JOURNAL OF MATHEMATICAL PSYCHOLOGY 10, 148–203 (1973)

Multidimensional Scaling of Measures of Distance between Partitions

PHIPPS ARABIE

Department of Psychology, Stanford University, Stanford, California 94305

AND

SCOTT A. BOORMAN

Society of Fellows and Department of Sociology, Harvard University, Cambridge, Massachusetts 02138

The techniques of multidimensional scaling were used to study the numerical behavior of twelve measures of distance between partitions, as applied to partition lattices of four different sizes. The results offer additional support for a system of classifying partition metrics, as proposed by Boorman (1970), and Boorman and Arabie (1972). While the scaling solutions illuminated differences between the measures, at the same time the particular data with which the measures were concerned offered a basis both for counterexamples to some common assumptions about multidimensional scaling and for some conjectures as to the nature of scaling solutions. The implications of the latter findings for selected examples from the literature are considered. In addition, the methods of partition data analysis discussed here are applied to an example using sociobiological data. Finally, an argument is made against undue emphasis upon interpreting dimensions in nonmetric scaling solutions.

INTRODUCTION

There is an extensive behavioral science literature dealing with the construction and use of numerical measures for quantitative description of relational structures. Structural measures of various kinds have been employed for describing and comparing sets, partitions, orderings, trees, graphs, groups and semigroups, grammars, and other kinds of complex relational entities. Any of a large number of information-theoretic measures fall in this category, as do various measures of grammatical complexity (Chomsky and Miller, 1963) and hierarchical structure (Landau, 1951; Boorman, 1970). Partly because of the substantive importance of structural equivalence concepts in many behavioral problems, see Lorrain (in press), structural measures dealing with partitions (their complexity, similarity, etc.) are probably the single

most important class of measures of relational structure. A systematic approach to structural measures on partitions and trees is developed from a lattice-theoretic viewpoint in previous papers of the present authors (Boorman, 1970; Boorman and Arabie, 1972; Boorman and Olivier, in press). This development is more algebraically oriented than earlier comparative and survey work for other kinds of structural measures, e.g., Restle (1959), for set metrics; Attneave (1959) for information measures; Flament, (1962, 1963) for measures of graphical structure; Dalrymple-Alford (1970), Frankel and Cole (1971), and Roenker, Thompson, and Brown (1971), for clustering in free recall.

In the absence of substantive or theoretical constraints, measures proliferate, and it is often unclear what is being captured by any particular one; see, for example, the critical survey of alternative measures of the "centrality" of a graph in Sabidussi (1967). To gain some kind of methodological control over this proliferation is the motivation of the comparative work just cited. A major problem, however, is that overwhelmingly the existing literature deals only with abstract formulas defining structural measures and highly qualitative deductions from the formulas. Because all relational structure is intrinsically complex, such qualitative discussions are frequently very weak and conclusions drawn from them are unreliable. A good example is the provocative but impractical discussion of Bavelas's (1950) centrality index from the standpoint of the automorphism group of the graph in Flament (1963, pp. 50–52). Empirically-oriented users, on the other hand, are interested in the details of the numerical behavior of a particular measure. Despite some distribution and estimation work on entropy and related information-theoretic measures, e.g., Kullback (1955), and detailed knowledge of the mean and variance of Landau's hierarchy index (Landau, 1951; Holland, 1971) under a variety of interesting conditions, relatively little is known about the numerical behavior of most structural measures. There is essentially no literature for any kind of structural measure which seriously attacks the problem of quantitatively comparing different measures and classifying them according to their numerical output.

The present study is one specialized attempt to study this problem of comparison from a particular viewpoint, namely, that of nonmetric multidimensional scaling and related techniques. Concretely, we will report the results of extensive application of scaling to representation of measures of distance between partitions of a finite set. It was hoped that scaling could elucidate differences among the distance geometries which distance measures can impose on partition spaces. For example, one of our findings (Section 2) was that by no means do information-theoretic measures necessarily have more desirable performance characteristics than do alternative measures of partition structure. This fact has very practical implications as far as the use of entropy-based structural measures is concerned. Although the influence of information-theoretic formalism in behavioral science investigations has clearly diminished in the past decade, there is still a strong tendency on the part of investigators to introduce in-

formation-theoretic measures without considering their appropriateness for the problem at hand. To some extent, this tendency derives from confidence in axiomatic justifications of the entropy measure, e.g., Khinchin (1957), which is frequently cited in the literature. The relevance and utility of these axiomatic derivations is effectively attacked by Mandelbrot (1966) from a theoretical standpoint, but his context (certain models in psycholinguistics) is very limited. One of the few investigators to consider empirically the need for a systematic comparison of the entropy measure with rival structural measures is Levins (1968a, 1968b), but his focus is on a particular set of ecological niche breadth data, and is, hence, again too restricted to be very informative.

Measurement problems related to partitions have intrinsic importance because partitions are a kind of mathematical structure underlying much behavioral data, for example, that generated by the method of sorting. Stimuli to be sorted have ranged from color patches (Rapoport and Fillenbaum, 1972) to common nouns (Miller, 1969), and the method of the Q-sort has been extensively employed by George Kelly and other clinicians (Block, 1961). The relative ease with which data can be collected by sorting methods has led possible questions to outrun systematic techniques in many aspects of the study of semantic and associative structures to which sorting may be applied. To a large extent, existing formal work on sorting methods, e.g., Miller (1969), has been confined to analyzing cooccurrence incidence matrices which can be derived from any family of partitions. This analysis poses a simpler problem, but one which involves discarding a large amount of the structure inherent in partition data. In particular, once the data are reduced to incidence matrix form in this manner, it no longer becomes possible to analyze the obtained family of partitions for homogeneous clusters or to discuss individual differences between sorters. Such problems naturally arise, for example, in the practical use of sorting to differentiate diagnosed schizophrenics from a control group of normals (Pavy, 1968).

Formally similar problems arise in the study of social systems consisting of freely forming small groups (White, 1962; Cohen, 1971). At any observation time, such a system may be described as a partition of its individual members among the groups they form, and we may ask questions about structural changes over time in terms of distance between partitions. Using scaling of partition distances, it is possible, for example, to differentiate systematic trends from random drift effects. An application of this type to a sociobiological example arising in primate behavior studies is developed in Section 7.

At the same time that a better understanding of partition distance measures is relevant for behavioral data analysis, the present study has methodological implications for the use of multidimensional scaling. It should be emphasized that there is no a priori guarantee that geometric modeling of the multidimensional scaling variety will provide meaningful information in any particular application to partition data. Shepard's original conjecture, that the interpoint distance orderings asymptotically define a given configuration of points in a Minkowski r-space up to the similarity

group (Shepard, 1962, 1966; Benzécri, 1964, 1965), has the nature of a unique recovery property, rather than an existence result. Moreover, prior to the present study, little systematic attempt appears to have been made to assess the possibility of fruitfully scaling discrete spaces, though there is a growing body of literature which performs such scalings on concrete empirical data of various kinds. For justification of the scaling of continuous spaces there is, by contrast, the well-known work of Beals, Krantz, and Tversky (1968). Although much of the early work using multidimensional scaling was for stimulus spaces which could at least in principle be treated as continuous, e.g., Ekman (1954), all semantic spaces are discrete, and it is only recently that the theoretical issues concerning the application of Shepard–Kruskal methods to such spaces have been investigated, e.g., Arnold (1971).

1. Procedure

In the present study, 12 distance measures were applied to partitions of a finite set and the obtained proximity matrices were scaled in Euclidean and city-block spaces of two and three dimensions, using Kruskal's (1964a, 1964b) scaling algorithm. The data on which our conclusions are based involved well over 500 runs of Kruskal's MDSCAL-4 and MDSCAL-5 programs, as well as various supplementary programs described later. As is clear from Table I, most of the scaling runs involved between 50 and 60 points, and consequently the present study is one of the most extensive tests of MDSCAL for large input configurations which has been undertaken to date. Because of space limitations, the present report reproduces only a limited sample of the obtained scalings; Table I gives summary statistics on the runs made (see also Table XI below for summary statistics on the obtained stresses for stress formula 1).

We first introduce some minimal formal background material on partition lattices; see Ore (1942), Birkhoff (1967), and Szász (1963) for more extensive treatments. A partition P of a finite set S is a division of that set into a collection of nonempty, pairwise disjoint, and collectively exhausitive subsets. We employ parenthetical notation to describe specific partitions; thus $P = ((ab)(cd))$ is a partition of $S = \{a, b, c, d\}$ with two components (ab) and (cd) which we will term *cells*.

Definition 1.1. A partition P is *finer than* a partition $Q (P \leqslant Q)$ if and only if each cell $c \in P$ is contained in some cell $d \in Q$ (dually, we say that Q is *coarser than* P). We write $P < Q$ if $P \leqslant Q$ and $P \neq Q$.

Definition 1.2. The intersection $P \cap Q$ of two partitions is defined as follows: two elements $a, b \in S$ are in the same cell of $P \cap Q$ if and only if they are in the same cell of P and also in the same cell of Q.

For example, if $P = ((ab)(cd))$ and $Q = ((abc)(d))$, then $P \cap Q$ is $((ab)(c)(d))$.

TABLE I

Inventory of Runs on Which the Present Study is Based

2 dimensions, Euclidean metric	Both $P_{splitter}$ and P_{lumper}	Without P_{lumper}	Without $P_{splitter}$	Without either
4-lattice (15 points)	6	5	10	14
5-lattice (52 points)	8	6	8	9
6-lattice sample (60 points)	16	24	22	14
10-lattice sample (60 points)	9	25	7	15
2 dimensions, City-block metric	**Both $P_{splitter}$ and P_{lumper}**	**Without P_{lumper}**	**Without $P_{splitter}$**	**Without either**
4-lattice	0	2	10	14
5-lattice	8	10	12	16
6-lattice sample	8	3	0	0
10-lattice sample	18	9	30	11
3 dimensions, Euclidean metric	**Both $P_{splitter}$ and P_{lumper}**	**Without P_{lumper}**	**Without $P_{splitter}$**	**Without either**
4-lattice	6	5	10	14
5-lattice	4	5	6	8
6-lattice sample	0	0	0	0
10-lattice sample	0	0	0	0
3 dimensions, City-block metric	**Both $P_{splitter}$ and P_{lumper}**	**Without P_{lumper}**	**Without $P_{splitter}$**	**Without either**
4-lattice	0	2	10	16
5-lattice	8	10	12	16
6-lattice sample	0	0	0	0
10-lattice sample	0	0	0	0

Note—Each entry includes all nondegenerate runs obtained summed over all 12 measures and all scalings repeated with different random initial configuration(s). A degenerate run is defined to be one which separates the partitions scaled into two clusters whose radii approach zero as iterations continue. If a rescaling of a run with a different initial configuration produced a degeneracy, neither run is counted in the tabulation.

There is a dual concept of the union of two partitions (this will be explicitly used in what follows only in connection with the discussion of alternatives to the measure TI2MINUS in Section 4).

DEFINITION 1.3. The *union* $P \cup Q$ of two partitions P and Q is defined as follows: two elements $a, b \in S$ are in the same cell of $P \cup Q$ if and only if there is a finite sequence $a_0 = a, a_1, \ldots, a_q = b$ of elements of S such that a_i, a_{i+1} cooccur either in a cell of P or in a cell of Q. In the above example, $P \cup Q = ((abcd))$.

We will refer in general to the partition which lumps all elements of S together as P_{lumper} and that which splits all elements of S into singleton cells as P_{splitter}. Thus if $S = \{a, b, c, d\}$, $P_{\text{lumper}} = ((abcd))$ and $P_{\text{splitter}} = ((a)(b)(c)(d))$. Because of the geometric position of these partitions in the Hasse diagram of a finite lattice, we will have occasion later to refer to P_{splitter} and P_{lumper} as (lattice) endpoints.

We now have the following basic fact:

THEOREM. *The set $\mathcal{P}(S)$ of all partitions of S forms a lattice under $\leqslant$, i.e., every pair of partitions P and Q will have a unique least upper bound, $P \cup Q$, and a unique greatest lower bound, $P \cap Q$, in the lattice ordering.*

Proof. See Szász (1963) for proof of this assertion, together with a much fuller characterization of the lattice structure.

By a *partition space* we mean simply a family of partitions constituting a subset of some partition lattice. Four basic partition spaces were endowed with metrics and submitted as inputs to multidimensional scaling algorithms: the full lattice of all 15 partitions of a four-element set (henceforth termed the 4-lattice); the full lattice of all 52 partitions of a five-element set (5-lattice); and two 60-element random samples from the partition lattices on six and ten objects (which will be termed, respectively, the 6- and 10-lattice samples). These last samples were obtained by a procedure for generating random partitions. This procedure (described in Appendix A) quite closely approximates the uniform distribution on the entire partition lattice over a reasonably small underlying set S. In the case of the 6-lattice the ratio of the sample to the full lattice size was

$$60/203 = 0.296,$$

while in the case of the 10-lattice it was

$$60/115{,}975 = 0.0005 = 5 \times 10^{-4}.$$

Tables II and III display the specific samples used.

Given any partition lattice $\mathscr{L}$, a semimetric on that lattice is a mapping m: $\mathscr{L} \times \mathscr{L} \rightarrow R$ which satisfies

(i) $m(P, Q) \geqslant 0$ and $= 0$ if and only if $P = Q$,

(ii) $m(P, Q) = m(Q, P)$ for all $P, Q \in \mathscr{L}$,

see Blumenthal (1953) for discussion of semimetrics and Shepard (1969, 1972). If, in addition, m satisfies

(iii) $m(P, R) \leqslant m(P, Q) + m(Q, R)$ (triangle inequality),

for all $P, Q, R \in \mathscr{L}$, then m is a metric. All measures of partition distance considered in the present study are semimetrics, and the majority are either metrics or derived from metrics in some simple way. Twelve such measures were studied in all. For the sake of easy reference, their defining formulas are collected in Tabel IV. The measures

TABLE II

6-lattice Sample of 58 Partitions with P_{lumper} and P_{splitter} Adjoined

(123456)	(13)(245)(6)	(1)(2)(36)(45)
(146)(235)	(13)(25)(46)	(145)(2)(3)(6)
(1236)(45)	(1)(235)(46)	(15)(24)(3)(6)
(126)(345)	(124)(3)(56)	(1)(245)(3)(6)
(1245)(36)	(14)(23)(56)	(135)(2)(4)(6)
(13)(2456)	(14)(25)(36)	(12)(35)(4)(6)
(16)(245)(3)	(12)(34)(56)	(1)(235)(4)(6)
(145)(26)(3)	(1)(23)(456)	(15)(2)(34)(6)
(15)(246)(3)	(1)(2)(3456)	(124)(3)(5)(6)
(12)(356)(4)	(1)(26)(35)(4)	(14)(2)(35)(6)
(123)(46)(5)	(15)(2)(36)(4)	(1)(24)(35)(6)
(13)(246)(5)	(1)(2)(356)(4)	(15)(2)(3)(46)
(1)(2346)(5)	(16)(24)(3)(5)	(14)(2)(3)(56)
(136)(2)(45)	(14)(26)(3)(5)	(1)(24)(3)(56)
(1345)(2)(6)	(1)(246)(3)(5)	(1)(24)(3)(5)(6)
(1245)(3)(6)	(126)(3)(4)(5)	(12)(3)(4)(5)(6)
(145)(23)(6)	(13)(26)(4)(5)	(1)(2)(3)(4)(56)
(135)(24)(6)	(12)(36)(4)(5)	(1)(2)(3)(4)(5)(6)
(125)(34)(6)	(14)(2)(36)(5)	
(1234)(5)(6)	(1)(24)(36)(5)	
(123)(45)(6)	(1)(23)(46)(5)	

TABLE III

10-lattice Sample of 58 Partitions with P_{lumper} and P_{splitter} Adjoined

(1234567890)	(1)(29)(345)(6)(7)(80)	(1269)(3)(45)(7)(8)(0)
(123)(45)(6)(780)(9)	(1890)(2)(3)(45)(67)	(1)(2479)(3680)(5)
(10)(23458)(679)	(179)(2)(30)(48)(56)	(1)(20)(3)(4)(58)(69)(7)
(1)(2689)(30)(4)(5)(7)	(147)(269)(350)(8)	(168)(27)(3)(450)(9)
(1)(2780)(356)(4)(9)	(1)(24)(3)(59)(6)(780)	(16)(2)(3)(459)(78)(0)
(1236)(49)(578)(0)	(1240)(36)(57)(8)(9)	(1)(2)(39)(45678)(0)
(170)(29)(35)(4)(6)(8)	(13480)(2679)(5)	(15)(2)(3890)(4)(67)
(1)(2)(37)(490)(5)(6)(8)	(1350)(29)(4)(68)(7)	(149)(258)(3)(60)(7)
(18)(2)(35)(4)(6)(7)(9)(0)	(1)(235)(47)(6)(890)	(1)(23)(48)(5679)(0)
(148)(20)(3679)(5)	(15)(2)(36)(4)(7)(89)(0)	(10)(2459)(3)(67)(8)
(157)(239)(46)(80)	(1578)(20)(349)(6)	(14789)(260)(3)(5)
(17)(235)(489)(6)(0)	(1358)(20)(49)(67)	(125)(3790)(46)(8)
(170)(2)(39)(48)(56)	(157)(290)(38)(46)	(148)(2)(35690)(7)
(1370)(2)(4)(5)(6)(8)(9)	(16)(2390)(4578)	(148)(20)(3)(57)(69)
(1350)(268)(49)(7)	(160)(247)(3)(589)	(126)(3)(4)(5)(7)(8)(9)(0)
(1)(28)(37)(46)(50)(9)	(19)(26)(358)(470)	(128)(3470)(59)(6)
(1)(25)(369)(40)(78)	(15)(20)(37)(4)(68)(9)	(1249)(37)(58)(60)
(146)(28)(3)(5)(7)(9)(0)	(1)(2)(3)(4)(560)(7)(89)	(1)(2)(3)(4)(5)(6)(7)(8)(9)(0)
(19)(26)(3)(457)(80)	(135)(26)(4)(7)(89)(0)	
(156)(2)(39)(470)(8)	(1)(2)(3)(468)(590)(7)	
(135)(2)(469)(780)	(1)(2)(3)(45)(60)(78)(9)	

PAIRBONDS, INFOTWO, and MULTINOM are all metrics derived from supervaluations on partition lattices and are discussed in Section 2; INFOTHRY, 1-NT/SQP, 1-NT/1 + 2, INVRHEIT, and NVRNTROP are essentially all normalizations of these metrics and are treated in Section 3; APPROX and TI2MINUS are characterized by a high proportion of interpoint distance ties and are analyzed in Section 4, while Section 5 handles 1-LAMDA and CHISQUAR, which are simple linear transformations of the two classical measures of association from which their names are taken.

