Journal of Experimental Psychology:
Human Perception and Performance
1984, Vol. 10, No. 2, 179-194

Combine and Conquer: A Joint Application of Conjoint and Functional Approaches to the Problem of Risk Measurement

Elke U. Weber
Harvard University

Functional measurement and conjoint measurement techniques were combined in a study of subjective risk designed as a partial check of the risk measures suggested by Luce (1980, 1981). The effect of multiplying all outcomes of a gamble by a constant (i.e., a change of scale) on subjective risk was of interest for this purpose. Ten Harvard undergraduates repeatedly rated the riskiness of two sets of gambles. Both sets of gambles allowed independent assessment of the effect of a gamble's expected payoff (mean), its skewness, and its scalefactor. Each of the 20 data sets was treated as a separate experiment. Employing both functional measurement and polynomial conjoint measurement techniques, we obtained the following results: All three variables showed significant main effects. For 6 of the 18 data sets analyzed, a combination rule that was multiplicative for the variables mean and skewness and additive for scale could be fit. Six additional data sets could be fit by a combination rule multiplicative in all three factors (or, equivalently, additive in all three factors for some monotonic transformation of risk, because there was no sign dependence). In the remaining six data sets simple independence was violated for skewness and/or scale, indicating that more complex combinations of these variables than were considered here might have been applied. Thus, for two-thirds of the data sets the effect of a change of scale on perceived risk was additive. Comparison of the effectiveness of conjoint and functional techniques suggests that the two methods should be used in a complementary way.

Risky choice, or decision making under risk, has been a longstanding interdisciplinary topic of interest and controversy. Research on risk as a variable in its own right has a relatively short history, probably because of the firm establishment of expected utility as the normative model of risky choice (Coombs, 1975). In the traditional expected utility approaches, risk is treated as byproduct of utility assessment. Risk attitudes with their characteristically shaped utility functions (concave for risk aversion, convex for risk seeking) are used to account for an individual's behavior under risk.

An early criticism of this approach came from Allais (1953), who found a positive correlation between risk of an alternative and the dispersion of possible payoffs and argued that even with the introduction of utility functions the expectation principle alone was insufficient to explain risky choice. A more recent critic of the expected utility model, for example, is Shanteau (1980, November), whose results draw both the existence of generalized utility curves and additive utility into question.

A brief review of the psychological literature on risk shows that progress in this area of research has been slow. It also helps to place this study of subjective risk into context. Slovic (1967) conducted the first explicit comparison of perceived riskiness and risk preference. He found that the two judgments had different determinants (i.e., different patterns of correlation with the risk dimensions amount-to-win, amount-to-lose, probability-of-winning, and probability-of-losing) and concluded that

This research was supported in part by a National Science Foundation Grant IST-7924019 to the University of California at Irvine (L. Narens, principal investigator) and in part by a postgraduate fellowship from the Natural Sciences and Engineering Research Council of Canada to the author.

I wish to thank my advisor, R. Duncan Luce, for many important comments and suggestions and for his interest and guidance throughout the project. Three anonymous reviewers also provided valuable comments on an earlier version of this article.

Requests for reprints should be sent to Elke U. Weber, Psychology and Social Relations, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

"riskiness is a meaningful concept, quite distinct from attractiveness" (p. 223).

Coombs' portfolio theory (1969, 1975; Coombs & Huang, 1970b) was designed to explain the relationship between perceived riskiness, risk preference, and risky choice. It does so by assuming an interaction between the expected value and the perceived risk of an option on the one hand and a subject's "ideal risk," a personality variable, on the other hand. Choice among risky decisions is a compromise between maximizing expected value and optimizing the level of risk. Portfolio theory does a significantly better job in accounting for risky choice than its competitor, subjective expected utility theory (Coombs, 1980; Coombs & Huang, 1976), but it has one serious shortcoming. Although risk is treated explicitly as a variable on a par with expected value, it is not explicitly defined. Coombs uses his intuitions about risk when manipulating the value of the risk value.

Pollatsek and Tversky (1970) attempted to incorporate some of these intuitions as axioms of risk judgments that yield a scale expressing the risk of a gamble as a linear combination of its mean and variance:

$$R(X) = t\mathrm{V}(X) - (1 - t)\mathrm{E}(X),$$

where the function R is defined for all random variables X with density functions for which the mean and variance exist. If the risk orderings satisfy the axioms of a risk system, defined by Pollatsek and Tversky in close analogy to an extensive structure, then (a) R is order preserving for some $0 < t < 1$, and (b) for independent random variables X and Y, $R(X \circ Y) = R(X) + R(Y)$, where $\circ$ denotes the binary operation of adding independent random variables, that is, the convolution of their density functions. The appeal of such a risk function is considerable. The only free parameter, t, is attainable from a single judgment of risk-equality between two distinct distributions. The risk of any option can then be easily computed.

However, Coombs and Bowen (1971) found Pollatsek and Tversky's risk measure inadequate. They showed that although perceived risk was a function of the expectation and variance of a gamble, it was also a function of a transformation that changed the skewness of a gamble without affecting either its variance or expectation.

Subsequent to the failure of Pollatsek and Tversky's risk measure, few new suggestions have been made. One alternative approach to investigating risk that has been taken since is to look at the effect of certain transformations of an alternative on its perceived riskiness. The experiments of Coombs and Bowen (1971), Coombs and Huang (1970a), and Coombs and Lehner (1981), for example, have followed this paradigm. Luce (1980, 1981) has taken this approach one step further by deriving some risk measures for which certain transformation effects on perceived risk form necessary and sufficient conditions. In particular, his theory has two choice-points. The first concerns the question of what happens to the risk of a gamble under a change of scale (i.e., when all outcomes are multiplied by a constant c). The two simplest possibilities, an additive or a multiplicative effect, are considered:

$$R(f_c) = R(f) + S(c) \tag{1}$$

or

$$R(f_c) = S(c)R(f), \tag{2}$$

where f is the original density function, f_c is the density function of the transformed gamble, and S is a strictly increasing function with $S(1) = 0$ for (Equation 1) and $S(1) = 1$ for (Equation 2).

The second choice-point concerns the question of how a density function aggregates into a single value for risk. Again two possibilities are considered by Luce. Either the function undergoes a pointwise transformation T and is then integrated,

$$R(f) = \int_{-\infty}^{\infty} T(f(x))dx, \tag{3}$$

for some nonnegative function T, $T(0) = 0$, or the random variable undergoes some transformation T, and R is the expectation of the resulting random variable,

$$R(f) = \int_{-\infty}^{\infty} T(x)f(x)dx = \mathrm{E}(T(X)). \tag{4}$$

The combination of options at the two choice-points yields four possible measures of risk. By testing the validity of assumptions in Equations 1, 2, 3, and 4 for a set of gambles, it can be decided which, if any, of the four suggested risk measures is correct.

