Acta Psychologica 69 (1988) 185-203
North-Holland

A DESCRIPTIVE MEASURE OF RISK *

Elke U WEBER

University of Chicago, USA

Accepted March 1988

A measure of the perceived riskiness of choice alternatives is introduced which will allow further assessment of the role of perceived risk in preference decisions The risk function suggested here (CER for conjoint expected risk) can predict a person's subjective risk judgments for risky choice alternatives (e g gambles) on the basis of a small number of easily estimated individual difference parameters The CER function is shown to be superior to other risk measures previously suggested in its descriptive fit as well as its agreement with qualitative results regarding subjective risk judgments A caveat against the exclusive use of two-outcome gambles in studies of judgment and decision making under risk is issued

Evidence that expected utility models are inadequate to account for choice behavior in many situations is steadily accumulating (e g, Schoemaker 1982) This has led researchers in the decision sciences to a search for additional variables responsible for preference One of the most promising candidates has been the familiar, but not well defined, concept of risk Theories of choice that incorporate risk as a central concept (e g, Coombs' (1969, 1975) portfolio theory of the preference) have not achieved wide acceptance because of the absence of a descriptively adequate measure of risk

Decision making under risk refers to choices among alternatives that can be described be probability distributions over possible outcomes (i e, the equivalent of lotteries) It other words, the outcome of the

* Some of the results reported in this paper first appeared in my Ph D dissertation submitted to Harvard University The work was supported by a National Science Foundation Grant IST-7924019 to the University of California at Irvine and by a postgraduate fellowship from the Natural Sciences and Engineering Research Council of Canada to the author I would like to thank R Duncan Luce for many important comments and discussions

Requests for reprints should be sent to Elke U Weber University of Chicago, Graduate School of Business Center for Decision Research 1101 East 58th Street, Chicago IL 60637 USA

chosen course of action is known only probabilistically at the time of decision. The importance of understanding such decisions at the personal, corporate, and social level in an increasingly unpredictable and risk-laden world cannot be overemphasized. In order to determine whether risk as a variable in its own right plays an explanatory role in risky choice, there has to be an independent method of *quantifying* the perceived riskiness of choice alternatives.

Measurement of subjective risk

When Coombs (1969, 1975) developed portfolio theory according to which a choice among risky prospects is a compromise between maximizing expected value and optimizing level of risk, he made some noncontroversial assumptions about the relative riskiness of lotteries but left the nature of risk essentially undefined. The history of subsequent attempts to develop an independent theory of risk is briefly reviewed in this section.

Pollatsek and Tversky (1970) incorporated some of Coombs' assumptions about subjective risk as axioms into a risk representation that was a linear combination of a choice alternative's mean and variance. Thus, the gamble 'Win $100 with probability 0.5, lose $100 otherwise' may be judged riskier than the gamble 'Win $10 with probability 0.5, lose $10 otherwise' (same mean, smaller variance) or than the gamble 'Win $50 with probability 0.5, lose $150 otherwise' (smaller mean, same variance). Coombs and Bowen (1971a), however, showed Pollatsek and Tversky's risk measure to be empirically inadequate because expected value and variance alone are insufficient to determine risk. In particular, by using transformations on gambles that left mean and variance unchanged, they found that risk varied systematically with the skewness of a gamble.

Subsequent to the failure of Pollatsek and Tversky's risk measure, Coombs and Bowen (1971a, b), Coombs and Huang (1970a, b) and Coombs and Lehner (1981) took the approach of considering how certain transformations of a gamble affected its perceived riskiness. Luce (1980, 1981) furthered this tradition by examining first the effect of a change of scale on risk (i.e., the effect of multiplying all outcomes of a gamble by a constant, e.g., transforming penny outcomes to dollar outcomes). He considered the two simplest possibilities, an additive effect (i e , a particular change of scale has the same effect on all gambles) and a multiplicative effect (i e , the effect of a change of scale depends on the riskiness of the gamble before transformation) 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 (a sum of crossproducts of probabilities and transformed outcomes) which results in an expected risk function as suggested by Huang (1971) The second possibility considered was that the probabilities undergo a transformation before integration From the combination of options considered at these two choice points, Luce derived four distinct possible risk measure but 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 using a conjoint and functional measurement analysis For Luce's second choice point regarding integration of probabilities and outcomes into a single risk value, the assumption that the probabilities undergo a transformation before integration leads to risk functions that are insensitive to change of origin transformations (i e , adding a constant amount to all outcomes of a lottery) which change the expected value of gambles This is contrary to the empirical fact that the subjective risk of a gamble *is* significantly affected by a change of origin (Coombs and Bowen 1971a, Keller et al 1985, Weber 1984a)

Further empirical work by Weber eventually led to a revision and expansion of Luce's original (1980, 1981) set of risk assumptions A description of the new axiom system can be found in Luce and Weber (1986) The present paper is a detailed account of the empirical research that provided the motivation for this reaxiomatization

Experiment 1

Effect of change of scale

The conjoint and functional measurement procedures of Weber (1984a) established that changing the scale of lotteries had either an additive or a multiplicative effect on perceived risk but were unable to distinguish between the two Yet, an additive effect of change of scale on risk appears very counterintuitive (i e it seems unlikely that a switch from gambling for pennies to gambling for dollars should result in an identical increase in perceived risk for every gamble thus transformed). This experiment addresses the issue with a different set of techniques.