Given the partition spaces we have described, the actual distance values for the 12 measures were computed for each space by means of a FORTRAN IV program written by the authors.[1] This program produced halfmatrices of distances which were

[1] Copies of this program are available from Phipps Arabie upon request.

then used as input data for the programs in the MDSCAL series. Finally, the resulting scaling configuration was rotated to congruence with a given target configuration by means of a program of D. C. Olivier [Olivier (1970) gives a description of the mathematical foundations of this least-squares procedure]. This rotation was performed only for Euclidean solutions and was designed to orient the natural lattice dimension from P_{lumper} to P_{splitter} along the x-axis for all plots and, hence, permitted comparison of metrics more easily than would otherwise have been the case.[2]

Computation of some of the 12 measures revealed degeneracies in the multidimensional scalings owing to their behavior on the lattice endpoints, P_{splitter} and P_{lumper}.

TABLE IV

Structural Measures of Partition Distance

In what follows, we employ the following abbreviations and conventions (where $|\ |$ is used to denote the cardinality of a set, not absolute value in the arithmetic sense):

$$n = |\ S\ | \text{ (size of } S);$$

$$P = \{c_1, c_2, \ldots, c_m\}, \text{ where the } c_i \text{ are the cells of } P;$$

$$Q = \{d_1, d_2, \ldots, d_p\}, \text{ where the } d_j \text{ are the cells of } Q;$$

$$P \cap Q = \{z_{ij}\};$$

$$m = |\ P\ |,\ p = |\ Q\ |;\ |\ P \cap Q\ | = \text{size of } P \cap Q \text{ (number of nonempty } z_{ij});$$

$$D(P) = \sum_i \binom{|\ c_i\ |}{2};$$

$$M(P) = \sum_i \log_2(|\ c_i\ |!);$$

$$E(P) = -\sum_i \frac{|\ c_i\ |}{n} \log_2 \frac{|\ c_i\ |}{n};$$

$$V(P) = \log_2 n - E(P);$$

$$|\ c\ |_{\max} = \max\{|\ c\ | : c \in P\} \text{ (largest cell size in } P);$$

$$|\ d\ |_{\max} = \max\{|\ d\ | : d \in Q\} \text{ (largest cell size in } Q);$$

$$|\ z_i \cdot\ | = \max\{|\ z_{ij}\ | : d_j \in Q\};$$

$$|\ z \cdot_j\ | = \max\{|\ z_{ij}\ | : c_i \in P\}.$$

Table continued

[2] Specifically, the one-dimensional target configuration used in the Olivier program was obtained by associating each partition in the lattice with its cardinality, which is a crude but, in this instance, effective quantification of the lattice dimension which orders partitions by fineness as in Definition 1.1.

TABLE IV (*continued*)

Name of measure	Defining formula
PAIRBONDS	$D(P) + D(Q) - 2D(P \cap Q)$
INFOTWO	$V(P) + V(Q) - 2V(P \cap Q)$
MULTINOM	$M(P) + M(Q) - 2M(P \cap Q)$
INFOTHRY[a]	INFOTWO/$E(P \cap Q)$
1 − NT/SQP[b]	$1 - [D(P \cap Q)]/[D(P)\, D(Q)]^{1/2}$
1 − NT/1 + 2[c]	$1 - \dfrac{2D(P \cap Q)}{D(P) + D(Q)} = \dfrac{\text{PAIRBONDS}}{D(P) + D(Q)}$
APPROX	$\lvert P \cap Q \rvert - \min(\lvert P \rvert, \lvert Q \rvert)$
1 − LAMBDA	$\left[\left(2n - \sum_i \lvert z_{i\cdot} \rvert - \sum_j \lvert z_{\cdot j} \rvert\right) \Big/ 2n - \lvert c \rvert_{\max} - \lvert d \rvert_{\max}\right]$
TI2MINUS	$2 \lvert P \cap Q \rvert - \lvert P \rvert - \lvert Q \rvert$
CHISQUAR[d]	$1 - [\chi^2(P, Q)/(n \min(\lvert P \rvert - 1, \lvert Q \rvert - 1))]$ $= 1 - \left[\left(\sum_{i,j} \dfrac{\lvert z_{ij} \rvert^2}{\lvert c_i \rvert \cdot \lvert d_j \rvert} - 1\right) \Big/ \min(\lvert P \rvert - 1, \lvert Q \rvert - 1)\right]$
INVRHEIT[a]	$\left[\text{PAIRBONDS} \Big/ \left(\binom{n}{2} - D(P \cap Q)\right)\right]$
NVRNTROP[e]	$[\text{INFOTWO}/(2 \log_2 n - E(P) - E(Q))]$

[a] The denominator vanishes if and only if $P = Q = P_{\text{lumper}}$, and in this case we define INFOTHRY and INVRHEIT to be 0.

[b] If $D(P)$ or $D(Q)$ is 0 then either P or Q is P_{splitter}. In this case, we define 1 − NT/SQP from P_{splitter} to itself to be 0, and between P_{splitter} and any other partition to be 1.

[c] The denominator vanishes if and only if $D(P) = D(Q) = 0$, and hence $P = Q = P_{\text{splitter}}$. In this case we define 1 − NT/1 + 2 to be 0.

[d] As in (b): CHISQUAR from P_{lumper} to itself is 0, and from P_{lumper} to any other partition is 1.

[e] The denominator vanishes if and only if $P = Q = P_{\text{splitter}}$, and in this case, we define NVRNTROP to be 0.

Specifically, various of the normalized measures place either P_{splitter} or P_{lumper} a maximum (unit) distance from any other point. The result is to make the designated endpoint behave somewhat like the point at infinity in standard metrizations of the Riemann sphere, for example, stereographic projection. This phenomenon leads to degeneracies when we try to scale the space with the given distance measure, since

the algorithm attempts to place the endpoint at a maximum (unit) distance from all other points, subject to the constraint (Kruskal, 1964b)

$$\sum_{\mathscr{P} \in \mathscr{L}} \| x_{\mathscr{P}} \|^2 = 1,$$

so that the remaining points collapse into a cluster whose radius approaches zero as iterations continue. In order to eliminate this difficulty, we considered various subspace scalings involving removal of one or both endpoints. A complete list of the degeneracies is given in Table V. Thus, in the case of a measure on a given lattice with no endpoint

TABLE V

Endpoint Degeneracies for the Distance Measures of Table IV

Name of measure	Both P_{lumper} and P_{splitter}	Without P_{lumper}	Without P_{splitter}	Without either
PAIRBONDS				
INFOTWO				
MULTINOM				
INFOTHRY	*		*	
1 − NT/SQP	*	*		
1 − NT/1 + 2	*	*		
APPROX				
1 − LAMBDA	*		*	
TI2MINUS				
CHISQUAR	*		*	
INVRHEIT	*		*	
NRVNTROP	*	*		

Note—A star (*) indicates that with the indicated endpoint(s) present the configuration degenerates because one or the other of the endpoints present is maximal (unit) distance from all other points in the space.

degeneracies, we computed up to four scalings for each choice of dimensionality and output metric (compare Table I). Even where there were no degeneracies, these alternative scalings provided a useful record of the extent to which the endpoints dominate the behavior of the scaling solutions; see Section 2.

Various MDSCAL parameter options were tried, involving choice of initial configuration (L-, random, or rational), dimension (2- or 3-), and metric of the scaling

solution (city-block or Euclidean).[3] In all cases scaled, the stress formula employed was stress formula 1, which is the original badness-of-fit function suggested by Kruskal (1964a, 1964b; see also Kruskal and Carroll, (1969)).[4] There is a published literature on the expected behavior of stress formula 1 which is lacking for stress formula 2; see, e.g., Klahr (1969) and Stenson and Knoll (1969). In all cases scaled, moreover, we employed the primary stress approach to ties, which does not treat tied interpoint distances as constraints, as opposed to the secondary stress approach option, which accepts ties as constraints (Kruskal, 1964b). In view of the large number of interpoint distance ties given by our distance measures (see Appendix B), this restriction to the primary stress approach should be emphasized. As far as input configurations are concerned, the standard L-configuration frequently led to local minima, and in certain cases to quite severe ones, with an arrangement of points which was considerably less interpretable than that produced by the algorithm from a random initial configuration. Consider a particular 5-tuple ⟨lattice, presence/absense of endpoint(s), input distance measure, output dimensionality, Minkowski r⟩, which completely describes a scaling run except for selection of the initial configuration. Of the 154 such 5-tuples initially scaled with the L-configuration and then rescaled with a random initial configuration (in some cases more than once), 73 achieved stress reduction in the best of the random solutions when compared with the L-configuration solution. In the 41 of such 5-tuples rescaled more than once with random initial configuration(s), the overwhelming majority (37) achieved stress reduction. Most of the repeated rescalings were done on the 6-lattice with the Euclidean metric or on the 10-lattice with the city-block and Euclidean metrics. Table VI describes relevant statistics of average stress reduction for these cases. Since, for the data considered here, random initial configurations give appreciable decrease in stress (typically an absolute

[3] The L-initial configuration for the MDSCAL gradient algorithm is obtained in the following way, where d is the dimensionality imposed on the scaling solution and r is the number of points being scaled. We form the following list of vectors in d-space:

.01 (1, 0,..., 0)
.01 (0, 1,..., 0)
⋮
.01 (0, 0,..., 1)
.01 (2, 0,..., 0)
.01 (0, 2,..., 0)
⋮

and select the first r vectors in the sequence as the starting configuration.

[4] In Kruskal's notation, stress formula 1 is $[\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2/\sum_{i<j} d_{ij}^2]$; stress formula 2 is $[\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2/\sum_{i<j}(d_{ij} - \bar{d})^2]$, where $\bar{d} = \sum_{i<j} d_{ij}/\binom{n}{2}$ and n is the number of points in the scaling.

TABLE VI

Reduction of Stress Values by Rescaling with Random Initial Configuration; Dimensionality = 2

	Number of 5-tuples[a]	Stress reduction[b]
6-lattice, Euclidean metric	22	3.7%
10-lattice, Euclidean metric	13	2.0%
10-lattice, city-block metric	10	1.2%

[a] This is the number of ordered pairs ⟨input distance measure, presence/absence of endpoint(s)⟩ which were repeatedly rescaled for given lattice, output dimensionality, and Minkowski r.

[b] This is given by the formula:

$$S = \frac{1}{K} \sum \max(0, S_L - \min\{S_{\text{random}}\}),$$

where summation is over all ordered pairs considered in the preceding note, K is the total number of such pairs (given by entries in the first column of the table), S_L is the L-configuration stress obtained for a given pair, and $\min\{S_{\text{random}}\}$ is the minimum stress obtained over all scalings with random initial configurations. For the 6-lattice with $r = 2$ the average number of rescalings over which this minimum was taken was 2.2; for the 10-lattice with $r = 2$ it was 2.6; for the 10-lattice with $r = 1$ it was 3.8.

reduction of 1–3%), it may well be the case that other data analyzed with the L-configuration, including that of the two most useful Monte Carlo studies, Klahr (1969) and Stenson and Knoll (1969), also have inflated stress values and fail to achieve the configuration associated with minimum stress.

In addition to random configurations, a version of MDSCAL-3 was run, which was designed so as to have inputs obtained from a Torgerson metric analysis of the original data; see Young and Torgerson (1967). The objective was to test the hypothesis that a metric principal components analysis of the proximity matrix might provide a starting configuration for the nonmetric gradient algorithm more nearly optimal than either the L-configuration or a random one. In the eight cases scaled, however, the obtained stress was considerably *higher* than that produced by the L-configuration and a random one; in consequence, for data of the type we are considering, a Torgerson metric input seems by no means necessarily better than the L- or random initial configurations.

2. Comparison of Three Basic Measures

The present section is devoted to detailed discussion of multidimensional scaling results for the three distance measures PAIRBONDS, INFOTWO, and MULTINOM;

see again Table IV for defining equations. These three measures are each derived from a particular supervaluation on the partition lattice $\mathscr{P}(n)$ on a set of n elements, i.e., a real-valued function v on $\mathscr{P}(n)$ satisfying the supervaluation inequality

$$v(P) + v(Q) \leqslant v(P \cup Q) + v(P \cap Q). \tag{1}$$

Any supervaluation v which is strictly increasing in the lattice ordering[5] can be shown (Boorman and Arabie, 1972) to induce a metric μ on $\mathscr{P}(n)$ by a distance formula which is formally reminiscent of the law of cosines

$$\mu(P, Q) = v(P) + v(Q) - 2v(P \cap Q). \tag{2}$$

The three distance measures under consideration (PAIRBONDS, INFOTWO, and MULTINOM) are each induced in this way from supervaluations which are labeled D, V, and M, respectively, in Table IV. Of the three measures, PAIRBONDS is the most frequently used in the literature and has been introduced in a wide variety of statistical applications, e.g., Johnson (1968); Bersted, Brown, and Evans (1970), Mirkin and Chorny (1970); and Rand (1971). INFOTWO is a standard distance metric of information theory (Csiszár, 1967; Parry, 1969; Jardine and Sibson, 1971, p. 12 ff; Boyd, 1972). MULTINOM is a variant information measure whose behavior may be compared with INFOTWO by results of Boulton and Wallace (1969). Both INFOTWO and MULTINOM are closely related to a variety of other information-theoretic measures connected with correlation, redundancy, and other structural concepts (Kotz, 1966; Pearson, 1966).

In order to examine the output configurations generated by the three metrics, we need to consider the structure of the partition lattice in somewhat greater detail.

DEFINITION 2.1. The *type* of a partition $P = \{c_1, c_2, \ldots, c_m\}$ is the distribution of its cell sizes $\{| c_i |\}$.

For example, the *type* of $((ab)(cd)(e))$ can be coded as 2-2-1, which is, of course, one partition of the integer 5, since $2 + 2 + 1 = 5$. As further examples, we note that the type of P_{lumper} for the n-lattice is just n, while the type of splitter is

$$\underbrace{1 - 1 - \cdots - 1}_{n \text{ times}}.$$

For the 5-lattice, we display all possible types in Table VII; for any positive integer n there are well-known algorithms for generating all corresponding types (Wells, 1971; see also Appendix B).

The significance of considering types is evident from Fig. 1, which displays the 5-lattice scaled with the PAIRBONDS measure with both P_{lumper} and P_{splitter}

[5] I.e., for which $P < Q \Rightarrow v(P) < v(Q)$.

TABLE VII

Partition Isomorphism Types for the 5-lattice

Type	Number of distinct partitions having that type	Code	Height ($D(P)$ in Table IV)
5	1	*A*	10
4–1	5	*B*	6
3–2	10	*C*	4
3–1–1	10	*D*	3
2–2–1	15	*E*	2
2–1–1–1	10	*F*	1
1–1–1–1–1	1	*G*	0

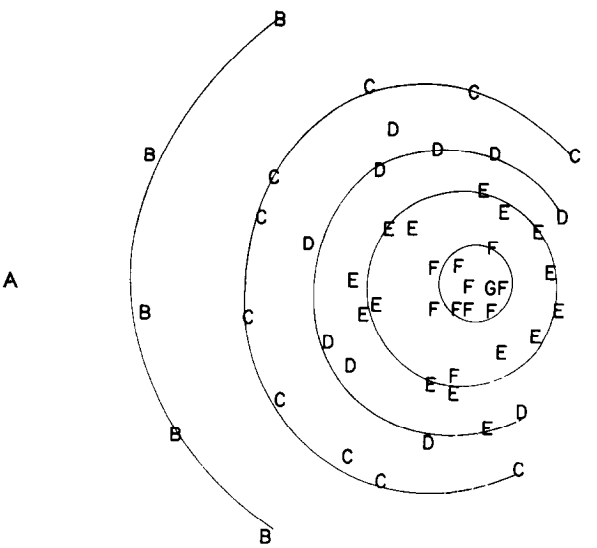

FIG. 1. Scaling of PAIRBONDS on 5-lattice with both endpoints present. 2 dimensions, Euclidean metric, *L*-configuration. Stress formula 1 = 23.9%. The points are labeled by type according to the coding in Table VII.

present. The points are coded by type, following the notation of Table VII. There is clearly good recovery of the lattice types as contour lines in the scaling representation.

The algorithm behaves much as though P_{splitter} and P_{lumper} were opposing attractors, each trying to organize the types around itself as circular isosimilarity contours.

