

# Scaling Distortion in Numerical Conjoint Measurement

Carol A. Nickerson and Gary H. McClelland  
University of Colorado

Proponents of numerical conjoint measurement generally assume that the technique's goodness-of-fit measure will detect an inappropriate composition rule or the presence of random response error. In this paper a number of hypothetical and real preference rank orderings are analyzed using both axiomatic conjoint measurement and numerical conjoint measurement to demonstrate that this assumption is not warranted and may result in a distorted scaling.

Axiomatic conjoint measurement and numerical conjoint measurement (often called conjoint analysis) have been described as two complementary approaches to the study of composition rules in psychology (Green & Wind, 1973; Krantz & Tversky, 1971). Axiomatic conjoint measurement begins with an ordering of the dependent variable and tests the properties or axioms that this ordering must satisfy if it is to be represented numerically according to a proposed composition rule. In contrast, numerical conjoint measurement does not test the ordinal properties of the data but instead searches (usually through an iterative computer algorithm) for an appropriate monotonic transformation of the dependent variable that best satisfies an assumed composition rule; then it evaluates the correspondence between the assumed rule and the scaled data with a goodness-of-fit (or badness-of-fit) measure. The functions of axiomatic conjoint measurement

and numerical conjoint measurement have been viewed as complementary because the former is primarily concerned with model testing and the latter focuses on scaling. Despite the emphasis on scaling, it is usually assumed in numerical conjoint measurement that the goodness-of-fit measure will detect an inappropriate composition rule or the presence of random response error. The present paper demonstrates that such an assumption is not warranted and may result in a distorted scaling.

Few papers exist that either compare the abilities of axiomatic conjoint measurement and numerical conjoint measurement to diagnose an appropriate composition rule or to assess the accuracy of the scaling produced by numerical conjoint measurement when the assumed composition rule is invalid or when there is error in the data. This is no doubt due to the widely held view that the two approaches are complementary. Krantz and Tversky (1971), proponents of axiomatic conjoint measurement, acknowledged this complementarity but argued (1) that the ordinal properties of the axiomatic approach allow a more powerful diagnostic test between alternative composition rules than does the goodness-of-fit criterion of the numerical approach; and (2) that although the goodness-of-fit criterion may indicate that a composition rule is invalid, it is unlikely to reveal the source of the discrepancy, whereas a direct test of the critical axioms may often do so. They did not, however, present any empirical evidence in support of their position. Such evidence was provided in a more general context

---

*APPLIED PSYCHOLOGICAL MEASUREMENT*

*Vol. 8, No. 2, Spring 1984, pp. 183-198*

© Copyright 1984 Applied Psychological Measurement Inc.  
0146-6216/84/020183-16\$2.05

by studies (Anderson & Shanteau, 1977; Birnbaum, 1973; Shanteau, 1977; Zeleny, 1976) which have decried the use of goodness-of-fit measures, particularly correlation coefficients, as indices of model fit. These studies have demonstrated rather compellingly that correlations between theoretical or mathematical models and simulated (error-free) or real (containing error) data can be very high, even when the model is incorrect. In fact, the correlation of the data with an incorrect model may be higher than the correlation of the data with the correct model, indicating that correlation coefficients should not be used to compare the fits of competing models.

The opposite position was taken by Emery and Barron (1979). In a comparative study of axiomatic conjoint measurement and numerical conjoint measurement, Emery and Barron examined the ability of the two approaches to identify the simple polynomial composition rules used to generate a number of data sets. They chose to use error-free synthetic data because there is no adequate error theory for axiomatic conjoint measurement (although it should be noted that no adequate error theory exists for numerical conjoint measurement, either). On the basis of this study, Emery and Barron concluded that numerical conjoint measurement performs at least as well as axiomatic conjoint measurement, both in diagnosing the data-generating composition rule and in distinguishing between possible composition rules. This finding contradicts Krantz and Tversky (1971) in that it implies that numerical conjoint measurement allows model tests which are at least as powerful as those of axiomatic conjoint measurement.

The results obtained by Emery and Barron are limited for several reasons. First, empirically obtained data sets typically contain error, and any evaluation of the validity of a proposed composition rule should take the possibility of error into account (Krantz & Tversky, 1971). Second, Emery and Barron compared the abilities of axiomatic conjoint measurement and numerical conjoint measurement to discriminate between three alternative classes of simple polynomials: additive, distributive, and dual-distributive. With few exceptions (e.g., Coombs & Huang, 1970; Hurvich & Jame-

son, 1955; Spence, 1956; Wallsten, 1972), however, there is little evidence that distributive and dual-distributive composition rules describe many psychological processes. A more common problem is to determine whether or not a data set can be considered consistent with a proposed or assumed additive composition rule (Green & Wind, 1973, p. 104; Krantz & Tversky, 1971). Third, Emery and Barron based their conclusions on the values of two goodness-of-fit measures, though, as has been noted previously, such measures are suspect as indices of model fit. Timmermans (1980) made similar observations in a study in which he examined the ability of numerical conjoint measurement to determine which of eight composition rules best described a number of sets of real data. He concluded that numerical conjoint measurement was ineffective for this purpose, primarily because of the robustness of the goodness-of-fit measure. He cautioned that the use of a goodness-of-fit measure for testing composition rules should be supplemented by the diagnostic axiom tests developed by Krantz and Tversky (1971) and by tests of predictive validity.

Proponents of numerical conjoint measurement other than Emery and Barron seem merely to assume that the technique performs well in the presence of a small amount of random error and that the goodness-of-fit measure will detect error or an inappropriate composition rule. There is only an occasional suggestion that this might not always be so. For example, Green and Wind (1973), who have used numerical conjoint measurement extensively in marketing studies, mentioned in a footnote that the goodness-of-fit measure may not indicate that "the model being fitted represents an incorrect specification" (p. 122). This caution, however, has not usually appeared in their empirical applications of the technique. Takane (1978) observed that it is difficult to evaluate the goodness-of-fit of a model to nonmetric data within the least squares framework. Standard numerical conjoint measurement algorithms such as ADDALS (de Leeuw, Young, & Takane, 1976), MONANOVA (Kruskal, 1965), and OPSCOLES (Emery, 1978) are within the least squares framework.