Birnbaum (1974, 1982) provides a way to discriminate between an additive and a nonadditive combination of two variables with the help of difference judgments. Assume that a gamble is described by two characteristics, F and G, each of which can take a range of values and which combine either in an additive or nonadditive way to determine risk. For the current application, let F_i stand for the riskiness of the original lottery i (prior to the change of scale transformation) and G_a for the contribution of the change of scale factor a. Gamble (F_i, G_a) thus is the lottery created by transforming (multiplying) all outcomes of lottery i by scale factor a. By comparing the judgment of difference in risk between gambles (F_i, G_a) and (F_i, G_b) with that of the difference in risk between gambles (F_j, G_a) and (F_j, G_b), that is,

$$\mathrm{Diff}_{F_iG_a, F_iG_b} = J(\psi_{F_iG_a} - \psi_{F_iG_b})$$

versus

$$\mathrm{Diff}_{F_jG_a, F_jG_b} = J(\psi_{F_jG_a} - \psi_{F_jG_b})$$

where J is a monotonic response function and ψ the impression of riskiness of the transformed lottery, we can decide between additivity and nonadditivity. If additivity holds, the two difference judgments should be equal. That is, if

$$\psi_{F_iG_a} = F_i + G_a$$

then

$$\begin{aligned}\mathrm{Diff}_{F_iG_a, F_iG_b} &= J(F_i + G_a - F_i - G_b)\\ &= J(G_a - G_b) = J(F_j + G_a - F_j - G_b)\\ &= \mathrm{Diff}_{F_jG_a, F_jG_b}\end{aligned}$$

Two-outcome versus higher-outcome alternatives

Experiment 1 serves several additional purposes. The first one is that of extending the generality of previous studies (Weber 1984a, b) which used two- and five-outcome gambles. Weber (1984b) found different pattern of correlation between risk judgments and risk dimensions (e.g., probability of winning or of losing) for two- than for five-outcome lotteries. Given the prevalence of two-outcome lotteries in laboratory studies of risky choice behavior and their relative rarity in real life, the possibility of qualitative differences in the way two-outcome lotteries are perceived and judged puts into question the generalizability of two-outcome lottery studies. Thus, differences between two-outcome and higher-outcome lotteries need to be further examined.

Method

Stimuli and design

Following Weber (1984a) stimulus lotteries were constructed by applying three independent transformations to the basic two-outcome gamble of winning \$1.50 with probability 1/2 and losing \$1.50 with probability 1/2: $g = (a = \$1.50, 1/2; -a = -\$1.50)$. The first transformation changed the skewness of the gamble while leaving the expectation and variance unchanged (Coombs and Bowen 1971a): $\alpha(g) = (y = a\sqrt{q/p}, p; -z = -a\sqrt{p/q})$. Five levels of p (1/5, 1/3, 1/2, 2/3, and 4/5) were used.

The second transformation, a change of origin, transforming the gamble $g = (y, p; -z)$ into $\beta(g) = (y + b, p; -z + b)$ employed two levels of b ($b = -\$0.60, +\$0.60$). For the third transformation, a change of scale, the outcomes of the previously generated gambles were multiplied by a constant c, using three levels of c = 1, 2, 4.

From this set, 30 three-outcome gambles were generated by the multiple play transformation $\delta(g) = (y, p; -z)^d$ with $d = 2$, where the integer d indicates that the resulting gamble has the outcomes that would be obtained if one played the original gamble d times independently.

A subset of twelve gambles (those generated by using $p = 1/3$ and $2/3$ in the skewness transformation) were combined in all possible $\binom{12}{2} = 66$ pairwise combinations to form the *risk difference judgment set*.

Gambles were represented as shown in fig. 1. The probabilities of outcomes were depicted by a proportionate number of X's as well as by their numerical values. Monetary outcomes (losses indicated by a minus sign) appeared above the respective probabilities.

Subjects

Six female and four male Harvard University students were paid to participate in the study. Subjects were tested individually, each participating in four one-hour sessions on consecutive days. Since all analyses are done at the individual subject level (i.e., every subject constitutes a replication of the experiment) only a small number of subjects are necessary (mainly to guarantee the generality of any results).

Risk instructions

Similar to Weber (1984a), the term risk was intentionally left undefined. Subjects were told to imagine themselves in the situation of playing a displayed gamble. They were told to study the gamble until they could decide with some confidence how risky it was. It was stressed that they were to rate the risk of each gamble, not whether they

10.80	−\$4.20	−\$19.20
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXX	XX
0.64	0.32	0.04

Fig. 1. Sample stimulus lottery from the three-outcome risk rating set. Outcomes are shown as dollar-amount gains or losses (amounts preceded by a minus sign). The corresponding probabilities are shown beneath the dollar-amounts and are represented both graphically (by a proportionate number of X's) and numerically.

would like to play it. Risk judgments were to be given 'intuitively', as a 'gut reaction' to each individual gamble as opposed to 'computing' the risk in some way. All subjects indicated that they clearly understood their task.

Procedure

Subjects came to four sessions. Each session constituted a complete replication of the following two tasks. The 30 stimulus gambles of the *risk rating set* were displayed one at a time on a CRT screen. Subjects were instructed to rate those gambles with respect to subjective riskiness on a graphic 'Risk' rating scale ranging from 'low' (computer coded as 0) to 'high' (computer coded as 700) using a light pen. Order of presentation of the gambles in the risk rating set was random and different for each session.