The terminology here is reminiscent of McGee (1966), but the motivation at present is purely geometrical. Hence, in the multidimensional scalings we have a clear conflict between the "attraction" exerted by P_{lumper} and the conflicting attraction exerted by P_{splitter}. The P_{splitter} attractor is more effective than the P_{lumper} one (we shall see the opposite effect for certain normalizations of PAIRBONDS in Section 3), though the effect of P_{lumper} is evident in rendering the outer contours essentially elliptical with the points on these contours contracted on a limited arc. These effects vanish when we remove P_{lumper} and the contours become very nearly circular with points of the same type spaced about equidistantly on each contour. Note, moreover, that the radial distances of the contours from G (P_{splitter}) as origin fall very nearly in the proportions 10 : 6 : 4 : 3 : 2 : 1. These proportions coincide with the ratios of the heights of the partitions of the different types (see Table 7). Algebraically, this invariance may be restated by saying that, for all partitions P and Q in the lattice, the scaling algorithm is attempting to produce a solution satisfying the equations

$$\text{PAIRBONDS}\,(P, P_{\text{splitter}})/\text{PAIRBONDS}\,(Q, P_{\text{splitter}}) = \| v\,(P)\|_2/\| v\,(Q)\|_2\,, \qquad (3)$$

where $v(P)$, $v(Q)$ are the images of P and Q in the two-dimensional Euclidean scaling.

The only essential errors in the contour recovery are one "E" type which is misplaced close to what should be the D-contour and one "F" type on the E-contour. It is also noteworthy that the recovery appears to become more elastic as we move toward P_{splitter} (F-contour), where there are many relatively small distances to be taken into account (it is a general characteristic of partition lattices that the lattices are more dense near P_{splitter} and, moreover, small distances are clearly less important in the stress function than are large ones). When P_{splitter} is removed, the basic scaling is similar but the contours near the missing P_{splitter} attractor tend to become somewhat more blurred, thus further confirming the structural role of the splitter partition.

As we shall see consistently later, this contour structure of the lattice scalings tends to be a general fact underlying the interpretation of the solutions. One consequence of this contour organization was that when hierarchical clustering [see Johnson (1967) for both the connectedness and the diameter method] was applied to the same input proximity data on which the scaling in Fig. 1 was based, as well as to numerous other lattices and distance measure combinations, the results were typically degenerate in the sense that the clusterings were weak and uninterpretable.

On page 158 we noted that for some of the normalized measures, the removal of the endpoint(s) is a prerequisite for obtaining nondegenerate solutions. This constitutes a dramatic example in support of Nelson Goodman's (1972) objections to the strong demands of Luce and Galanter (1963) for stability of a spatial representation when additional points must be incorporated. The present analysis of the obtained scalings in terms of the structural role of the endpoints is a further justification of the Goodman position.

Figures 2 and 3 display additional PAIRBONDS plots for the 10-lattice sample for

two different initial configurations. Again, the basic observation to make is that contour recovery in both cases is surprisingly good (we have indicated roughly where the "complete" contours should fall if we increased the sample size), especially since the random nature of the partitions scaled eliminates possible peculiarities owing to the specialized symmetries possessed by the full n-lattice, which as an *algebraic* structure has an automorphism group containing at least the full symmetric group and, hence, $n!$ elements; this algebraic structure is related in detail to the metric structure of the full lattice in Boorman (1970). It is probable, however, that the general phenomenon

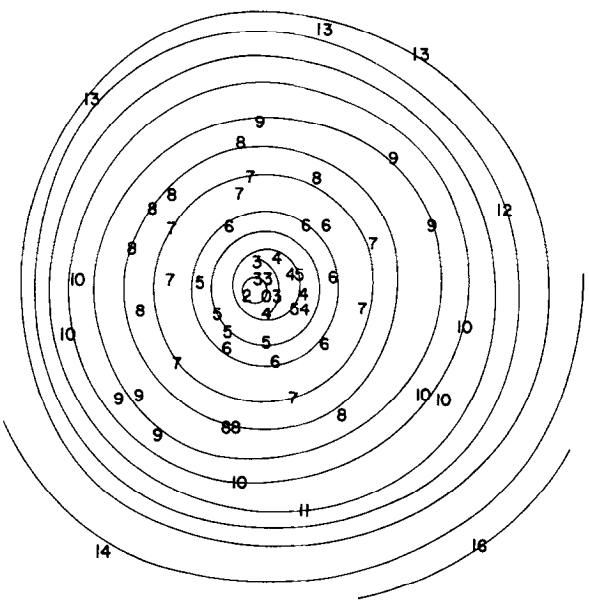

FIG. 2. Scaling of PAIRBONDS on 10-lattice sample with P_{lumper} removed and P_{splitter} present. 2 dimensions, Euclidean metric, L-configuration. Stress formula 1 = 28.4%. The points are labeled by the height $D(P)$ of the corresponding partition (compare Table IV for definition of D).

which we are observing here is a consequence of the fact that the metrics PAIRBONDS, INFOTWO, and MULTINOM (and, in fact, all the measures we consider in the present study) are *permutation-invariant* in the sense that any relabeling of the elements of S leaves distances in the partition lattice on S invariant; see Boorman (1970). In fact, it can be shown that in the case of any of the three metrics, the orbits of the

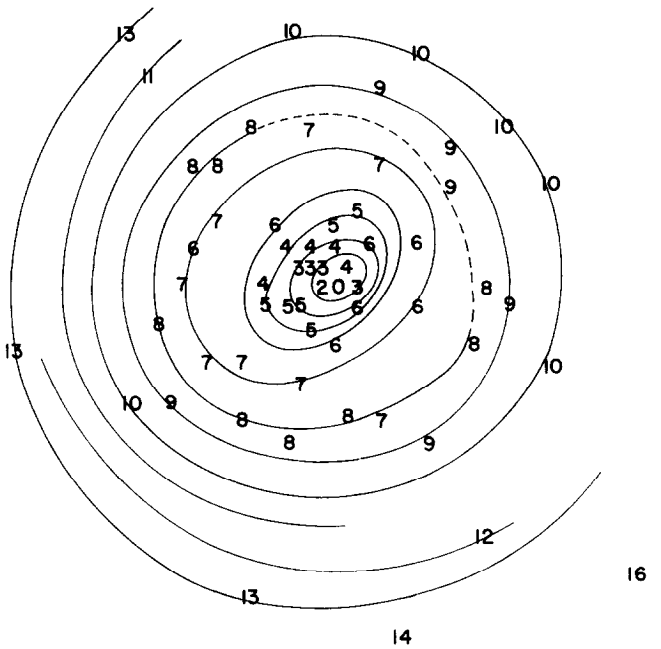

FIG. 3. Scaling of PAIRBONDS on 10-lattice sample with P_{lumper} removed and P_{splitter} present. 2 dimensions, Euclidean metric, random initial configuration. Stress formula 1 = 27.5 %. Points labeled as in Fig. 2.

autometry group of the metric will be precisely partitions of the same type.[6] One way of describing Figs. 2 and 3 would be to say that the scaling algorithm is mapping partitions of a given type into orbits of the autometry group of Euclidean space with a fixed origin (namely, circles).[7]

Juxtaposing the two plots in Figs. 2 and 3, an additional feature is clear from direct inspection and can be formally verified by rotating the two plots to maximum congruence by the algorithm of Olivier (1970). Both plots recover the contour structure of partition types to about the same degree; but within contour lines the recoveries are

[6] For any metric space M with metric μ, an autometry $\mathcal{A}$ of M is a bijection $\mathcal{A} : M \to M$ which is distance-preserving, i.e., for which $\mu(x, y) = \mu(\mathcal{A}x, \mathcal{A}y)$, for all $x, y \in M$. The set of all autometries clearly forms a group under composition. That the autometry group of all our supervaluation-induced metrics is *precisely* given by relabeling the elements of the underlying set being partitioned is not completely trivial, and was proved by Jónsson (personal communication).

[7] Another study of the present authors suggests that a general feature of the MDSCAL algorithm in Euclidean space appears to be this observed priority given to the recovery of symmetries possessed by the input proximity matrix.

very different. The stresses obtained are very similar, to about 0.9% discrepancy in favor of the latter initial configuration. Thus, unless we want to take stress values very seriously indeed—which goes against much additional evidence that has accumulated in the literature—we must conclude that the MDSCAL solutions are essentially nonunique for the data considered here, being interpretable only up to recovery of contours [compare remarks on multiple solutions in Torgerson (1965, pp. 381–382)]. Despite the restricted context of the present investigation, this finding is worth emphasizing in view of the tendency of many users of multidimensional scaling to take their scaling solutions seriously at the level of individual points rather than merely at the level of clusters and contours [compare the similar emphasis of Shepard (1969, pp. 50 ff.)].

As compared with the PAIRBONDS measure, of which we have been considering various scalings, INFOTWO does not recover lattice types as contours to nearly such a perfect extent. One particular plot for the 10-lattice sample is shown in Fig. 4,

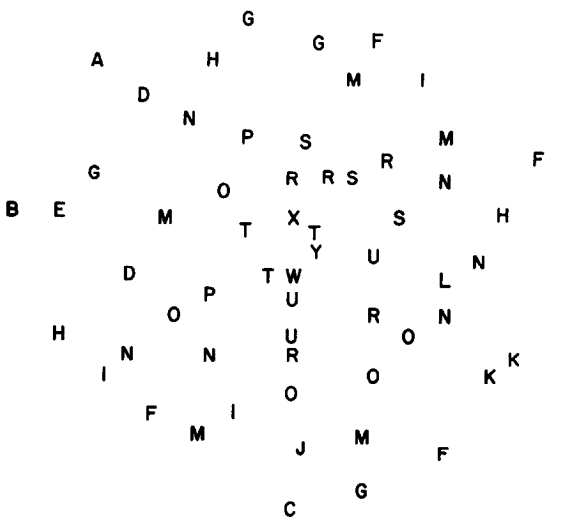

FIG. 4. Scaling of INFOTWO on 10-lattice sample with P_{lumper} removed and P_{splitter} present. 2 dimensions, Euclidean metric, random initial configuration. Stress formula 1 = 32.2%. Points are labeled by entropy according to the following code:

$A = 1.36$	$F = 1.85$	$K = 1.97$	$P = 2.32$	$U = 2.72$
$B = 1.49$	$G = 1.87$	$L = 2.07$	$Q = 2.37$	$V = 2.82$
$C = 1.52$	$H = 1.90$	$M = 2.12$	$R = 2.45$	$W = 2.85$
$D = 1.69$	$I = 1.92$	$N = 2.17$	$S = 2.52$	$X = 2.92$
$E = 1.72$	$J = 1.96$	$O = 2.25$	$T = 2.65$	$Y = 3.32$

where each value represents the entropy $E(P)$ [$= -\Sigma(|\, c_i \,|/n) \log_2(|\, c_i \,|/n)$] of the corresponding partition.

which labels partitions in the scaling according to the entropy. The recovery of contours is clearly fuzzy at best, and similar fuzziness prevails in other parameter cases scaled with INFOTWO. For example, Fig. 5 illustrates the 5-lattice scaling analogous to that for PAIRBONDS in Fig. 1. Once again, recovery of contour structure is considerably inferior to that of PAIRBONDS. This inferior recovery, which is quite a general phenomenon over lattice sizes and parameter configurations, suggests that INFOTWO may be considerably less useful for scaling purposes than is PAIRBONDS.

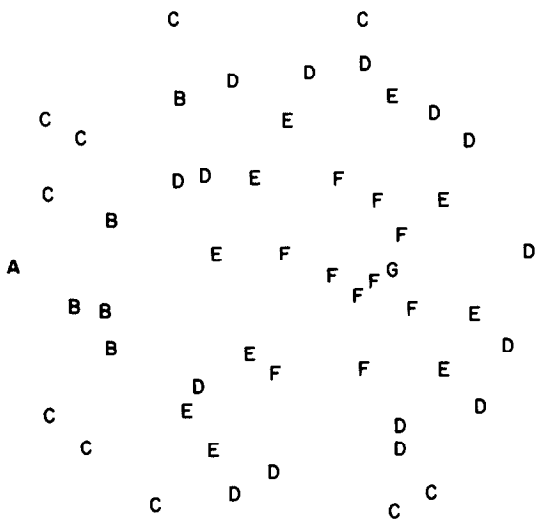

FIG. 5. Scaling of INFOTWO on 5-lattice with both endpoints present. 2 dimensions, Euclidean metric, L-configuration. Stress formula 1 = 31.8%. Points are labeled by type according to Table VII.

A more theoretical analysis sheds additional light on the comparative behavior of the two metrics. Following Boorman and Arabie (1972), consider the PAIRBONDS distance generated by moving a subset of size s from a cell of size k to another cell of size k'. The three parameters s, k, and k' are sufficient to compute this distance, which is

$$2((k - s)s + k's), \tag{4}$$

and, for s small relative to k, we have (4) reducing to

$$2(k + k')s, \tag{5}$$

which is linear in all three parameters taken separately. By contrast, in the case of the same problem for INFOTWO, the distance generated is

$$\frac{1}{n}\left[\log_2 \frac{k^k(k'+s)^{k'+s}}{k'^{k'}(k-s)^{k-s}s^{2s}}\right]. \tag{6}$$

It can be shown that if s is small relative to k and k', then (6) is asymptotic to

$$(1/n)(\log_2 k + \log_2 k') \tag{7}$$

for large k and k'. Comparison of (5) with (7) shows that, by contrast to PAIRBONDS, INFOTWO shows convex behavior (decreasing returns to scale) in both k and k'. Consequently, it can be expected to discriminate less clearly than PAIRBONDS when k and k' are large, i.e., in the region of the lattice near P_{lumper} and the most important region from the point of view of the way in which distances enter into the target function. This blurring effect is quite clear in the outer contours of Figs. 4 and 5.[8]

It can also be shown (Boulton and Wallace, 1969; Boorman and Arabie, 1972) that MULTINOM behaves asymptotically like INFOTWO where expressions analogous to (5) and (7) are concerned. More generally, MULTINOM is asymptotically ordinally similar to INFOTWO, in the sense that as lattice size increases the induced interpoint distance orderings of the two measures become arbitrarily similar in the rank-order correlation sense. Hence, on mathematical grounds we can expect that MULTINOM will give a blurring of contours similar to INFOTWO for lattices on perhaps ten or more elements. In fact, however, the failure of MULTINOM to discriminate is already clear for the 5-lattice, as is demonstrated by Fig. 6.

We next consider Euclidean solutions in three dimensions. For even the most interpretable of the 2-dimensional cases for example, PAIRBONDS on the 5-lattice as shown in Fig. 1, the corresponding 3-dimensional solutions were not easily interpretable in spite of the fact that Olivier's rotation procedure was employed to orient the solutions along the natural lattice axis from P_{lumper} to P_{splitter} see p. 156 and footnote 2. Discrimination of contour surfaces formed by different partition types ranged from poor to totally unrecognizable. Specifically, after the 3-space solution was rotated by Olivier's algorithm, none of the three 2-dimensional projections perpendicular to an axis proved interpretable. In consequence, it appears that an escalation of dimensionality in this case actually decreases the effectiveness of the procedure. This supports an observation of Shepard (1969) to the effect that if an MDSCAL configuration is interpretable, then it is almost certainly interpretable in two dimensions. The only

[8] It should also be noted, however, that precisely the linear behavior (5) which effectively discriminates types also causes the PAIRBONDS scalings to degenerate on the 10-lattice sample where P_{lumper} is present, while INFOTWO does not degenerate in this case. If, however, P_{lumper} is removed, the behavior of PAIRBONDS is more satisfactory than that of INFOTWO, as earlier figures demonstrate.

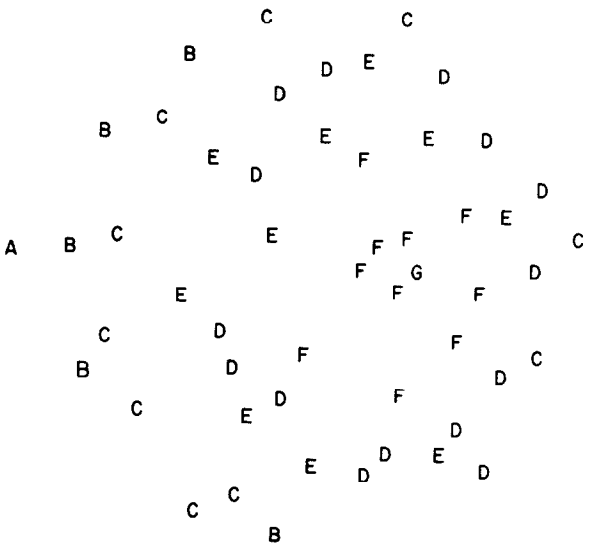

FIG. 6. Scaling of MULTINOM on 5-lattice with both endpoints present. 2 dimensions, Euclidean metric, L-configuration. Stress formula 1 = 31.1%. Points are labeled by type according to Table VII.

exception to this failure of interpretability was in the 4-lattice case, where there were at most only 15 points in the scalings and the 3-dimensional solutions took approximately the form of an ice-cream cone, with P_{splitter} at the center of the hemisphere and P_{lumper} at the point of the cone. This result did not, however, generalize to the 5-lattice, where only in the PAIRBONDS case was a provisionally acceptable solution of ice-cream cone type obtained, and this solution manifested considerable noise.

Given the current revival of interest in non-Euclidean solutions, e.g., Arnold (1971), Fischer and Micko (1972) and Shepard, and Cermak (1973), we were interested in exploring such representations for the partition metrics. In spite of the early interest in the city-block metric (Attneave, 1950), it is probably fair to say that most psychological investigations have been restricted to Euclidean spaces. Shepard (1969, p. 34) argued that this limitation (for continuous spaces) was not objectionable:

> As long as there is a continuous underlying space of well-defined dimensionality that is at all isotropic and uncontorted, anyway, the data so over-determine the representation that the erroneous assumption of a Euclidean metric will still permit a satisfactory recovery of the true underlying structure and, indeed, even a determination of the nature of the unknown metric.