This article concerns itself with the first choice-point in Luce's theory, that is, the effect of a change of scale on perceived riskiness. In addition, the other two transformations that Coombs and Bowen (1971) found to be affecting riskiness, change of origin and change of skewness, are considered in an attempt to find an algebraic combination rule for these three effects that could describe perceived riskiness and would provide subjective scale values for these variables.

The two analytical approaches that allow one to find and test algebraic combination rules of independent variables from the judgment of a dependent variable are conjoint measurement (Krantz, Luce, Suppes, & Tversky, 1971) and functional measurement (Anderson, 1981, 1982). Both approaches have their advantages as well as problems that will be fully discussed later on. For the purpose of obtaining an algebraic model of the effects of changes in mean, skewness, and scale on riskiness, the two approaches are used here in a complementary way. When functional or conjoint measurement encounters problems in diagnosing an algebraic model, such difficulty can mean one of two things. Either no simple model is appropriate for the data under investigation or the methods of analysis are inadequate to detect it. Using two approaches with very different techniques makes it possible to diagnose a given model on the basis of converging evidence even when each approach by itself would not allow us to do so (Wallsten & Budescu, 1981). Furthermore, a comparison of the two measurement techniques on the same set of data will help to demonstrate strengths and weaknesses in either approach, which may benefit future applications of either conjoint or functional measurement.

In our investigation, the functional measurement approach was originally chosen because of the experimental ease of obtaining the direct risk ratings required by it. To collect paired comparison or rank order response measures usually used with conjoint measurement is, in comparison, time consuming and tedious and may produce idiosyncratic results because subjects are bored as observed by Slovic, Lichtenstein, and Edwards (1965). Consequently, the number of levels of each factor is often restricted to keep the total number of items in the factorial design small. This limitation is much less severe for functional measurement. When functional measurement techniques proved insufficient to establish an algebraic model for the data in this investigation, conjoint measurement procedures were added to the analysis. Although not without their own problems, these provided further insight into the existence and nature of underlying composition rules and reasons for the difficulties encountered by the functional measurement analysis.

Method

Design

Three independent transformations were applied to the basic two-outcome gamble (g) of winning \$2.25 with probability $p = 1/2$ and losing \$2.25 with probability $1 - p = 1/2$:

$$g = (\$2.25, 1/2, -\$2.25).$$

The first transformation changed the skewness of the gamble while leaving the expectation and variance unchanged (as used by Coombs & Bowen, 1971). If $g = (a, 1/2, -a)$, then $\alpha(g) = (a\sqrt{q/p}, p, -a\sqrt{p/q})$, $0 < p < 1$. Three levels of skewness were used ($p = 1/4, 1/2, 3/4$).

The second transformation, a change of origin, transformed the gamble $g = (y, p, -z)$ into $\beta(g) = (y + b, p, -z + b)$, employing five levels of b ($b = -\$1.20, -\$0.60, \$0.00, \$0.60, \$1.20$). This transformation changed the expectation of the gamble by the amount of b while leaving the variance and skewness unchanged.

For the third transformation, a change of scale, the outcomes of the previously generated gambles, $g = (y, p, -z)$, were multiplied by a constant c, $\gamma(g) = (cy, p, -cz)$, using four levels of c ($c = 1, 2, 4, 8$). This transformation altered the expectation of the gamble by the factor c and the variance by c^2 but left the skewness unchanged; thus 60 two-outcome gambles were generated.

A set of five-outcome gambles was produced by the multiple play transformation $\delta(g) = (y, p, -z)^d$, with $d = 4$, where the integer d indicates that the gamble g is played d times independently. The multiple play transformation was applied to the 15 two-outcome gambles created by the transformations $\alpha(g)$ and $\beta(g)$. Only three levels of c ($c = 1, 2, 4$) were used for the subsequent change of scale transformation $\gamma(g)$, for a total of 45 five-outcome gambles.[1]

---

[1] A complete list of both sets of stimulus gambles and the mean risk judgments given for them by each of 9 subjects can be obtained from the author.

```
              Two-Outcome Set

 $49.20                     -$1.20

xxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
 .25                        .75


              Five-Outcome Set

 $9.40      $7.10      $0.60      -$44.30     -$74.00

 .004        X          X            X           X
            .047        X            X           X
                        X            X           X
                        X            X           X
                        X            X           X
                       .211          X           X
                                     X           X
                                     X          .316
                                     X
                                     X.
                                     X
                                   .422
```

Figure 1. Typical stimulus item for the two-outcome gamble set and five-outcome gamble set.

Stimuli

The gambles were displayed one at a time on a CRT screen. The probabilities of outcomes were depicted by a proportionate number of Xs as well as by their numerical values. Monetary outcomes (losses indicated by a minus sign) appeared above the respective probabilities (see Figure 1). A 20-cm long horizontal line, the risk rating scale labeled *LO, MED,* and *HI,* appeared underneath the gamble.

Procedure

In Session 1 subjects rated a block consisting of 60 two-outcome (2-o) gambles and a block consisting of 45 five-outcome (5-o) gambles; the order of gambles was randomized within each block.

A graphic rating response was chosen to avoid number preferences that seem to occur even with well-practiced subjects (Shanteau & Anderson, 1969). To avoid endpoint biases and to familiarize subjects with the range of probabilities and outcomes of the stimulus set, several practice gambles that had the same densities but more extreme outcomes than the stimulus items were presented at the beginning of each block. It was pointed out to the subjects that these gambles presented extremes relative to the stimulus gambles they could expect. The practice set included one gamble that offered a chance for a gain and the status quo otherwise—presumably a very low risk according to most definitions of risk that emphasize the downside potential of choices and assume that a zero outcome and its associated probability does not have a direct effect on risk (e.g., Coombs & Lehner, 1983). Another practice gamble offered the chance for a loss and the status quo otherwise, a high-risk gamble if one assumes that the negative impact of a possible loss on risk usually is counteracted by the positive impact of the possibility for gain. All experimental gambles involved both a chance of winning and of losing. We hoped that such gambles would establish the status quo for each subject as the reference point for the risk judgment.

Subjects rated one gamble at a time as it appeared on the screen. A response terminated the display of the current item and initiated the next display. Subjects were encouraged to proceed at their own speed (even though a record of response time measured from stimulus onset was kept).

The next four sessions were replications of Session 1, except that the stimulus sets were presented in a different random order at each session. Order of presentation (2-o block vs. 5-o block) was also random for each session.

