

An Examination of Conditional Violations of Axioms for Additive Conjoint Measurement

Thomas E. Nygren
The Ohio State University

Axiomatic conjoint measurement methodology offers a useful approach for evaluating different composition rules as potential models for fitting the components of multidimensional stimuli. The usefulness of this methodology has been somewhat hindered in applied settings because of a lack of an adequate error theory for testing the fit of data to the axioms. This paper presents the results of an attempt to provide a basis for an examination of errors of the conjoint measurement axioms. Specifically, this paper describes a means of evaluating the fit of an additive conjoint measurement model to a three-factor design. For each of the critical axioms of axiomatic conjoint measurement, the proportions of errors that would be expected by chance for different levels of satisfaction of the simple independence property are examined. The results indicate that violations of these axioms occur much less often than intuitively might be expected. Error proportion tables based on monte carlo analyses are presented to aid in comparisons with empirically obtained results. It is shown that two types of violations of the axioms can be defined and used to differentiate between systematic and unsystematic errors in fallible data.

Subjective scaling techniques are an integral part of much of social science research. In many situations it is assumed that the variable of interest in a study is a complex phenomenon that is multidimensional in nature, and that the ordering of scores produced by an individual on this variable may actually be based on the joint effects of two or

more independent variables. Often a researcher is interested in whether the composition rule by which the independent variables combine to produce the joint effect on the dependent variable is additive, and whether it might be possible to obtain numerical scale values for these independent variables themselves as well as for their resultant joint effects. That is, it is desired that the independent and dependent variables be scaled simultaneously by way of an additive composition rule in such a way that the order of the joint effects as found in the data is preserved.

Many composition rules might be hypothesized in such conjoint measurement models. The additive model is the simplest such rule and the one that pervades much of behavioral science research. Let A_1 , A_2 , and A_3 be labels for three factors and let a_1 be a level of factor A_1 , a_2 be a level of factor A_2 , and a_3 be a level of factor A_3 . It might be hypothesized that the joint effects of these three factors can be described as:

$$f(a_1, a_2, a_3) = f_1(a_1) + f_2(a_2) + f_3(a_3) \quad , \quad (1)$$

where f , f_1 , f_2 , and f_3 are separate and identifiable numerical functions.

The general additive model, like the three-factor model illustrated in Equation 1, is one of four simple polynomials in three factors that are represented in a variety of psychological theories (cf. Krantz, 1968). In addition to the additive model ($A_1 + A_2 + A_3$), there is the multiplicative model ($A_1 \times A_2 \times A_3$), the distributive model ($A_1 \times [A_2 + A_3]$), and the dual-distributive model ($A_1 + [A_2 \times A_3]$).

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 9, No. 3, September 1985, pp. 249-264
© Copyright 1985 Applied Psychological Measurement Inc.
0146-6216/85/030249-16\$2.05

Except for the notable work of Falmagne (1976), most theoretical work on these models has been in the framework of algebraic measurement theories. Krantz, Luce, Suppes, and Tversky (1971) have presented a number of diagnostic ordinal properties that are necessary though not sufficient for these four models to hold in a set of data. Despite the usefulness of these axioms for conjoint measurement methodology, the very algebraic nature of the measurement theory poses an important problem with fallible empirical data. Without a specific error theory, it is difficult to assess the fit of any of the models to data. In particular, when the axioms are violated, the crucial question for the applied researcher is how many violations are enough to cause rejection of the model.

Recent Research

Several studies have attempted to examine how well the conjoint measurement axioms could, in fact, be used to differentiate among the simple additive, distributive, and dual-distributive models. In particular, Emery and Barron (1979) were interested in the issue of misdiagnosis. They examined whether it was possible for a set of data to come from one simple polynomial model (e.g., distributive) but not be rejected as coming from a different model (e.g., additive). This could occur if the rank order associated with a set of data does not violate any of the axioms associated with either its own generating model or some other model. In such a case, the conjoint measurement axioms would be unable to reject the false model.

Emery and Barron (1979) generated 92 sets of data in three factors that perfectly fit either an additive model, a distributive model, or a dual-distributive model. They examined the axioms with the rank orders in each data set and found that, though the 20 additive data sets were diagnosed correctly, of the 36 sets from distributive models, only 23 were diagnosed as coming from a unique distributive model, and none of the 36 dual-distributive data sets were diagnosed correctly as coming from a dual-distributive model.

These results do not appear to be very encouraging for those who would like to use the conjoint

measurement axioms as diagnostic tools. However, it is important to note that the data used by Emery and Barron (1979) were error-free. Because the axioms as outlined by Krantz and Tversky (1971) are necessary but not sufficient for any of the four models, it is not surprising that Emery and Barron obtained their non-optimal results. An alternative approach that might prove to be more useful to users of conjoint measurement methodology, particularly in situations with fallible data, is to begin with a monte carlo approach to determine the likelihoods of violating the axioms by chance.

Arbuckle and Larimer (1976) used such a monte carlo approach to investigate the likelihoods associated with satisfying the conjoint measurement axioms in two-factor matrices of different sizes. In particular, they attempted to estimate the number of possible rankings in an $r \times c$ table that satisfy both independence and double cancellation, and that satisfy additivity. Although their study was extremely enlightening, it was faced with one rather difficult problem—their procedure simply was too costly, so that in many of their examples the sample sizes were small, perhaps too small to give accurate estimates of the probabilities. Nevertheless, their results seemed to indicate that as r and c increased, the probability of satisfying double cancellation or additivity by chance becomes very small. The proportion of $r \times c$ tables satisfying independence and double cancellation that are also additive decreases rapidly as r and c increase.

McClelland (1977) carried the work of Arbuckle and Larimer (1976) one step further in terms of accuracy by finding exact probabilities for those $r \times c$ tables small enough to allow for exact enumeration. In addition, McClelland's work is very interesting in that he attempted to find in greater detail some of the conditional probabilities for satisfying the additive conjoint measurement axioms. For example, the conditional probabilities of satisfying double cancellation given independence, and of satisfying additivity given independence or given independence and double cancellation, were found. As expected, results similar to those of Arbuckle and Larimer (1976) were obtained.