For the *risk difference judgment task*, subjects were presented with pairs of gambles, with each gamble printed on a separate index card. They were asked to study each gamble of the pair until they had formed an impression of its riskiness. They were then to judge the difference in risk between the two gambles on an integer scale between 0 and 9. If the two gambles seemed equally risky, a rating of 0 was to be given; a very large difference in risk was to be given a rating of 9. Subjects made risk difference judgments verbally. The experimenter recorded their judgment and then presented the next pair of the 66 pair stimulus set. Order of presentation of the pairs within the stimulus set was random and different for each of the four sessions.

For both tasks, several practice items preceded the stimulus items. The practice items were designed to familiarize subjects with the range of gambles in the stimulus set. The order of tasks was random for every session and every subject. Presentation of the items was self-paced (i.e., only completion of the previous rating triggered presentation of the next item) and subjects were under no time pressure.

Results

Effect of change of scale

Separate ANOVAs (three repeated within-factors: Skewness, Mean, and Scale) were computed for each subject's risk ratings. The results, which are summarized in table 1, are similar to those of Weber (1984a) and a detailed discussion of their significance can be found there. For the analysis of additivity or nonadditivity of the effect of change of scale, it is important to note that the factor Scale was not significant for five of the ten subjects. (This indicates, with the benefit of hindsight, that the change of scale factors ($c = 1$, 2, and 4) were not large enough to significantly alter the risk of the set of low-stake lotteries employed. However, these lotteries and change of scale factors had been chosen to keep the lotteries within a financial range that would be realistic to the average college student.) Those subjects whose risk ratings are not significantly affected by a change of scale transformation are more likely to satisfy the requirements of the Birnbaum (1982) difference test of additivity in a trivial way. (If risk is not significantly affected by particular changes in scale, one may not be able to discriminate between an additive or nonadditive combination of the two.) It will thus be the other five subjects who show a significant effect of Scale who will provide the more crucial data to decide between an additive or multiplicative effect of change of scale on risk.

Table 1
ANOVAs for risk ratings of experiment 1 with three repeated within-factors (Skewness, Mean, Scale) computed separately for each subject. F-ratios and significance level. (S = Skewness, M = Mean, C = Scale.)

Subject	Main effects			Interaction effects			
	S	M	C	$S \times M$	$S \times C$	$M \times C$	$S \times M \times C$
ANN	5.05 [a]	118.11 [b]	7.07 [a]	0.46	2.07	1.49	0.54
BAM	0.86	113.33 [b]	51.76 [c]	2.64	1.84	0.74	2.00
BAR	3.34 [a]	91.45 [b]	3.26	4.15 [a]	1.26	0.53	0.55
BRI	4.66 [a]	352.05 [c]	2.89	3.13	0.94	0.96	0.56
DAN	3.98 [a]	160.00 [b]	6.24 [a]	9.55 [c]	3.01 [a]	1.23	1.14
JOS	9.99 [c]	309.69 [c]	1.16	3.58 [a]	1.46	4.10	2.21
LEO	2.84	86.36 [b]	0.65	1.42	1.61	1.40	0.64
SCO	0.50	147.60 [b]	4.04	0.73	2.37	3.22	0.51
SUZ	0.06	69.84 [b]	16.22 [b]	2.57	1.80	2.32	0.45
ZAN	1.06	247.62 [b]	14.60 [b]	7.44 [b]	2.38	2.83	2.77

[a] $p < 0.05$ [b] $p < 0.01$ [c] $p < 0.001$

The 12 gambles of the risk difference set constitute a factorial combination of three levels of factor G (Scale) and four levels of factor F (different Mean–Skewness combinations). The conditions for additivity in the combination of factors F and G using Birnbaum's (1982) difference test translate into a series of planned orthogonal contrasts that test for equality of difference judgments in particular cells of a one-way ANOVA analysis of the difference judgments of the 66 stimulus pairs. For additivity to hold these contrasts have to be statistically nonsignificant. For example, the risk difference judgement of all cells involving a judgment between Scale levels G_1 and G_2 should be equal regardless of the level of F in these cells. For every subject there are 36 out of a total of $\binom{66}{2} = 2145$ possible pairwise comparison that provide this information. (There are three times six comparisons of the type 'difference between (F_i, G_a) and (F_i, G_b) vs. difference between (F_j, G_a) and (F_j, G_b)' and six times three comparisons of the type 'difference between (F_i, G_a) and (F_j, G_a) vs. difference between (F_i, G_b) and (F_j, G_b)'.)

Table 2 summarizes the results of these contrasts computed separately for every subject. Each of the 36 contrasts is a statistical test of the null hypothesis that the population means of the two risk difference judgments are the same (and that thus change of scale plays an additive role) versus the alternative hypothesis that the two difference judgments are *not* the same (and that thus change of scale plays a nonadditive role). Statistical tests are typically biased towards the null hypothesis by maintaining the probability of a Type-I error (the probability of falsely rejecting the null hypothesis) below a low conventional level (e.g., five percent) without usually worrying about the inversely related probability of a Type-II error (the probability of falsely failing the reject the null hypothesis). In this sense, the paired comparisons reported here are biased in favor of the null hypothesis (i.e., additivity). If the null

Table 2
Percentage of planned Birnbaum contrasts (out of 36) that were statistically significant. Degrees of freedom of all contrast F-ratios are 1 and 195.

Subject	ANN	BAM	BAR	BRI	DAN	JOS	LEO	SCO	SUZ	ZAN
Percentage of significant contrasts [a]	31	39	19	14	28	14	31	19	42	36

[a] Significance level was at most 0.05, but usually smaller.

hypothesis (i.e., additivity) were true, we would expect not more than 5 percent of the paired comparisons to be statistically significant by chance alone.