In support of Shepard's conjecture, Arnold (1971) has pointed out that Henley's (1969) data on semantic (discrete) spaces evince the ($r = 1$, $r = \infty$) unit disc quite

strikingly, despite the fact that she imposed a Euclidean metric on the solution. The same observation holds for data from a study by Nummenmaa (1964) on the perceived similarity of 21 facial expressions. For technical reasons, Nummenmaa included only eight of the stimuli in his analysis; a reconsideration of these eight stimuli by Ekman (1970), whose analysis assumed linearity and a Euclidean space, is reproduced here as Fig. 7 and again suggest that an underlying non-Euclidean metric may be appropriate.

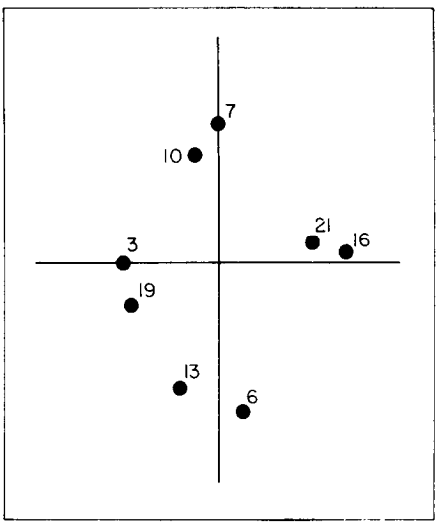

FIG. 7. Scaling solution for a study of the similarity of eight facial expressions (Nummenmaa, 1964). The analysis in the figure is from Ekman (1970).

The fact that a scaling solution shows a pattern of squares or diamonds symmetrically placed around some central point does not, of course, constitute a rigorous demonstration that the "natural" metric of the space being scaled is non-Euclidean; in fact, as is the case with partition lattices, there is no sense in which we can characterize the behavior of a distance measure from a unit ball, and, hence, the algorithm for recovering the unit ball for Euclidean scalings of Minkowski spaces described in Shepard (1969) is inapplicable. Despite these caveats, we may again focus on the contour lines defined by partition types in the scaling representations and argue from the geometry of these contours that the partition data conform quite nicely to the preferences of an investigator who has the choice of a Euclidean or a city-block solution.

Specifically, in Fig. 5 the choice of $r = 2$ gives slightly noisy circles evident of an underlying Euclidean space; in the typical city-block plot shown in Fig. 8 the choice of $r = 1$ gives for the same data a clear pattern of diamonds, indicative of the city-block metric, even though considerable confusion of contours occurs. This pattern of

diamond-like contour recovery occurred again and again in the $r = 1$ scalings undertaken in the course of the present study, and usually allowed for identification of the imposed metric even before points were labeled. (The pattern could also be representative of the supremum metric [$r = \infty$], since the supremum and city-block metrics cannot be distinguished in imposed Euclidean 2-dimensional solutions, as pointed out by Arnold (1971)).

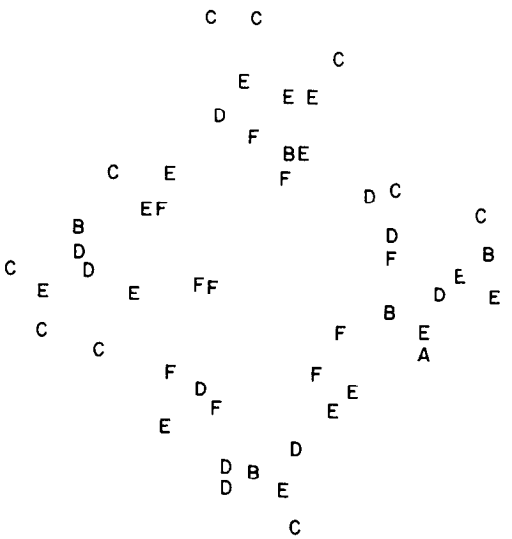

FIG. 8. Scaling of INFOTWO on 5-lattice with P_{lumper} present and P_{splitter} removed. 2 dimensions, city-block metric, random configuration. Stress formula 1 = 36.8%. Points are labeled by type according to Table VII.

In general, we conjecture that for any given measure of partition distance which is permutation invariant, the scaling algorithm will recover the partition space in such a way that the orbits of the automorphism group of the lattice are mapped as well as possible into isometric contours in the output metric. In short, the artificial data considered here adapt to whatever output metric the investigator chooses to impose. Thus, it would seem that a representation theorem is needed to decide when the data allow recovery of the underlying metric, as conjectured by Shepard and others, e.g., Shepard (1969) and Wender (1969). The theoretical significance of such a recovery has been demonstrated by the elegant work of Cross (1965a, 1965b) in multidimensional stimulus generalization.

3. Alternative Normalizations

Normalizations of structural measures can be used for essentially two different purposes: to provide a uniform range permitting meaningful comparison between measures of distance on lattices of different sizes; or to factor out the effects of an unwanted dimension of variability, thus, in effect, providing a conditional measure of the remaining structure (the classic case being conditional probability). In the present study, we considered five different normalizations of PAIRBONDS and INFOTWO. The measures $1 - NT/1 + 2$ and $1 - NT/SQP$ are essentially suggested in Johnson (1968). Johnson (personal communication) has proved by a Euclidean embedding argument that $1 - NT/SQP$ is a metric; the metric status of $1 - NT/1 + 2$ remains uncertain, though the behavior of the two measures is highly similar, and they are closely related as arithmetic to geometric mean, see again Table IV. The distance measure INFOTHRY is developed in some detail by Rajski (1961), who proves the triangle inequality by an intricate argument, see also Kotz (1966); Boorman (1970) gives a proof of the triangle inequality by a lattice-theoretic argument. INVRHEIT and NVRNTROP do not seem to have appeared in the literature.

These alternative normalizations are all designed to study the effect of controlling for the natural (lumper–splitter) dimension which orients the partition lattice. The five measures studied are organized in Table VIII, which classifies them by underlying measure (PAIRBONDS or INFOTWO) and by the normalization strategy employed.

TABLE VIII

Alternative Normalizations of Supervaluation-based Partition Metrics

	Underlying supervaluation (θ)	
Strategy	Entropy (V)	Height (D)
Divide by $(\theta(P_{\text{lumper}}) - \theta(P \cap Q))$	INFOTHRY	INVRHEIT
Divide by $\theta(P) + \theta(Q)$	NVRNTROP	(arithmetic mean) $1 - NT/1 + 2$; (geometric mean) $1 - NT/SQP$

The main result of scaling these derived measures is verifications of the fact that the strategy of normalization may be very important in controlling the behavior of the derived measure. In some cases (INFOTHRY and INVRHEIT), the effect of

an apparent normalization is essentially negligible in the scaling output. In other cases, the result of normalization is dramatic, and the resulting measure must be interpreted as measuring a quantity fundamentally different from the original one (NVRNTROP, 1 — NT/SQP, 1 — NT/1 + 2). This is already evident from the rank-order correlations of Table IX, where these measures provide the only *negative* value in the rank-order correlation matrices. We will henceforth call normalizations of the first kind *mild* and those of the second kind *strong*.

At the level of the MDSCAL representations, we find that the strong normalizations make a kind of inversion of the data, turning the MDSCAL plots inside out (compare Fig. 9). The effect is similar to the conformal mapping $\zeta \to 1/\zeta$ in the complex plane,

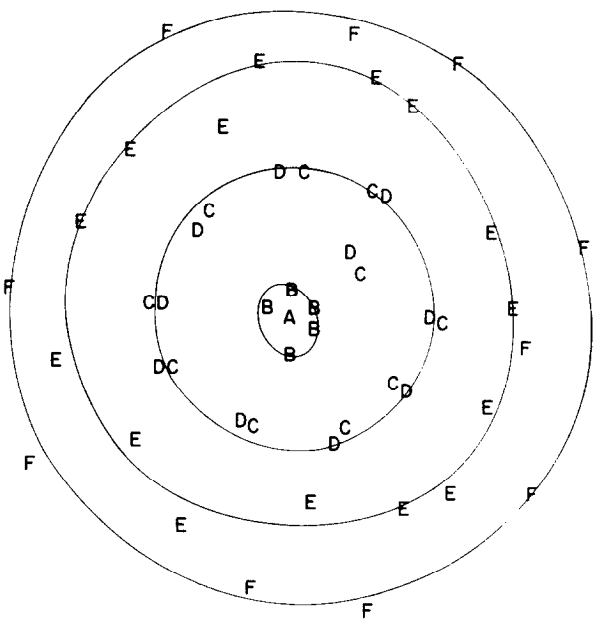

FIG. 9. Scaling of 1 — NT/1 + 2 on 5-lattice with P_{lumper} present and P_{splitter} removed. 2 dimensions, Euclidean metric, L-configuration. Stress formula 1 = 20.0%. Points are labeled by type according to Table VII.

which again suggests an analogy between the lattice endpoints and the point at infinity or, dually, zero. The result is that the contour lines nearest P_{splitter} are now outermost and those nearest P_{lumper} are innermost; again (as previously) the outer contours are regularly recovered.

The explanation of this effect may be obtained by analyzing the normalization procedure in more detail. In the case of the first strategy in Table VIII, which leads

to mild normalizations, as P and Q increase toward P_{lumper} in the lattice ordering, $P \cap Q$ likewise approaches P_{lumper}. The function θ is monotone increasing, hence $\theta(P_{\text{lumper}}) - \theta(P \cap Q)$ is monotone decreasing toward 0 as $P \cap Q \to P_{\text{lumper}}$. Consequently, distances between partitions at the P_{lumper} end of the lattice will be expanded relative to distances at the P_{splitter} end in comparison with the unnormalized metric. This expansion in the region of P_{lumper} is, however, merely an intensification of a similar dilation already present in the unnormalized distances (compare again Figs. 1–6), and again P_{splitter} will be placed at the origin and P_{lumper} at the periphery of the plot. The result is that we would expect INFOTHRY to recover contour lines better than INFOTWO or MULTINOM for the same reason that PAIRBONDS does: namely, it inflates distances more than do INFOTWO and MULTINOM as we move up in the lattice; in fact, the scaling results strongly bear out this presumption. In the case of the strong normalizations, by contrast, the denominator will *increase* as P and Q approach P_{lumper}. Thus, a countervailing influence to the dilation of distances at the upper end of the lattice is established, and, in fact, is sufficiently strong so that the P_{lumper} region is actually contracted relative to the P_{splitter} region.

4. The Effect of Interpoint Distance Ties

We now focus on the behavior of the two measures which gave the highest percentage of interpoint distance ties, namely, APPROX and TI2MINUS (see Table IV for defining formulas). A comment on the motivation of these measures is in order.

The measure APPROX is a closed-form approximation to the following least-moves measure proposed in Boorman and Arabie (1970). Define a *set move* transforming one partition into another to be the transfer of a subset of a cell of the first partition to another cell or to form a new cell. The metric B may be defined as the least number of set moves needed to transform any partition into any other given partition. Clearly, the range of B is contained in the set of integers $\{0, 1, \ldots, n-1\}$, and it can be shown that B has no closed-form representation but is computable as a Hitchcock problem of optimal assignment type (Ford and Fulkerson, 1962). The formula defining APPROX, namely,

$$\text{APPROX}(P, Q) = |P \cap Q| - \min(|P|, |Q|). \tag{8}$$

is derived from the definition of B by assuming essentially no interaction among allocation decisions in the optimal assignment problem associated with its computation; see Boorman (1970) for details.

The measure TI2MINUS can be expressed by a formula very similar to APPROX:

$$\text{TI2MINUS} = 2(|P \cap Q| - (|P| + |Q|)). \tag{9}$$

TI2MINUS is a distance concept dual to the metric C discussed in Boorman and

Arabie (1972), which is a standard lattice metric often employed in pure mathematics (Birkhoff, 1967) and which may be defined by

$$C(P, Q) = |P| + |Q| - 2|P \cup Q|. \tag{10}$$

Given the apparent similarity of their defining formulas (8) and (9) respectively, the remarkable feature of the two measures APPROX and TI2MINUS is the striking and consistent difference between the qualitative nature of their MDSCAL representations, using the primary approach to ties as previously indicated. The APPROX plots are roughly linear along the lumper–splitter dimension, with types arranged in linear segments perpendicular to this axis (Fig. 10). The TI2MINUS plots attempt to

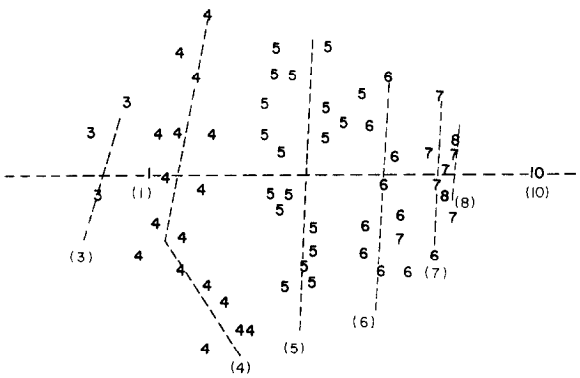

FIG. 10. Scaling of APPROX on 10-lattice sample with both endpoints present. 2 dimensions, Euclidean metric, random initial configuration. Stress formula 1 = 19.1%. Points are labeled by the cardinality $|P|$ of the corresponding partition. Numbers in parentheses classify segments of the plot by the cardinality of the partitions which they contain.

replicate contours as closed curves, but customarily become badly confused as to whether P_{lumper} or P_{splitter} should be dominant, and sometimes end up placing these two partitions on top of one another (Fig. 11).

It remains somewhat unclear why APPROX gives the extreme contour structure evinced by Fig. 10 while TI2MINUS gives a contour structure which can be explained by reasoning as in the last section. This question would not be interesting were it not for the fact that the behavior of APPROX makes it discriminate cardinalities as linear segments with such accuracy that a visual scan of an unlabeled output plot can usually differentiate partitions by size, which is not possible for PAIRBONDS, INFOTWO, or any of the less "degenerate" metrics. The success of MDSCAL in handling this tie-bound metric is all the more surprising if one considers the assertions of Roskam (1969) that Kruskal's algorithm is particularly vulnerable to degenerate solutions for

data containing ties. Since APPROX clearly will give even more interpoint ties than does TI2MINUS, in going from the latter measure to the former we have apparently achieved a situation where confusion of cardinalities (in the TI2MINUS plots) is replaced by a superficial clarity in the APPROX scaling, at least as long as the primary stress approach is used. In fact, however, all that APPROX is really doing is decoupling partitions of different sizes, so that classes of partitions of a fixed size can be scaled

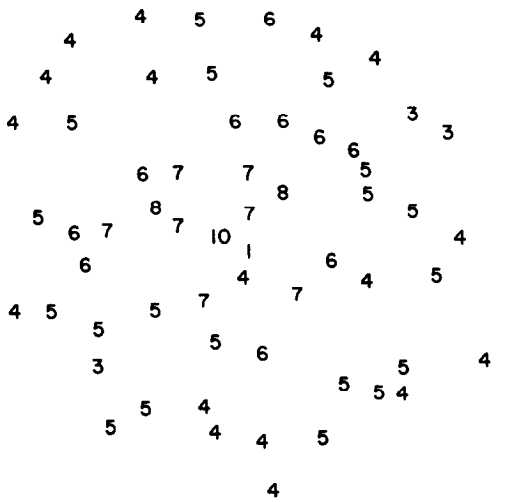

FIG. 11. Scaling of TI2MINUS on 10-lattice sample with both endpoints removed. 2 dimensions, Euclidean metric, random initial configuration. Stress formula 1 = 29.8%. Points are labeled by cardinality.

essentially as line segments, with little emphasis being given by the algorithm to the relation between different segments as long as they are kept distinct (compare again Fig. 10).

The implication is that scaling representations may be highly sensitive to the exact form of the input proximity measure if an excessive number of interpoint distance ties is permitted and the primary stress approach is employed. Since many structural measures suggested solely on the basis of mathematical considerations tend to produce many interpoint distance ties [compare Hays (1965)], there is need for considerable caution in practical employment of such measures in scaling or scaling-related applications. In fact, as Appendix B demonstrates, there are powerful constraints which force *any* permutation-invariant partition distance measure to have many interpoint distance ties for large lattices.

5. Measures of Association

The remaining measures to be discussed are both derived from standard measures of association by linear transformations. CHISQUAR (in Table IV) is based on a suggested normalization of χ^2 proposed by Cramér (1946); see also Goodman and Kruskal (1954). The measure 1 — LAMBDA (again see Table IV) is the obvious conversion to a distance measure of the measure of association λ, which is discussed at length in Goodman and Kruskal (1954) and which has various desirable invariance and normalization properties.

MDSCAL solutions were computed for both CHISQUAR and 1 — LAMBDA for both the 6- and the 10-lattices, with $r = 2$ (Euclidean metric) and $r = 1$ (city-block metric). Figures 12 and 13 show the obtained solutions for 1 — LAMBDA and CHISQUAR, respectively, in two typical cases. In both cases the points are labeled according to the number of cells of the corresponding partition.

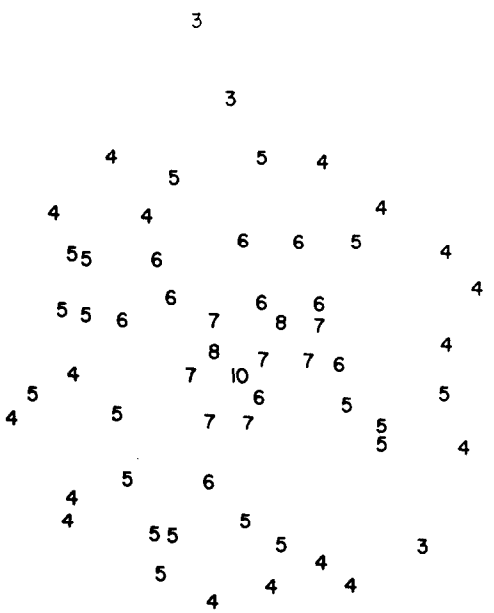

FIG. 12. Scaling of 1 — LAMBDA on 10-lattice sample with P_{lumper} removed and P_{splitter} present. 2 dimensions, Euclidean metric, random initial configuration. Stress formula 1 = 32.1%. Points are labeled by cardinality.

