Rational Distance Functions for Multidimensional Scaling[1]

DAVID H. KRANTZ

The University of Michigan, Ann Arbor, Michigan 48104

A rational distance function is a numerical measure of psychological distance whose geometric properties are deducible from psychological truths about some particular judgmental task. In this paper, we review two theoretical analyses that have led to proposed rational distance functions. These analyses are based on two different tasks: paired-associate learning and similarity judgments. A generalization of the theory on similarity judgments is presented.

Empirical results concerning similarity judgments seriously conflict with the basic psychological assumptions in the generalized treatment of similarity judgments. We conclude from these results that the construction of valid psychologically-based distance functions from analysis of choice probabilities in similarity judgments requires, as an initial step, the development of scaling models that take into account the influence of "irrelevant" dimensions on choice probability.

I. INTRODUCTION

In the first stage of multidimensional scaling, some sort of standard measurement procedure is used to evaluate the pairwise differences among the objects that are being scaled. In the second stage, the objects are represented by points in an Euclidean space of low dimension, in such a way that the interpoint distances reflect as closely as possible the corresponding pairwise differences.

There are two senses in which the Euclidean distances can reflect pairwise differences. Either the distances may reproduce some actual numerical evaluations of the pairs, obtained in stage 1, or they may merely reproduce the order of the pairs. Some of the methods for carrying out stage 2 of multidimensional scaling assume that numerical evaluations are to be reproduced (Torgerson, 1952, 1958), while other methods seek a representation that reflects only the order (Bennett and Hays, 1960; Coombs, 1952; Shepard, 1962). Shepard's impressive reductions of complex data

to spaces of low dimensionality have elevated the "ordinal" methods to the forefront of multidimensional scaling techniques.

The validity of either "numerical" or "ordinal" methods requires some strong theoretical suppositions. The purpose of the present paper is to examine some of the possible theories within which "numerical" methods may be partially justified. A theoretical analysis of "ordinal" methods will be presented elsewhere (Beals, Krantz, and Tversky, 1967; Tversky, 1966).

The assumption that the numbers obtained from some measurement procedure behave as distances in an Euclidean space is itself a strong theoretical statement about that particular procedure. Even the weaker assumption that the numbers satisfy the triangle inequality is a theoretical statement about the procedure. Such statements seem unnatural as *starting points* for theory, since they shed little light on the *psychology* involved in the measurement procedure under discussion. Moreover, several different measurement procedures have been employed to obtain numerical evaluation of pairwise differences for stage 1 of multidimensional scaling, including application of Thurstonian models (Messick, 1956; Torgerson, 1952, 1958) and direct estimation techniques (Indow and Uchizono, 1960). It would seem natural to start with a theoretical, psychological analysis of a particular procedure, and within such an analysis, to search for a *rational distance function*, whose distance-like behavior is a consequence of psychological truths about the procedure in question. Such an approach presupposes a serious interest in the psychological processes underlying the judgmental task, as well as in the geometric representation of object differences. But without this sort of analysis, it is hard to see what logical justification there can be for trying to reproduce numerical evaluations by means of distances.

As will be seen, the difficulties encountered in the development of rational distance functions are severe, because the judgmental tasks are psychologically complicated. In particular, the probability or the confidence with which one difference is judged smaller than another depends on factors other than the "sizes" of the two differences; this produces violations of a property that I shall call *simple scalability*.

Sections II and III below review two recent attempts to obtain rational distance functions, by Luce (1961) and Shepard (1958a). Shepard's theoretical analysis of stimulus confusions in paired-associate learning leads to a particular transformation of the stimulus confusion probabilities that ought to behave as a distance measure. Luce applies his general theory of choice to the case of similarity judgments, deriving a transformation of choice probabilities with distance properties. In Section IV, theoretical results similar to Luce's are presented, obtained under more general assumptions about choice behavior. Section V presents data analyses that show clearly the great difficulties encountered in using choice models to obtain distance functions; in particular, failure of simple scalability is demonstrated and discussed. Section VI gives a brief discussion of possibilities for revisions of theory that could lead to satisfactory rational distance functions.

II SHEPARD'S ANALYSIS OF STIMULUS CONFUSIONS

Shepard (1957) analyzed paired-associate response probabilities into a convolution of two theoretical probability distributions, one governing stimulus confusions, the other, response confusions. A second paper (Shepard, 1958a) proposed a relation between the theoretical stimulus confusion probabilities and the "psychological distance" between stimuli, in the context of paired-associate learning.