The Emery and Barron (1979), Arbuckle and Larimer (1976), and McClelland (1977) studies

suffered from several major limitations, however. The research presented here attempts to meet these limitations by providing further insight into the specific properties inherent in the axioms. First, the data in the present study are not error-free. The approach taken here began with completely random data and added structure to the data in several steps. Second, individual tests of the axioms were examined in detail. In particular, the conditional effects of satisfaction of perhaps the most critical axiom, simple independence, on the occurrence of violations of the other axioms were examined. Third, Arbuckle and Larimer (1976) and McClelland (1977) examined the axioms at a more global level than might be needed by the applied researcher. Their results were presented in terms of whether an axiom was or was not violated in a data set. In the present study, the emphasis was on the *proportions of violations* of individual tests of the axioms. Finally, and perhaps most important, the data sets from the Arbuckle and Larimer and McClelland studies were very small, coming from either 3×3 , 3×4 , 3×5 , or 4×4 two-factor designs. Hence, interesting relations in three factors (e.g., joint independence and distributive cancellation) were not and could not have been systematically investigated.

Method

The first four sets of values in Table 1 are examples of rankings for each of four experimental conditions, in which data were generated to represent first the random rankings of the 27 stimulus combinations from a $3 \times 3 \times 3$ design. The random or unconditional data matrix is labeled *Uncond* in Table 1. One thousand such random data sets were generated using an adaptation of the Uniform function used in the SAS statistical package (Helwig & Council, 1979). In addition, a second set of 1,000 random data sets were generated for 64 stimulus combinations in a $4 \times 4 \times 4$ design. For each of the total 2,000 data sets, the rows were next permuted so as to satisfy simple independence perfectly on the first of the three factors (factor A_1). An example of one of these modified data sets is shown in the second section of Table 1. This matrix

will be described as coming from the *Single* condition. Next, each of the 2,000 data sets were permuted so as to satisfy simple independence in two factors (factors A_1 and A_2), and finally in all three of the factors. These are the *Double* and *Triple* conditions shown in the middle two sections of Table 1, respectively.

Procedure

The testing of the axioms in these data matrices required the use of an efficient algorithm that would allow for enumeration of all possible tests of each axiom. A computer program called CMSCAL was developed as a hybrid of four widely used algorithms. Two of these are from general diagnostic programs for testing the conjoint measurement axioms, CONJOINT (Holt & Wallsten, 1974) and PCJM (Ullrich & Cummins, 1973), and the other two are nonmetric scaling procedures based on Kruskal's (1965) MONANOVA program and Johnson's (1973) monotone regression program. The axiom testing procedure in CMSCAL is based on the most useful parts of the CONJOINT and PCJM programs. It provides a more detailed analysis of violations of the axioms than does the CONJOINT program, especially for the critical axioms of simple independence and joint independence, and it employs corrected versions of some of the same very efficient algorithms introduced by Ullrich and Cummins (1973).

Results

Simple Independence

The fundamental property of independence, which can be checked separately for each of the three factors, says that:

$$\begin{aligned}
 &A_1 \text{ is independent of } A_2 \text{ and } A_3 \text{ whenever} \\
 &(a_1, a_2, a_3) > (b_1, a_2, a_3) \text{ if and only if} \\
 &(a_1, b_2, b_3) > (b_1, b_2, b_3) \quad . \quad (2)
 \end{aligned}$$

The independence of factor A_1 asserts that if $a_1 > b_1$ for some combination of levels of factors A_2 and A_3 , then this relation will hold for any other combination of levels of A_2 and A_3 .

Table 1
 Examples of 3x3x3 Data Sets for Ranked Observations

	a ₃			b ₃			c ₃		
	a ₁	b ₁	c ₁	a ₁	b ₁	c ₁	a ₁	b ₁	c ₁
Unconditional (Random)									
a ₂	21	13	24	11	16	23	15	5	26
b ₂	22	27	8	3	7	4	12	20	19
c ₂	14	2	9	10	1	17	25	6	18
Simple Independence, Factor A (Single)									
a ₂	13	21	24	11	16	23	5	15	26
b ₂	8	22	27	3	4	7	12	19	20
c ₂	2	9	14	1	10	17	6	18	25
Simple Independence, Factors A and B (Double)									
a ₂	2	9	14	1	4	7	5	15	20
b ₂	8	21	24	3	10	17	6	18	25
c ₂	13	22	27	11	16	23	12	19	26
Simple Independence, Factors A, B, and C (Triple)									
a ₂	1	4	7	2	9	14	5	15	20
b ₂	3	10	17	6	18	24	8	21	25
c ₂	11	16	23	12	19	26	13	22	27
Subject 1									
a ₂	1	8	10	3	9	18	6	14	22
b ₂	2	11	19	4	15	24	7	17	25
c ₂	5	16	23	12	20	26	13	21	27
Subject 2									
a ₂	1	10	19	3	13	22	9	16	25
b ₂	2	11	20	8	14	23	5	17	26
c ₂	7	12	21	4	15	24	6	18	27

Note: a₁, b₁, and c₁ are the levels of factor A₁; a₂, b₂, and c₂ are the levels of factor A₂; and a₃, b₃, and c₃ are the levels of factor A₃.

As was described above, an attempt was made to investigate violations of the axioms under several different conditions when the degree to which simple independence was satisfied in the data was varied. Table 2 presents the results of the extensive

search for violations of simple independence for each of the four types of data matrices for the 3 × 3 × 3 and 4 × 4 × 4 designs. Several points can be made from the reported proportions in this table.