As can be seen in table 2, subjects show a much higher rate of violation of additivity than the chance rate of five percent. (It should be noted that the figures in table 2 do *not* refer to the percentage of *judgments* that violated additivity but to the percentage of *statistical tests* for which the additivity hypothesis was rejected.) The result holds for *all* subjects, and the effect is, as expected, even stronger for those subjects who showed significant effects for Scale in the ANOVA of the risk ratings reported in table 1 (ANN, BAM, DAN, SUZ, ZAN). Thus, evidence from the risk difference judgments strongly points to a nonadditive and, in combination with the results reported in Weber (1984a), thus multiplicative effect of change of scale on perceived risk.

Modeling of perceived risk

The result of Weber (1984a) that a change of scale has either an additive or multiplicative effect on perceived risk, together with the outcome of experiment 1 indicating that the effect is nonadditive (and hence multiplicative) can serve as a starting point for deriving a descriptive risk function, i.e., a function of a gamble's outcomes and probabilities with which one is able to describe and predict an individual's rating of perceived risk. Based on the results discussed above, the function that derives from the assumption of a multiplicative effect of scale and a transformation of outcomes before integration of outcomes and probabilities was the most viable of the risk functions suggested by Luce (1980). If we let X stand for a particular gamble and R for the function that determines the perceived riskiness of that gamble, then

$$R(X) = E(|X|^k), \tag{1}$$

i.e., risk is equal to the expected value, E, of the absolute value of outcomes raised to power k.

The adequacy of this measure can be described with a goodness-of-fit statistic like R^2, the coefficient of determination or proportion of variance accounted for, after estimating the value of the parameter k using, for example, Chandler's (1969) parameter estimation subroutine STEPIT. This was done for the risk ratings of the three-outcome gambles collected in experiment 1. The proportion of the variance in individual subject's risk ratings accounted for by risk function (1) ranged between 0.11 and 0.33 for the ten subjects, with a mean R^2 of 0.26.

Thus function (1) clearly is not very successful in describing subjects' risk ratings. Further analysis revealed some reasons for this and suggested modifications. For two of the transformations employed to create the stimulus gambles of experiment 1, namely change of skewness and change of origin, $E(|X|^k)$ is insensitive to the direction of change. Transformation of a symmetric gamble into one that is positively skewed, for example, has exactly the same effect on $E(|X|^k)$ as a transformation into an equally but negatively skewed gamble. This property makes $E(|X|^k)$ a poor predictor of subjective risk because perceived risk appears to be extremely sensitive to the direction of such a transformation. It is this insensitivity to the direction of transformations that has contributed to the rejection of moments as useful variables in preferential choice among gambles (e.g. Coombs and Lehner 1984; Payne and Braunstein 1971).

These problems can be resolved by modifying Luce's (1980) aggregation rule as follows. Let us assume that the gamble or random variable X is split into positive and negative outcomes which are transformed separately and possibly differently and then integrated.

There is plenty of evidence in the risky choice literature that supports the assumption that people treat positive and negative outcomes differently. One example is Coombs and Lehner's (1984) thought experiment which showed that a change of \$10 in the amount to win reduces risk less than a change of \$10 in the amount to lose. Another example is Tversky's (1967) and Kahneman and Tversky's (1979) hypothesis of differently shaped value functions for gains and losses, a hypothesis empirically supported by Payne et al. (1980). As discussed by Lopes (1984), a separate consideration of gains and losses is often found in applied work (Fishburn 1977; Holthausen 1981).

Details of the modifications made to the assumptions about risk judgments (primarily the allowance for differences in the perception of positive and negative outcomes) can be found in Luce and Weber (1986) and will not be repeated here. The risk function that can be derived from these assumption is called CER for conjoint expected risk and has seven free parameters. Using the same notation as above, the riskiness of gamble X can be expressed as

$$R(X) = A(0)\Pr(X=0) + A(+)\Pr(X>0) + A(-)\Pr(X<0) + B(+)E\left[X^{k(+)} \mid X>0\right]\Pr(X>0) + B(-)E\left[|X|^{k(-)} \mid X<0\right]\Pr(X<0) \tag{2}$$

Thus, risk 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 of the respective components. It should be noted that probabilities enter into the equation twice, once by themselves and once as weights on the effect of outcomes.

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 out-

Table 3
Values of R^2 for three measures obtained from risk judgments for three-outcome gambles of experiment 1

Subject	R–D [a]	V–E [b]	CER [c]	Average reliability [d]
ANN	0.30	0.45	0.56	0.49
BAM	0.47	0.65	0.77	0.77
BAR	0.24	0.34	0.45	0.42
BRI	0.50	0.63	0.71	0.69
DAN	0.39	0.60	0.82	0.69
JOS	0.55	0.72	0.80	0.82
LEO	0.46	0.66	0.75	0.71
SCO	0.40	0.57	0.64	0.41
SUZ	0.39	0.55	0.65	0.44
ZAN	0.43	0.61	0.70	0.64

[a] R–D is the risk dimension measure of Payne (1973) with six free parameters for three-outcome gambles.
[b] V–E is the Pollatsek and Tversky (1970) risk measure $aV + bE$, where E is the expected value and V the variance of the gamble, and a and b are two free parameters.
[c] Although the general CER measure involves seven parameters, the gambles of experiment 1 all had $\Pr(X = 0) = 0$, so $A(0)$ was not relevant.
[d] Average squared Pearson product moment correlations between replications of the ratings of the 30 gambles.

comes a property that is not shared by risk dimension models of the kind suggested by Payne (1973). The assumptions made to that end seem plausible but will have to undergo empirical testing. (Should the assumptions of the expectation model be violated, the CER theory could be generalized along the lines of Luce and Narens' (1985) dual-bilinear model.)