Busemeyer (1980), who is not a proponent of

numerical conjoint measurement, noted that procedures that scale a dependent measure by maximizing the fit of an additive model tend to attenuate violations of additivity, especially with small factorial designs. Srinivasan and Shocker (1973) stated that in trying to minimize a badness-of-fit measure or maximize a goodness-of-fit measure, the non-metric value estimation procedures typical of numerical conjoint measurement computer algorithms often tie the scaled values of pairs of alternatives so that the number of tied pairs is likely to be larger than the number of pairs between which the individual is truly indifferent. Carroll (1973) and Kruskal (1965) also briefly noted that under certain conditions numerical conjoint measurement is likely to produce degenerate scaling solutions in which the scaled values of alternatives are tied, and that this will not be apparent from the value of the goodness-of-fit measure. As demonstrated below, the tying of alternatives can lead to a very distorted scaling.

Despite these caveats, most researchers using numerical conjoint measurement techniques to study applied problems appear to be satisfied with its performance, probably because of an interest in prediction rather than explanation. The belief is that numerical conjoint measurement provides good prediction even when the assumed composition rule is invalid, because linear compensatory composition rules can approximate the outcomes of other composition rules quite well (Dawes & Corrigan, 1974; Green & Srinivasan, 1978; Yntema & Torgerson, 1961), even though better prediction might actually be obtained using nonlinear noncompensatory composition rules (Einhorn, 1970).

The present study compares the abilities of axiomatic conjoint measurement and numerical conjoint measurement (1) to identify correctly an additive composition rule when the data contain error and (2) to reject a proposed additive composition rule when the composition rule is not in fact additive. First, a number of hypothetical rank orderings representative of those typically found in conjoint measurement research were analyzed using both the axiomatic and the numerical techniques. The hypothetical rank orderings are described as "preference" rank orderings because many appli-

cations of numerical conjoint measurement techniques are concerned with the quantification of preference. However, the problem treated here can obviously occur with other types of rank orderings. The set of preference rank orderings was meant to be illustrative rather than exhaustive; that is, it was not meant to be a systematic sampling from the set of all possible orderings. Emphasis is given to demonstrating that the goodness-of-fit measure used in numerical conjoint measurement may not indicate the presence of error or an inappropriate composition rule, and to determining the effect that this has on the resulting scaling. Then the results of an axiomatic conjoint measurement analysis and a numerical conjoint measurement analysis performed on a collection of 100 empirically obtained data sets are compared in order to show that the hypothetical situations illustrated can occur in real data sets and that they ought to be cause for concern in empirical applications of conjoint measurement techniques.

## Analysis of the Hypothetical Data

### Data

Each hypothetical preference rank ordering consisted of a set of 27 choice alternatives formed by the factorial combination of three variables A, P, and U, each with three levels. The set of rank orderings is shown in Table 1. A rank of 27 represents the most preferred alternative; a rank of 1 represents the least preferred alternative.

### Analysis

An axiomatic conjoint measurement analysis and a numerical conjoint measurement analysis were performed on each hypothetical preference rank ordering. The axiomatic analysis consisted of testing the axioms of independence, joint independence, and double cancellation for the variables A, P, and U. A rank ordering must satisfy these three axioms to be consistent with an additive composition rule. *Independence* means that the ordering of the levels of each variable must be invariant over the levels of each of the other two variables. *Joint independence* means that the ordering of every

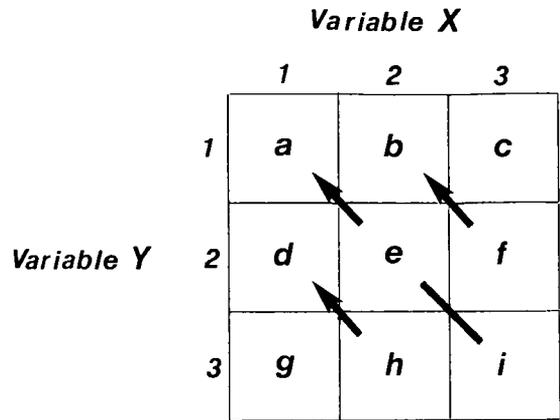
Table 1  
Hypothetical Preference Rank Orderings

	U1			U2			U3		
	A1	A2	A3	A1	A2	A3	A1	A2	A3
(a) Adjacent Pair Reversals									
P1	9	8	7	6	5	4	3	2	1
P2	27	26	25	23	24	22	15	14	13
P3	21	20	19	18	17	16	11	12	10
(b) Outlier									
P1	9	8	7	6	5	4	3	2	1
P2	27	26	25	24	23	22	14	13	12
P3	20	19	18	17	16	15	11	10	21
(c) Conjunctive/Additive									
P1	15	14	13	12	11	10	1	7	9
P2	27	25	23	21	19	17	3	8	2
P3	26	24	22	20	18	16	6	4	5
(d) Interaction 1									
P1	9	8	7	27	26	25	21	19	17
P2	6	5	4	24	23	22	20	18	16
P3	3	2	1	15	14	13	12	11	10
(e) Interaction 2									
P1	27	26	25	24	23	22	21	20	19
P2	16	17	18	13	14	15	10	11	12
P3	9	8	7	6	5	4	3	2	1
(f) One Inessential Variable									
P1	8	9	7	6	5	4	1	2	3
P2	25	26	27	24	23	22	14	13	15
P3	20	19	21	16	17	18	11	12	10
(g) Two Inessential Variables									
P1	27	22	21	17	16	18	9	4	6
P2	24	25	20	13	15	10	3	7	5
P3	19	23	26	12	11	14	1	2	8
(h) Additive									
P1	9	7	4	8	5	2	6	3	1
P2	27	25	22	26	23	20	24	21	19
P3	18	16	13	17	14	11	15	12	10

combination of the levels of any two of the three variables must be invariant over the levels of the third variable. *Double cancellation* is best illustrated by an example. Suppose that the level of any one of the three variables is held constant, and that the cells of the resulting  $3 \times 3$  factorial combination of the other two variables are labeled as shown in Figure 1. If the rank of cell *h* is greater than or equal to the rank of cell *d*, and if the rank of cell *f* is greater than or equal to the rank of cell *b*, then the rank of cell *i* must be greater than or equal to the rank of cell *a* if double cancellation is to be satisfied. For a rank ordering to be consistent with an additive composition rule, double cancellation must be satisfied for all possible combinations of two of the three variables at every level of the third variable. (See Coombs, Dawes, & Tversky, 1970; Krantz, Luce, Suppes, & Tversky, 1971; or Krantz & Tversky, 1971 for mathematical formulations and illustrations of these three axioms.) The axioms were tested using the computer program PCJM2 (Ulrich, Cummins, & Walkenbach, 1978).