Instructions

Subjects were told to imagine themselves in the situation of playing the displayed 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 would like to play it. Risk ratings were to be given "intuitively," a "gut reaction" to each individual gamble, as opposed to "computing" risk in some way or deriving hard and fast rules.

Apparatus

The stimulus items were presented on a Digital Equipment Corporation Type 340 CRT on a PDP 4, which was controlled by a PDP 1144. Rating responses made with a lightpen on the CRT were scored and stored by the PDP 1144.

Subjects

Five female and 5 male Harvard University students participated in the study. Each was run in five individual sessions of ½-hour duration within a 10-day period, for a total pay of $12.00.

Results

Response Scale

Risk ratings were determined from the graphic responses and could take any integer value between 0 and 635. Ratings actually ranged from 1–615, which seems to indicate that the attempt of inducing subjects to use the full range of the rating scale by providing practice gambles as end-anchors was successful.

Consistency of Ratings

The Pearson product-moment correlations between the risk ratings at each replication were computed. The average correlation for each subject ranged between .60 and .94 with a mean of .80 and a standard deviation of 0.12 for both the 2-o and the 5-o set, excluding Subject EL in the 2-o set and Subject RO in the 5-o set, who both had average replication correlations of .18. RO had reported that he got confused, especially in the more complex 5-o situation, between calling a gamble a "bad" risk (i.e., high on the scale) or a "low" risk (i.e., low on the scale) response. Subject EL's inconsistency should probably be attributed to carelessness because her response times for the 2-o set were less than half as long as those of the other subjects. Subject EL's 2-o set and RO's 5-o set were excluded from further analysis.

Functional Measurement

Statistical analysis. An analysis of variance of the individual subject risk ratings gives a first indication of the type of algebraic model that might fit the data.[2] If the risk judgments given by subjects were additive in all three variables—mean, skewness, and scale—we would expect significant main effects for these three factors, but none of the interactions should reach significance. As can be seen in Table 1, this is clearly not the case. The main effects of the variables skewness (S), mean (M), and scale (C) are highly significant for almost every subject. In only three cases was skewness not significant, and in four cases scale did not reach significance. Yet, many of the two-way and three-way interactions also reach statistical significance, thus excluding the three-factor additive model. The prevalence of the Skewness × Mean interaction and the absence of other interactions for many of the subjects suggests the dual-distributive model (S × M) + C. In other cases the interaction pattern suggests the three-factor multiplicative model S × M × C.

To conduct goodness-of-fit tests for these models the Linear × Linear (trilinear) component and the residual component were computed for all significant two-factor (three-factor) interactions using Shanteau's (1977) POLYLIN program. The results of the tests are also reported in Table 1, which shows the *F*-ratios of the residual components over the error term for the overall interaction. The interactions are consistent with multiplicative integration of the factors involved to the extent that these residual *F*-ratios fail statistical significance (Anderson & Shanteau, 1970). As can be seen in Table 1, the Skewness × Mean interaction can be explained by the Linear × Linear component in seven cases, and all but two of the Skewness × Mean × Scale interactions are accounted for by the trilinear component.

Where does this leave model diagnosis? The dual-distributive model requires a Linear × Linear Skewness × Mean interaction with no other interaction reaching significance. This condition is satisfied for three data sets (Subjects RO, SI, in the 2-o set and Subject SI in the 5-o set). The three-factor multiplicative model requires that all interactions be accounted for by the bilinear or trilinear component. This condition is satisfied by only one data set (Subject DA in the 2-o set). For the remaining 14 sets the goodness-of-fit tests do not allow us to diagnose any simple algebraic model.

Graphic analysis. Anderson (e.g., 1981) recommends the use of visual inspection to check the adequacy of an algebraic model for a set of data. A plot of mean responses against the subjective scale values of one of the component factors for every level of the other factor (in a two-factor design) might, for example, reveal a pattern of parallel lines. Such parallelism is diagnostic of an additive integration. In comparison, a pattern of diverging lines, or "linear fan," is diagnostic of multiplicative integration of the two factors. Even if the statistical analysis failed to provide a clear diagnosis of an integration rule for many of the data sets, visual inspection of the factorial plots may reveal the reasons for this failure and some remedy. Unfortunately, graphic analysis quickly becomes cumbersome when more than two factors are involved because it becomes necessary to inspect several plots for each subject to assess the adequacy of an algebraic model. Furthermore, visual inspection alone

[2] A between-subjects ANOVA showed significant interactions between the subject variable and all main effects as well as all interaction effects except for the M × C interaction in the 2-o set and the S × C and the M × C interaction in the 5-o set. For this reason, all subsequent analyses were carried out on the individual subject data.

Table 1
ANOVA With Three Repeated Within-Factors (Skewness, Mean, Scale) for Each Subject: F-Ratios and Significance Level

	Main effects			Interaction effects			
Subject	S	M	C	S × M	S × C	M × C	S × M × C
				Two-outcome set			
AL							
F	24.63***	175.40***	16.47**	2.96*	4.26**	1.75	1.68*
F^a				(1.62)	5.05		(1.29)
AU							
F	314.40***	355.70***	101.30***	16.93***	2.31	3.19**	4.18***
F^a				4.07		3.40	2.09
BI							
F	103.44***	436.80***	6.14***	17.89***	8.69***	2.66**	4.98***
F^a				6.32	(1.62)	(1.95)	4.72
CH							
F	2.65	115.62***	12.51***	21.57***	1.31	2.88**	2.17**
F^a				20.28		2.52	(2.07)
DA							
F	2.87	38.77***	5.09*	1.87	1.32	0.82	1.64*
F^a							(1.52)
DO							
F	366.31***	329.78***	51.21***	3.75**	0.17	1.65	2.23**
F^a				2.92			(0.99)
JA							
F	18.62***	131.22***	4.94*	15.67***	0.30	1.91	1.85
F^a				12.28			(2.54)
RO							
F	209.13***	124.35***	14.71***	32.70***	1.32	0.55	1.13
F^a				(2.29)			
SI							
F	14.52**	166.47***	8.28**	4.66***	0.75	0.79	1.49
F^a				(2.21)			
df	(2, 8)	(4, 16)	(3, 12)	(8, 32)	(6, 24)	(12, 48)	(24, 96)
				Five-outcome set			
AL							
F	67.64***	225.5***	53.01***	6.12***	0.34	3.34**	0.84
F^a				(2.06)		3.23	
AU							
F	73.91***	339.45***	20.00***	19.22***	9.31***	2.81*	1.73
F^a				3.94	3.87	(1.92)	
BI							
F	12.28**	479.14***	5.75*	12.66***	1.55	1.44	1.69
F^a				14.3			
CH							
F	9.33**	445.04***	5.78*	17.24***	1.99	4.48***	0.93
F^a				132.4		5.12	
DA							
F	33.32***	266.56***	1.04	6.30***	1.95	0.91	0.67
F^a				3.78			
DO							
F	75.51***	232.89***	46.29***	3.22**	1.60	1.31	1.53
F^a				21.5			