The CER function (2) was fit to the risk ratings of experiment 1 using the least-squares parameter estimation procedure STEPIT (Chandler 1969). The parameter estimates are for conjoint expected risk expressed as deviation from the grand mean of risk judgments, b (i.e., judged risk = CER + b). Had the risk judgments been standardized, b would of course be equal to zero. The CER function accounted for a significantly larger percent of the variance than risk function (1): the z-statistic of the test of difference between two dependent correlations (Steiger 1980) was highly significant for every subject. For comparison purposes, the coefficient of determination, R^2, was also computed for the variance-expectation (V–E) measure of Pollatsek and Tversky (1970) and for the risk dimensions (R – D) regression of Payne (1973). The results, along with the average reliability (r^2) of the rating judgments between replications, are shown in table 3. Without exception, the R^2 values of the regressions (which are adjusted for the different numbers of independent variables in the models) are ordered

R – D < V – E < CER ≅ Reliability.

Table 4
Values of the parameters in the CER model of risk for ten subjects and three-outcome stimulus set of experiment 1

Subject	$A(+)$	$A(-)$	$B(+)$	$B(-)$	$k(+)$	$k(-)$	Grand mean	R^2
ANN	321.2	−449.5	−347.1	372.2	0.11	0.20	324.4	0.56
BAM	185.3	−375.4	−332.8	515.3	0.18	0.25	386.6	0.77
BAR	21.3	−103.2	−62.8	237.8	0.30	0.28	454.2	0.45
BRI	263.8	−505.7	−431.7	487.2	0.25	0.30	301.5	0.75
DAN	220.5	−434.4	−167.7	453.1	0.34	0.44	188.0	0.82
JOS	176.3	−331.2	−315.4	390.1	0.24	0.29	357.4	0.80
LEO	175.8	−321.4	−231.9	357.5	0.29	0.24	435.8	0.75
SCO	164.2	−304.9	−320.1	352.9	0.21	0.35	424.6	0.64
SUZ	141.2	−268.7	−154.5	415.2	0.20	0.24	390.7	0.65
ZAN	241.2	−468.7	−372.4	316.8	0.19	0.39	358.6	0.71

Given the reliability of the data, much improvement over the CER fit is not to be anticipated.

Table 4 shows the values of the six risk parameters of the CER function and the value of the grand mean of risk judgments for the ten subjects.[1] The values of $k(+)$ and $k(-)$ are both very similar and rather small for this homogeneous group of undergraduate students. Small values of $k(+)$ and $k(-)$ indicate, of course, that the size of wins or losses is less salient for the risk judgments of an individual than the probability of winning or losing.

The reader may be puzzled to note that $A(+)$, the coefficient of the probability of positive outcomes, takes on positive values and that $A(-)$, the coefficient of the probability of negative outcomes, takes on negative values for all subjects. This seems to imply the counterintuitive prediction that risk would *increase* with the probability of positive outcomes. The explanation to the puzzle lies in the fact that probabilities enter into the CER equation twice, as pointed out above. Thus it is the marginal effect of probability on risk and not the signs of $A(+)$ and $A(-)$ that are relevant. To verify that the parameter estimates make the intuitively correct prediction that risk should *decrease* as the probability of winning increases (all other things being equal), one can compute the CER predictions for two lotteries differing only in the probability of winning. For example, using the parameter estimates of subject ANN, the predicted risk judgment for the lottery of winning \$12 with probability 0.1 and losing \$4 with the remaining probability 0.9 is equal to 348. The predicted risk judgment for the lottery of winning \$12 with probability 0.5 and losing \$4 with the remaining probability 0.5 is equal to 278, a value that is smaller (i.e., less risky) than 348.

[1] These parameter estimates are different from those reported in Luce and Weber (1986). The latter estimates were incorrect as the result of an error in the parameter estimation program.

Discussion

Experiment 1 helped to establish that the effect of a change of scale on perceived risk was multiplicative. It also pointed out a major shortcoming of the Luce (1980, 1981) axiomatizations of risk, namely the implicit assumption of equivalent weight of positive and negative outcomes for risk. When an assumption allowing for differential weights was added, the resulting CER risk function accounted for subjects' risk ratings significantly better than other competitors.

The CER function does well not only with respect to goodness-of-fit measures but also in an evaluation against more diagnostic 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 functions as well as eliminating 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 of origin (i.e., 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 et al. (1980), on the other hand, appear to have brought about a reversal in relative risk by increasing the expected value of two gambles by the same amount. It is not difficult to construct examples where, with the right choice of parameters, the CER function would predict a reversal in relative risk.

It should be noted that for a particular set of parameter values, the CER function could take on negative values when the positive outcome contributions outweigh the negative outcome contributions. (Positivity or negativity of CER values carries no special meaning. Being measured on an interval scale, it is only the relative distance between CER values that is interpretable.) This is in direct contrast to Axiom B4 of Fishburn (1982) which restricts risk functions to positive values and postulates that gambles without losses have zero risk. This axiom, which is part of all risk measures axiomatized by Fishburn (1982, 1984), in fact rules out all models that are additively separable in gains and losses (e.g., the CER model or Coombs and Lehner's (1981, 1984) bilinear risk model). Instead, Fishburn considers multiplicatively separable representations which allow for an effect of gain on risk without changing his assumption that risk is zero when no loss is possible. Weber and Bottom (1989) report data suggesting that perceived risk is *additively* separable in gains and losses, thus putting into question all of Fishburn's risk functions on empirical grounds.