Shepard assumed that when a stimulus S_i is presented on one trial of a paired-associate learning experiment, a large number of stimulus trace elements are conditioned from S_i to an internal representation of the stimulus, denoted S_i^*. He further assumed that with passage of time, some of the trace elements from that trial *diffuse* in the psychological space of the internal representations, becoming conditioned instead to representations S_k^* of stimuli S_k, for values of $k \neq i$. In addition to this diffusion process, he assumed a *decay* process, over time, for the trace elements from a given trial, the two stochastic processes operating independently.

By making some simple, qualitative assumptions about these two processes, Shepard was able to deduce a relation between psychological distance and stimulus confusion probabilities. His key postulate, which contains an implicit *definition* of psychological distance, is the following: the probability that, in some small time interval, a trace diffuses from S_j^* to S_k^*, depends on the *psychological distance* from S_j^* to S_k^* (or from S_j to S_k) but not upon the previous history of that trace, nor upon the "absolute locations" of S_j^* and S_k^* in psychological space. From this, and other simple qualitative postulates, Shepard rigorously deduced that *generalization* is an exponential decay function of (Euclidean) psychological distance. More precisely, he proved that

$$\left[\frac{P_{ij}^S \cdot P_{ji}^S}{P_{ii}^S \cdot P_{jj}^S} \right]^{1/2} = e^{-KD_{ij}} \tag{1}$$

where P_{ij}^S denotes the probability that, when stimulus S_i is presented, it is taken to be S_j ($i = j$ is possible). P^S is exactly the theoretical stimulus confusion probability mentioned above. The symmetric psychological distance between S_i and S_j (which must be assumed Euclidean to justify the derivation of Eq. 1 above) is denoted D_{ij}. The constant K depends on the diffusion rate and the decay rate of the stimulus traces. The expression on the left of the equation is a symmetrized, normalized measure of stimulus generalization, in terms of the stimulus confusion probabilities P^S.

Essentially the same equation relating psychological distance to stimulus confusions was proposed by Luce (1963) on the basis of a quite different psychological theory, involving a choice-theory analysis of complete identification experiments.

To test the above theory, Shepard compared empirical values of the stimulus generalization measure, obtained from several experiments in different laboratories, with the predicted exponential decay curve. For the most part, these tests were based

on the assumption that the distance measures D_{ij} are proportional to *physical* difference measures. He attempted to justify this by pointing out that the stimuli employed were evenly spaced at small intervals along the physical dimension, and that the property of even spacing would be preserved by any continuous, differentiable transformation of the physical measure.

Shepard's test of his theory is incorrect, however. The argument that differentiable transformations preserve even spacing of small intervals is correct only if the psychological distance D_{ij} is a function of physical distance $|x_i - x_j|$ on the continuum, i.e.,

$$D_{ij} = f(|x_i - x_j|) \tag{2}$$

for some differentiable function f. If this were the case, then, indeed, one would have D_{ij} approximately proportional to $|x_i - x_j|$, where the constant of proportionality is $f'(0)$, the derivative of f at 0. However, the very assumption that there is *any* single-valued functional relation between physical and psychological distance begs the entire question, since it might easily be that equal physical distances, in different parts of the scale, go over into grossly unequal psychological distances (suppose, for example, that physical distance was measured in decibels, psychological distance was difference in mean magnitude estimates, and Stevens' power law held). Rather more plausible than Eq. 2 is the assumption

$$D_{ij} = |g(x_i) - g(x_j)| \tag{3}$$

where g is a "psychophysical function" mapping the physical values x_i into psychological space. Equation 3 is *logically equivalent* to additivity of psychological distance, and is thus inherent both in Shepard's theoretical analysis (since D_{ij} is Euclidean, hence, in one dimension, it is additive) and in his method of plotting data (since he assumes that two equal steps represent double the psychological distance of one step). However, Eq. 3 does *not* imply approximate proportionality of psychological distance and physical difference; in fact, D_{ij} is approximately $|g'(x_i)| \cdot |x_i - x_j|$, where the proportionality constant, $|g'(x_i)|$, can vary from one part of the physical dimension to another.