Table 1 (*continued*)

	Main effects			Interaction effects			
Subject	S	M	C	S × M	S × C	M × C	S × M × C
JA							
F	58.76***	191.34***	4.00	5.39***	3.25*	2.71*	0.88
F^a				5.12	(0.90)	(2.31)	
EL							
F	76.17***	51.75***	0.03	4.99***	1.70	1.88	0.81
F^a				2.81			
SI							
F	264.80***	98.74***	7.70***	8.16*	1.73	0.61	
F^a				(2.06)			1.03
df	(2, 8)	(4, 16)	(2, 8)	(8, 32)	(4, 16)	(8, 32)	(16, 64)

Note. S = skewness; M = mean, C = scale.
[a] For significant interaction effects, the *F*-ratios for the test of fit of multiplicative combination are reported. Parentheses indicate that the multilinear component accounts for the interaction (i.e., the *F*-ratio of the residual component fails significance at the .05 level).
* $p < .05$. ** $p < .01$. *** $p < .001$.

may be insufficient to discriminate between different algebraic models for "noisy" data sets like the ones at hand. Figure 2 shows the factorial plots of three subjects selected because they are representative of three groups of subjects identified by the conjoint measurement analysis discussed below. It is clear that neither parallelism nor linear-fan divergence characterizes the Skewness × Mean interaction in any case. Beyond that, one may observe that the DO data seem most orderly, followed by the AL data, with the BI data being least interpretable. It seems difficult to tell more about the nature of the interactions or to judge on the basis of visual inspection alone for which of the data sets monotone transformation of the risk data could improve model fit.

Thus, in summary, functional measurement indicates that the dual-distributive integration model characterizes three data sets and the three-factor multiplying model one other set. It seems likely that other data sets could conform to one or the other of these models after some appropriate rescaling, but there is little indication for which data sets such rescaling might be successful. To gain a better understanding of this issue the data were reanalyzed in terms of the conjoint measurement axioms.

Conjoint Measurement

Checking a small number of conditions is, theoretically, sufficient to diagnose any of the four classes of simple composition rules based on addition and multiplication that exist for three variables. The main characteristic of the conjoint measurement approach is that only the ordinal aspects of the data are required to test compatibility with a particular composition rule. Different patterns of violation of simple and joint independence, double cancellation, and distributive and dual-distributive cancellation should allow one to distinguish between an additive, distributive, dual-distributive, and multiplicative rule. Simple independence requires that the ordering of the dependent variable (risk) with respect to one of the independent variables (e.g., skewness) is independent of the levels of the other two variables. Joint independence requires that the joint ordinal effect of the simultaneous variations of two variables is the same regardless of the level of the third variable. The cancellation conditions require that the effect of replacing one level of a variable by another level can be cancelled by a similar replacement on another variable (Krantz, Luce, Suppes, & Tversky, 1971; Krantz & Tversky, 1971). The data were analyzed for compliance with these axioms. The results for the tests of independence, double cancellation, and joint independence appear in Table 2.

Three complications in this analysis deserve some discussion. One concerns the conversion of the numerical—presumably interval scale—data into the ordinal form used by the mea-

Figure 2. Factorial plots of subjective risk judgment as a function of mean, skewness, and scale of a gamble for Subjects DO, AL, and BI in the 2-o set.