A similar analysis to that performed earlier for PAIRBONDS and INFOTWO (pp. 167–168 previously) provides insight into the scaling behavior of these measures of association. Specifically, consider the 1 — LAMBDA distance produced by moving a

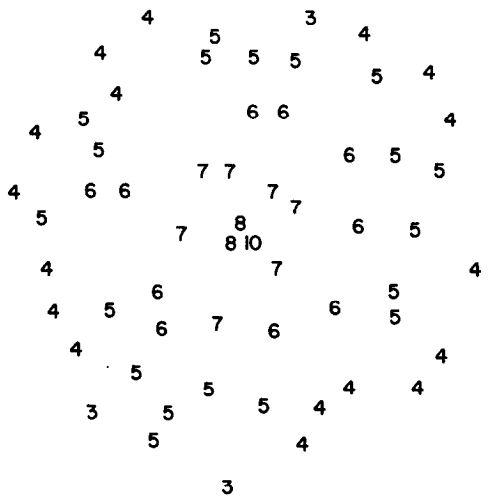

FIG. 13. Scaling of CHISQUAR on 10-lattice sample with P_{lumper} removed and P_{splitter} present. 2 dimensions, Euclidean metric, L-configuration. Stress formula 1 = 36.2%. Points are labeled by cardinality.

single element from a cell of size k to one of size k'. Then 1 — LAMBDA may straightforwardly be computed:

$$1 - \text{LAMBDA} = \frac{2n - [(k-1) + k' + (n-k-k') + (k-1) + k' + (n-k-k')]}{2n - |\, c \,|_{\max} - |\, d \,|_{\max}}, \tag{11}$$

where the notation in the denominator is as in Table 4 and we are assuming $k, k' > 1$. This expression reduces to

$$\frac{1}{n - \frac{1}{2}(|\, c \,|_{\max} + |\, d \,|_{\max})} \sim \frac{1}{n}\left\{1 + \frac{1}{2n}(|\, c \,|_{\max} + |\, d \,|_{\max})\right\}, \tag{12}$$

for $|\, c \,|_{\max}$, $|\, d \,|_{\max}$ small compared with n. Hence, *if* k is the largest cell size in the original partition we have a 1 — LAMBDA distance linearly dependent on k. If, however, we have the polar case where $|\, c \,|_{\max} = |\, d \,|_{\max} = n - k - k'$, then the 1 — LAMBDA distance (12) reduces to $(1/(k+k'))$ and, unlike either the corresponding computation for PAIRBONDS ((5) for $s = 1$), or that for INFOTWO (7), this expression is monotone *decreasing* as a function of k and k'. This behavior is not sufficient to cause inversion to occur, though contour recovery is considerably confused.

In the case of CHISQUAR, consider again the effect of moving a single element from

a non-singleton cell of size k to a cell of size k'. Then the cardinality of the partition remains invariant. Let this cardinality be Z. Then the distance generated is

CHISQUAR

$$= 1 - \left[\left(-1 + \frac{(k-1)^2}{k(k-1)} + \frac{1}{k(k'+1)} + \frac{k'^2}{k'(k'+1)} + Z - 2\right)\Big/(Z-1)\right]. \tag{13}$$

Equation (13) may be simplified to obtain

$$\text{CHISQUAR} = 1 - \left[\left(-\frac{1}{k} + \frac{1}{k(k'+1)} + \frac{k'}{k'+1} + Z - 2\right)\Big/(Z-1)\right]. \tag{14}$$

The relevant (k- and k'-dependent) part of this last expression may now be rewritten

$$\Psi(k, k') = \frac{1}{Z-1}\left[\frac{1}{k} - \frac{1}{k(k'+1)} - \frac{k'}{k'+1}\right] = -\frac{1}{Z-1}\left(\frac{k'}{k'+1}\right)\left[\frac{k-1}{k}\right]. \tag{15}$$

This expression is again monotone decreasing in both k and k'. If k' is large, then, for fixed k, (15) is approximately a linear function of $1/k'$; and for fixed k', (15) is a linear function of $1/k$ in any case. From an empirical scaling standpoint, as is clear from Fig. 13, CHISQUAR rather effectively differentiates contours according to cardinality. In fact, CHISQUAR typically gives better scalings by cardinality than does TI2MINUS, which is defined explicitly in terms of cardinality. At the same time, since cardinality is a much cruder structural property than height, as used to define PAIRBONDS, there is no question that CHISQUAR is less sensitive to the fine structure of types than is PAIRBONDS; this is not surprising, since PAIRBONDS is defined explicitly with reference to the lattice structure of partitions while CHISQUAR is not.

6. Comparison of the Behavior of the Measures: Rank-order Correlation, INDSCAL, Stresses

Several approaches were employed to the problem of classifying the measures we have been considering with reference to their scaling and other numerical behavior. The motivation for this comparison originally derived from abstract typology of partition distance measures (Boorman, 1970; Boorman and Arabie, 1972) based on purely lattice-theoretic considerations. In view of the current interest in classifying different kinds of metrics in various psychological applications (Cross, 1965a, 1965b, Shepard, 1969, 1972; Micko and Fischer, 1970; Fischer and Micko, 1972), it is relevant to see how well an abstract classification correlates with the concrete numerical

behavior of distance measures, and what further classificatory information comes out of such numerical output.

Table IX gives rank-order correlation coefficients between each pair of distinct partition distance measures on each of the spaces considered (4-lattice, 5-lattice, 6-lattice sample, 10-lattice sample). The metrics on each particular space induce a (nonstrict) linear ordering of unordered pairs of partitions (P, Q) according to their assigned metric distance; this linear ordering involves ties, usually a substantial number (see Appendix B). Kendall's rank-order correlation coefficient τ_b (one way of extending the standard τ to take care of ties [see Kendall(1962)]) was then employed to give each entry of Table IX. A τ_b value of $+1$ indicates that the two distance measures are ordinally similar in the strict sense (i.e., there is a monotone transformation taking one into the other, and by definition of Kruskal's algorithm their Shepard–Kruskal scaling behavior is identical). A τ_b value near 0 means little or no correlation between the metrics under the τ_b measure of rank-order correlation.

Because Table 9 is difficult to synthesize as a whole, Shepard–Kruskal scalings of the entries in the table were made. Figures 14 and 15 illustrate, respectively, the Euclidean and city-block scalings of the τ_b matrices for each of the four lattices. In order to minimize the possibility of achieving a seriously distorted solution corresponding to a local minimum of stress formula 1, each of the scalings was rerun eight times with different random initial configurations and the solutions with lowest obtained stresses were selected for inclusion in Figs. 14 and 15. The final stresses obtained were excellent. The four Euclidean solutions were then rotated to maximum congruence using the algorithm of Olivier (1970); the four city-block solutions were eyeballed to maximum congruence by reflections, since rotation invariance does not hold for city-block spaces.

The eight solutions contained in Figs. 14 and 15 are seen to be remarkably similar and correspond closely to the qualitative analysis of Boorman and Arabie (1972) and the comparative MDSCAL analysis of the present paper. The strong normalizations 1 − NT/1 + 2, 1 − NT/SQP, and NVRNTROP (which may also be defined as those distance measures placing P_{lumper} at the center of the scaling solution) form a tight and isolated cluster in all cases. The measures APPROX and TI2MINUS fall between this outlying cluster and the rest of the distance measures, though they are typically not close enough to one another to be designated a cluster. The remaining seven measures (PAIRBONDS, INFOTWO, MULTINOM, INFOTHRY, 1 − LAMBDA, CHISQUAR, and INVRHEIT) form a loose cluster, in which CHISQUAR often appears as an outlying point. All other measures in this loose cluster except 1 − LAMBDA are either supervaluation-based (PAIRBONDS, INFOTWO, MULTINOM) or derived from these metrics by a mild normalization (INFOTHRY, INVRHEIT). There is a strong, though not universal, tendency for the scalings to place PAIRBONDS and INVRHEIT close to one another, as also INFOTWO and MULTINOM. It is also noteworthy that 1 − LAMBDA tends consistently to fall

between CHISQUAR and the lattice-based measures, and itself falls close to PAIRBONDS and INFOTHRY in most of the plots.

A more formal analysis of the same data is shown in Fig. 16, which displays the result of applying Johnson's (1967) diameter method of hierarchical clustering to the τ_b matrix for the 10-lattice and then superimposing the obtained clusterings on the relevant MDSCAL solution in Fig. 14. It is clear that the Johnson algorithm replicates our major qualitative observations and supports Shepard's (1969) argument for the combined use of hierarchical clustering and multidimensional scaling.

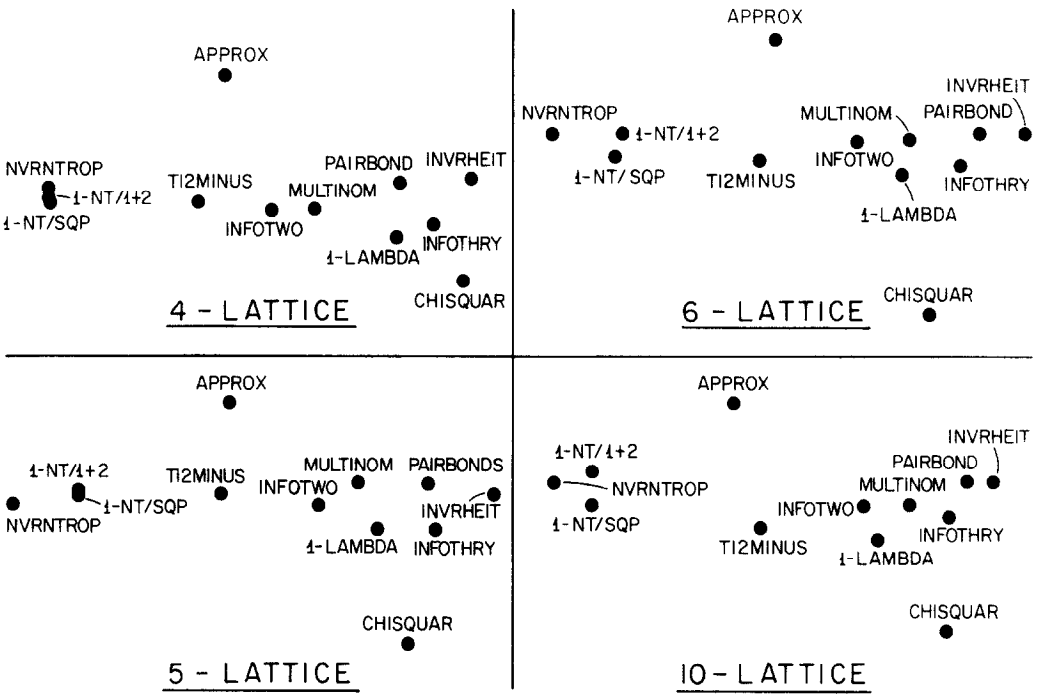

FIG. 14. Euclidean scaling of τ_b matrices for each of the four lattice sizes (4-lattice, 5-lattice, 6-lattice sample, 10-lattice sample). Respectively, stress formula 1 values were: 1%, 1.6%, 1.4%, 1%.

A different approach to the classification problem was also tried, employing the INDSCAL algorithm for the metric analysis of individual differences via canonical decomposition of 3-way tables (Carroll and Chang, 1970; Carroll, 1972; Schönemann, 1972). The objective of INDSCAL is to take proximity data from several subjects, pool this data and to analyze it by an iterative least-squares algorithm (thus, giving a stimulus space), and concurrently to analyze differences between subjects (producing a subject space).

TABLE IX

VALUES OF TAU FOR 4 SAMPLES

(1) 15 DISTINCT PARTITIONS FROM THE 4 LATTICE (S. D. = 0.06615)

(2) 52 DISTINCT PARTITIONS FROM THE 5 LATTICE (S. D. = 0.01833)

(3) 60 DISTINCT PARTITIONS FROM THE 6 LATTICE (S. D. = 0.01586)

(4) 60 DISTINCT PARTITIONS FROM THE 10 LATTICE (S. D. = 0.01586)

	PAIRBOND	INFOTWO	MULTINOM	INFOTHRY	1-NT/1+2	1-NT/SQP	1-LAMBDA	APROXMAT	TI2MINUS	CHISQUAR	INVRHEIT	NVRNTROP
PAIRBOND	1.00000	0.72375	0.83411	0.86695	0.12899	0.08665	0.88388	0.52613	0.52585	0.75385	0.87850	0.10362
	1.00000	0.73798	0.84508	0.88229	0.21041	0.17693	0.86981	0.50429	0.49336	0.68001	0.90871	0.00260
	1.00000	0.75970	0.85892	0.89366	0.25409	0.22337	0.84228	0.54027	0.49970	0.63146	0.94263	0.05634
	1.00000	0.72702	0.85588	0.86006	0.07811	0.05113	0.74425	0.34610	0.42015	0.61807	0.96574	-0.05602
INFOTWO	0.72375	1.00000	0.88978	0.61498	0.48326	0.44666	0.71854	0.66561	0.86241	0.49303	0.48960	0.45696
	0.73798	1.00000	0.89818	0.72425	0.47078	0.45423	0.81970	0.68149	0.79008	0.60620	0.57567	0.27369
	0.75970	1.00000	0.90941	0.80692	0.47607	0.45453	0.87499	0.69654	0.76923	0.63123	0.64645	0.29429
	0.72702	1.00000	0.88144	0.80745	0.30181	0.29022	0.86171	0.55554	0.72081	0.61948	0.66387	0.20310
MULTINOM	0.83411	0.88978	1.00000	0.73873	0.34335	0.30568	0.83431	0.60816	0.75055	0.60694	0.61641	0.31810
	0.84508	0.89818	1.00000	0.81070	0.37555	0.36301	0.87264	0.60875	0.69044	0.66682	0.67580	0.17437
	0.85892	0.90941	1.00000	0.87251	0.40126	0.38014	0.90233	0.63949	0.67904	0.66587	0.73600	0.21021
	0.85588	0.88144	1.00000	0.89530	0.20138	0.18555	0.85052	0.46636	0.59086	0.65073	0.78885	0.08919
INFOTHRY	0.86695	0.61498	0.73873	1.00000	0.02522	-0.01447	0.88334	0.35878	0.43523	0.87714	0.87771	0.00257
	0.88229	0.72425	0.81070	1.00000	0.16368	0.14624	0.87627	0.45756	0.49373	0.72618	0.84345	-0.00558
	0.89366	0.80692	0.87251	1.00000	0.25957	0.23851	0.87365	0.55060	0.56406	0.67958	0.82486	0.09565
	0.86006	0.80745	0.89530	1.00000	0.09412	0.08373	0.81057	0.38075	0.52937	0.69299	0.82041	0.00487
1-NT/1+2	0.12899	0.48326	0.34335	0.02522	1.00000	0.96714	0.13978	0.52517	0.64875	-0.05677	-0.10439	0.98221
	0.21041	0.47078	0.37555	0.16368	1.00000	0.95812	0.29159	0.61557	0.68760	0.15308	0.03862	0.88294
	0.25409	0.47607	0.40126	0.25957	1.00000	0.96140	0.38272	0.61192	0.69821	0.25484	0.13034	0.87867
	0.07811	0.30181	0.20138	0.09412	1.00000	0.95451	0.25575	0.58148	0.52509	0.04964	0.01137	0.89473
1-NT/SQP	0.08665	0.44656	0.30568	-0.01447	0.96714	1.00000	0.11683	0.48230	0.65303	-0.09852	-0.14529	0.95608
	0.17693	0.45423	0.36301	0.14624	0.95812	1.00000	0.28681	0.57320	0.70115	0.16433	0.00753	0.87114
	0.22337	0.45453	0.38014	0.23851	0.96140	1.00000	0.36979	0.57321	0.70578	0.25444	0.10142	0.88229
	0.05113	0.29022	0.18555	0.08373	0.95451	1.00000	0.25092	0.54364	0.54912	0.05488	-0.01444	0.90104