Table 2
Observed Unconditional and Conditional Error Proportions for the
Simple Independence Axiom for 3x3x3 and 4x4x4 Designs

Analysis	Test Factors	All Failures		Dominant Failures:				Tradeoff Failures:			
				Uncond'l		Cond'l		Uncond'l		Cond'l	
				3	4	3	4	3	4	3	4
Uncond	A ₁ of A ₂ A ₃	.500	.500	.374	.350	.749	.700	.126	.150	.251	.300
	A ₂ of A ₃ A ₁	.499	.501	.374	.350	.750	.700	.125	.150	.250	.300
	A ₃ of A ₁ A ₂	.498	.501	.374	.350	.751	.700	.124	.150	.249	.300
	(Expected)	.500	.500	.375	.350	.750	.700	.125	.150	.250	.300
Single	A ₁ of A ₂ A ₃	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	A ₂ of A ₃ A ₁	.453	.468	.330	.317	.728	.678	.123	.151	.272	.322
	A ₃ of A ₁ A ₂	.457	.468	.332	.317	.728	.678	.124	.151	.272	.322
Double	A ₁ of A ₂ A ₃	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	A ₂ of A ₃ A ₁	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	A ₃ of A ₁ A ₂	.384	.392	.288	.274	.750	.698	.096	.118	.250	.302

Note: Each mean proportion is based on 108,000 tests (108 tests for each of 1000 data sets) or 720,000 tests for the 3x3x3 and 4x4x4 designs, respectively. The Triple condition is omitted since all values must equal .000.

CMSICAL allows a test for simple independence among all pairs of factors. Independence for factor A₁ of A₂ would be checked by comparing the rank order of the cells for the levels of factor A₁ at each level of factor A₂. Similarly, a separate check can be made for the independence of A₂ at each level of A₁. There are 108 and 720 such tests for each factor pair in the 3 × 3 × 3 and 4 × 4 × 4 designs, respectively. In addition, Kendall's coefficients of concordance (W) can be obtained across rows or across columns at each level of the third factor. Thus, if independence is satisfied, then a data matrix should look like the matrix in the Triple condition with all rows and columns being in the same rank order, yielding W values for all pairs of factors to be 1.0. This is clearly not always the case with real data. For the Double condition data, it must be the case that the "A₁ of A₂," "A₁ of A₃," "A₂ of A₁," and "A₂ of A₃" coefficient values are all 1.0. This is not necessarily the case for "A₃ of A₁" or "A₃ of A₂." To the extent that some of the W values are near zero, there may be either

nonindependence of factors or irrelevance of a factor or a level of the factor.

The last possibility is particularly interesting from an empirical standpoint. Suppose that W values of 1.0 were found for two of the factors when simple independence is tested, but values near zero were found for the third factor. It might be tempting to conclude that no simple conjoint rule can be applied to the data. However, violations of independence would be restricted to factor A₃. These violations may have occurred here because this factor carried no importance or weight for the individual, and hence, that person did not differentiate among the levels of factor A₃. In this case, the person's judgments or rankings of alternatives would be based on the combination of only two independent factors A₁ and A₂.

In viewing Table 2 it is found that, as expected, the "Failures" column indicates that the observed proportions of violations of the simple independence axiom are very close to .5, which would be expected for random data. These values indicate

that the random number generator used in the study appears to be very good. They also indicate that the nonindependence of subsets of comparisons within a complete set of axiom tests has very little effect on the overall error proportions, an important point that is examined in detail later. It is interesting to note, however, that the proportion of failures in the Single and Double conditions for the remaining random factors are somewhat less than .5.

When the levels of the various factors are assumed to have at least ordinal properties and can be ordered a priori (e.g., level 1 < level 2 < level 3), then a further breakdown of the violations is possible. The remaining columns in Table 2 divide the failures into two types, *dominant* and *tradeoff*. Recall from Equation 2 that in simple independence two levels of one factor (a_1 and b_1) are being compared at two combinations of the second and third factors ($[a_2, a_3]$ and $[b_2, b_3]$). A violation occurs when $a_1 > b_1$ for one combination and $a_1 < b_1$ in the other. A *dominant failure* is defined then as one for which both the a_2 and a_3 levels in the $[a_2, a_3]$ combination dominate or are dominated by their respective counterparts in $[b_2, b_3]$. *Tradeoff failures* are defined as those that occur between stimuli when one stimulus does not dominate the other on both of the combined factors. For example, suppose $(1, [1, 1]) > (2, [1, 1])$ but $(1, [2, 2]) < (2, [2, 2])$. This would result in a dominant violation since for the two outside factors, A_2 and A_3 , $[1, 1]$ is dominated by $[2, 2]$. The test $(1, [2, 1]) > (3, [2, 1])$ but $(1, [1, 3]) < (3, [1, 3])$ is a tradeoff violation since combination $[2, 1]$ does not dominate and is not dominated by $[1, 3]$.

From Table 2 it is clear that under random data assumptions, the results approximate what, through combinatorial analysis, the proportions of dominant errors would be expected to be—.375 for a $3 \times 3 \times 3$ design and .350 for a $4 \times 4 \times 4$ design. These proportions are valuable in evaluating the fit of an additive model in an empirical situation. First, these values give a benchmark to indicate whether or not an individual's data are being fit significantly better than would be expected by chance. Although evaluating the fit of data to a random model may seem initially to be a weak null hypothesis, it will be shown later to be very important in evaluating

whether an individual treats a factor as inessential in the judgment process. Second, the conditional proportions of failures indicate some interesting results. These values are estimates of the conditional probability of a dominant violation given that a violation has occurred and of a tradeoff violation given that a violation has occurred. Although the unconditional proportions of violations decrease as the number of factors satisfying simple independence goes from zero to one and to two, the conditional values of $p[\text{dominant violation} | \text{violation}]$ and $p[\text{tradeoff violation} | \text{violation}]$ remain constant at .750 and .250 for the $3 \times 3 \times 3$ design, and at .700 and .300 for the $4 \times 4 \times 4$ design, respectively.

These latter results suggest a means of testing between two theoretically interesting possible sources of violations in an individual's data. Violations may occur either because the individual ignores the factor completely (unsystematic errors), or because he/she uses the factor but in a way that does not conform to the simple additive model (systematic errors). It seems reasonable that in the former case when the factor is ignored, the data would exhibit violations similar to those found in Table 2, with random data for dominant *and* tradeoff pairs. In the latter case, if the individual is in fact attending to the factor, then an overall reduction in failures would be expected, at least to a moderate degree.

The important aspect here, however, would be that this reduction would show up to a greater extent in the *dominant* tests. It is clear that in many judgment situations, regardless of the combination rule used by the individual, comparisons among stimuli that dominate others on relevant ordinal factors are easier to make, and are more likely to satisfy the independence axiom, than are those comparisons involving tradeoffs among levels of different factors. This suggests an important reason for examining individual tests of the axioms in more detail. The test of a null hypothesis of violations of an axiom against a random model may, in this case, be an appropriate hypothesis to test, if the researcher wants to examine whether a given factor is being ignored by an individual or is being used in conjunction with a nonadditive combination rule.