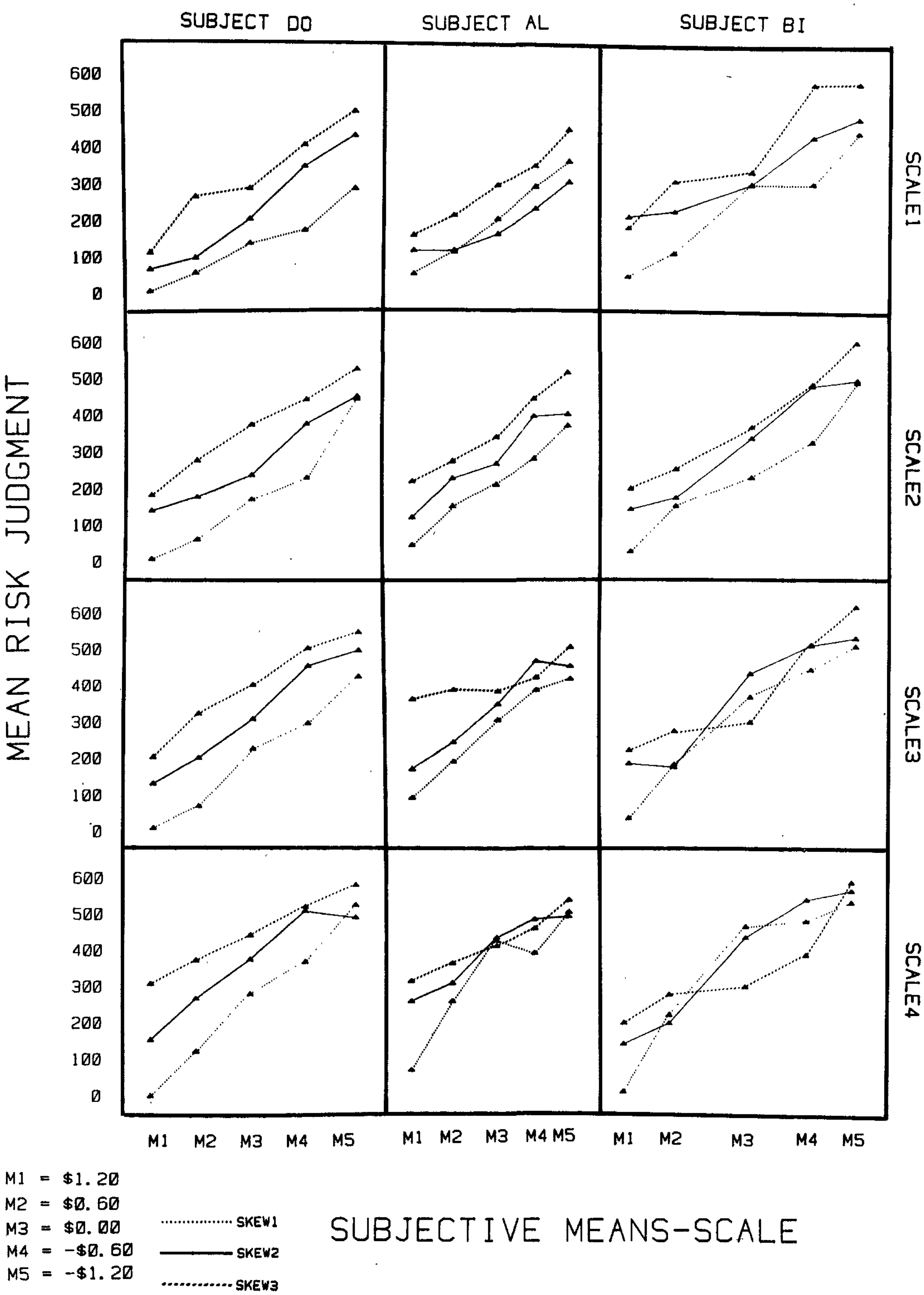

Table 2
Proportion of Violations of the Independence, Double Cancellation, and Joint Independence Axioms of Conjoint Measurement

Subject	Independence[a] Skew	Independence[a] Scale	Double cancellation (S × C)[b]	Joint independence (S × C)[b]
		Two-outcome set		
AL				.056
AU				
BI	.088	.131	.126	.104
CH	.176			.146
DA	.070			.062
DO				
JA				
RO				.056
SI	.070			.076
		Five-outcome set		
AL				.097
AU				
BI	.143	.119	.089	.181
CH	.143	.071	.125	.167
DA	.119	.071		
DO				
JA	.071	.142		.069
EL				
SI	.071			.056

Note. Only violations in excess of 5% are listed. M = mean, S = skewness, C = scale.
[a] There were no violations in excess of 5% for Mean.
[b] There were no violations in excess of 5% for M × S or M × C.

surement axioms. The induced order of two data points may violate one of the axioms, yet the two points may be so close to each other on the risk scale that in a paired-comparison or ranking task an indifference response would have been given. To prevent such inflation of the violation count, a *t* test was performed for each apparent violation. Only cases in which the null hypothesis of no difference between the data pair values was rejected were counted as violations. To ensure that this test had sufficient power, an alpha-level of .10 (one-tailed) was used.

The second and more fundamental concern is the question of whether risk ratings and choices lead to the same rank orderings. There are well-documented response mode effects involving bids versus choices (e.g., Lichtenstein & Slovic, 1971). Such effects are generally attributed to changes in salience of different risk dimensions brought about by the response mode (i.e., payoffs being salient in money bids, and chance-to-win being salient in preference choices), and not to the response mode as such. Although this issue deserves further investigation, it is irrelevant to our analysis, which used the induced rank ordering simply to determine the existence of monotone transformations on the rating data that would allow the fit of one of the four algebraic models.

The third complication arises from the major shortcoming of conjoint measurement, that is, its lack of an appropriate error theory. How many violations can reasonably be attributed to random error? The following ad hoc rule was used in the present analysis. We accepted a certain percentage of violations (i.e., attributed it to random error) such that the results of the functional measurement analysis and the conjoint measurement analysis would agree. Setting violation tolerance to 5% fulfilled this purpose with only one exception (violation of skewness independence was .07 for Subject SI).

As can be seen in Table 2, all three subjects who by functional analysis were diagnosed as using the dual-distributive integration rule (S × M) + C satisfied simple independence for all three factors, double cancellation for all pairs of factors, and joint independence for two pairs of factors. Ideally, joint independence should have held for only one pair of factors (Mean × Skewness). The narrow range of scale values (especially true for the scale and the skewness variable) may have been responsible for this irregularity (Krantz & Tversky, 1971). Possibly the same reason lies behind distributive cancellation holding for these subjects even though a dual-distributive model had been fit to their data. In fact, for both distributive and dual-distributive cancellation no subject exceeded the 5% violation limit, which is the reason these two axioms do not appear in Table 2. In addition, the axiom tests implicated the dual-distributive model for Subjects AL and DA in the 2-o set and for Subject AL in the 5-o set. This means that for these cases all but the Linear × Linear Skewness × Mean interactions can be removed by appropriate rescaling. The one case for which functional diagnosis (multiplicative) and conjoint diagnosis (dual-distributive) disagreed (Subject DA in the 2-o set) can be considered a degenerate instance of both the dual-distributive and the multiplicative model in that the variable skewness was not significant in the original ANOVA and the variable scale barely so (see Table 1). Sign dependence occurred only for Subject DA in the 2-o set, where skewness was sign dependent on mean.

The pattern of axiom compliance implicated the multiplicative rule S × M × C for Subjects DO, EL, and AU in the 5-o set, and Subjects JA and AU in the 2-o set. The fact that this was not diagnosed by the functional measurement analysis means that a monotone transformation of the risk judgments is necessary to make the data multiplicative.

Because none of the sets identified as conforming to the multiplicative rule showed sign dependence, the risk judgments can be rescaled to comply with a three-factor additive integration rule. Both functional measurement (Birnbaum, 1982) and conjoint measurement (Krantz & Tversky, 1971) are incapable of distinguishing between an additive and a multiplicative combination function in the absence of sign dependence. Because monotone transformation of the risk ratings was necessary anyway, a transformation to fit the three-factor additive model (Young, 1972) was applied for reasons of parsimony.

Two points are worth noting about the scale values shown in Table 3. First, in three cases the scale estimates for skewness are not monotonic with the objective values; instead, the symmetric distribution (S2) is judged less risky than either the positively (S1) or the negatively skewed (S3) distribution. Individual differences in the subjective evaluation of changes in risk for different levels of skewness are consistent with Leopard's (1978) observation that there is no a priori risk ordering over gambles differing in skewness (but not in expected value and variance) because skewness differences are the result of simultaneous and compensatory changes in both the probabilities and the amounts to win and lose. Thus, for example, by selectively attending to the increased prob-

Table 3
Standardized Estimates of the Subjective Scale Values for Skewness, Mean, and Scale for Subjects Using the Three-Factor Additive Model (S + M + C) and Subjects Using the Dual-Distributive Model (S × M) + C