	PAIRBOND	INFOTWO	MULTINOM	INFOTHRY	1-NT/1+2	1-NT/SQP	1-LAMBDA	APROXMAT	TI2MINUS	CHISQUAR	INVRHEIT	NVRNTROP
1-LAMBDA	0.88388	0.71854	0.83431	0.88334	0.13978	0.11683	1.00000	0.44807	0.56882	0.79435	0.78746	0.11513
	0.86981	0.81970	0.87264	0.87627	0.29159	0.28681	1.00000	0.52988	0.63040	0.74279	0.74633	0.09587
	0.84228	0.87499	0.90233	0.87365	0.38272	0.36979	1.00000	0.60499	0.68177	0.71093	0.73042	0.19684
	0.74425	0.86171	0.85052	0.81057	0.25575	0.25092	1.00000	0.48242	0.67425	0.66921	0.68178	0.14858
APROXMAT	0.52613	0.66561	0.60816	0.35878	0.52517	0.48230	0.44807	1.00000	0.70645	0.20259	0.32169	0.52093
	0.50429	0.68149	0.60875	0.45756	0.61557	0.57320	0.52988	1.00000	0.76568	0.26677	0.34630	0.44642
	0.54027	0.69654	0.63949	0.55060	0.61192	0.57321	0.60499	1.00000	0.76271	0.35098	0.43571	0.44205
	0.34610	0.55554	0.46636	0.38075	0.58148	0.54364	0.48242	1.00000	0.67297	0.18608	0.28581	0.48079
TI2MINUS	0.52585	0.86241	0.75055	0.43523	0.64875	0.65303	0.56882	0.70645	1.00000	0.34140	0.28743	0.64352
	0.49336	0.79008	0.69044	0.49373	0.68760	0.70115	0.63040	0.76568	1.00000	0.48229	0.32300	0.48978
	0.49970	0.76923	0.67904	0.56406	0.69821	0.70578	0.68177	0.76271	1.00000	0.53203	0.38017	0.52274
	0.42015	0.72081	0.59086	0.52937	0.52509	0.54912	0.67425	0.67297	1.00000	0.46530	0.35894	0.45323
CHISQUAR	0.75385	0.49303	0.60694	0.87714	-0.05677	-0.09852	0.79435	0.20259	0.34140	1.00000	0.78216	-0.07961
	0.68001	0.60620	0.66682	0.72618	0.15308	0.16433	0.74279	0.26677	0.48229	1.00000	0.64558	0.00368
	0.63146	0.63123	0.66587	0.67958	0.25484	0.25444	0.71093	0.35098	0.53203	1.00000	0.57438	0.11147
	0.61807	0.61948	0.65073	0.69299	0.04964	0.05488	0.66921	0.18608	0.46530	1.00000	0.59593	-0.02892
INVRHEIT	0.87850	0.46960	0.61641	0.87771	-0.10439	-0.14529	0.78746	0.32169	0.28743	0.78216	1.00000	-0.12619
	0.90871	0.57567	0.67580	0.84345	0.03862	0.00753	0.74633	0.34630	0.32300	0.64558	1.00000	-0.14164
	0.94263	0.64645	0.73600	0.82486	0.13034	0.10142	0.73042	0.43571	0.38017	0.57438	1.00000	-0.04861
	0.96574	0.66387	0.78885	0.82041	0.01137	-0.01444	0.68178	0.28581	0.35894	0.59593	1.00000	-0.11574
NVRNTROP	0.10362	0.45696	0.31810	0.00257	0.98221	0.95608	0.11513	0.52093	0.64352	-0.07961	-0.12619	1.00000
	0.00260	0.27369	0.17437	-0.00558	0.88294	0.87114	0.09587	0.44642	0.48978	0.00368	-0.14164	1.00000
	0.05634	0.29429	0.21021	0.09565	0.87867	0.88229	0.19684	0.44205	0.52274	0.11147	-0.04861	1.00000
	-0.05602	0.20310	0.08919	0.00487	0.89473	0.90104	0.14858	0.48079	0.45323	-0.02892	-0.11574	1.00000

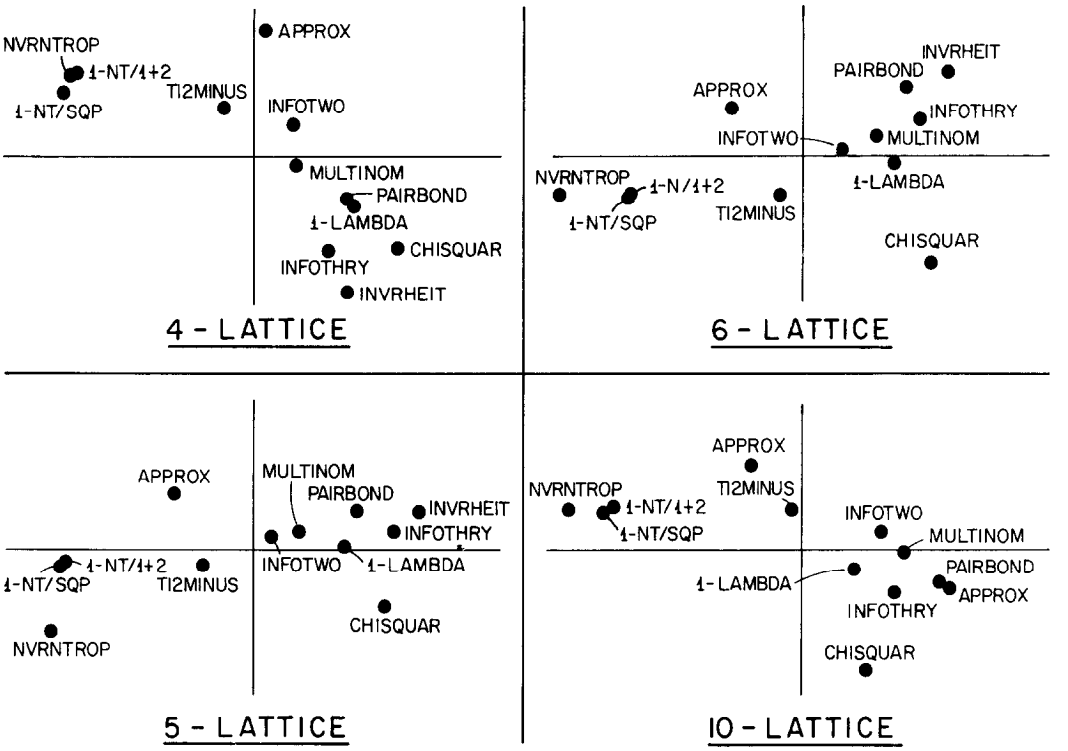

FIG. 15. City-block scaling of τ_b matrices for each of the four lattice sizes (4-lattice, 5-lattice, 6-lattice sample, 10-lattice sample). Respectively, stress formula 1 values were: 2.5%, 3.6%, 1.7%, 2.3%.

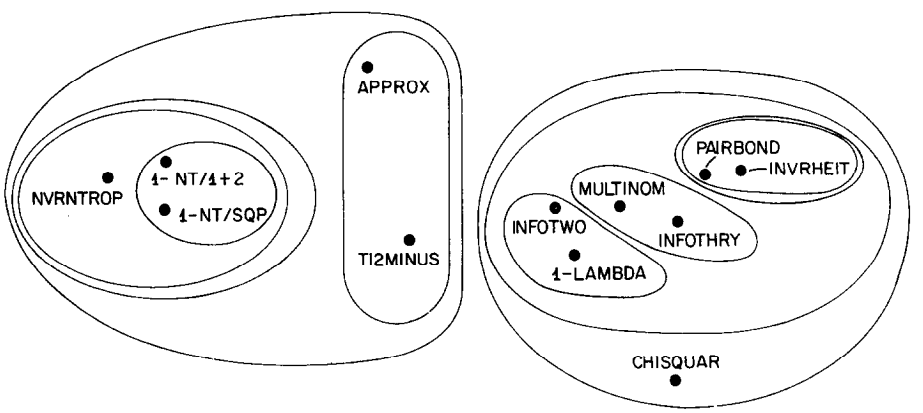

FIG. 16. Superposition of hierarchical clustering of the τ_b matrix for the 10-lattice sample, Johnson's diameter method (1967), on MDSCAL solution in Fig. 15.

In the present application, the points of a given partition lattice corresponded to stimuli in the standard interpretation of the INDSCAL model, while the partition distance measures corresponded to the subjects. Following the procedure described by Carroll and Chang (1970, p. 288), each partition distance measure μ was normalized so as to equate the variance for each subject (measure). Using the INDIFF data structure, INDSCAL solutions in each of dimensions 5, 4, 3, and 2 were then determined for the partition spaces obtained by truncating both P_{lumper} and P_{splitter} from the 5-, 6- and 10-lattice samples. The obtained stimulus spaces were, in general, interpretable with very clear differentiation of partition types as contour lines and the partitions near P_{splitter} in the center of the plot. The subject space solutions, however, were degenerate for the 5- and 6-lattices in the sense that the solutions fell very nearly on a ray through the origin and, hence, corresponded to an attempt on the part of the model to treat all the input distance measures as if they were identical up to a positive constant. By contrast, a meaningful 10-lattice INDSCAL solution was obtained for both the stimulus and subject spaces. Figure 17 gives the two-

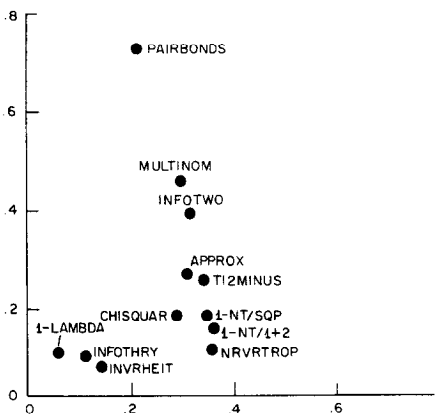

FIG. 17. Two-dimensional INDSCAL subject space representation for the 12 partition distance measures.

dimensional subject space representation. As in Figs. 14 and 15, 1 — NT/1 + 2, 1 — NT/SQP, and NVRNTROP appear in a cluster, as do APPROX and TI2MINUS. The supervaluation-based lattice metrics (PAIRBONDS, MULTINOM and INFOTWO), however, become strongly separated from 1—LAMBDA, INFOTHRY, and INVRHEIT in the INDSCAL solution, whereas these measures form a loose but single cluster in the MDSCAL solution based on the τ_b matrix.

A third kind of numerical comparison of the distance measures is more indirect and simply involves ordering the obtained stress values (all computed by means

of Kruskal's stress formula 1). Although various recent investigators, e.g., Zinnes (1969) and Young (1970), have de-emphasized the importance of stress as a summary statistic, except to indicate obviously degenerate solutions, various results of interest were obtained from stress comparisons. Table X records stress values obtained by the following averaging procedure designed to give a reasonable number of replications in each cell (on the order of 10). A given partition measure, a given dimensionality (=2 or 3), and a given Minkowski r (Euclidean or city-block) were fixed, and the obtained stresses were first minimized over the initial configuration with presence/absence of endpoint(s) held fixed, and then averaged over the latter. Thus if μ is a partition distance measure (on a particular lattice), r is the Minkowski r, and d is the scaling dimensionality, then the corresponding entry in Table X is: stressav(μ; r, d) = (1/4) ((stress minimum with both endpoints)

$$+ \text{ (stress minimum without } P_{\text{lumper}})$$

$$+ \text{ (stress minimum without } P_{\text{splitter}})$$

$$+ \text{ (stress minimum without } P_{\text{lumper}} \text{ and } P_{\text{splitter}})). \tag{16}$$

The averaging over presence/absence of endpoint(s) was done in order to provide greater stability in the entries. Missing entries in the table indicate that no scalings were done for the particular $\langle \mu, r, d \rangle$ configuration corresponding to the cell. Only eight metrics were scaled with enough different configurations to make reporting useful.

The following interpretation of Table X is of interest. Let us partially order the stress vectors in this table (for the eight metrics) in the standard product ordering: $\mathbf{x} \leqslant \mathbf{y}$ if and only if $x_n \leqslant y_n$ for all coordinate dimensions i. Fig. 18 shows that a remarkably close approximation to a *linear* ordering emerges. This figure indicates that stress may, in fact, be a more fundamental statistic than commonly thought, for not only is the ordering very close to being linear (which indicates that stress is essentially an ordinal invariant over lattice size, Minkowski r, and dimensionality), but also the ordering is highly interpretable in terms of our previous results. The information-theoretic lattice metrics and their mild normalization INFOTHRY are at the top of Fig. 18. PAIRBONDS is next in sequence, followed by its strong normalizations 1 − NT/SQP and 1 − NT/1 + 2. 1 − LAMBDA is an outlying point in the ordering but is contained in the segment (INFOTWO, PAIRBONDS) and, hence, is closer to this cluster than to APPROX or to 1 − NT/1 + 2 and 1 − NT/SQP (compare again Figs. 14 and 15). APPROX is a minimal element in the ordering, as is to be expected since the stress approach is primary and, hence, APPROX has the fewest operative constraints.

This analysis in terms of comparative stress behavior hence integrates closely with the classification of different measures by MDSCAL and INDSCAL. Normalization appears in all cases to reduce stress, especially severe normalization (1 − NT/1 + 2

TABLE X

Stress Formula 1 Values (in Percent) Obtained for Given Measures of Partition Distance

	4-lattice, $r = 1$, dim = 2	5-lattice, $r = 1$, dim = 2	6-lattice, $r = 1$, dim = 2	10-lattice, $r = 1$, dim = 2	4-lattice, $r = 2$, dim = 2	5-lattice, $r = 2$, dim = 2	6-lattice, $r = 2$, dim = 2	10-lattice, $r = 2$, dim = 2
PAIRBONDS	17	25	25	28	16	24	21	28
INFOTWO	28	35	35	37	23	32	31	34
MULTINOM	27	34	29	35	23	31	28	33
INFOTHRY	24	32		35	21	29	28	32
1 − NT/SQP	13	23		31	12	21	23	28
1 − NT/1 + 2	13	22		32	12	20	22	28
APPROX	*X*	25	16	21	*X*	19	13	20
1 − LAMBDA	24	31		37	20	27	29	32
CHISQUAR							31	36
TI2MINUS				32			25	30
INVRHEIT				30			23	30
NVRNTROP				31			22	28

	4-lattice, $r = 1$, dim = 3	5-lattice, $r = 1$, dim = 3	6-lattice, $r = 1$, dim = 3	10-lattice, $r = 1$, dim = 3	4-lattice, $r = 2$, dim = 3	5-lattice, $r = 2$, dim = 3	6-lattice, $r = 2$, dim = 3	10-lattice, $r = 2$, dim = 3
PAIRBONDS	9	20			9	18		
INFOTWO	18	26			16	24		
MULTINOM	19	26			16	23		
INFOTHRY	15	27			12	22		
1 − NT/SQP	6	18			7	16		
1 − NT/1 + 2	6	18			7	15		
APPROX	2	10			*X*	9		
1 − LAMBDA	16	25			12	20		
CHISQUAR								
TI2MINUS								
INVRHEIT								
NVRNTROP								

Note—An *X* indicates that runs were made but produced no nondegenerate solutions.

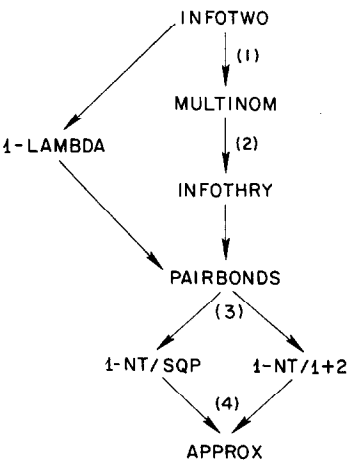

FIG. 18. Partial ordering of eight measures of partition distance on the basis of obtained stress formula 1 values. One measure is above another in the ordering if its stresses are higher, component by component, in Table X, with the following exceptions: (1) Except for 1% discrepancy for 4-lattice, $r = 1$, $d = 3$; (2) Except for 1% discrepancy for 5-lattice, $r = 1$, $d = 3$; (3) Except for 3% discrepancy for 10-lattice, $r = 1$, $d = 2$; (4) Except for 2–3% discrepancy for 5-lattice, $r = 1$, $d = 2$.

and 1 — NT/SQP. PAIRBONDS-derived measures give consistently lower stress than information-theoretic ones, and—if stress is taken seriously as a badness-of-fit function—this situation reinforces our earlier comments to the effect that PAIRBONDS is better adapted to multidimensional scaling applications than are information-theoretic measures.

7. Application to a Sociobiological Example

In a study which provides the data base of a recent stochastic model by Cohen (1970), Struhsaker (1967a, 1967b) has given detailed reportings of the sleeping groups formed by vervet monkeys (*Cercopithecus aethiops*) over a six month observation period. These sleeping groups are recorded in Table XI and constitute ideal data for partition analysis because of the comparative clarity of their operational definition, which Struhsaker gives as follows (1967b, p. 110): "Nearly every evening, just before sunset (between 1830 and 1900 hours), the vervet groups divided into sleeping subgroups that spent the night separated from one another by at least one impassable break in the tree canopy."

Cohen (1970) has analyzed the size distribution of these sleeping groups in terms of a family of stochastic processes (which he terms LOST dynamics), and has found that

TABLE XI

Sleeping Groups Formed by a Vervet Monkey Troop on Successive Nights[a]

	I	adult male SG
	II	the older adult male (left the troop between 17–18 April and 27–28 April 1964)
	III	adult male LP
	IV	adult female DK
	V	juvenile male LYA
	VI	adult female TK
	VII	young juvenile female D
	VIII	young juvenile female B
	IX	young juvenile female DR
	X	juvenile female NN
	XI	subadult female N
	XII	adult female TBW
	XIII	two young indistinguishable juvenile males
	XIV	infant male DK (son of IV)
	XV	infant female TBW (born to XII between 4–5 Jan. and 29–30 Jan. 1964)
	XVI	infant male TK (born to VI between 29–30 Jan and 4–5 Feb. 1964)
(1)	4–5 Jan. 1964	(IV, XIV, XIII), (I, III, V, XII, XIII), (XI, IX, X), (II), (VIII, VII, VI)
(2)	29–30 Jan. 1964	(VI, X, VIII, VII, IX), (I, II, III, IV, XIV, XII, XV, V, XI, XIII, XIII)
(3)	4–5 Feb. 1964	(V, XIII, XIII, IV, XIV, II, XII, XV, XI), (I), (III), (VIII, IX, VII, X, VI, XVI)
(4)	9–10 Feb. 1964	(III, II, V), (I, XI, XIII, XIII), (X, IX), (VIII, VII, VI, XVI, IV, XIV, XII, XV)
(5)	23–24 Feb. 1964	(XII, XV, XI), (other 14)
(6)	24–25 Feb. 1964	(I, III, XI), (other 14)
(7)	1–2 March 1964	(XII, XV, XI, III), (other 13)
(8)	9–10 March 1964	(X), (XI, III, XIII, XII, XV), (V, I, IX, VII, VIII, VI, XVI, IV, XIV, II, XIII)
(9)	31 March 1964–1 April 1964	(III, XIII, X), (XI, IX), (II, XII, XV), (I, VII, VIII, V, IV, XIV, VI, XVI, XIII)
(10)	1–2 April 1964	(II, XII, XV), (XI, XIII, IX), (IV, XIV, XIII), (VI, XVI, VII, VIII, III, I, V, X)

Table continued

TABLE XI (*continued*)

(11)	4–5 April 1964	(I, II), (VI, XVI), (III, X), (V), (VII, VIII, IX, XII, XV, XIII, XIII, XI, IV, XIV)
(12)	6–7 April 1964	(IV, XIV, XIII, XIII, XI, IX), (III, XII, XV, VI, XVI, VIII, X), (II, I, VII, V)
(13)	17–18 April 1964	(X), (IV, XIV, XIII, XIII, IX, XI), (III, XII, XV), (I, V, VI, XVI, VII, VIII), (II)
(14)	27–28 April 1964	(I, IX, V, X, XIII, III), (XII, XV), (XI), (IV, XIV, XIII, VIII, VII, VI, XVI)
(15)	30 April 1964–1 May 1964	(IV, XIV, XIII, IX, X), (I, XI, XII, XV), (VI, XVI, VII, VIII, XIII, V), (III)
(16)	4–5 May 1964	(VI, XVI), (IV, XIV, XIII), (III, XII, XV), (I, XI), (X, XIII, VII, VIII, IX), (V)
(17)	9–10 May 1964	(XIII, IX, XI), (III, I, V, XIII), (XII, XV), (VI, XVI, VII, VIII), (IV, XIV), (X)
(18)	11–12 May 1964	(XII, XV), (IX, III, XIII, X), (XI), (I, V, IV, XIV, XIII), (VI, XVI, VII, VIII)
(19)	19–20 May 1964	(XI, IX, I), (III), (V, XII, XV, VII, VIII, X, VI, XVI, XIII), (XIII, IV, XIV)
(20)	21–22 May 1964	(VI, XVI, VII, VIII), (III), (XI, IX), (XII, XV), (I), (IV, XIV, XIII, XIII, X, V)
(21)	1–2 June 1964	(XII, XV), (XI), (XIII, V, X, IV, XIV, IX, I), (XIII, VI, XVI, VII, VIII), (III)
(22)	4–5 June 1964	(XII, XV, V), (XI), (VIII, IV, XIV), (I, X, IX, XIII), (VI, XVI, VII, VIII), (III)

[a] Taken from Cohen (1970). Reproduced by permission of Harvard University Press.

these processes predict well the size distribution found (which is essentially truncated negative binomial), by contrast to a variety of other possible random processes.