Table 3 summarizes the means of the Kendall's W values that were found in the data sets in each of the four conditions. These values are also useful in providing a benchmark to which empirical results can be compared. It is important to note that in no case are these mean coefficients close to zero. In particular, these means increase dramatically for a nonindependent, random factor, when one or more of the other factors do satisfy simple independence. Since these values are means, it is clear that an impressively large W value could occur in any given case very easily by chance alone. Hence, it is important not to use these observed W s exclusively but in conjunction with the information found in Table 2.

Joint Independence

A second form of independence can also be examined in the three factor models. The property, known as *joint independence*, states that:

A_1 and A_2 are jointly independent of A_3 whenever $(a_1, a_2, a_3) > (b_1, b_2, a_3)$ if and only if $(a_1, a_2, b_3) > (b_1, b_2, b_3)$. (3)

Joint independence of A_1 and A_2 with respect to A_3 indicates that if one combination of A_1 and A_2 is greater than another at a fixed level of A_3 (i.e., $[a_1, a_2] > [b_1, b_2]$ at a_3), then the ordering should be preserved for any other level of the third factor (b_3). If joint independence holds for all pairs of factors, then this implies that simple independence holds for each factor. However, the converse is not necessarily true. Two other forms of the joint independence property can be stated for A_1 and A_3 of A_2 , and A_2 and A_3 of A_1 .

CMSCAL allows the testing of joint independence for all pairs of factors by making all possible comparisons (108 and 720 for each pair of factors in a $3 \times 3 \times 3$ and $4 \times 4 \times 4$ design, respectively). In addition, coefficients of concordance can again be computed. These coefficients are somewhat difficult to interpret in and of themselves. To understand the implications of and the differences between the test of simple independence and joint independence, it is important to follow how the W values are computed. First, simple independence

is illustrated with the Uncond data from Table 1. A value of $W = .333$ was obtained from the check of independence for factor A_2 of A_1 at level 1 of factor A_3 . It was obtained by comparing the rank orders of the following three sets $(a_1 - a_3)$ of three numbers $(b_1 - b_3)$:

1. 21, 22, 14
2. 13, 27, 2
3. 24, 8, 9

In a comparable manner, independence for factor A_1 of A_2 at level 1 of factor A_3 , where W was found to be .111, was obtained by comparing the rank orders of the three sets $(b_1 - b_3)$ of three numbers $(a_1 - a_3)$:

1. 21, 13, 24
2. 22, 27, 8
3. 14, 2, 9

A joint independence value of $W = .346$ for A_3 independent of A_1, A_2 was obtained by comparing rank orders of the following nine sets $([a_1, b_1], \dots, [a_3, b_3])$ of three numbers $(c_1 - c_3)$:

1. 21, 11, 15
2. 22, 3, 12
3. 14, 10, 25
- .
- .
9. 9, 17, 18

Finally, $W = .578$ for A_1, A_2 independent of A_3 as found from the ranks of three sets of nine numbers:

1. 21, 13, 24, 22, 27, 8, 14, 2, 9
2. 11, 16, 23, 3, 7, 4, 10, 1, 17
3. 15, 5, 26, 12, 20, 19, 25, 6, 18

In a manner comparable to that discussed above for the simple independence axiom, mean proportions of violations were computed for joint independence. The results of these tests are summarized in Table 4. The "Failures" column indicates the mean error proportions for tests of joint independence when either zero, one, two, or all three factors satisfy simple independence. Several important results are shown in this column. First, for random data the probability of observing a violation of joint independence is near .5, the theoretical value expected by chance. When even one factor satisfies simple independence, however, the expected proportion of violations drops to about .245

Table 3
 Mean Kendall's Coefficient of Concordance Values for Simple
 Independence and Joint Independence in the 3x3x3 and 4x4x4 Designs

		Simple Independence Factor Combinations					
Analysis	Factor Design	Factor					
		A ₁ /A ₂	A ₁ /A ₃	A ₂ /A ₁	A ₂ /A ₃	A ₃ /A ₁	A ₃ /A ₂
Uncond	3	.327	.338	.329	.338	.327	.329
	4	.252	.248	.251	.248	.249	.250
Single	3	1.000	1.000	.578	.342	.571	.338
	4	1.000	1.000	.537	.248	.539	.248
Double	3	1.000	1.000	1.000	1.000	.519	.571
	4	1.000	1.000	1.000	1.000	.557	.536

		Joint Independence Factor Combinations					
Analysis	Factor Design	Factor					
		A ₁ /A ₂ A ₃	A ₂ /A ₃ A ₁	A ₃ /A ₁ A ₂	A ₂ A ₃ /A ₁	A ₃ A ₁ /A ₂	A ₁ A ₂ /A ₃
Uncond	3	.110	.113	.114	.326	.279	.415
	4	.063	.061	.062	.247	.249	.250
Single	3	1.000	.202	.196	.618	.757	.753
	4	1.000	.131	.133	.577	.760	.761
Double	3	1.000	1.000	.341	.831	.898	.950
	4	1.000	1.000	.301	.839	.925	.956

Note: Each mean value in the table is based on 1000 data sets. The η -values for the Triple condition are all 1.000's in each case and are omitted.

for tests involving that factor. Second, the error proportions again appear to be about the same for either the $3 \times 3 \times 3$ or $4 \times 4 \times 4$ design. Finally, the Triple data indicate an important finding that is often overlooked when examining real data. The proportions of violations in the Triple data are about .05 despite the fact that simple independence is satisfied perfectly for all three factors. Simple independence does not imply joint independence.