In summary, the CER function appears to describe perceived risk quite well. It also has the qualitative attributes required to account for a wide variety of empirical results. The same function allows predictions for gambles that differ in their number of outcomes. Its seven parameters are easily estimated for an individual from a set of his or her risk judgments. The parameters allow comparisons between subjects with respect to differences in risk perception. They also reflect the relative influence of different components of the gambles on risk (in particular, the size of $k(+)$ and $k(-)$ indicates the importance attributed to the size of wins or losses as opposed to the probability of winning or losing).

The parameter estimates for the group of ten college students of experiment 1 were quite homogeneous It would be of interest to see in which way the estimates of subjects that differed in age and/or disposable income would differ Another question of interest is the generality of the CER parameters for gambles that differ in their number of outcomes That is do parameter estimates derived from a set of three-outcome gambles also predict risk judgments for a set of five-outcome gambles? Such generality would obviously increase the practical appeal of the CER model Experiment 2 was designed to answer these questions

Experiment 2

Method

Stimuli and design
Stimulus lotteries were constructed as in experiment 1 Starting with the basic two-outcome gamble of winning or losing $5 00 with probability 1/2, the first transformation changed the skewness of the gamble Five levels of p (1/9 1/3 3/5 3/4 and 8/9) were used The second transformation a change of origin employed three levels of b ($b = -\$2\ 10$, $+\$0\ 30$ and $+\$1\ 70$) For the third transformation, a change of scale the outcomes of the previously generated gambles were multiplied by a constant c, using two levels of c = 1 and 3 This generated the 30 items of the *two-outcome set of gambles (02-set)* The expected value of lotteries in this set ranged from −$6 30 to +$5 10 Individual outcomes took values between +$47 55 and −$39 90

A set of 30 *four-outcome gambles (04-set)* was produced by applying the multiple play transformation to the 02-set, $\delta(g) = (x_i\ p_i, -z)^d$ with $d = 3$ and dividing the resulting outcomes by three to keep the expected values and the range of outcomes comparable to those of the 02-set

Two sets of three-outcome gambles were created by applying the skewness transformation as for the 02-set, followed by a change of origin transformation with $b = -\$1\ 50$ $0 00, and +$1 30 and a change of scale transformation with c = 1 and 3 The resulting 30 two-outcome lotteries were transformed into the *small range three-outcome set (03s-set)* by applying the multiple play transformation with $d = 2$ and dividing the resulting outcomes by two The expected value of lotteries in this set ranged from -$4 50 to +$3 90 Individual outcomes took values between +$37 86 and −$31 38 The second set of three-outcome gambles was constructed like the 03s-set but used only $b = -\$1\ 50$ and +$1 30 for the change of origin transformation and added c = 11 to the set of c for the change of scale transformation The expected value of lotteries in this *larger range three-outcome set (03l-set)* ranged from −$16 50 to +$14 30 Individual outcomes took values between +$138 82 and −$115 06

Subjects
Two female and two male Canadian highschool teachers volunteered to participate in the study They were between 33 and 45 years of age and naive to the risk model under study

Procedure

Subjects came to four session on consecutive days Each session constituted a complete replication of the experiment The gambles of the four sets of gambles were displayed one at a time on an IBM PC/XT computer screen Subjects were instructed to rate those gambles with respect to subjective riskiness on a graphic rating scale ranging from 'low risk' (computer coded as 0) to 'high risk' (computer coded as 80) using a light pen The order in which the four stimulus sets were rates as well as the order of gambles within each stimulus set was random and different for each session As in experiment 1, presentation of the stimuli was self-paced and subjects were under no time pressure They were encouraged to take breaks between the different stimulus sets Each session lasted between 1 and 1½ hours

Results and discussion

Table 5 shows the values of the six CER parameters and the risk rating grand mean estimated from the risk ratings of the four highschool teachers for each of the four different sets of gambles These judgments were for different sets of gambles than the one used in experiment 1, using a rating scale with a smaller range, so that the parameters are not directly comparable to those in table 5 It is however clear that again there is considerable similarity in parameters within the group and high similarity within subjects in parameters estimated separately from ratings given for sets

Table 5
Values of the parameters in the CER model of risk for four subjects and four stimulus sets of experiment 2

Subj	Set	$A(+)$	$A(-)$	$B(+)$	$B(-)$	$k(+)$	$k(-)$	Grand mean	R^2
ROB	02	22 8	−47 3	−6 0	32 0	0 70	0 42	42 7	0 63
	03s	11 3	−36 1	−12 6	47 4	0 73	0 39	43 9	0 63
	03l	91 2	−132 4	−32 8	156 7	0 31	0 15	5 7	0 67
	04	3 3	−23 0	−24 9	39 4	0 41	0 41	46 2	0 65
CIN	02	52 1	−98 2	−46 8	74 9	0 17	0 17	58 5	0 72
	03s	75 0	−131 6	−109 4	111 2	0 16	0 18	74 3	0 68
	03l	92 5	−140 4	−29 5	181 3	0 30	0 09	11 3	0 56
	04	75 8	−127 8	−176 0	55 3	0 09	0 24	97 9	0 54
KEN	02	23 3	−48 4	−5 0	67 3	0 20	0 22	14 0	0 54
	03s	19 6	−49 0	−10 2	76 1	0 20	0 18	16 9	0 49
	03l	57 4	−88 7	−2 0	101 6	0 21	0 22	4 8	0 61
	04	31 4	−63 8	−8 6	104 2	0 18	0 12	10 5	0 49
ALB	02	24 3	−50 1	−34 1	16 3	0 20	0 60	51 3	0 85
	03s	2 8	−24 0	−3 3	14 1	0 40	0 78	31 1	0 76
	03l	15 5	−27 6	−9 8	29 4	0 28	0 40	13 9	0 77
	04	32 4	−65 7	−63 4	28 8	0 07	0 54	59 8	0 74