	Skewness			Mean					Scale			
Subject	S1	S2	S3	M1	M2	M3	M4	M5	C1	C2	C3	C4
					Three-factor additive model							
Two-outcome set												
AU	−1.0	−1.6	1.0	2.0	1.0	0.0	−6.0	−19.9	*ns*			
DO	−1.0	−1.6	1.0	1.8	1.0	0.0	−2.2	−6.7	1.0	2.0	1.7	2.0
JA	−1.0	0.0	0.3	*ns*					*ns*			
Five-outcome set												
AU	−1.0	−1.8	1.0	1.5	1.0	0.0	−1.6	−6.3	1.0	0.0	0.4	
DO	−1.0	0.0	1.0	2.7	1.0	0.0	−3.4	−0.9	*ns*			
EL	−1.0	0.0	3.6	1.4	1.0	0.0	−1.1	0.0	*ns*			
					Dual-distributive model							
Two-outcome set												
AL	−1.0	0.3	1.0	1.9	1.0	0.0	−1.2	−2.2	1.0	2.0	3.0	3.9
DA	*ns*			1.7	1.0	0.0	−0.9	−2.3	1.0	2.0	8.5	21.0
RO	−1.0	0.1	1.0	1.6	1.0	0.0	−0.9	−2.6	1.0	2.0	3.5	4.7
SI	−1.0	0.7	1.0	2.0	1.0	0.0	−1.6	−2.8	1.0	2.0	3.3	4.1
Five-outcome set												
AL	−1.0	0.1	1.0	2.9	1.0	0.0	−4.2	−7.8	1.0	2.0	3.0	
SI	−1.0	−0.8	1.0	2.0	1.0	0.0	−1.2	−1.9	1.0	2.0	3.1	

Note. M = mean, S = skewness, C = scale, *ns* = not significant.

ability of losing in a positively skewed gamble and to the increased dollar loss in the negatively skewed gamble, a subject could perceive both gambles as riskier than the corresponding symmetric gamble. Second, there is no significant difference between the scale value estimates for the different values of the scale variable in four cases; in the remaining two cases the relationship of the estimates to the objective values is nonmonotonic.

Because both methods of analysis essentially agreed on the dual-distributive integration rule $(S \times M) + C$ for six of the data sets, we assumed that for these data sets the ANOVA marginal mean estimates of the subjective scales for skewness and mean were satisfactory approximations to the optimal scale values, and no further rescaling was attempted. These scale values are also reported in Table 3.

For the remaining cases for which functional measurement could not fit an algebraic model, the reason for this failure also became apparent. Simple independence of skewness and/or scale was violated for Subjects BI and CH in the 2-o set and for Subjects BI, CH, DA, and JA in the 5-o set. These violations were not the result of sign dependence. Thus it seems that these subjects either used variables other than skewness and scale in making their risk judgments or that they combined them according to a mathematically more complex form than we have examined.

Discussion

Recall the dual purpose of this study. The first one was a substantive investigation of the effect of changes in mean, skewness, and scale on a gamble's perceived risk. This was to provide a partial check on the appropriateness of the risk measures suggested by Luce (1980, 1981). The second purpose was a methodological one, a comparison of the performance of functional and conjoint measurement analysis on a "difficult" set of data. As the title of this paper in a paraphrase of the "divide and conquer" decomposition approach to decision aiding (Slovic, Fischhoff, & Lichtenstein, 1977) suggests, data encountered in the area of risk judgment may be sufficiently noisy and plagued by individual differences that the investigator needs all the analytic help he or she can collect from various camps. We will first discuss the methodological issues brought up in this study and then turn to a discussion of the substantive insights about subjective risk perception.

Lack of an error theory is the major shortcoming of conjoint measurement and, according to Falmagne (1976), probably the reason why so few empirical studies have made use of this measurement technique. In the form presented by Krantz et al. (1971), conjoint measurement assumes error-free rank-order responses, so that the results of diagnostic tests must be interpreted algebraically rather than statistically. A critical review of the various ways that have been suggested for dealing with random error in a conjoint measurement analysis may perhaps increase the use of this powerful analytic technique. Falmagne, in an attempt to develop "random analogues" of the classical theories of fundamental measurement, introduced an additive random conjoint measurement model in 1976 that has random variables as its primitives and develops necessary and/or sufficient conditions on the medians of these random variables that guarantee the existence of an additive model. In 1979, a probabilistic conjoint measurement model followed that used probability measures (the probability of picking object ax in the set [ax, by]) as its primitives and developed conditions on these probabilities that guaranteed the existence of some classes of models. The problem with Falmagne's suggestions is that they have only a narrow range of applicability. The random conjoint theory (Falmagne, 1976) uses a matching paradigm instead of the usual comparison paradigm and covers only the additive case. The probabilistic conjoint measurement theory (Falmagne, 1979) in its most general form has the desired generality and can accommodate noisy data but puts unreasonably weak constraints on the data. Additional constraints that would strengthen the representation have been developed for only some classes of combination rules.

Another strategy of dealing with random error for replicated designs has been to code the noisy cell samples into an error-free framework prior to diagnosis (e.g., Coombs & Huang, 1970a), using cell means or medians as the coded rank representing the central tendency of the cell population. Problems with this approach, pointed out by Person and Barron (1978), are that the diagnostic results likely depend on the particular coding method applied and that the fundamental difficulty in interpreting diagnostic failures (i.e., model inadequacy or noisy data) remains. The latter problem is usually dealt with by using an ad hoc heuristic that defines an acceptable number of failures (e.g., Coombs & Bowen, 1971; Coombs & Huang, 1970a).

The third strategy, demonstrated by Person and Barron (1978), involves a nonparametric test on the response ranks in each pair of cells (Person, 1977b). The upper-tail probabilities, P_{ij}, for the rank sums thus computed are then used to judge whether a cell is statistically larger than another in the conjoint measurement diagnosis (Person, 1977a). After deciding on an alpha level for P_{ij}, we attributed any violations of the conjoint axioms to model failure, but the number of diagnostic failures is, of course, a function of the alpha level employed. The procedure that was used in this study is the parametric equivalent of Person's strategy. In our case, an alpha level of .10 still resulted in violations of single-factor independence up to 7% for data set for which a particular combination rule had already been diagnosed by functional measurement. Thus it seems that alpha levels in excess of .10 may have to be accepted if a zero violation rate is to be the criterion of model adequacy.

The latest suggestion for dealing with random error comes from Coombs and Lehner (1981, 1983). The basic idea is that low-consistent (LC) subjects should show a higher percentage of random error than more consistent (MC) subjects in the conjoint axiom checks. High and low consistency is defined by the degree of correlation between replications of the experimental ratings. If MC subjects show a higher violation rate than LC subjects, Coombs and Lehner conclude that model inadequacy rather than random error is responsible for the axiom violation. Problems with their line of argument are discussed in the Appendix.

In summary, the most appropriate and most practical of the available approaches for dealing with random error within conjoint measurement seems to be the one suggested by Person (1977a, b) and its parametric equivalent in this article. The comparison of functional and conjoint measurement results here has thrown some light on an appropriate level of alpha for this procedure.