For purposes of this study, numerical conjoint measurement was equated with the MONANOVA (Kruskal, 1965) computer program, although other programs exist (e.g., ADDALS, de Leeuw et al., 1976; OPSCOLES, Emery, 1978; POLYCON, Young, 1972; UNICON, Roskam, 1974). MONANOVA was chosen because it is the most widely used algorithm for numerical conjoint measurement, and because it was the algorithm used by Emery and Barron (1979). MONANOVA performs an iterative search for a numerical estimate or scale value for each level of each variable so that when the appropriate scale values are added together to obtain the scaled values of the input data, the rank ordering of the scaled data points corresponds to the rank ordering of the input data points as closely as possible. The degree of difference between the two orderings is defined by a goodness-of-fit (actually a badness-of-fit) measure called "stress." A stress value of zero or nearly zero—the usual criterion of less than .10 or 10% (Green, Carmone, & Wind, 1972) was adopted here—was interpreted to mean that the orderings of the original data and the scaled data are so close that an additive

Figure 1  
The Double Cancellation Axiom



composition rule can be assumed without appreciable loss in predictive accuracy. In the analysis of the hypothetical rank orderings, MONANOVA was allowed to iterate until stress equaled zero or until a maximum of 50 iterations was reached. Output from MONANOVA includes the stress value reported as a percentage, a scatterplot of the input data points plotted against the scaled data points, and a set of numerical estimates or scale values, one for each level of the variables A, P, and U. The range of the scale values for each variable divided by the range of scale values across all variables constitutes the relative importance of that variable. The relative importance percentages of the three variables therefore sum to 100%.

**Results**

Examples of these commonly occurring types of preference rank orderings are analyzed below: (1) the rank ordering contains random response error, (2) the rank ordering contains one or more outliers, (3) the rank ordering is not consistent with an additive composition rule, and (4) one or more variables are inessential. Comparison of the results of the axiomatic and the numerical conjoint measurement analyses is undertaken from two perspectives: (1) a strict or mathematical perspective, and (2) a more practical, psychological perspective.

*Random response error.* One type of random response error consists of a pair of adjacent choice

alternatives whose ordering is reversed from the ordering expected if an additive composition rule is assumed (Coombs, Coombs, & McClelland, 1975). For example, in the preference rank ordering shown in Table 1a, the alternatives with ranks 23 and 24 and with ranks 11 and 12 are reversed. These adjacent pair reversals cause violations of independence and joint independence for axiom tests involving variable A. In the strict sense, the proposed additive composition rule must be rejected because no adequate error theory has been developed for axiomatic conjoint measurement performed on single-replication data sets. In contrast, the stress value computed by MONANOVA for this rank ordering equals zero despite the errors, incorrectly indicating that the data set is perfectly consistent with an additive composition rule. From a mathematical perspective, then, the performance of axiomatic conjoint measurement is superior to that of numerical conjoint measurement because it is able to detect the presence of the adjacent pair reversals.

From a psychological perspective, the practice of rejecting the proposed composition rule in the presence of any axiom violations is unrealistically harsh, for there is no allowance for chance variation. In the presence of even a single pair reversal, it must be concluded that the rank ordering is not consistent with the proposed additive composition rule. However, a researcher would probably not want to conclude that the psychological process generating the rank ordering is therefore not additive. Krantz and Tversky (1971) stated that because real-world data are typically fallible, any evaluation of the validity of the proposed composition rule must take the possibility of error into account. Accordingly, it seems reasonable to adopt the strategy used by others (Coombs & Bowen, 1971; Coombs et al., 1975; Fischer, 1976; Nickerson, McClelland, & Kegeles, 1981; Ullrich & Painter, 1974) of allowing a small, prespecified number of nonsystematic errors before rejecting the proposed composition rule. Since correction of only two errors in this data set (i.e., reversing to their expected ordering the alternatives with ranks 23 and 24 and with ranks 11 and 12) eliminates all axiom violations, the rank ordering can be consid-

ered consistent with an additive composition rule. From a psychological perspective, then, axiomatic conjoint measurement and numerical conjoint measurement performed about equally well in identifying the proposed additive composition rule in the presence of one or a few adjacent pair reversals.

Although the relative importance percentages of the three variables computed for the original data set ( $A = 5\%$ ,  $P = 60\%$ ,  $U = 35\%$ ) and for the corrected data set ( $A = 2\%$ ,  $P = 60\%$ ,  $U = 38\%$ ) are nearly identical, examination of the scale values presented in Table 2a for these two data sets shows some very mild distortion in the scaling produced by numerical conjoint measurement. Most importantly, the scale values for variable A in the original data set imply that there is no preference ordering across the first and second levels of A, although examination of that data set (Table 1a) shows that in 7 of the 9 orderings across A in which the levels of P and U are fixed the first level of A is preferred to the second. In the remaining two orderings, which contain the adjacent pair reversals, the second level of A is preferred to the first. The scaling produced on the corrected data set more accurately reflects this preference pattern, since the scale values for the first and second levels of A indicate that the first is preferred to the second.

*Outliers.* An outlier is a choice alternative that is quite far from where it is expected to be in a rank ordering. The alternative with rank 21 shown in Table 1b is an example of an outlier. Outliers are often due to random response error. For example, in many preference studies, respondents are asked to rank order a set of choice alternatives, each of which is printed on a separate card. The accidental misplacement of a card can produce an outlier such as that shown in the data set in Table 1b. An axiomatic conjoint measurement analysis of this data set shows a substantial number of violations of independence and joint independence. From a mathematical perspective, the data are not consistent with the proposed additive composition rule. However, because deletion of only the alternative with rank 21 eliminates all axiom violations, it would not be unreasonable to consider the psychological process underlying the preference ordering to be consistent with an additive rule.

Table 2  
Scale Values Produced by MONANOVA for the  
Hypothetical Data Sets of Table 1