The tests of joint independence were divided as before into dominant and tradeoff tests. Here, *strict dominant* tests are defined as those for which the levels of the joint factors are both strictly dominant in one of the stimulus pairs. *Tradeoff* tests are defined as those for which the levels of the joint factors are strictly dominant in one direction for

one factor and in the opposite direction for the other factor. Finally, *weakly dominant* tests are defined as those for which there is equality of levels on one of the joint factors and dominance on the other. Thus, there can be weak dominance on the first or the second factor of the pair of joint factors. Examples of a violation of a strict dominance test, a weak dominance test on factor A₁, a weak dominance test on factor A₂, and a tradeoff test for joint independence of A₁ and A₂ from A₃ are:

1. *Strict dominance*: ([2,2],2) > ([1,1],2), but ([2,2],3) < ([1,1],3)
2. *Weak dominance, A₂*: ([2,2],2) > ([2,1],2), but ([2,2],3) < ([2,1],3)
3. *Weak dominance, A₁*: ([2,2],2) > ([1,2],2), but ([2,2],3) < ([1,2],3)

Table 4
Observed Dominant and Tradeoff Error Proportions for the
Joint Independence Axiom for 3x3x3 and 4x4x4 Designs

Analysis	Tested Factors	All Failures		Dominant Failures		Tradeoff Failures		Weak Dom Failures 1st Factor		Weak Dom Failures 2nd Factor	
		3	4	3	4	3	4	3	4	3	4
		Uncond	A ₁ A ₂ of A ₃	.495	.500	.123	.150	.123	.150	.124	.100
	A ₂ A ₃ of A ₁	.500	.502	.124	.151	.125	.150	.125	.100	.126	.099
	A ₃ A ₁ of A ₂	.503	.501	.125	.150	.125	.150	.125	.100	.125	.100
	(Expected)	.500	.500	.125	.150	.125	.150	.125	.100	.125	.100
Single	A ₁ A ₂ of A ₃	.245	.233	.061	.067	.061	.066	.000	.000	.123	.100
	A ₂ A ₃ of A ₁	.332	.335	.082	.100	.082	.101	.083	.068	.084	.066
	A ₃ A ₁ of A ₂	.243	.234	.058	.066	.060	.067	.124	.100	.000	.000
Double	A ₁ A ₂ of A ₃	.079	.076	.000	.000	.079	.076	.000	.000	.000	.000
	A ₂ A ₃ of A ₁	.190	.183	.048	.052	.048	.054	.000	.000	.094	.077
	A ₃ A ₁ of A ₂	.140	.118	.027	.025	.028	.026	.085	.067	.000	.000
Triple	A ₁ A ₂ of A ₃	.053	.047	.000	.000	.053	.047	.000	.000	.000	.000
	A ₂ A ₃ of A ₁	.075	.071	.000	.000	.075	.071	.000	.000	.000	.000
	A ₃ A ₁ of A ₂	.046	.040	.000	.000	.046	.040	.000	.000	.000	.000

Note: Each mean proportion is based on 108,000 tests (108 tests for each of 1000 data sets) or 720,000 tests for the 3x3x3 and 4x4x4 designs, respectively.

4. Tradeoff: $([2,1],2) > ([1,2],2)$, but $([2,1],3) < ([1,2],3)$

Table 4 presents a breakdown of the violations of joint independence into the four categories illustrated above. These mean proportions are clearly stable for the unconditional data matrices. When simple independence is satisfied by one or more factors, however, it is possible to further differentiate among the tests. For the Single data where simple independence is satisfied by one factor (A_1) but the data are otherwise still random, tests of joint independence result in markedly fewer violations, especially for A_1 and A_2 from A_3 and for A_3 and A_1 from A_2 .

These results as well as those in the last two columns of Table 4 indicate the theoretical as well as empirical significance of the approach of examining individual violations of the axioms. As

independence begins to hold in the data, the breakdown of errors indicates important relationships useful to the applied researcher. The weak dominant tests of joint independence are equivalent to tests of simple independence. Note in Table 4 that weakly dominant failures *cannot* occur in the joint independence axiom on factor A_1 for either A_1A_2 jointly independent of A_3 (A_1 is the first factor in the joint pair) or A_3A_1 jointly independent of A_2 (A_1 is now the second factor in the pair), if simple independence holds in A_1 . When simple independence holds for two factors (A_1 and A_2), this implies that *all* violations of joint independence for A_1 and A_2 of A_3 must be tradeoff violations; there can be no dominant errors, either strong or weak. Also there can be no weak dominant errors for A_2A_3 of A_1 on factor A_2 and none for A_3A_1 of A_2 on factor A_1 .

When simple independence holds for all three

factors, then all violations for A_1A_2 of A_3 , A_2A_3 of A_1 , and A_3A_1 of A_2 are tradeoff violations; no dominant errors are possible. If the error proportions in the Triple condition of Table 4 had been obtained with real data and had been viewed at a global level by the researcher, they would probably be so few in number, an average of about 5%, that they would appear to support an additive model. However, the pattern exhibited at the bottom of Table 4 indicates that all of the errors must be tradeoff errors and that the successful tests of joint independence are expected to be at about the 95% level by *chance alone*, if simple independence holds in the data.

With nonadditive data, it may also be possible with this axiom to determine whether an individual is ignoring a particular factor or whether he or she is using it, but joint independence with a third factor simply does not hold. Clearly, if a third factor (A_3) is being ignored, then dominant and tradeoff error proportions comparable to those in the Double condition in Table 4 might be expected. If, however, the individual attends to all three factors and simple independence holds in the data, but the judgmental process is a nonadditive one, then results similar to those in the Triple condition of Table 4 would be expected.

Finally, the mean Kendall's W values for the tests of joint independence are presented in the lower half of Table 3. Since these W values are usually based on either a greater number of rankings or more extensive rankings than are the values in the upper half of Table 3, they tend to be closer to zero for random data. However, as simple independence is satisfied in one or more factors, these W values again increase rather dramatically. Individual observed values from empirical data can be impressively high, even when simple independence holds in only one factor. The mean W values in Table 3 are very valuable for comparison purposes and again indicate that high values obtained with empirical data must be interpreted cautiously.

Double Cancellation and Distribution Cancellation

The third property to be examined is the property usually referred to as *double cancellation* or *Luce-*

Tukey cancellation and is stated for factors A_1 and A_2 as:

$$\begin{aligned} \text{If } (a_1, b_2, a_3) > (b_1, c_2, a_3) \text{ and,} \\ (b_1, a_2, a_3) > (c_1, b_2, a_3), \text{ then,} \\ (a_1, a_2, a_3) > (c_1, c_2, a_3) \end{aligned} \quad (4)$$

Double cancellation requires at least three levels of each of the factors A_1 and A_2 and deals with only two such factors at a time. It must be satisfied for all pairs of factors in an additive model.