A related problem of conjoint measurement, but one less frequently discussed, is the dependence of diagnostic power of some of the axiom tests on the appropriate spacing of scale values (Krantz & Tversky, 1971, p. 157). If the range of values is too narrow, joint independence, for example, may be satisfied for more than one pair of factors even when the dual-distributive rule holds, as was demonstrated in this study for several data sets. Without the confirming evidence from the functional measurement analysis that a particular combination rule does in fact fit the data, it may be difficult to detect such restricted range problems that might lead to an incorrect model diagnosis.

On the positive side of conjoint measurement, the axiom tests helped us to detect the cases for which monotone rescaling allowed the fit of an integration rule, that is, cases where the functional measurement assumption that subjects' responses are a linear function of their judgments (e.g., Anderson, 1981, p. 9) was inappropriate. Anderson recommends the use of visual inspection of the factorial plots to detect such cases. Subject DO in Figure 2 is a typical example of the three-factor additive group, Subject AL is a typical dual-distributive case, and Subject BI is one of the subjects who violated independence of skewness and of scale. Although it may be obvious that the DO data can be rescaled to three-factor additivity, it is less clear that this should not be the case for the AL data. To infer violation of skewness and scale independence from the BI plot would require a long experience of interpreting factorial plots.

Budescu and Wallsten (1979) recommend a multiple-step procedure for dealing with monotone rescaling that is in some respects similar to the procedure used in this study. They suggest the use of a distribution-free expected normal scores test (Marascuilo & McSweeney, 1977) instead of the less powerful regular nonparametric ANOVA tests usually used in functional measurement (Anderson, 1977, p. 203) to detect instances that might profit from monotone rescaling, but they recommend the use of the conjoint measurement axioms in those cases where not even monotone rescaling could lead to acceptance of a model "in order to reveal the exact reasons for its failure" (Budescu & Wallsten, 1979, p. 309).

The fact that the conjoint and functional measurement analyses essentially agreed on some combination rule for most data sets gives some preliminary encouragement to the use of rating responses (and the induction of rank order from them) in conjoint measurement studies. If indeed it was found that the rank order induced from ratings differed from that induced from pairwise comparisons, it would not be clear which of the two more correctly reflected the subject's "true" ordering. As already mentioned, the use of rank order or pairwise comparison responses in conjoint measurement studies restricts the number (and thus often range) of factor values used. Coombs and Lehner (1983), for example, develop a new methodology to test the bilinear model despite the existence of Huang's (1975) theory for that model because "testing the conditions [of Huang's model] by conjoint measurement methods requires that a subject rank order 64 stimuli, an excessive burden that could degrade the data" (Coombs & Lehner, 1983, p. 5). Having the subject rate the stimulus items and inducing rank orders from these ratings eliminates most of these concerns and the problems associated with restrictions in range. It also allows the joint application of functional measurement techniques with the immediate benefit of subjective scale values if an algebraic model can be fit to the ratings without any further transformations.

What, then, has the "combine and conquer" strategy contributed to our understanding of subjective risk? Our results are a replication of Coombs and Bowen (1971) in that skewness was a significant determinant of subjective risk. In their discussion (p. 27), Coombs and Bowen mention that their results are compatible with either an additive or a multiplicative combination of the variables labeled skewness (S) and mean (M) in our study. Our results show that both combination rules probably occur. Furthermore, 6 of the 8 subjects for whom both stimulus sets were analyzed combined S and M in the same way in both sets, indicating that the type of rule used is determined more by the individual subject than by task demands.

The effect of a change of scale on subjective risk turned out to be more consistent across subjects. For all 12 subjects for whom a simple algebraic model could be fit, the effect of scale was additive. This immediately narrows the set of possible risk measures suggested by Luce (1980, 1981) from four to two.

Finally, it should be emphasized that the combination rules for the three factors (skewness, mean, and scale) suggested by this study are not intended to be descriptive of the actual cognitive integration process that produces a risk judgment nor of the actual dimensions used in such a process. But since the combination rules are descriptive of risk judgments in the descriptive-predictive sense, they will constrain any process model in that such a model will have to be able to account for our results.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine [The behavior of the rational man under risk: Critique of the postulates and axioms of the American school]. *Econometrica, 21,* 503–546.

Anderson, N. H. (1977). Note on functional measurement and data analysis. *Perception & Psychophysics, 21,* 201–215.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory.* New York: Academic Press.

Anderson, N. H., & Shanteau, J. C. (1970). Information integration in risky decision making. *Journal of Experimental Psychology, 84,* 441–451.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.

Budescu, D., & Wallsten, T. S. (1979). A note on monotonic transformations in the context of functional measurement and analysis of variance. *Bulletin of the Psychonomic Society, 14,* 307–310.

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

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

Coombs, C. H. (1980). Risk preferences and the theory of risk. In E. D. Lautermann & H. Feger (Eds.), *Similarity and choice: Papers in honour of Clyde Coombs* (pp. 346–363). Bern: Hans Huber.

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

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

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

Coombs, C. H., & Huang, L. (1976). Tests of the betweenness property of expected utility. *Journal of Mathematical Psychology, 13,* 323–337.

Coombs, C. H., & Lehner, P. E. (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., & Lehner, P. E. (1983). *Conjoint design and analysis of the bilinear model: An application to judgments of risk.* Manuscript submitted for publication.

Falmagne, J. C. (1976). Random conjoint measurement and loudness summation. *Psychological Review, 83,* 65–79.

Falmagne, J. C. (1979). On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology, 19,* 73–88.

Huang, L. C. (1975). A nonsimple conjoint measurement model. *Journal of Mathematical Psychology, 12,* 437–448.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement,* (Vol. 1). New York: Academic Press.

Krantz, D. H., & Tversky, A. (1971). Conjoint measurement analysis of composition rules in psychology. *Psychological Review, 78,* 151–169.

Leopard, A. (1978). Risk preference in consecutive gambling. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 521–528.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89,* 46–55.

Luce, R. D. (1980). Several possible measures of risk. *Theory and decision, 2,* 217–228.

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

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences.* Monterey, CA: Brooks/Cole.

Person, H. B. (1977a). SCJM: A stochastic conjoint-measurement program. *Journal of Marketing Research, 14,* 98–99.

Person, H. B. (1977b). WMW: A Wilcoxon Mann-Whitney cell comparison program. *Journal of Marketing Research, 14,* 97–98.

Person, H. B., & Barron, F. H. (1978). Polynomial psychophysics of group risk perception. *Acta Psychologica, 4,* 421–428.

Pollatsek, A., & Tversky, A. (1970). A theory of risk. *Journal of Mathematical Psychology, 7,* 540–553.