Variable	Original Ordering				Corrected Ordering			
	Level			Importance	Level			Importance
	1	2	3		1	2	3	
(a) Adjacent Pair Reversals								
A	.60	.60	.51	5%	.61	.59	.58	2%
P	.00	1.00	.70	60%	.00	1.00	.77	60%
U	.82	.64	.24	35%	.85	.71	.21	38%
(b) Outlier								
A	.67	.67	.67	0%	.62	.55	.52	5%
P	.00	1.00	1.00	100%	.00	1.00	.69	59%
U	.67	.67	.67	0%	.83	.64	.22	36%
(c) Conjunctive/Additive								
A	.67	.67	.67	0%	.66	.57	.48	11%
P	.67	.67	.67	0%	.25	.76	.71	30%
U	1.00	1.00	.00	100%	1.00	.71	.00	59%
(d) Interaction 1								
A	.57	.57	.57	0%	.61	.59	.58	2%
P	.73	.73	.24	33%	.85	.71	.21	38%
U	.00	1.00	.71	67%	.00	1.00	.77	60%
(e) Interaction 2								
A	.50	.50	.50	0%	.56	.50	.45	8%
P	1.00	.50	.00	75%	1.00	.50	.00	69%
U	.67	.50	.33	25%	.67	.50	.33	23%
(f) One Inessential Variable								
A	.56	.56	.56	0%	.56	.56	.56	0%
P	.00	1.00	.69	64%	.00	1.00	.69	64%
U	.81	.63	.25	36%	.81	.63	.25	36%
(g) Two Inessential Variables								
A	.49	.49	.48	0.6%	.50	.50	.50	0%
P	.49	.49	.49	0.3%	.50	.50	.50	0%
U	1.00	.46	.00	99.1%	1.00	.50	.00	100%
(h) Additive								
A	.65	.50	.35	20%	.65	.50	.35	20%
P	.00	1.00	.50	68%	.00	1.00	.50	68%
U	.59	.50	.41	12%	.59	.50	.41	12%

MONANOVA yields a stress value of zero for the rank ordering which contains an outlier (Table 1b), incorrectly suggesting that the data are perfectly consistent with an additive composition rule. In a mathematical sense, axiomatic conjoint measurement is superior to numerical conjoint measurement because the latter failed to detect the presence of the outlier. From a psychological perspective, however, both approaches appear to perform equally well in that they both diagnose the composition rule as additive if a small number of errors is allowed. However, comparison of the relative importance percentages of the three variables computed for the data sets with the outlier present ( $A = 0\%$ ,  $P = 100\%$ ,  $U = 0\%$ ) and with it deleted ( $A = 5\%$ ,  $P = 59\%$ ,  $U = 36\%$ ) shows that although numerical conjoint measurement is able to identify the composition rule as additive, it does so at the expense of producing a very distorted scaling. Specifically, the scale values for the three levels of variable A and the three levels of variable U are identical, with the result that the relative importance percentages of A and U are computed to be zero. This implies that there are no consistent preferences across the levels of A and U and that these variables have no effect on the rank ordering of the alternatives. Examination of the data in Table 1b, however, shows that there are indeed consistent preferences across the levels of A and U and that these variables have a definite impact on the rank ordering.

*Nonadditive composition rules.* One type of nonadditive composition rule is the conjunctive/additive composition rule. In a rank ordering consistent with a conjunctive/additive rule, the choice alternatives characterized by the least desirable level (or levels) of one or more of the variables are randomly ordered, whereas the ordering of the remaining alternatives is consistent with an additive composition rule. This type of ordering might be obtained in the situation where one level of a variable is completely unacceptable so that all alternatives described by that level are immediately eliminated from further consideration, regardless of the levels of the other variables describing those alternatives. An example of such a rank ordering is shown in Table 1c: the 9 least-preferred alter-

natives, characterized by the third level of the variable U, are randomly ordered, whereas the remaining 18 are consistent with an additive composition rule. This data set is one example of the type of data noted by Carroll (1973) and by Kruskal (1965) as producing a degenerate scaling. An axiomatic conjoint measurement analysis of this rank ordering shows large numbers of axiom violations, all of which disappear if the 9 least-preferred alternatives are deleted or are reordered to be consistent with the remainder of the rank ordering. Thus, axiomatic conjoint measurement correctly rejected the proposed additive composition rule and, because it pinpointed the source of the axiom violations, suggested the correct composition rule (Krantz & Tversky, 1971).

In contrast, the MONANOVA stress value for the rank ordering in Table 1c equals zero, incorrectly implying that the ordering is perfectly consistent with the proposed additive composition rule. Since stress is zero, it is unlikely that the user of MONANOVA would examine the data set and discover that the composition rule is not in fact additive. The scaling would no doubt therefore be accepted as accurate. However, a comparison of the relative importance percentages computed on the original rank ordering ( $A = 0\%$ ,  $P = 0\%$ ,  $U = 100\%$ ) and on the rank ordering in which the 9 least-preferred alternatives have been reordered to be consistent with the remainder of the rank ordering ( $A = 11\%$ ,  $P = 30\%$ ,  $U = 59\%$ ), shows that the scaling is seriously distorted. The identical scale values for the three levels of variable A and variable P and the resulting relative importance percentages of zero for A and P in the scaling for the uncorrected data set (Table 2c) imply that there are no consistent preferences over the levels of A and P, and that A and P have no effect on the ordering of the alternatives. It is clear from examination of the data set in Table 1c, however, that at the first and second levels of U, there are perfectly consistent preferences across the levels of A and P and that A and P have a definite effect on the ordering of the 18 most-preferred alternatives. The conclusion is that when the composition rule is conjunctive/additive, numerical conjoint measurement can inaccurately identify the com-

position rule as additive and produce a distorted scaling, whereas axiomatic conjoint measurement correctly rejects the proposed composition rule.

A second type of nonadditive composition rule often found is a composition rule characterized by interaction between the independent variables. These interactions can take many forms; two illustrative examples are shown in Tables 1d and 1e. In the preference rank ordering shown in Table 1d, the rank ordering of the 9 choice alternatives at the first and at the second level of variable U is A1P1, A2P1, A3P1, A1P2, A2P2, A3P2, A1P3, A2P3, A3P3. At the third level of U, however, the ordering is A1P1, A1P2, A2P1, A2P2, A3P1, A3P2, A1P3, A2P3, A3P3. For a composition rule to be additive, the ordering of the alternatives described by all possible combinations of A and P must be identical at each level of U; the systematic deviation of the ordering at the third level of U represents an interaction. An axiomatic conjoint measurement analysis of this rank ordering shows (1) no violations of independence or double cancellation for any of the three variables, and (2) a small number of violations of joint independence showing that the combination of variables A and P is not independent of variable U. The conclusion is that the rank ordering is not consistent with an additive composition rule. The MONANOVA analysis, however, yields a stress value of zero, incorrectly implying that the rank ordering is perfectly consistent with an additive rule.