The last property, *distributive cancellation* is said to be satisfied if and only if

$$\begin{aligned} (a_1, b_2, a_3) > (d_1, c_2, c_3), \\ (b_1, a_2, a_3) > (c_1, d_2, c_3), \text{ and} \\ (d_1, d_2, c_3) > (b_1, b_2, a_3), \text{ then} \\ (a_1, a_2, a_3) > (c_1, c_2, c_3). \end{aligned} \quad (5)$$

It can be shown that this property is a necessary condition for the distributive model to hold. However, distributive cancellation must also hold in an additive representation. Hence, although this property can be used to support a distributive representation, it cannot be used to reject additivity.

Tables 5 and 6 present the mean error proportions obtained from the analyses of the double and distributive cancellation axioms. Several points are of interest here. It is important to note that both of these cancellation axioms have several antecedent conditions that need to be met before a test is even possible. For double cancellation there are two such antecedent conditions and for distributive cancellation there are three. Hence, Tables 5 and 6 first present the proportion of all tests that were actually possible in the data; that is, tests that met the antecedent conditions. For double cancellation this is not a trivial matter in practice, since, as Table 5 illustrates, for random data only one-third of the tests can be expected to meet the antecedent conditions.

As more order becomes present in the data, the proportion of possible tests increases rather quickly. It appears from Table 5 that if either two or all three factors satisfy simple independence, then nearly two-thirds or more of all tests are possible. These proportions suggest that the number of possible tests in the data may be as important as the number of actual violations of these tests in using the double cancellation axiom as a diagnostic tool, since ad-

Table 5
 Observed Unconditional and Conditional Error Proportions for
 the Double Cancellation Axiom for 3x3x3 and 4x4x4 Designs

Analysis	Tested Factors	Possible Tests		Unconditional Failures		Conditional Failures	
		3	4	3	4	3	4
		<hr/>					
Uncond	A ₁ xA ₂ plane	.356	.331	.269	.247	.757	.748
	A ₂ xA ₃ plane	.325	.332	.249	.248	.766	.747
	A ₃ xA ₁ plane	.341	.330	.253	.247	.743	.748
	(Expected)	.333	.333	.250	.250	.750	.750
<hr/>							
Single	A ₁ xA ₂ plane	.475	.509	.068	.049	.144	.097
	A ₂ xA ₃ plane	.326	.339	.250	.250	.768	.750
	A ₃ xA ₁ plane	.490	.509	.052	.051	.106	.100
<hr/>							
Double	A ₁ xA ₂ plane	.623	.643	.095	.064	.152	.100
	A ₂ xA ₃ plane	.485	.509	.065	.050	.135	.098
	A ₃ xA ₁ plane	.639	.718	.021	.014	.032	.020
<hr/>							
Triple	A ₁ xA ₂ plane	.654	.654	.101	.053	.155	.081
	A ₂ xA ₃ plane	.642	.639	.097	.063	.154	.099
	A ₃ xA ₁ plane	.790	.837	.038	.018	.047	.022

Note: Mean proportions in the table are based on 3000 tests (3 tests for 1000 data sets) and 64,000 tests for the 3x3x3 and 4x4x4 designs, respectively.

ditivity would imply that *all* tests are possible as well as correct.

The relationship between simple independence and double cancellation is seen more clearly in the proportions of failures in Table 5. When simple independence is satisfied by one of the factors (A_1), the proportion of violations of double cancellation drops greatly for tests in both the $A_1 \times A_2$ and $A_1 \times A_3$ planes. When simple independence is satisfied by all three factors, the proportions of conditional violations (i.e., a violation given that a test is possible) drops to about 10% to 15%. Hence, the researcher using conjoint scaling methodology should not be overly optimistic about an additive model when conditional error rates for double cancellation are around 10% to 15%. These values can be obtained for many nonadditive data sets as long as simple independence is not violated by only two of the three factors.

The last axiom, distributive cancellation, is examined in Table 6. It is clear from the results summarized in this table that the distributive cancellation axiom is not a useful diagnostic tool for the applied researcher for several reasons. First, the axiom is a very weak one. Unlike the double cancellation axiom, the actual proportions of tests that meet the antecedent criteria are so high that they are not useful diagnostic values. It is interesting and somewhat surprising that for even random data, the proportion of tests that meet the antecedent conditions exceeds 90% in all cases. Second, even for random data only about 25% of the tests result in violations. When independence is satisfied by one or more of the factors, these proportions of violations are reduced considerably. These error proportions are, in fact, so small that the data almost look as though they nearly perfectly support an additive model. The error proportions are in the

Table 6
 Observed Unconditional and Conditional Error
 Proportions for the Distributive Cancellation
 Axiom for 3x3x3 and 4x4x4 Designs

Analysis	Tested Outside Factor	Possible Tests		Unconditional Failures		Conditional Failures	
		3	4	3	4	3	4
Uncond	A ₁ outside	.971	.971	.254	.255	.261	.262
	A ₂ outside	.971	.972	.257	.255	.257	.262
	A ₃ outside	.971	.971	.249	.255	.265	.263
Single	A ₁ outside	.996	.996	.044	.053	.044	.053
	A ₂ outside	.932	.922	.069	.053	.073	.057
	A ₃ outside	.930	.922	.068	.053	.074	.058
Double	A ₁ outside	.989	.984	.012	.012	.012	.012
	A ₂ outside	.956	.930	.021	.018	.022	.020
	A ₃ outside	.907	.900	.030	.018	.033	.020
Triple	A ₁ outside	.984	.979	.005	.004	.005	.004
	A ₂ outside	.945	.918	.008	.006	.008	.005
	A ₃ outside	.923	.910	.010	.005	.010	.006

Note: Mean proportions in the table are based on 243,000 tests (243 tests for 1000 data sets) and 7,776,000 tests for the 3x3x3 and 4x4x4 designs, respectively.

5% range when only one factor satisfies simple independence and are in the 1% to 2% range when two or three factors satisfy simple independence.