Employing scaling methods for partition spaces described and analyzed above, another analysis of the Struhsaker data was undertaken, with an end to seeing whether any systematic temporal variation would emerge from a partition distance analysis. The only technical problems were that the underlying set of individuals varied slightly with time, and two specific individuals were indistinguishable to the observer (both labeled XIII in Table XI). The first problem can be handled for each of our measures simply by computing c_i, d_j, and z_{ij} *as if* the sets underlying two given partitions P and Q were always identical. For each given sleeping group observation we took the two possible ways of assigning individual identity to the indistinguishables and identified the observation with a *pair* (P_1, P_2) of partitions. Given two such pairs,

an assigned distance for any partition distance measure m may be computed by:

$$m^*((P_1, P_2), (Q_1, Q_2)) = \min[(m(P_1, Q_2), m(P_1, Q_1)]; \tag{17}$$

m^* can be shown to be a metric if m is.

Scalings were then performed for each one of the 12 measures in Table IV, in two and three dimensions with the Euclidean metric on the output configuration. (Several different random initial configurations were employed as a control on the result.) Figure 19 gives the obtained PAIRBONDS configuration in two dimensions; as

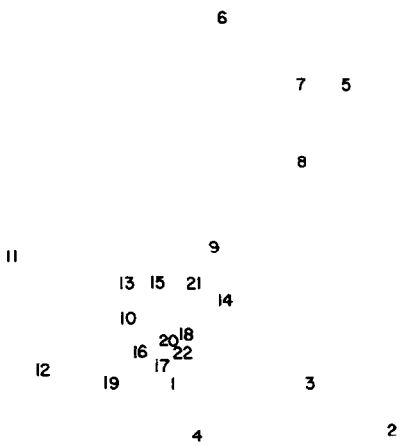

FIG. 19. PAIRBONDS scaling of Struhsaker's data on vervet sleeping groups over a six-month observation period. 2 dimensions, Euclidean metric, random initial configuration. Stress formula 1 = 18.1 %. Points are labeled according to the observation numbers in Table XI.

previous analysis has indicated, this metric gives more information on lattice structure than various of the other measures. A similar picture was obtained for city-block scaling of PAIRBONDS.

Figure 19 clearly shows that the points corresponding to early observations tend to fall on the outside of the plot, while points corresponding to late observations tend to cluster at the center. The distance between successive observations tends to be large at first, but rapidly diminishes as the data points near the compact central cluster. These trends are somewhat noisy, since the observation point 1 is very close to the central cluster, as is the observation 9, while 11 and 12 are clearly outlying points. Despite this fluctuation, the partitions in periods 1–8 have an average height (lumpiness):

$$\bar{D} = \overline{\sum_i \binom{|c_i|}{2}} = 63.75,$$

while those in the subsequent observation periods have an average height of only 32.0 (see Table IV for definition of D). A finer breakdown of more minor temporal fluctuations shows several clusters among the outlying points, especially (2, 3,4) and (5, 6, 7, 8). The greater tendency of the group to sleep in larger units in the early time periods may have been connected with the onset of the copulating season (in May) and the termination of the normal birth season (the last actual birth in this particular troop was in the period between 29–30 January and 4–5 February, 1964); see Struhsaker (1967a) for more detailed discussion of seasonal variation in reproductive behavior.

Although Cohen (1971, pp. 26–27) makes a brief examination of temporal variation in mean group size, which is proportional to the reciprocal of the number of cells in the observed partitions and is hence a simple cardinality measure, he discovers no systematic variation of this quantity in time. In fact, however, this is probably in large part owing to the crudeness of cardinality as a measure, which we have already seen can lead to degeneracies and instabilities in connection with attempted scaling. Moreover, Cohen's emphasis is on an attempt to capture temporal variation by a time-varying Poisson process with a parameter possessing a gamma distribution. Since the Poisson parameter is assumed to be separately generated at each discrete time point, it is not surprising that the resulting process fails to capture systematic diachronic variation.

Struhsaker (1967a, 1967b) undertakes the technically easier task of searching for nonrandom entries in the incidence matrix defined by the frequency with which particular pairs slept in the same group [an analysis similar to that performed for a semantic space by Miller (1969)]. It is not surprising that he found such nonrandomicities (mother–infant being the single most important). The present results indicate that the data analyzed by Struhsaker and Cohen possess yet another level of structure which is important in the *medium* run, on a time scale of perhaps several months. This structure may be important in confronting the inadequately understood question—derived ultimately from evolutionary sociobiology—as to what adaptive function the propensity to form sleeping groups may serve. In this particular application, the MDSCAL algorithm reveals itself as a method of extracting trend from fluctuation in a structural space (the partition lattice) considerably wilder than the standard n-dimensional random walk familiar from statistical physics.

8. Discussion

Most existing methodological studies of multidimensional scaling deal either with Minkowski r-space configurations (at most perturbed by noise), which it is the job of the program to recover, or with highly general proximity matrices without much structure. The present study is an attempt to enter the problem at an *intermediate* level, where we are scaling proximity matrices about whose structure we can say

something analytically, but which are not even close to being Minkowskian for any r. The partition lattice is an ideal starting point for such an analysis, because of its very rich combinatorial structure, e.g., Whitman (1946), combined with the simplicity of basic partition algebra.

Our major qualitative conclusions for our particular data have already been stated. They are (1) the difficulty of specifying a unique combination rule (i.e., the value of r in the Minkowski metric) associated with a particular partition space; (2) the utility of stress as ordinal evidence of badness-of-fit, and (3) the nonuniqueness of MDSCAL solutions at the level of individual points. We believe that the last conclusion, combined with the results of other studies, suggests a particular way of viewing spatial representations of the stimulus domain.

In discussing the nonunique nature of scaling solutions for the partition lattice under a given metric, it was argued that such solutions are interpretable only to the level of contours and clusters, not necessarily down to the location of single points. Similarly, we would like to suggest that the traditional emphasis on the dimensions of *nonmetric* multidimensional scaling solutions may be unjustified; see parallel arguments in Shepard (1969, pp. 52 ff.).

For partition metrics of the kind considered in the present study, we have argued that a three-dimensional solution is largely uninterpretable and that a two-dimensional solution is governed by the location along the lumper–splitter axis of the family of partitions being scaled (i.e., the range of types (Definition 2.1) represented in the family) so that psychological "explanations" of such data can reflect little more than differential placement of the given sorter along this lumper–splitter axis.

One consequence of this finding is that the investigator's attempt to attach psychologically appropriate attributes to orthogonal axes in such a space becomes a matter of little importance, contrary of the legacy of factor analysis. It may be countered that the artificial data considered here constitute a very special case. However, our position is reinforced by the author of a particularly competent investigation using semantic stimuli (Arnold, 1971, p. 367), who has denigrated the naming of dimensions: "The writer knows that an attempt to assign names to dimensions descriptive of verbal concepts can amount to no more than a self-administered word-association test."

Perhaps the strongest emphasis on dimensions has come from Beals, Krantz, and Tversky (1968) who have argued that metric and dimensional representations should be viewed as psychological theories of similarity, rather than as methods of organizing, summarizing, or displaying data, in contrast to the viewpoint of Shepard (1962) and Shepard and Carroll (1966, p. 566). Beals, et al. (1968) have stated that "...if multidimensional scaling models are regarded as useful data reduction models rather than as theoretical models one may ask what is the source of their usefulness" (p. 141). Two studies, in both of which the psychological dimensions were of minimal importance, may be cited to answer this question.

The first finding comes from an investigation of the space of color names carried

out by Rapoport and Fillenbaum (1972). On the basis of prior analyses, e.g., Shepard (1962), these authors had expected to find a circular representation, and their results confirmed that expectation. Now at this point, the traditional emphasis would demand names for two attributes underlying the circular representation (an especially difficult task, given the rotation invariance of MDSCAL solutions).

Instead, Rapoport and Fillenbaum suggested the use of polar coordinates, in which the "intrinsic" dimensionality is now equal to one, and the topic of interest is the ordering of colors along this one "dimension." Considering the result in polar coordinates renders useless most of the elegant theoretical tests for dimensional representation, e.g., Beals et al. (1968) and Tversky and Krantz (1969). Nonetheless, we think most investigators would agree that the imaginative approach taken by Rapoport and Fillenbaum renders the data much more meaningful psychologically than would an insistence upon questions of dimensional representations. The advantage obtained from using polar coordinates for this particular set of data suggests that strict adherence to Cartesian coordinate systems may be unwise; see Coolidge (1963, p. 171), for a survey of alternatives.

The second investigation is a superb study by Rumelhart and Abrahamson (1971) of the multidimensional space of animal names obtained by Henley (1969). If one asks the names of the dimensions Henley obtained, the answer—size, ferocity, "humanness"—can hardly be surprising. The important result from her experiments is surely not the names given to the dimensions; rather, the replicated spatial representation *per se* and the underlying non-Euclidean metric, pointed out by Arnold (1971), are the important findings. The validity of the space Henley obtained was dramatically confirmed in a series of experiments by Rumelhart and Abrahamson, using an "analogical paradigm." It is noteworthy that their experimental task made no reference to the three dimensions listed above. Instead, the subjects made analogical judgments of the form: Animal A is to animal B as animal C is to $[x_1, x_2, x_3, x_4]$ (four-alternative forced choice). By means of this experimental design, Rumelhart and Abrahamson verified the results of Henley (1969), but it should be noted that the data upon which their study was based were the relative locations of points (animals) within the spatial representation, rather than orientation with respect to any particular set of psychological dimensions.

Finally, perhaps the strongest emprirical evidence against the traditional emphasis upon dimensions has come from Lockhead's (1970) study of multidimensional discrimination space. Lockhead argues that S's task is one of locating the multidimensional stimulus in a psychological space without the requirement that the stimulus be analyzed according to its separate components. An implication of this theory is that, in the design of a discrimination experiment, the set of stimuli that will be identified most accurately is that for which the sum of the distances between each stimulus and its immediate neighbors is a maximum and the variance of these distances is a minimum. Lockhead offers data to support this implication for cases in which the underlying physical

dimensions of the stimuli are "analyzable" as well as other "unanalyzable" cases (Shepard, 1964). Lockhead further argues that the (physical) dimension of importance is actually the relative location of the information in a multidimensional space, and he again offers experimental results in support of this thesis.

The studies by Rapoport and Fillenbaum and by Rumelhart and Abrahamson were cited as evidence that a better understanding of psychological processes may be obtained by foregoing the usual emphasis upon psychological dimensions. Lockhead's results suggest that the traditional approach to discrimination learning places undue emphasis upon the physical dimensions of the stimuli, since it is the relative locations of the stimuli within the space that are important. If the underlying (physical) dimensions are of as little consequence as his study suggests, then it would seem that the usual emphasis given to the possible recovery of corresponding psychological dimensions in multidimensional scaling solutions may also be ill-spent.

APPENDIX A: Procedure for Random Generation of Partitions

We describe here a procedure for generating random partition P of a set S of fixed size n.

Enumerate the elements of S, $\{S = e_i\}_{i=1}^{n}$, and consider the class C of $n \times n$ matrices whose entires are drawn from the set $\{-1, 0, +1\}$. We may code any partition P of S as a matrix $M = [a_{jk}]_{n \times n}$ in C by setting:

$$a_{jk} = \begin{cases} -1 & e_j, e_k \text{ not in the same cell of } P \\ +1 & e_j, e_k \text{ in the same cell of } P. \end{cases}$$

The converse, however, is not true, since most of the matrices in C do not correspond to partitions. The intuitive idea underlying our random generation procedure will be to start with the zero matrix $0 = [0]_{n \times n}$ and to fill in -1's and $+1$'s according to a probabilistic procedure consistent with the requirement that we will eventually produce a matrix M corresponding to some partition Q of S.

To simplify the problem somewhat, we henceforth restrict attention to the half region above the main diagonal of the matrices we are considering. It will be notationally convenient to refer to the entries of an upper halfmatrix by *unordered* subscripts. Thus, if $g \neq h$, $a_{\{h,g\}} \equiv a_{\{g,h\}}$ $(= a_{\min(g,h),\max(h,g)}$ in the ordered subscript notation traditional for matrices). The restriction to upper halfmatrices clearly sacrifices no generality, since all matrices M corresponding to partitions of S are symmetric about the main diagonal ($a_{ij} = a_{ji}$) and have main diagonal elements identically $= +1$. We hence start with the upper halfmatrix whose entires are all 0 and consider a procedure for randomly changing the 0's to -1's and $+1$'s without violating the remaining constraint that the final result be consistent with the transitivity of cell membership.

For present purposes, we may define this consistency requirement as follows:

DEFINITION. An upper halfmatrix H is *consistent* with transitivity if the matrix K which results when

(i) All 0 entires in H are changed to -1, producing an upper halfmatrix G;

(ii) The upper halfregion of K is set to be G;

(iii) The lower halfregion of K is defined by symmetry (so that $k_{ij} = k_{ji}$); and

(iv) Diagonal elements of K are set $= +1$

is the incidence matrix of some partition Q.

We are in a position to define the random generation procedure. Starting with the upper halfmatrix, all of whose entires are 0, we select an entry at random and fill it with a $+1$ (with probability p) or a -1 (with probability $q = 1 - p$). In the first contingency, we refer to the change as a p-decision and, in the second contingency, as a q-decision. Then we select another entry at random and repeat the process. In general, however, after the second p-or q-decision, the procedure becomes slightly more complicated because we must in general fill one or more additional 0 entires following each p- or q-decision in order to preserve consistency in the above sense. In general, suppose that we are given an upper halfmatrix H consistent with transitivity. If H has no 0 entries, H corresponds to the completely defined incidence matrix of a partition. If H still has 0 entries, select a 0 entry at random and fill it with a $+1$ (with probability p) or a -1 (with probability $q = 1 - p$). Suppose the entry in question is $a_{\{g,h\}}$. We now fill in further entries as follows, depending upon whether $a_{\{g,h\}}$ was set equal to $+1$ or to -1.

If $a_{\{g,h\}} = +1$, consider all entries $a_{\{g,i\}}$ which are nonzero. If a specific $a_{\{g,i\}} = 1$ but $a_{\{h,i\}} = 0$ in H, we set $a_{\{h,i\}} = 1$. If $a_{\{g,i\}} = -1$ but $a_{\{h,i\}} = 0$ in H, we set $a_{\{h,i\}} = -1$.

We fill in entries similarly with the roles of g and h interchanged.

If $a_{\{g,h\}} = -1$, consider again all entries $a_{\{g,i\}}$ which are nonzero. If a specific $a_{\{g,i\}} = 1$ but $a_{\{h,i\}} = 0$, we set $a_{\{h,i\}} = 1$. If $a_{\{g,i\}} = -1$, we make no change in $a_{\{h,i\}}$.

Again, we repeat the procedure with the roles of g and h interchanged.

It is easy to see that the resulting halfmatrix H, when these substitutions are completed, is consistent with transitivity in the sense of the preceding definition. Since

H will have fewer 0 entries than H, the procedure we have specified will always terminate in a halfmatrix, defining the incidence matrix of a specific partition after no more than $\binom{n}{2}$ steps.

For any choice of $p \in (0, 1)$, each partition in the n-lattice has some positive probability of occurrence, and hence the procedure defines a one-parameter probability measure on the n-lattice as a function of p. The choice of p gives the investigator control over the centering of the obtained distribution in the lattice ordering: $p \to 0$ corresponds to centering of the distribution in the neighborhood of P_{splitter}, while $p \to 1$ corresponds to centering near P_{lumper}.