Rank orderings with true interactions cannot be scaled with MONANOVA, because the data are not consistent with an additive composition rule. The regularity of the interaction in Table 1d suggests that it is a true interaction, rather than one caused by random response error. It is worthwhile noting, however, that the correction of only three errors (i.e., converting ranks 19, 17, 20, 18 to ranks 20, 19, 18, 17 by performing a series of three adjacent pair reversals) renders the data set consistent with an additive composition rule. The strategy of allowing the correction of a small number of nonsystematic errors before rejecting the proposed additive composition rule has been suggested as a means for dealing with error. Although the errors in the rank ordering are systematic, there are

very few of them, so that it seems worthwhile to compare the MONANOVA scaling for the original data set with the scaling for the data set in which the interaction has been removed by performing the three adjacent pair reversals. These scalings, as shown in Table 2d, are not very discrepant. The relative importance percentages for A, P, and U in the uncorrected data set are 0%, 33% and 67%, respectively, while in the corrected data set they are 2%, 38%, and 60%. Most of the distortion occurs in the first, second, and third levels of A and the first and second levels of P. The scale values for these levels indicate there are no preferences across the first, second, and third levels of A and the first and second levels of P, although the data show that there are.

A different type of interaction is shown in Table 1e. At every combination of the first and the third level of P with all levels of U, the first level of A is preferred to the second, which in turn is preferred to the third. At every combination of the second level of P with all levels of U, however, the third level of A is preferred to the second, which in turn is preferred to the first. This preference reversal represents an interaction. An axiomatic conjoint measurement analysis of the rank ordering in Table 1e shows a large number of axiom violations for tests of the independence of A and for tests of the joint independence of A and U from P, leading to the conclusion that the rank ordering is not consistent with an additive composition rule.

The MONANOVA analysis, however, yields a stress value of zero, incorrectly implying that the data are perfectly consistent with an additive composition rule. Nine adjacent pair reversals would be required to remove this interaction, so that it would probably be unreasonable to conclude that this ordering is consistent with an additive composition rule except for random error. Nevertheless, it is interesting to compare the MONANOVA scalings for the original rank ordering and for the rank ordering in which the interaction has been removed (i.e., by reversing the nine adjacent pairs to convert ranks 16, 17, 18, 13, 14, 15, 10, 11, 12 to 18, 17, 16, 15, 14, 13, 12, 11, 10). The two scalings, shown in Table 2e, are similar. The relative importance percentages of A, P, and U for

the data set with interaction are 0%, 75%, and 25%, respectively, and for the corrected data set they are 8%, 69% and 23%. The distortion present occurs in the first, second, and third levels of A. The identical scale values for these levels indicate that there is no preference across these levels. However, in the original data set, the first level of A is preferred to the second, which in turn is preferred to the third in six of the nine orderings of these three levels. As A is the least important variable, this distortion is not too serious. The important point to note, however, is that the axiomatic conjoint measurement analysis indicates the presence of the interaction, whereas the numerical conjoint measurement analysis does not. The inability of numerical conjoint measurement to detect an interactive composition rule has been noted before (Busemeyer, 1980; Carmone, Green, & Jain, 1978; Green & Wind, 1973; Nygren, 1980).

*Inessential variables.* If an individual has no preference across the levels of one of the three variables, the ordering of the levels of that variable will be random. In the data set shown in Table 1f, there is no preference across the levels of variable A. A variable with levels to which an individual is indifferent is termed an "inessential" variable. (See Coombs et al., 1975, p. 285, and Krantz et al., 1971, p. 256, for formal definitions of essential and inessential variables.) It is also possible for two of the three variables to be inessential. In this situation, the rank ordering shows the set of 27 choice alternatives divided into three subsets, each characterized by one level of the important (essential) variable, with random ordering within the subsets. In the rank ordering shown in Table 1g the variables A and P are both inessential.

An axiomatic conjoint measurement analysis of the rank ordering in Table 1f, in which the variable A is inessential, shows violations of independence and joint independence for axiom tests involving A. If A is eliminated—either by assigning the same rank to each set of three alternatives described by the same levels of P and U and different levels of A, or by removing A from the design and reducing the set of alternatives to nine—all axiom violations disappear. The rank ordering thus corrected is then perfectly consistent with an additive composition rule over the two variables P and U.

An axiomatic conjoint measurement analysis of the rank ordering in Table 1g, in which two variables are inessential, shows violations of independence for variables A and P, a violation of double cancellation, and violations of joint independence for all three variables. If A and P are eliminated—either by assigning the same rank to each set of nine alternatives characterized by the same level of U, or by deleting A and P from the design and reducing the set of alternatives to three—all axiom violations disappear. The corrected rank ordering is then consistent with a one-variable composition rule. (One-variable composition rules cannot be additive in the strict sense; two or more variables are required for additivity.) Thus, in situations where a variable or variables are inessential, axiomatic conjoint measurement correctly rejects the proposed additive composition rule in three variables. The pattern of axiom violations suggests the reason for this rejection and results in identification of the correct composition rule.

MONANOVA performed on the rank ordering with one inessential variable (Table 1f) and on the corrected rank ordering, in which the same rank is assigned to each set of three choice alternatives described by the same levels of P and U and different levels of A, yields stress values of zero and identical scale values and relative importance percentages ( $A = 0\%$ ,  $P = 64\%$ ,  $U = 36\%$ ) for all three variables (Table 2f). A MONANOVA analysis of the rank ordering with two inessential variables (Table 1g) yields a stress value of .6% and relative importance percentages of .6%, .3%, and 99.1% for variables A, P, and U, respectively (Table 2g). MONANOVA performed on the corrected rank ordering, in which the same rank is assigned to each set of nine alternatives characterized by the same level of U, yields a stress value of zero and relative importance percentages of 0%, 0%, and 100% for variables A, P, and U, respectively. Comparisons of the MONANOVA results for the original and corrected rank orderings of the alternatives suggest that numerical conjoint measurement performs very well in situations in which one or two of the variables are inessential, because it identifies the composition rule as additive and provides a nearly accurate scaling. A completely accurate scaling of a rank ordering in which one or

two of the variables are inessential can be obtained by assigning the same rank to all levels of the inessential variable or variables.

### Analysis of the Empirical Data

#### Data

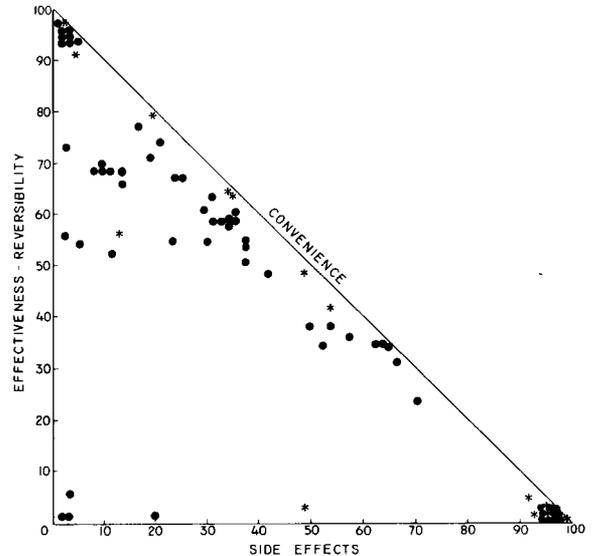
The data for this analysis consisted of a collection of 100 empirically obtained preference rank orderings of a set of 27 hypothetical contraceptive methods formed by the factorial combination of three variables, each with three levels. The three variables and their levels were (1) convenience (very convenient, somewhat convenient, inconvenient), (2) effectiveness-reversibility (99% effective and irreversible, 95% effective and reversible, 65 to 85% effective and reversible), and (3) side effects (rare, minor, major). Each choice alternative was printed on a card and a preference rank ordering was obtained by instructing an individual to order the cards from "most preferred contraceptive" to "least preferred contraceptive". (See Nickerson et al., 1981, for a complete description of the study from which these data sets are taken.)