Empirical Example

To illustrate the axiom testing procedure, rankings of perceived workload were obtained from six students taking a graduate course in scaling theory. Each student was asked to perform the standard factorial ranking task that is used in the Subjective Workload Assessment Technique (SWAT, cf. Nygren, 1985; Reid, Shingledecker, & Eggemeier, 1981). Here, the 27 combinations of the three ordered levels of each of three factors assumed to be components of perceived workload—*time load*, *effort load*, and *stress load*—were varied. The levels of the factors are presented in sentence form as descriptions of low, moderate, and high levels of time, effort, and stress load. Thus, the levels of

the factors present a clear ordering, forcing every individual who uses the three factors to have, at a minimum, combinations (1,1,1) and (3,3,3) as the least and greatest perceived workload, respectively. The ordering of the other 25 combinations would depend on the individual's decision rule and on the importance he/she placed on each factor. A complete description of the levels of each of the factors is presented in Reid et al. (1981).

The students were asked to rank the 27 combinations of the three factors according to their perception of workload. Three of the students had perfectly additive data, and one had additive data except for one pair reversal. Rank order data from the other two students with axiom violations are presented in the lower part of Table 1. Following completion of the task, Subject 1 indicated that she made her rankings on the basis of all three factors, but that the time load factor was most important. Subject 2 indicated that she had made her rankings

on the basis of all three factors also, but that at a low level of time load, there was no difference in workload, regardless of stress or effort levels.

Initial analysis of Subject 1 and Subject 2's data might lead to the quick endorsement of an additive model. Across all levels of the factors, the overall proportions of violations of simple and joint independence, and of double and distributive cancellation, for Subject 1 were .000, .049, .000, and .005, respectively. For Subject 2 these values were not quite as good but were still low, being .145, .093, .000, and .005, respectively.

The monte carlo results presented in Tables 2 through 6 and the detailed breakdown of errors presented in this paper, however, allow a much more accurate picture of each subject's decision rule. For Subject 1 the data reveal that simple independence holds for all three factors. That is, level 1 was always perceived as being less than level 2, and level 2 less than level 3 for each factor, regardless of the levels of the other factors. This can be easily seen either by plotting the data as would be done if these were ANOVA means or by noting that the rows, columns, and planes are all monotonic. This implies a situation similar to that of the Triple condition in Table 4—any violation of joint independence for this subject must be a tradeoff error. The actual tradeoff error proportions found for this subject for joint independence for factor pairs A_1A_2 , A_2A_3 , and A_3A_1 were .056, .074, and .019, respectively. These values are no better than what would be expected by chance from Table 4, indicating that these errors are more likely due to real interactions among all three factors in a nonadditive model, rather than simply being "true" errors made by a subject using an additive rule. Finally, the satisfaction of simple independence in the data virtually forces very small error proportions for double and distributive cancellation, in agreement with the Triple condition results of Tables 5 and 6. Subject 1's rankings and these detailed examinations of the axioms reveal, in agreement with her perception of the task, that she was making her judgments using a process that, though closely approximated by a three-factor additive model with time load being the most important factor, had some consistent interactions among levels of factors.

A more detailed examination of Subject 2's data revealed that simple independence held perfectly for factor A_1 , but the error proportions were .222 for both factors A_2 and A_3 . Further, an examination of the Kendall's coefficient values indicated that these errors were restricted to *only* level 1 of factor A_1 . This can be seen at a more global level by noting that the rankings for this subject in Table 1 show monotonicity across rows, columns, and planes *except* at level 1 of factor A_1 .

Note that in Table 2 it was found that for any factor, .5 (54/108) of the tests should result in errors. This implies that .167 (18/108) would be expected at each of the three levels of the factor. This compares very closely with the .222 values that were found for A_2 independent of A_3 at A_1 , and A_3 independent of A_2 at A_1 . A z test indicated that the .222 and .167 values were not significantly different ($p > .10$), nor were the observed dominant/tradeoff error proportions, .583 and .417, significantly different from their expected values as shown in Table 2, .750 and .250 ($p > .05$).

Since simple independence held for factor A_1 , the errors for joint independence would be compared with the values in the Single condition of Table 4. For the three pairs of factors, the observed error proportions were .056 (6/108), .167 (18/108), and .056 (6/108), respectively. Note how they are much smaller than, but fit the same pattern as, the values from Table 4 (.245, .332, and .243, respectively). In addition, these observed errors were virtually all weak dominant errors (6/6, 12/18, and 6/6, respectively). Recall that weak dominant errors for joint independence are really the result of simple independence errors. These error patterns, then, strongly suggest that Subject 2 ordered the stimuli in a perfectly additive way except at level 1 of time load, which was ordered in an irrelevant, random manner. This error pattern fits with the subject's own assertion that, for her, stress and effort levels had no effect on workload when time demands were minimal.

Discussion

The research presented above attempted to examine violations of the essential properties of additive conjoint measurement. These properties have had limited usefulness in applied research at least

partly because their inherent subproperties and interrelationships have not been well understood. Without an adequate error theory for the axiomatic conjoint measurement model, assessing the goodness-of-fit of data to the model remains a subjective process. Recent literature (Nickerson & McClelland, 1984) suggested that using only the numerical conjoint measurement approach may in fact produce highly misleading values of goodness-of-fit measures. It is argued here in concurrence with the findings of Nickerson and McClelland (1984) that the additive conjoint measurement procedure should always begin with an analysis of the axioms.

This axiom testing procedure, however, should go beyond the simple step-by-step procedure first outlined by Krantz and Tversky (1971). It is not enough to use the overall error rates associated with each axiom to assess nonadditivity. At such global levels, it is clear that the axioms are relatively weak as diagnostic tools and may imply a better fit to additivity than is warranted. When errors are examined by type, either dominant or weak, and by location within levels of a particular factor, a stronger test of additivity or of a particular nonadditive model may be found. The results presented here indicated that it is not surprising that at a global level, the misdiagnosis problem illustrated by Emery and Barron (1979) could occur.