Table 6
Values of R^2 the coefficient of determination for predicting data set 03s using the parameters of the CER risk model estimated from either the original set 03s or sets 02 03l or 04

Subject	Parameters estimated from			
	03s	02	03l	04
ROB	0 63	0 32	0 53	0 56
CIN	0 68	0 32	0 60	0 62
KEN	0 49	0 45	0 43	0 45
ALB	0 76	0 65	0 70	0 73

of gambles differing in their number of outcomes (02, 03s, and 04) as well as range of outcomes (03l) Informal inspection seems to suggest that the range of outcomes plays a larger role in determining the CER parameters (especially $k(+)$ and $k(-)$) than the number of outcomes, as the parameters estimated for the 02, 03s and 04 sets which have the same range of outcomes are more similar to each other than those of the 03l set which has a wider range of outcomes In addition, in comparison to student subjects, the risk judgments for at least two of the highschool teachers can be seen to be more sensitive to the magnitude of wins (higher $k(+)$) and especially of losses (higher $k(-)$) (With the caveat that outcomes and expected values of the 02 03s and 04 gambles differed approximately by a factor of two from those used in experiment 1)

The similarity of CER parameters estimated from different sets of risk judgments implies that CER parameters estimated for an individual from a set of say three-outcome gambles can be used to predict his of her risk judgments for gambles with different numbers of outcomes provided that the range of outcomes is not too different This would add considerable practical appeal to the measure The generality of the CER parameter estimates was put to test by computing the extent to which CER predictions using the parameter estimates from different sets of gambles predicted the actual risk judgments for the small-range three-outcome set 03s Table 6 reports the values of R^2 the coefficient of determination for the four CER prediction functions (differing in parameter values) and each of the four subjects Not surprisingly the set of parameter values that were estimated from the three-outcome 03s set provides the best fit for the risk judgments of that set However the parameter values estimated from the four-outcome 04 set provide a fit for the 03s risk judgments that is almost as good A z-test of the difference between dependent correlations (Steiger 1980) reveals that for two subjects (KEN and ALB) the fit with the 04 parameters is not significantly worse than that with the 03s parameters For the other two subjects the difference is significant, but p-values are relatively large (ROB, $p < 0.02$ CIN $p < 0.05$) Fits of the 03s risk judgments employing the 03l parameter estimates are somewhat worse in all cases than those employing the 04 parameter estimates but not significantly so They are, however, significantly different from the fits employing the benchmark 03s parameter estimates for all subjects except KEN This is in line with the impression obtained from table 6 that CER parameter estimates generalize well across gambles that differ in their number of outcomes (and perhaps less so for gambles that differ widely in their range of outcomes)

Most interesting, however is the poor fit of the CER function for the three-outcome O3s risk judgments when parameters estimates from the two-outcome O2 set are employed For three of the four subjects the difference in fit is highly significant ($p < 0.001$) These poor fits stand in contrast to the uniformly good ones that these parameters provide for the two-outcome gambles from which they were estimated (see table 6) Thus, the claim of generalizability of CER parameters should probably be restricted to three- and higher-outcome gambles

General discussion

The empirical results described above show the CER model to be a measure of perceived risk with good descriptive and predictive ability Its parameters allow one to account for individual differences in subjective risk, which is precluded by such measures as the variance or negative semivariance of outcomes commonly equated with 'risk' in applied settings Through a comparison of parameter values, similarities and differences in risk perception can be conveniently diagnosed The parameters are interpretable The size of the exponents $k(+)$ and $k(-)$, for example, give information about the relative importance of the size of gains or losses in an individual's assessment of risk

The CER function incorporates and accounts for the body of empirical results about perceptions of riskiness of unidimensional risky choice alternatives Thus, unlike the admittedly more elegant Pollatsek and Tversky (1970) expectation–variance risk model, the CER model accounts for effects of the shape (or skewness) of the outcome distribution on risk Some empirical observations (e.g., the asymmetric effect of gains and losses on risk) were incorporated into the behavioral assumptions or axioms that give rise to the CER function Goodness-of-fit tests and comparisons of numbers of parameters used to fit a particular set of risk judgments which are documented to be problematic (e.g., Birnbaum 1973) can fortunately play a minor role in developing an axiomatic measurement model A more informative and more reliable test of the fit of a function is the adherence of empirical data to the simple and qualitative postulates of the axioms from which it derives On these grounds, the CER function clearly exceeds other, seemingly 'simpler', models

The CER parameters, which are easily estimated for a particular individual from a set of his or her risk judgments, seem to generalize well to other sets of choice alternatives, including alternatives that differ in the number of outcomes per alternative. One exception appear to be parameters estimated from two-outcome alternatives which are not very successful in predicting risk judgments for higher-outcome lotteries.