#### Analysis

A MONANOVA analysis was performed on each of the individual preference rank orderings. In order to be considered consistent with an additive composition rule, the obtained stress value had to be less than the generally accepted criterion of .10 or 10% (Green et al., 1972). It was found, however, that all stress values were either less than 2% or greater than or equal to 14%, so that the actual criterion value was 2%. According to this criterion, 74 of the 100 preference rank orderings were judged consistent with an additive composition rule.

The pattern of the relative importance percentages of the three variables for each of the 74 preference rank orderings judged consistent with an additive composition rule is marked by a star or point in the triangular diagram shown in Figure 2. (The meaning of the stars and points will be clarified later.) Distance along the abscissa of the diagram represents the relative importance of the side effects variable, and distance along the ordinate

**Figure 2**  
Scaling of Contraceptive Data Produced By  
Numerical Conjoint Measurement Alone



represents the relative importance of the effectiveness-reversibility variable. Distance from the diagonal line into the interior of the triangle represents the relative importance of the convenience variable because knowledge of the relative importance of two of the three variables determines the relative importance of the third.

An axiomatic conjoint measurement analysis was also performed on the 100 preference rank orderings. Three or fewer errors were allowed before rejection of the proposed additive composition rule. The analysis classified the composition rules as follows: 32 additive rules with three variables essential, 2 additive rules with two variables essential, 9 one-variable rules, 21 conjunctive/additive rules with three variables essential, and 2 conjunctive/additive rules with two variables essential. Of the remaining 34 preference rank orderings, 6 exhibited substantial regularity but contained more than three errors and 28 exhibited little or no regularity.

The 66 preference rank orderings for which composition rules could be classified by axiomatic conjoint measurement were scaled by MONANOVA,

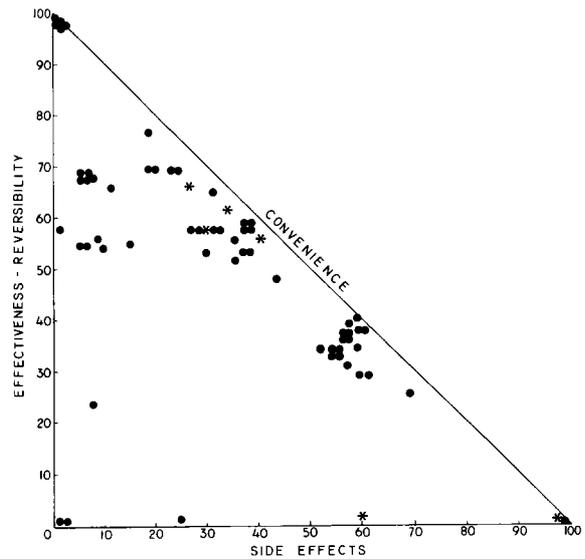
after random response errors were corrected and the appropriate adjustments were made for conjunctive/additive rules, inessential variables, etc., as described previously. The pattern of relative importance percentages of the three variables for each of the preference rank orderings is marked by a star or a point in the triangular diagram shown in Figure 3.

Comparison of Figures 2 and 3 shows that the scalings produced by axiomatic conjoint measurement and numerical conjoint measurement are quite different. For example, in Figure 3 there are only 2 relative importance percentage patterns in the lower right corner of the triangle, which represent an extremely high value for side effects, while in Figure 2 there are 15 patterns in this corner. Similarly, in Figure 3 there are 5 relative importance percentage patterns in the upper left corner of the triangle, which represent an extremely high value for effectiveness-reversibility, whereas in Figure 2 there are 10 patterns in this corner. Other less severe discrepancies between the two figures are also evident.

There were 74 data sets that the numerical conjoint measurement analysis judged consistent with an additive composition rule and 66 data sets that axiomatic conjoint measurement judged consistent with a three- or two-variable additive or conjunctive/additive composition rule or a one-variable composition rule. The points in Figures 2 and 3 represent those 60 data sets that passed both the numerical conjoint measurement and axiomatic conjoint measurement criteria (regardless of whether or not both techniques classified them as *additive*). The stars in Figure 2 represent the 14 data sets that the numerical analysis judged consistent with an additive composition rule but the axiomatic analysis was unable to classify. The stars in Figure 3 represent the 6 data sets that numerical conjoint measurement judged inconsistent with an additive composition rule but axiomatic conjoint measurement was able to classify. Discrimination is made between these different kinds of data sets in order to show that the differences in scaling in the two figures are not due solely to the different numbers of data sets in the two scalings.

The scaling in Figure 2 resulted from the use of

**Figure 3**  
Scaling of Contraceptive Data Produced By  
Axiomatic Conjoint Measurement Followed By  
Numerical Conjoint Measurement



a badness-of-fit measure, stress, as an index of fit to an assumed composition rule. The scaling in Figure 3 resulted from using critical axioms to test the fit of a proposed composition rule. The distortion of the scaling in Figure 2 is primarily due to the existence of a large number of conjunctive/additive composition rules which MONANOVA mistakenly diagnosed as additive. As was demonstrated previously, this misdiagnosis resulted in a distorted scaling. The distorted scaling in Figure 2 results in an inaccurate depiction of the contraceptive preferences of the individuals who participated in the study.

## Discussion

The comparison of the axiomatic and the numerical conjoint measurement analyses of the hypothetical preference rank orderings presented above clearly demonstrates that numerical conjoint measurement was unable to detect common violations of the assumed additive composition rule. It is to be expected that slight deviations from a rank or-

dering perfectly consistent with an additive composition rule—a few random pair reversals, an outlier, one or more inessential variables, an interaction, or a conjunctive/additive data set—will produce a stress value *near zero*. However, the stress values for the illustrated rank orderings were not usually just near zero, but exactly zero, and always below the usual criterion of 10%. Such stress values give the misleading impression that there were no important deviations from the assumed additive composition rule when in fact there were. The inability of numerical conjoint measurement to detect small deviations from an additive composition rule would not be serious if it had little or no impact on the scaling produced. Unfortunately, the scaling can be extraordinarily sensitive to such deviations. For example, when the rank ordering contains a single outlier, the MONANOVA algorithm produced an extremely distorted scaling.