If one makes the unlikely assumption that both (2) and (3) hold, then the resulting functional equation implies that both f and g are *linear* functions of the physical measures, so that psychological distance and physical distance are essentially indistinguishable. This certainly is wrong for arbitrarily chosen physical measures. In view of the incorrectness of the logic behind Shepard's test, it is very difficult to interpret the fact that he obtains excellent fits to an exponential decay function. Certainly, this cannot be regarded as evidence for the theoretical analysis that led to Eq. 1, unless the various experimenters whose data are used all happened to choose physical measures that closely represent psychological distances.

How ought Shepard's theory be tested? An alternative approach is to compute the psychological distances D_{ij} by means of Eq. 1 and to see whether the distances in fact

satisfy the constraints assumed in the model—i.e., whether they are well fit by Eucli-
dean distances, in a suitable number of dimensions. (Of course, this only tests a
consequence of the theory, rather than the entire theory.) In particular, the quantities
D_{ij} must satisfy the triangle inequality: for any three stimuli S_i, S_j, S_k,

$$D_{ij} + D_{jk} \geqslant D_{ik}.$$

The triangle inequality, however, follows if the following condition is true for
all i, j, k:

$$\frac{P_{ij}^S}{P_{jj}^S} \leqslant \frac{P_{ik}^S}{P_{jk}^S}. \tag{4}$$

Although (4) is not quite mathematically necessary, it is hard to see how the triangle
inequality can be satisfied unless it holds.

When (4) is satisfied nontrivially (i.e., the two expressions are both smaller than one)
its content can be phrased as follows: for any three stimuli, S_i, S_j, S_k, the confusions
of S_i with S_k and of S_j with S_k are most unequal when S_k coincides with S_j or S_i.
In other words, the generalization gradient between S_j and S_i is steepest when one
of them is the "training stimulus."

The preceding condition on generalization gradients, when combined with a theory
of discrimination learning based on absolute responding, predicts failure of transposi-
tion. Suppose that a rat in a dark-light discrimination situation is reinforced
for approach to a gray card (S_2) and for not approaching a white card (S_1). Suppose
that, asymptotically, errors are entirely due to stimulus confusions. Then on a trans-
position test, with a black card (S_3) and the same gray card (S_2), the tendency to
approach S_3 is due to generalization from S_2, while the tendency to avoid S_3 is due to
generalization from S_1. The tendency is similar for approach and avoidance for S_2.
Thus the ratio of approach probability to avoidance probability is P_{32}^S/P_{31}^S for S_3 and is
P_{22}^S/P_{21}^S for S_2. By (4), the latter ratio is larger; a simple Markov model of choice-point
behavior predicts that the rat will choose S_2 more often than S_3, i.e., transposition
fails.

Of course, the fact that transposition is observed cannot be interpreted as evidence
against Shepard's analysis, since (among other things) the assumption of absolute
responding is suspect: transposition may depend on genuine relational learning.

Shepard's theory provides a solid rationale for applying Euclidean multidimensional
scaling models to the distance measures D_{ij} derived by Eq. 1 from paired-associate
confusion probabilities: the goodness-of-fit of the Euclidean model in such applications
provides the most adequate test of the theory. Shepard (1958b) has applied this
kind of test to data from small one-dimensional and two-dimensional experiments with
excellent results.

III. LUCE'S ANALYSIS OF SIMILARITY JUDGMENTS

The analysis of choice behavior by Luce (1959) is based on the postulate that relative values of nonzero choice probabilities are unaffected by "irrelevant" alternatives. This analysis was applied by Luce (1961) to a generalization of the method of triads, in which several stimuli are compared to a standard, and the subject chooses the one most similar to the standard.

Let T denote the set of comparison stimuli and let $P_T(x; z)$ denote the probability that comparison stimulus x is chosen as most similar to the standard, z. For $T = \{x, y\}$, we abbreviate $P_{\{x, y\}}(x; z)$ as $P(x, y; z)$. For simplicity, Luce assumes that all the choice probabilities are nonzero. The basic postulate then guarantees the existence of numerical scale values $v(x, z)$, defined for all comparison stimuli x and all standards z, such that

$$P_T(x; z) = \frac{v(x, z)}{\sum_{y \in T} v(y, z)}. \tag{5}$$

Intuitively, $v(x, z)$ is a measure of the similarity of x to z, since the larger it is, the more likely is x to be chosen as most similar to z. The question of symmetry of this similarity measure is a tricky one, since the v-scale is unique only up to multiplication of the values of $v(\cdot, z)$ by constants $k(z)$. Thus, for two specific stimuli x and y, one can always obtain $v(x, y) = v(y, x)$ by suitable adjustment of the units of one of the two v-scales, $v(\cdot, x)$ or $v(\cdot, y)$. However, as Luce points out, under certain conditions the units for all the scales $v(\cdot, z)$ can be chosen so as to guarantee $v(x, y) = v(y, x)$ for all x and y simultaneously.