The lack of generalizability of the two-outcome parameter estimates is yet another example of results involving two-outcome gambles that do not readily generalize to higher-outcome gambles. Subjects may be using special strategies when confronted with two-outcome gambles that are not employed in other situations. Researchers in the area of decision theory should be aware and on the lookout for any potential lack of generalizability of results obtained with two-outcome choice alternatives. Given the prevalence of choice alternatives with more than two outcomes in natural settings, it seems advisable that researchers should more routinely include higher-outcome alternatives in their experiments and pay closer attention to possible strategy differences in dealing with two-outcome versus higher-outcome alternatives.

The CER model was developed as a descriptive model of risk, i.e., with the goal of providing a measure that would capture individual differences as well as similarities in people's perception of risk. This, however, does not preclude its use as a normative measure of risk. One could imagine a company establish a set of CER parameters that agree with its corporate goals and objectives and subsequently test its employees to determine the extent to which their risk perceptions deviate from the corporate norm.

Finally, the reader should take note of the limited domain of current theories about risk such as the CER model or theories by other researchers such as Coombs, Fishburn, Lopes, and Payne. These theories are designed to account for the perceived riskiness of choice alternatives that can be described as probability distributions of *uni-dimensional* outcomes (e.g., monetary gambles). In natural settings, choice alternatives more often than not are multidimensional. To account for perceptions of risk for such multidimensional alternatives, extensions of current models of risk (possibly along similar lines as those of multiattribute utility theory) will be necessary, which constitutes a non-trivial task. If successful, such extensions will allow work on risk measurement of the type described in this paper to make contact with scaling studies (e.g., Fischhoff et al. 1982, Slovic et al. 1984, Vlek and Stallen 1981) designed to identify the psychological dimensions (e.g., degree of voluntariness, disaster potential) underlying people's risk perception for social or technological choices Knowledge of the dimensions (both physical and psychological) that are salient and used in judgments of the riskiness of complex stimuli is crucial Equally important (and in many ways complementary) is knowledge of the way in which particular values on these dimensions influence risk (as, for example, the relative contributions and particular functional combination of probability and outcome information in the CER function) or of the way in which different risk dimensions (e g , number of fatalities and voluntariness of exposure) combine into an overall index of perceived riskiness

References

Birnbaum M H 1973 The devil rides again correlation as an index of fit Psychological Bulletin 79 239–242

Birnbaum M H 1974 The nonadditivity of personality impressions Journal of Experimental Psychology 102 543–561

Birnbaum, M H , 1982 Controversies in psychological measurement' In B Wegener (ed) Social attitudes and psychophysical measurements Hillsdale NJ Lawrence Erlbaum

Chandler J P 1969 STEPIT finds local minima of a smooth function of several parameters Behavioral Science 14 81–82

Coombs C H 1969 Portfolio theory a theory of risky decision making La Decision Paris Centre National de la Recherche Scientifique

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

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

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

Coombs C H and J N Bowen 1971b Additivity of risk in portfolios Perception and Psychophysics 10 43–46

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

Coombs C H and L Huang 1970b Tests of a portfolio theory of risk preference Journal of Experimental Psychology 85 23–29

Coombs C H and P E Lehner 1981 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 1110–1123

Coombs C H and P E Lehner 1984 Conjoint design and analysis of the bilinear model an application to judgments of risk Journal of Mathematical Psychology 38 1–42

Fischhoff B P Slovic and S Lichtenstein 1982 Lay foibles and expert fables in judgments about risk American Statistician 36 240–255

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
Holthausen D M 1981 A risk-return model with risk and return measured as deviations from a target return American Economic Review 71 182–188
Huang L C 1971 Experiments on the measurement of risk Michigan Mathematical Psychology Program Report 71-7
Kahneman D and A Tversky 1979 Prospect theory an analysis of decision under risk Econometrika 47 263–291
Keller L R R K Sarin and M Weber, 1985 Empirical investigation of some properties of the perceived riskiness of gambles Organizational Behavior and Human Decision Processes 38 114–130
Lopes L L 1984 Risk and distributional inequality Journal of Experimental Psychology Human Perception and Performance 10 465–485
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 and L Narens 1985 Classification of concatenation measurement structures according to scale type Journal of Mathematical Psychology 29 1–72
Luce R D and E U Weber 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 and M L Braunstein 1971 Preferences among gambles with equal underlying distributions Journal of Experimental Psychology 87 13–18
Payne J W D J Laughhunn and R Crum 1980 Translation of gambles and aspiration level effects in risky choice behavior Management Science 26 1039–1060
Pollatsek A and A Tversky 1970 A theory of risk Journal of Mathematical Psychology 7 540–553
Schoemaker P J H 1982 The expected utility model its variants purposes evidence and limitations Journal of Economic Literature 20 529–563
Slovic P B Fischhoff and S Lichtenstein 1984 Behavioral decision theory perspectives on risk and safety Acta Psychologica 56 183–203
Steiger J H 1980 Tests for comparing elements of a correlation matrix Psychological Bulletin 87 245–251
Tversky A 1967 Utility theory and additivity analysis of risky choices Journal of Experimental Psychology 75 27–36
Vlek C and P-J Stallen 1981 Judging risks and benefits in the small and in the large Organizational Behavior and Human Performance 28 235–271
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 Cambridge MA Harvard University Unpublished Doctoral Dissertation
Weber E U and W P Bottom 1989 Axiomatic measures of perceived risk some tests and extensions Journal of Behavioral Decision Making (in press)