Although the stress value did not detect small deviations from the additive composition rule, the scatterplot of the original data points plotted against the scaled data points may indicate that a distorted scaling has been produced (Carroll, 1973; Kruskal, 1965). The scatterplot for a rank ordering that contains no error and is perfectly consistent with an additive composition rule, such as that in Table 1h, will be a relatively smooth monotonically increasing function similar to that shown in Figure 4h. If the scatterplot obtained is of this form, the model-testing and scaling aspects of numerical conjoint measurement are likely to be relatively valid. The scatterplot for the rank ordering with adjacent pair reversals (Table 1a) shown in Figure 4a is a function of this type. As noted above, the existence of one or a few pair reversals has little practical impact.

If, on the other hand, the scaling produces a step function, it should be suspected that either the model-testing or the scaling aspect of numerical conjoint measurement is seriously in error. The scatterplots in Figures 4b through 4g are all step functions, indicating that the MONANOVA results should be viewed with suspicion. Step functions—also called “degenerate solutions”—occur because, in trying to minimize a badness-of-fit measure, MONANOVA and other nonmetric scaling programs tend

to tie the scaled values of pairs of alternatives in order to attenuate violations of additivity (Srinivasan & Shocker, 1973). It is important to note, however, that the presence of a step function can only suggest the possibility of a distorted scaling; it does not indicate the cause of the distorted scaling or even that a distorted scaling has occurred. For example, the scatterplots in Figures 4b and 4c are identical, but the one in Figure 4b is produced by an outlier and the one in Figure 4c is caused by a conjunctive/additive composition rule. The scatterplots in Figures 4e and 4f are also quite similar, but the one in Figure 4e is caused by an interactive composition rule and the one in Figure 4f is caused by an inessential variable.

The existence of error and nonadditive composition rules in empirically obtained data sets (in the contraception study discussed above only 11 of the 100 preference rank orderings were error-free and perfectly consistent with an additive composition rule) and the possibility of distorted MONANOVA scalings in the presence of these situations strongly suggest that numerical conjoint measurement ought *not* to be used routinely for model testing. Axiomatic conjoint measurement and numerical conjoint measurement should be viewed not as complementary but as sequential analyses, with axiomatic conjoint measurement providing the model testing and numerical conjoint measurement providing the scaling. The recommended course of analysis:

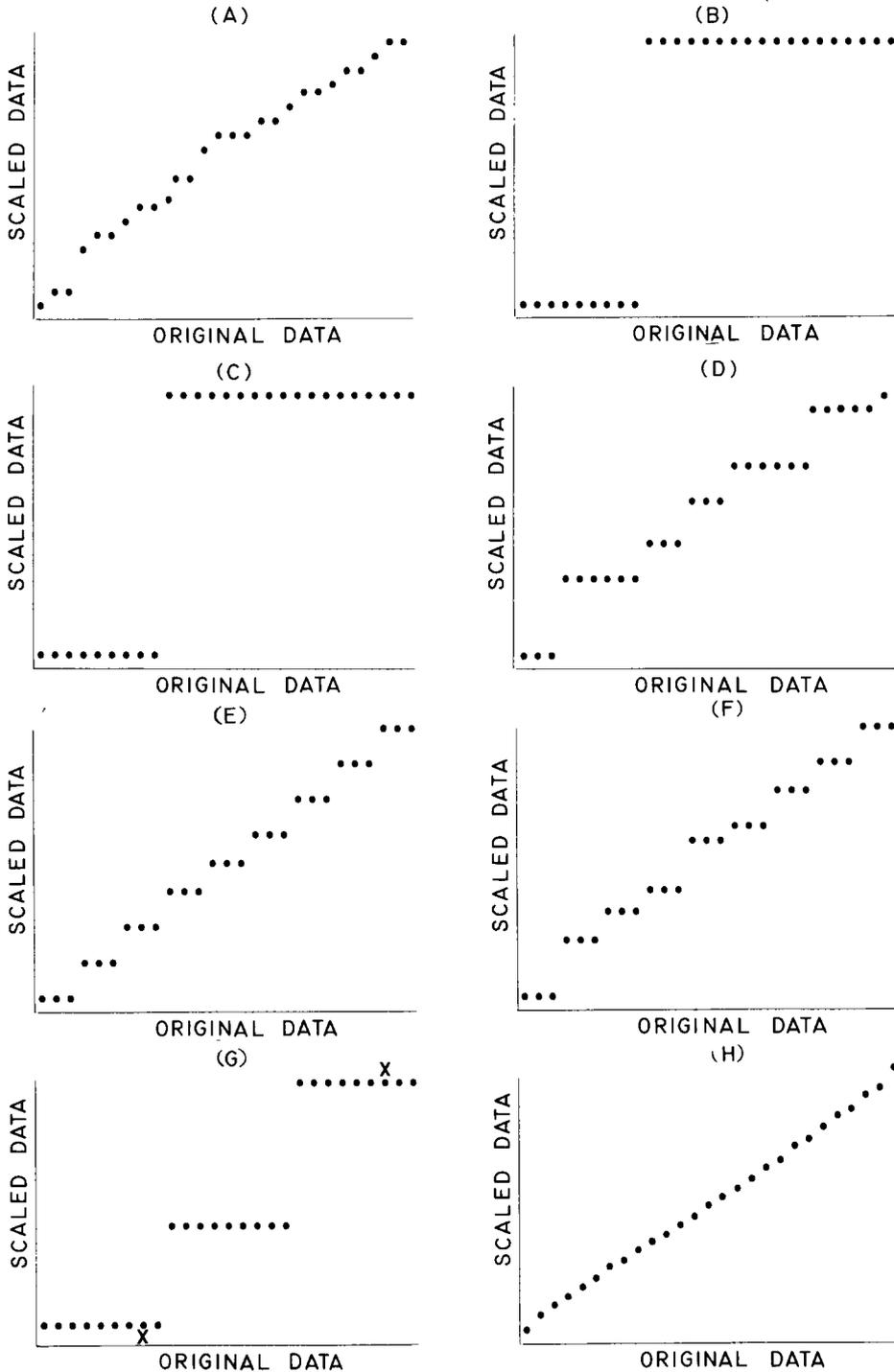
1. Axiomatic conjoint measurement is used first to locate the source of axiom violations, if any, and to identify the composition rule.
2. The data are corrected, if possible, on the basis of the axiomatic conjoint measurement analysis, allowing some small prespecified number of nonsystematic errors to be corrected, deletion of inessential variables, etc.
3. The corrected data are input to a numerical conjoint measurement program such as MONANOVA (Kruskal, 1965) or ORDMET (Lehner & Noma, 1980; McClelland & Coombs, 1975) to produce the scaling. Noncorrectable data sets should not be scaled.

