C. *Utility theory for decision making*. New York: Wiley, 1970.
Fishburn, P. C. 'Mean-risk analysis with risk associated with below-target returns', *American Economic Review*, 67 (1977), 116–126.
Fishburn, P. C. 'Foundations of risk measurement. II. Effects of gains on risk', *Journal of Mathematical Psychology*, 25 (1982), 226–242.
Hershey, J. C. and Schoemaker, P. J. H. 'Prospect theory's reflection hypothesis: A critical examination', *Organizational Behavior and Human Performance*, **25** (1980), 395–418.
Huang, L. C. 'The expected risk function', Michigan Mathematical Psychology Program Report 71-6, 1971a.
Huang, L. C. 'Experiments on the measurement of risk', Michigan Mathematical Psychology Program Report 71-7, 1971b.
Johnson, E. and Tversky, A. 'Representations of perceptions of risk', *Journal of Experimental Psychology: General*, 113 (1984), 55–70.
Kahneman, D. and Tversky, A. 'Prospect theory: An analysis of decision under risk', *Econometrika*, 47 (1979), 263–291.
Kaplan, S. and Garrick, B. J. 'The quantitative definition of risk', *Risk Analysis*, 1 (1981), 11–27.
Keller, L. R. 'The effect of problem representation on the sure-thing and substitution principles', *Management Science*, 31 (1985), 738–751.
Keller, L. R., Sarin, R. K. and Weber, M. 'Empirical investigation of some properties of the perceived riskiness of gambles', *Organizational Behavior and Human Decision Processes*, 38 (1986), 114–130.
Libby, R. and Fishburn, P. 'Behavioral models of risk taking in business decisions: A survey and evaluation', *Journal of Accounting Research*, 1977, 272–292.
Luce, R. D. 'Several possible measures of risk', *Theory and Decision*, 12 (1980), 217–228.
Luce, R. D. Correction to 'Several possible measures of risk', *Theory and Decision*, 13 (1981), 381.
Luce, R. D. and Narens, L. 'Classification of concatenation measurement structures according to scale type', *Journal of Mathematical Psychology*, 29 (1985), 1–72.
Luce, R. D. and Weber, E. U. 'An axiomatic theory of conjoint, expected risk', *Journal of Mathematical Psychology*, **30** (1986), 188–205.
MacCrimmon, K. R. and Larsson, S. 'Utility theory: Axioms versus "paradoxes,"' in Allais, M. and Hagen, O. (eds.), *Expected utility and the Allais paradox* (pp. 333–409). Dordrecht, Holland: Reidel Publishing, 1979.
Markowitz, H. *Portfolio selection: Efficient diversification of investment*. New York: Wiley, 1959.
Payne, J. W. 'Approaches to decision making under risk: Moments versus risk dimensions', *Psychological Bulletin*, 80 (1973), 439–453.
Payne, J. W., Laughhunn, D. J. and Crum, R. 'Translation of gambles and aspiration level effects in risky choice behavior', *Management Science*, 26 (1980), 1039–1060.
Payne, J. W., Laughhunn, D. J. and Crum, R. 'Further tests of aspiration level effects in risky choice behavior', *Management Science*, 27 (1981), 953–959.
Pollatsek, A. and Tversky, A. 'A theory of risk', *Journal of Mathematical Psychology*, 7 (1970), 540–553.
Pratt, J. W. 'Risk aversion in the small and in the large', *Econometrica*, 32 (1964), 122–136.
Rothschild, M. and Stiglitz, J. E. 'Increasing risk: I. A definition', *Journal of Economic Theory*, 2 (1970), 225–243.
Schneider, S. L. and Lopes, L. L. 'Reflection in preferences under risk: Who and when may suggest why', *Journal of Experimental Psychology: Human Perception and Performance*, 12 (1986), 535–548.
Schoemaker, P. J. H. 'The expected utility model: Its variants, purposes, evidence, and limitations', *Journal of Economic Literature*, 20 (1982), 529–563.
Schwartz, S. and Griffin, T. *Medical thinking: The psychology of medical judgment and decision making*. New York: Springer Verlag, 1986.
Vlek, C. and Stallen, P.-J. 'Judging risks and benefits in the small and in the large', *Organizational Behavior and Human Peformance*, 28 (1981), 235–271.
Weber, E. U. '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 (1984a), 179–194.
Weber, E. U. 'Risk as an independent variable in risky choice'. Cambridge, MA: Harvard University, Unpublished Doctoral Dissertation, 1984b.
Weber, E. U. 'A descriptive measure of risk', *Acta Psychologica*, in press, 1988.

Weber, E. U., Birnbaum, M. H. and Anderson, C. 'Subjective risk and attractiveness: One construct or two?' Manuscript in preparation, 1988. (Available from first author.)

Weber, E. U. and Bottom, W. P. 'An empirical evaluation of the transitivity, monotonicity accounting, and conjoint axioms for perceived risk.' Manuscript submitted for publication, In: *Organizational Behaviour and Human Decision Processes* 1988. (Available from first author.)

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

*Authors' biographies:*

**Elke U. Weber** is an Assistant Professor of Behavioral Science at the Center for Decision Research in the Graduate School of Business at the University of Chicago. Her research interests are in the area of descriptive decision models, in particular the measurement of perceived risk and its role in preference and choice with recent applications in the domain of medical decision making.

**William P. Bottom** is an Assistant Professor of Organizational Behavior in the John M. Olin School of Business at Washington University. His current research interests include decision making and negotiation.