Monte Carlo studies indicate that the obtained probability measure is most nearly uniform for $p = p_u$, which equates the likelihood of P_{lumper} and P_{splitter}, i.e., for which

$$p_u^{n-1} = (1 - p_u)^{\binom{n}{2}}. \tag{1-1}$$

For large n the solution of this equation can be shown to be well approximated by

$$p_u \cong (\ln n)/(n - 1). \tag{1-2}$$

The 6-lattice and 10-lattice samples in Tables II and III were genrated by solving (1 — 1) numerically to obtain $p_u = 0.31767$ for the 6-lattice and $p_u = 0.24512$ for the 10-lattice.

An alternative method for generating random partitions has been used by Miller (1969, pp. 190–191) in connection with an application to the method of sorting. The present method is of independent interest owing to the role of transitivity constraints in our inductive generation procedure and, in particular, the interaction of p- and q-decisions with these constraints. This interaction reveals a surprising degree of nonapparent complexity in the process of sorting a set of stimuli.

APPENDIX B: Interpoint Distance Ties for Certain Partition Measures

Let us first note that *all* our measures except 1 — LAMBDA and CHISQUAR obviously have the general form

$$d(P, Q) = f(P, Q, P \cap Q).$$

Moreover, as we see from inspection of these measures, if we denote by $t(P)$ the type of P, we in fact have

$$d(P, Q) = f(t(P), t(Q), t(P \cap Q)).$$

Thus, $d(P, Q)$ can assume at most ${T_n}^3$ values, where T_n is the number of types in the n-lattice.

But by a result initially due to Hardy and Ramanujan (1918), see also Gupta et al. (1958) we have

$$T_n \sim (1/(4n\sqrt{3}))\, e^{\pi\sqrt{n/6}} \qquad \text{as} \quad n \to \infty. \tag{2-1}$$

Now if the size of the n-lattice is denoted by E_n, we can use an unpublished result of Joyal, Labelle, and Lorrain (n.d.) to simplify a formula owing to Moser and Wyman (1955) and obtain

$$\ln E_n \sim n \ln n, \tag{2-2}$$

where in fact for $10 \leqslant n \leqslant 100$ the ratio $(\ln E_n/(n \ln n))$ is very close to $\frac{1}{2}$; see table in Joyal, Labelle, and Lorrain (n.d.). Now the quantity we want to show approaches zero as $n \to \infty$ is

$$|\text{ Range } d\ |\Big/\binom{E_n}{2} \leqslant {T_n}^3\Big/\binom{E_n}{2}. \tag{2-3}$$

By (2 − 1) and (2 − 2) we see that:

$$\begin{aligned} \ln\left[|\text{ Range } d\ |\Big/\binom{E_n}{2}\right] &\leqslant 3 \ln T_n - \ln \binom{E_n}{2} \\ &\cong 3\pi(n/6)^{1/2} - 3 \ln[4n(3)^{1/2}] - n \ln n \to -\infty \end{aligned} \tag{2-4}$$

and hence

$$|\text{ Range } d\ |\Big/\binom{E_n}{2} \to 0.$$

This result says that the number of interpoint distance ties becomes necessarily very large as n gets large as a consequence of the symmetries of the lattice and the invariance of our measures. Stronger results may be obtained by more detailed analysis of | Range d |. For example, if d is APPROX or TI2MINUS, then | Range | $= n$ and we have in the range $n = 10$ to $n = 100$

$$\ln\left[|\text{ Range } d\ |\Big/\binom{E_n}{2}\right] \simeq \ln n - 2 \ln E_n \simeq \ln n - n \ln n = (1 - n) \ln n. \tag{2-5}$$

ACKNOWLEDGMENTS

We are most indebted to Paul W. Holland, Donald C. Olivier, and David V. Cross, without whose practical and moral support this study would have been impossible. We must also thank Nelson Goodman, David H. Krantz, Joseph B. Kruskal, Paul R. Levitt, G. R. Lockhead,

Roger N. Shepard, and Amos Tversky for discussion of conceptual and technical aspects of the present study. The information in Table XI is reproduced from unpublished data of Thomas T. Struhsaker; we are indebted to him and to Harvard University Press for permission to use this material in the form in which it is published in Cohen (1971), as well as to Joel E. Cohen for discussion of relevant substantive issues. Figure 7, originally published by the late Gösta Ekman, is reproduced here with the kind permission of Dr. Teodor Künnapas. Funding for the computation on which the present work is based was obtained from several sources. Thanks are most especially owing to Ann E. Moskol of the Division of Engineering and Applied Physics of Harvard University; to the Cambridge Project for a research grant during the academic year 1970–71; to Richard C. Atkinson of Stanford University, through NSF Grant GJ-443X2; to Harrison C. White of Harvard University, through NSF Grant GS-2689; and to the Society of Fellows of Harvard University, which authorized use of computer funds for the present project through the generosity of its chairman Wassily Leontief. Finally, we are grateful to A. P. Dempster for administration of Cambridge Project funds through the Harvard Department of Statistics, and to D. B. Yntema for his cooperation as director of the Cambridge Project in the logistically difficult task of transcontinental coordination.

REFERENCES

ARNOLD, J. B. A multidimensional scaling study of semantic distance. *Journal of Experimental Psychology Monograph*, 1971, **90**, 2, 349–372.

ATTNEAVE, F. Dimensions of similarity. *American Journal of Psychology*, 1950, **63**, 516–556.

ATTNEAVE, F. *Applications of information theory to psychology*. New York: Holt, Rinehart and Winston, 1959.

BAVELAS, A. Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 1950, **22**, 725–730.

BEALS, R., KRANTZ, D. H., AND TVERSKY, A. Foundations of multidimensional scaling. *Psychological Review*, 1968, **75**, 127–142.

BENZÉCRI, J. P. Analyse factorielle des proximités. *Publications de l'Institute de Statistique de l'Université de Paris, I.* 1964, **13**; II. 1965, **14**.

BERSTED, C. T., BROWN, B. R., AND EVANS, S. H. Evaluation of unconstrained sorting data. *Behavioral Research Methodology and Instrumentation*, 1970, **2**, 108–110.

BIRKHOFF, G. *Lattice theory*, Revised edition. Providence, R. I.: American Mathematical Society, 1967.

BLOCK, J. *The Q-sort method in personality assessment and psychiatric research*. Springfield, Ill.: Charles C. Thomas, 1961.

BLUMENTHAL, L. M. *Theory and applications of distance geometry*. Oxford: Clarendon Press, 1953.

BOORMAN, S. A. Metric spaces of complex objects. Unpublished senior honors thesis, Harvard College, Division of Engineering and Applied Physics, 1970.

BOORMAN, S. A. AND ARABIE, P. Structural measures and the method of sorting. In R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. 1, New York: Seminar Press, 1972.

BOORMAN, S. A. AND OLIVIER, D. C. Metrics on spaces of finite trees. *Journal of Mathematical Psychology*, 1973, 10, **7**, 26–59.

BOULTON, D. M. AND WALLACE, C. S. The information content of a multistate distribution. *Journal of Theoretical Biology*, 1969, **23**, 269–278.

BOYD, J. Information distance for discrete structures. In R. N. Shepard, A. K. Romney, and

S. Nerlove (Eds), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. 1, New York: Seminar Press, 1972.

CARROLL, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. 1, New York: Seminar Press, 1972.

CARROLL, J. D. AND CHANG, J. J. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart–Young" decomposition. *Psychometrika*, 1970, **35**, 283–319.

CHOMSKY, N. AND MILLER, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2, New York: Wiley, 1963.

COHEN, J. E. *Casual groups of monkeys and men.* Cambridge, Mass.: Harvard University Press, 1970.

COOLIDGE, J. L. *A history of geometrical methods.* New York: Dover, 1963.

CRAMÉR, H. *Mathematical methods of statistics.* Princeton, N. J.: Princeton University Press, 1946.

CROSS, D. V. Multidimensional stimulus control of the discriminative response in experimental conditioning and psychophysics. Technical Report No. 05613-4-F (78[d]), University of Michigan, 1965a.

CROSS, D. V. Metric properties of multidimensional stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization.* Stanford: Stanford University Press, 1965b.

CSISZÁR, I. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicum Hungaricae*, 1967, **2**, 299–318.

DALRYMPLE-ALFORD, E. C. Measurement of clustering in free recall. *Psychological Bulletin*, 1970, **74**, 32–34.

EKMAN, G. Dimensions of color vision. *Journal of Psychology*, 1954, **38**, 467–474.

EKMAN, G. Comparative studies on multidimensional scaling and related techniques. Reports from the Psychological Laboratories, University of Stockholm, Supplement 3, November, 1970.

FISCHER, W. AND MICKO, H. C. More about metrics of subjective spaces and attention distributions. *Journal of Mathematical Psychology*, 1972, **9**, 36–54.

FLAMENT, C. La mesure en psychologie sociale. In J. M. Faverge *et al.* (Eds.), *Les problèmes de la mesure en psychologie.* Paris: Presses Universitaires de France, 1962.

FLAMENT, C. *Applications of graph theory to group structure.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

FORD, L. R., JR. AND FULKERSON, D. R. *Flows in networks.* Princeton, N. J.: Princeton University Press, 1962.

FRANKEL, F. AND COLE, M. Measures of category clustering in free recall. *Psychological Bulletin*, 1971, **76**, 39–44.

GOODMAN, L. A. AND KRUSKAL, W. H. Measures of association for cross-classifications. *Journal of the American Statistical Association*, 1954, **49**, 732–764.

GOODMAN, N. Order from indifference. *Problems and projects.* Indianapolis: Bobbs-Merrill, 1972.

GUPTA, H., GWYTHER, C. E., AND MILLER, J. C. P. *Tables of partitions.* Royal Society Mathematical Tables, Vol. 4, Cambridge: Cambridge University Press, 1959.

HARDY, G. H. AND RAMANUJAN, S. Asymptotic formulae in combinatory analysis. *Proceedings of the London Mathematical Society*, 1918, **17**, 75–115.

HAYS, W. L. Lattice models in psychological scaling. Unpublished, University of Michigan, 1965.

HENLEY, N. M. A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 176–184.

HOLLAND, P. W. Reexpressing Landau's h in terms of triad frequencies. Unpublished, Department of Statistics, Harvard University, 1971.

JARDINE, N. AND SIBSON, R. *Mathematical taxonomy* London: Wiley, 1971.

JOHNSON, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, **32**, 241–254.

JOHNSON, S. C. Metric clustering. Unpublished, Bell Laboratories, 1968.

JOHNSON, S. C. Personal communication, May, 1968.

JÓNSSON, B. Personal communication, January, 1971.

JOYAL, A., LABELLE, G., AND LORRAIN, F. Equivalence relations and linear operations. Unpublished, Université de Montréal, no date.

KENDALL, M. G. *Rank correlation methods*, 3rd edition. London: Griffin, 1962.

KHINCHIN, A. I. *Mathematical foundations of information theory*. New York: Dover, 1957.

KLAHR, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. *Psychometrika*, 1969, **34**, 319–330.

KOTZ, S. Recent results in information theory. *Journal of Applied Probability*, 1966, **3**, 1–93.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964a, **29**, 1–28.

KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964b, **29**, 115–129.

KRUSKAL, J. B. AND CARROLL, J. D. Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis II*. New York: Academic Press, 1969.

KULLBACK, S. *Information theory and statistics*. New York: Wiley, 1955.

LANDAU, H. G. On dominance relations and the structure of animal societies: I. Effect of inherent characteristics. *Bulletin of Mathematical Biophysics*, 1951, **13**, 1–18.

LEVINS, R. Toward an evolutionary theory of the niche. In E. T. Drake (Ed.), *Evolution and environment*. New Haven: Yale University Press, 1968a.

LEVINS, R. *Evolution in changing environments*. Princeton, N. J.: Princeton University Press, 1968b.

LOCKHEAD, G. R. Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology*, 1970, **85**, 1–10.

LORRAIN, F. *Réseaux sociaux et classifications sociales*. Paris: Éditions du Hermann, in press; English-language version: doctoral thesis, Department of Sociology, Harvard University, 1972.

LUCE, R. D. AND GALANTER, E. Discrimination. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1, New York: Wiley, 1963.

LUCE, R. D., BUSH, R. R., AND GALANTER, E., Eds. *Handbook of mathematical psychology*. 3 vol., New York: Wiley, 1963.

MANDELBROT, B. Information theory and psycholinguistics: A theory of word frequencies. In P. F. Lazarsfeld and N. W. Henry (Eds.), *Readings in mathematical social science*. Chicago: Science Research Associates, 1966.

MCGEE, V. E. The multidimensional scaling of "elastic" distances. *The British Journal of Mathematical and Statistical Psychology*, 1966, **19**, Part 2, 181–196.

MICKO, H. C. AND FISCHER, W. The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. *Journal of Mathematical Psychology*, 1970, **7**, 118–143.

MILLER, G. A. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 1969, **6**, 169–191.

MIRKIN, C. G. AND CHORNY, L. B. Ob izmerenii bliznosti mezhdu razlichnymi konechnovo mnozhestva ob'ektov. (On measurement of proximity between various partitions of a finite set.) *Avtomatika i Telemekhanika*, No. 5, 1970, 120–127.

MOSER, L. AND WYMAN, M. An asymptotical formula for the Bell numbers. *Transactions of the Royal Society of Canada*, Section 3 (Series III), 1955, **49**, 49–53.

NUMMENMAA, T. The language of the face. Jyväskylä Studies in Education Psychology and Social Research, University of Jyväskylä (Finland), 1964, No. 9.

OLIVIER, D. C. Metrics for comparison of multidimensional scalings. Unpublished, Harvard University, 1970.

ORE, O. Theory of equivalence relations. *Duke Mathematical Journal*, 1942, **9**, 573–627.

PARRY, W. *Entropy and generators in ergodic theory*. New York: Benjamin, 1969.

PAVY, D. Personal communication, January, 1968.

PEARSON, W. H. Estimation of a correlation coefficient from an uncertainty measure. *Psychometrika*, 1966, **31**, 421–433.

RAJSKI, C. A metric space of discrete probability distributions. *Information and Control*, 1961, **4**, 371–377.

RAND, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 1971, **66**, 846–850.

RAPOPORT, A. AND FILLENBAUM, S. Experimental studies of semantic structure. In R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. 2, New York: Seminar Press, 1972.

RESTLE, F. A metric and an ordering on sets. *Psychometrika*, 1959, **24**, 207–220.

ROENKER, D. L., THOMPSON, C. P., AND BROWN, S. C. Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 1971, **76**, 39–44.

ROSKAM, E. E. A comparison of principles for algorithm construction in nonmetric scaling. Michigan Mathematical Psychology Program Report 692. Ann Arbor, Mich.: 1969.

RUMELHART, D. E. AND ABRAHAMSON, A. A. Toward a theory of analogical reasoning. Technical Report 18, Center for Human Information Processing, University of California, San Diego, 1971.

SABIDUSSI, G. The centrality index of a graph. In *Theory of Graphs, International Symposium* (Rome, 1966). New York: Gordon and Breach, 1967.

SCHÖNEMANN, P. H. An algebraic solution for a class of subjective metrics models. *Psychometrika*, 1972, 37, **4**, 441–451.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, **27**, 125–140; II. *Psychometrika*, 1962, **27**, 219–246.

SHEPARD, R. N. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1964, **1**, 54–87.

SHEPARD, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 1966, **3**, 287–315.

SHEPARD, R. N. Some principles and prospects for the spatial representation of behavioral science data. Paper presented at the Mathematical Social Science Board Advanced Research Seminar, Irvine, California, June 13–18, 1969.

SHEPARD, R. N. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. 1, New York: Seminar Press 1972.

SHEPARD, R. N. AND CARROLL, J. D. Parametric representation of nonlinear data structures. In P. R. Krishnaiah (Ed.), *International symposium on multivariate analysis, Dayton, Ohio*, 1965. New York: Academic Press, 1966.

SHEPARD, R. N. AND CERMAK, G. W. Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, in press.

SHEPARD, R. N., ROMNEY, A. K., AND NERLOVE, S. (Eds.). *Multidimensional scaling: Theory and applications in the behavioral sciences*. 2 vol., New York: Seminar Press, 1972.

STENSON, H. H. AND KNOLL, R. L. Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychological Bulletin*, 1969, **71**, 122–126.

STRUHSAKER, T. T. *Behavior of vervet monkeys* (Cercopithecus aethiops). University of California Publications in Zoology, Vol. 82, Berkeley and Los Angeles: University of California Press, 1967a.

STRUHSAKER, T. T. Social structure among vervet monkeys *(Cercopithecus aethiops)*. *Behavior*, 1967b, **29**, 83–121.

SZÁSZ, G. *Introduction to lattice theory*, 3rd revised edition. New York: Academic Press, 1963.

TORGERSON, W. S. Multidimensional scaling of similarity. *Psychometrika*, 1965, **30**, 379–393.

TVERSKY, A. AND KRANTZ, D. H. Similarity of schematic faces: A test of inter-dimensional additivity. *Perception & Psychophysics*, 1969, **5**, 124–128.

WELLS, M. B. *Elements of combinatorial computing*. Oxford: Pergamon Press, 1971.

WENDER, K. Die psychologische Interpretation nichteuklidischer Metriken in der multidimensionalen Skalierung. Doctoral dissertation, Technical Univ. Darmstadt, 1969.

WHITE, H. C. Chance models of systems of casual groups. *Sociometry*, 1962, **25**, 153–172.

WHITMAN, P. Lattices, equivalence relations and subgroups. *Bulletin of the American Mathematical Society*, 1946, **52**, 507–522.

YOUNG, F. W. Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, 1970, **35**, 455–473.

YOUNG, F. W. AND TORGERSON, W. S. TORSCA, A FORTRAN IV program for Shepard–Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.

ZINNES, J. L. Scaling. In *Annual Review of Psychology*, Vol. 20. Palo Alto, Calif.: Annual Reviews, Inc., 1969.

RECEIVED: July 1, 1972