## References

Anderson, N. H., & Shanteau, J. C. (1977). Weak in-

**Figure 4**

Scattergrams of the Original Data Plotted Against the Scaled Data for the Hypothetical Preference Rank Orderings (An x Represents Deviation from the Best-Fitting Scaling)



- ference with linear models. *Psychological Bulletin*, 84, 1155–1170.
- Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, 79, 239–242.
- Busemeyer, J. R. (1980). Importance of measurement theory, error theory, and experimental design for testing the significance of interactions. *Psychological Bulletin*, 88, 237–244.
- Carmone, F. J., Green, P. E., & Jain, A. K. (1978). Robustness of conjoint analysis: Some Monte Carlo results. *Journal of Marketing Research*, 15, 300–303.
- Carroll, J. D. (1973). Appendix B: Models and algorithms for multidimensional scaling, conjoint measurement, and related techniques. In P. E. Green & Y. Wind, *Multiattribute decisions in marketing: A measurement approach* (pp. 299–371). Hinsdale IL: Dryden Press.
- Coombs, C. H., & Bowen, J. N. (1971). A test of VE-theories of risk and the effect of the central limit theorem. *Acta Psychologica*, 35, 15–28.
- Coombs, C. H., Coombs, L. C., & McClelland, G. H. (1975). Preference scales for number and sex of children. *Population Studies*, 29, 273–298.
- Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs NJ: Prentice Hall.
- Coombs, C. H., & Huang, L. C. (1970). Polynomial psychophysics of risk. *Journal of Mathematical Psychology*, 7, 317–338.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95–106.
- de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, 471–503.
- Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 73, 221–230.
- Emery, D. R. (1978). *Optimal scaling via ordinary least squares* (Working Paper No. WP-31-78). Alberta, Canada: University of Calgary, Faculty of Business.
- Emery, D. R., & Barron, F. H. (1979). Axiomatic and numerical conjoint measurement: An evaluation of diagnostic efficacy. *Psychometrika*, 44, 195–210.
- Fischer, G. W. (1976). Multidimensional utility models for risky and riskless choice. *Organizational Behavior and Human Performance*, 17, 127–146.
- Green, P. E., Carmone, F. J., & Wind, Y. (1972). Subjective evaluation models and conjoint measurement. *Behavioral Science*, 17, 288–299.
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5, 103–123.
- Green, P. E., & Wind, Y. (1973). *Multiattribute decisions in marketing; A measurement approach*. Hinsdale IL: Dryden Press.
- Hurvich, L. M., & Jameson, D. (1955). Some quantitative aspects of an opponent-colors theory: II. Brightness, saturation, and hue in normal and dichromatic vision. *Journal of the Optical Society of America*, 45, 602–616.
- 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.
- Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B*, 27B, 251–263.
- Lehner, P., & Noma, E. (1980). A new solution to the problem of finding all numerical solutions to ordered metric structures. *Psychometrika*, 45, 135–138.
- McClelland, G. H., & Coombs, C. H. (1975). ORD-MET: A general algorithm for constructing all numerical solutions to ordered metric structures. *Psychometrika*, 40, 269–290.
- Nickerson, C. A., McClelland, G. H., & Kegeles, S. M. (1981). *A decision-making approach to contraceptive choice* (Report No. 234). Boulder CO: University of Colorado, Center for Research on Judgment and Policy.
- Nygren, T. E. (1980). Limitations of additive conjoint scaling procedures: Detecting nonadditivity when additivity is known to be violated. *Applied Psychological Measurement*, 4, 367–383.
- Roskam, E. E. (1974). *Unidimensional conjoint measurement (UNICON) for multi-faceted designs*. The Netherlands: University of Nijmegen, Psychologisch Laboratorium.
- Shanteau, J. C. (1977). Correlation as a deceiving measure of fit. *Bulletin of the Psychonomic Society*, 10, 134–136.
- Spence, K. W. (1956). *Behavior theory and conditioning*. New Haven CT: Yale University Press.
- Srinivasan, V., & Shocker, A. D. (1973). Linear programming techniques for multidimensional analysis of preferences. *Psychometrika*, 38, 337–369.
- Takane, Y. (1978). A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—Theory. *Japanese Psychological Research*, 20, 7–17.
- Timmermans, H. (1980). Unidimensional conjoint measurement models and consumer decision-making. *Area* (Publication of the Institute of British Geographers), 12, 291–300.
- Ullrich, J. R., Cummins, D. E., & Walkenbach, J. (1978). PCJM2: A program for the axiomatic conjoint mea-

- surement analysis of polynomial composition rules. *Behavior Research Methods and Instrumentation*, 10, 89–90.
- Ullrich, J. R., & Painter, J. R. (1974). A conjoint-measurement analysis of human judgment. *Organizational Behavior and Human Performance*, 12, 50–61.
- Wallsten, T. S. (1972). Conjoint-measurement framework for the study of probabilistic information processing. *Psychological Review*, 79, 245–260.
- Yntema, D. B., & Torgerson, W. S. (1961). Man-computer cooperation in decisions requiring common sense. *IRE Transactions of the Professional Group on Human Factors in Electronics*, HFE-2(1), 20–26.
- Young, F. W. (1972). A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K., Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1). New York: Academic Press.
- Zeleny, M. (1976). On the inadequacy of the regression paradigm used in the study of human judgment. *Theory and Decision*, 7, 57–65.

### Acknowledgments

*The authors gratefully acknowledge the support of this research by grants HD-10802 and HD-14403 from the National Institute of Child Health and Human Development, and by grant RR0 7013-14 from the Biomedical Research Support Grant Program, Division of Research Resources, National Institutes of Health; the technical assistance of Doreen Victor; and the many helpful substantive and editorial suggestions of Curtis Brown, Alice Healy, Robert Quinn, and two anonymous reviewers.*

### Author's Address

Send requests for reprints or further information to Gary H. McClelland, Center for Research on Judgment and Policy, Campus Box 344, University of Colorado, Boulder CO 80309–0344, U.S.A.