Shanteau, J. (1977). POLYLIN: A Fortran IV program for the analysis of multiplicative (multilinear) trend components of interactions. *Behavior Research Methods & Instrumentation, 9,* 381–382.

Shanteau, J. (1980, November). *Where's the psychology in economic utility theory?* Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.

Shanteau, J. C., & Anderson, N. H. (1969). Test of a conflict model of preference judgment. *Journal of Mathematical Psychology, 6,* 312–325.

Slovic, P. (1967). The relative influence of probabilities and payoffs upon perceived risk of a gamble. *Psychonomic Science, 9,* 223–224.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

Slovic, P., Lichtenstein, S., & Edwards, W. (1965). Boredom-Induced changes in preference among bets. *American Journal of Psychology, 78,* 208–217.

Wallsten, T. S., & Budescu, D. V. (1981). Additivity and nonadditivity in judging MMPI profiles. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1096–1109.

Young, F. W. (1972). A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and application in the behavioral sciences* (Vol. 1, Theory, pp. 69–104). New York: Academic Press, 69–104.

Appendix

The triple cancellation test for additivity of two variables extended to the multiplication of two variables, with one of the variables sign-dependent on the other (Coombs & Lehner, 1981, p. 1119), consists of three antecedent inequalities and one conclusion inequality. A violation of the model occurs if all the antecedent inequalities and the conclusion inequality have the same direction. The model is confirmed if the conclusion inequality has the direction opposite from the antecedent inequalities. The case of the conclusion agreeing in direction with one antecedent but not with the other two is compatible with the model. (That is, it neither violates nor strongly confirms the model.)

Coombs and Lehner argue that the fact that the MC (most consistent) subjects show a higher proportion of violations of this test than the LC (least consistent) subjects rules out the possibility of attributing the violations to inconsistency. Although this logic clearly holds for axiom tests that involve only a single antecedent condition, it needs qualification for more complex tests. There a higher degree of inconsistency leads to a higher rate of apparent violations due to inconsistencies of the response function, but this may be more than offset by apparent confirmations, that is, judgments that violate the test but appear as confirmations because of inconsistencies of the response function. This appendix shows that in the case of the triple cancellation test, the percentage of violations due to inconsistency can be less, equal, or greater for the MC group than for the LC group depending on the proportions of violation, confirmation, and compatibility in the data without the superimposed random error.

Let p' and p'' be the probability of reversing the direction of an inequality due to inconsistency for the MC and the LC group, respectively (i.e., $p' < p''$). Let V, C, and M be the proportion of violations, confirmations, and compatibilities, respectively, in the error-free data. Assume that V, C, and M have the same values for MC and LC subjects. Let V', C', and M' (V'', C'', and M'') be the proportions of violations, confirmations, and compatibilities, respectively, observed in the MC group (LC group). It will be shown that the relationship between V' and V'' for fixed values of p' and p'' is a function of V, C, and M.

By considering the various opportunities for the reversal of inequalities due to inconsistency in the antecedents and the conclusion of the triple cancellation test it can easily be verified that

$$V' = (1 - 4p' + 6p'^2 - 3p'^3)V + (p' + p'^3)C + 2p'^2M \quad \text{(A1)}$$

$$C' = (1 - 4p' - p'^3)C + (4p' + 3p'^3)V + (3p' + p'^3)M \quad \text{(A2)}$$

$$M' = (1 - 3p' - 2p'^2 - p'^3)M + 6p'^2V + 3p'C. \quad \text{(A3)}$$

The values for V'', C'', and M'' are obtained by substituting p'' for p' in equations A1 to A3.

Thus, V', the observed proportion of violations in the MC group will be greater than V'', the observed proportion in the LC group, $V' > V''$

$$\text{iff } (1 - 4p' + 6p'^2 - 3p'^3)V + (p' + p'^3)C + 2p'^2M > (1 - 4p'' + 6p''^2 - 3p''^3)V + (p'' + p''^3)C + 2p''^2M$$

or, equivalently

$$\text{iff } V > \frac{(p'' + p''^3 - p' - p'^3)C + (p''^2 - p'^2)2M}{(4p'' - 6p''^2 + 3p''^3 - 4p' + 6p'^2 - 3p'^3)}. \quad \text{(A4)}$$

The conditions for $V' < V''$ and $V' = V''$ are obtained by substituting $<$ or $=$ for $>$ in Equation A4.

For Coombs and Lehner's data, $C = 2M$ (approximately). This means that $V' > V''$ iff

$$V > \frac{(p'' + p''^2 + p''^3) - (p' + p'^2 + p'^3)}{(4p'' - 6p''^2 + 3p''^3 - 4p' + 6p'^2 - 3p'^3)}\,C.$$

For $p' = .05$ and $p'' = .20$, for example, $V' > V''$ iff $V > (.189/.3985)C = 0.47C$. It appears from Coombs and Lehner's data that $V = (1/2)C$ (approximately) so that the fact MC subjects show a higher proportion of violations than LC subjects is in fact quite consistent with an equal proportion of violations due to model inadequacy (V) in the two groups.

Further evidence for this possibility comes from considering the relationship between M' and M'' as well as C' and C'' in the case where $V = M = (1/2)C$. In this case,

$$M' = (1 + 3p' + 4p'^2 - p'^3)C$$

and

$$M'' = (1 + 3p'' + 4p''^2 - p''^3)C.$$

Because $p' < p''$, it is clear that $M' < M''$. This is true for the data of Coombs and Lehner.

Similarly,

$$C' = (2 - p' + 2p'^3)C$$

and

$$C'' = (2 - p'' + 2p''^3)C.$$

Because $p' < p''$, $C' > C''$, which is also true (at least for Matrix 1) for the Coombs and Lehner data.

Equation A1 makes it clear that V' will always be less than V'' when there are no violations due to model inadequacy (i.e., when $V = 0$). This does not imply that all violations observed are due to inconsistency when $V' < V''$. We also demonstrated above that when V is not equal to zero, any relationship between V' and V'' is possible. Thus Coombs and Lehner's strategy in its present form is insufficient for distinguishing between systematic and random error.

Received August 1, 1983
Revision received September 27, 1983 ■

Roediger Appointed Editor, 1985–1990

The Publications and Communications Board of the American Psychological Association announces the appointment of Henry L. Roediger III, Purdue University, as Editor of the *Journal of Experimental Psychology: Learning, Memory, and Cognition* for a 6-year term beginning in 1985. As of February 1, 1984, manuscripts should be directed to

Henry L. Roediger III
Department of Psychology
Purdue University
West Lafayette, Indiana 47907