The diagnostic value of the axioms can be enhanced at least three ways. These include (1) establishing the baseline error rates expected from a random response model; (2) establishing conditional error rates expected in data when varying amounts of structure are present in the data; and (3) dividing, when possible, the violations into two types—dominant and tradeoff—and establishing expected error rates for these classes of violations.

When structure is added, even minimally on one factor, to a set of random data, error proportions for the four axioms described here can be expected to decrease markedly. This decrease causes no serious concern in and of itself as long as baseline or expected values are available for comparison purposes. The means in Tables 2 through 6 provide benchmark values against which empirically obtained proportions can be compared and tested statistically. Without appropriate comparison values,

it is clear from the results of the study and the data from the two subjects that rankings satisfying simple independence but not additivity may produce misleadingly small error rates for the other axioms. For example, a conclusion of additivity based on error rates in the 1% to 5% range for distributive cancellation could be quite erroneous. Such seemingly impressive results could be easily obtained when the model is not at all additive, and, in fact, when only one factor satisfies simple independence.

One potential problem with using the binomial distribution to test hypotheses about the error proportions expected under different conditions is the violation of the assumption of independence of observations. The 108 tests of joint independence for a pair of factors in a $3 \times 3 \times 3$ design are clearly not completely independent of one another. Each comparison is actually used in eight different tests. In addition, the conversion of data to ranks imposes constraints on the comparisons. Nevertheless, the ratio of the number of dependent tests (8) to independent tests (100) for any given stimulus combination is probably small enough in terms of the statistical model, so that the violation of the independence assumption is of little practical significance when using the normal approximation to the binomial to test hypotheses about proportions. The actual error proportions found in this study match very closely the expected values obtained under the assumption of complete independence of all tests. These results are comparable to those found by Nygren (1979) in a study of the axioms associated with the additive difference model for multidimensional scaling. Testing hypotheses about observed W values is more problematic, however, since no clear sampling distribution exists for W s other than zero.

The examination of conditional violations of one axiom with respect to another yields several advantages over the unconditional approach. Subclasses of axiom tests are effectively produced, each with the potential of evaluating fit to an additive model. For example, with double cancellation, the proportion of tests actually possible in a data set can be as useful as a diagnostic tool as actual violations themselves. It is clear that even

if simple independence holds for all three factors, not all tests of double cancellation will be possible. This may seem somewhat counterintuitive at first, since both axioms are necessary for an additive model. However, the two axioms are examining properties of the data that are, though clearly related, somewhat unique. Conditionalizing errors leads to a better understanding of the interrelationships among the axioms. Finally, the method used in this study also suggests a possibly useful means of allowing for a distinction between violations due to randomness and violations due to a nonadditive model.

This study was limited to an examination of what happens to violations of the conjoint measurement axioms when simple independence is satisfied by none, one, two, or all three of the factors. The results are certainly encouraging with respect to being able to set expected violation proportions. The approach began with random data and added order to it systematically by satisfying simple independence in one to three factors. Another approach might be to work in somewhat of an opposite direction. It might start with data sets that satisfy different known interaction models and systematically reduce the interaction effects to produce nearly additive models. Then, the method would test for violations of the axioms. Such a monte carlo study might lead to an axiomatization of the, as yet, unsolved interaction models. What this would do for the applied researcher is to allow a determination of the expected violations of the axioms for commonly found interactions models.

References

- Arbuckle, J., & Larimer, J. (1976). The number of two-way tables satisfying certain additivity axioms. *Journal of Mathematical Psychology, 13*, 89–100.
- Emery, D. R., & Barron, F. H. (1979). Axiomatic and numerical conjoint measurement: An evaluation of diagnostic efficacy. *Psychometrika, 44*, 195–210.
- Falmagne, J. (1979). Random conjoint measurement and loudness summation. *Psychological Review, 83*, 65–79.
- Helwig, J. T., & Council, K. A. (Eds.). (1979). *SAS user's guide, 1979 edition*. Raleigh NC: SAS Institute.
- Holt, J. O., & Wallsten, T. S. (1974). *A user's manual for CONJOINT: A computer program for evaluating certain conjoint-measurement axioms* (Technical Report 42). Chapel Hill NC: University of North Carolina, L. L. Thurstone Psychometric Laboratory.
- Johnson, R. M. (1973). Pairwise nonmetric multidimensional scaling. *Psychometrika, 38*, 11–18.
- Krantz, D. H. (1968). A survey of measurement theory. In G. B. Dantzig & A. F. Veinott, Jr. (Eds.), *Mathematics of the decision sciences, Part 2* (pp. 314–350). Providence RI: The American Mathematical Society.
- 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. *Psychology 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, 27*, 251–263.
- McClelland, G. H. (1977). A note on Arbuckle and Larimer, "The Number of Two-way Tables Satisfying Certain Additivity Axioms." *Journal of Mathematical Psychology, 14*, 292–295.
- Nickerson, C. A., & McClelland, G. H. (1984). Scaling distortion in numerical conjoint measurement. *Applied Psychological Measurement, 8*, 183–198.
- Nygren, T. E. (1979). A theoretical framework for testing the additive difference model for dissimilarities data: Representing gambles as multidimensional stimuli. *Journal of Mathematical Psychology, 20*, 53–77.
- Nygren, T. E. (1985). Axiomatic and numeric conjoint measurement: A comparison of three methods for obtaining subjective workload (SWAT) rankings. In *Proceedings of the IEEE 1985 National Aerospace and Electronics Conference, NAECON '85* (pp. 878–883). New York: Institute of Electrical and Electronics Engineers.
- Reid, G. B., Shingledecker, C. L., & Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 522–526). Santa Monica CA: Human Factors Society.
- Ullrich, J. R., & Cummins, D. E. (1973). PCJM: A program for conjoint measurement analysis of polynomial composition rules. *Behavioral Science, 18*, 226–227.

Acknowledgments

The author gratefully acknowledges the support of this

project through Grant 82-0175 from the U.S. Air Force Office of Scientific Research. The author also acknowledges the support of Gary Reid and the Workload and Ergonomics Group, AFAMRL, Wright-Patterson AFB OH, U.S.A.

Author's Address

Send requests for reprints or further information to Thomas E. Nygren, Department of Psychology, The Ohio State University, 404C W. 17th Avenue, Columbus OH 43210, U.S.A.