Luce also examines the psychological plausibility of the functions $1/v(x, y)$ and $\log[1/v(x, y)]$ as distance measures. The former is unreasonable since it requires $v(x, x) = \infty$, which implies that for $v(y, x)$ finite, $P(y, x; x) = 0$ no matter how similar y is to x. For $\log[1/v(x, y)]$ to act as a distance measure, the units of the v-scales must be chosen so that $v(x, x) = 1$ for all x. If the units are to be adjusted so that v is also symmetric, then $P(x, y; y) = P(y, x; x)$ for all x and y. Furthermore, the triangle inequality is equivalent to $v(x, y) \leqslant v(x, z)/v(y, z)$, i.e.,

$$P(x, y; y) \leqslant P(x, y; z). \tag{6}$$

The property expressed by (6) is very similar to the one expressed by (4) in Shepard's analysis, and similar intuitions underlie the plausibility of either. Suppose the comparison stimuli, x and y, are fixed, and the standard z is varied. Then it is plausible that discrimination will be sharpest (choice probability farthest from $\frac{1}{2}$) when z coincides with one of the comparison stimuli. This is equivalent to property (6) above. If, in addition, the two extreme choice probabilities $P(x, y; y)$ and $P(x, y; x)$ are equally far from $\frac{1}{2}$, then $P(x, y; y) = P(y, x; x)$ as required above.

Thus, the basic metric axioms, at least, have a fairly transparent and plausible psychological meaning for the proposed rational distance function $\log [1/v(x, y)]$. The meaning of more sophisticated geometric properties, including Euclidean ones, is much less transparent.

Critical discussion of Luce's analysis is deferred to Section V, since all the points that I shall raise apply equally to the generalizations of Luce's work that are presented in Section IV.

IV. A SCALING MODEL FOR THE METHOD OF TRIADS

The analysis of similarity judgments presented in Section III can be generalized considerably. The assumption that all probabilities are different from 0, even if true, is impractical, since one easily obtains relative frequencies of zero that are useless for scaling. It can be replaced by a much weaker "density" assumption. The scaling Eq. 5 based on Luce's postulate can be replaced (for pairwise choices only) by any equation of form

$$P(x, y; z) = F[w(y, z) - w(x, z)] \tag{7}$$

where F is a known function of one variable (embodying much of one's theory about the choice process) and $w(x, z)$ etc. are appropriate disimilarity scale values. For Luce's model, $F(t) = (1 + e^{-t})^{-1}$ and $w(x, z) = \log [1/v(x, z)]$; in Thurstone's Case V,

$$F(t) = \int_{-\infty}^{t} (2\pi)^{-(1/2)} e^{-(1/2)s^2} ds.$$

It is assumed that F will be known explicitly in any particular application. The gain in generality by leaving F unspecified lies in the presentation of a unified analysis, applicable to any particular scaling formula satisfying (7).

The basic scaling axiom can be formulated as follows:

(A) *There exist functions w and F such that w assigns a real number to each pair of stimuli and F is a strictly increasing function mapping the real numbers into themselves, with $F(0) = \frac{1}{2}$, and such that whenever $P(x, y; z) \neq 0$ or 1, then*

$$P(x, y; z) = F[w(y, z) - w(x, z)].$$

To maneuver around 0 and 1 probabilities, we require a density assumption somewhat stronger than the one used by Luce (1959) for a similar purpose. The following is a preparatory definition.

DEFINITION. A pair of stimuli (x, y) is *confusable* if for every standard z, $P(x, y; z) \neq 0, 1$.

The density assumption can now be formulated.

(B) *If $P(x, y; z) = 0$, then a finite sequence of stimuli $x_0, ..., x_n$ can be chosen, such that $x_0 = x$, $x_n = y$, and for $i = 1, ..., n$, (x_{i-1}, x_i) is a confusable pair and $P(x_{i-1}, x_i; z) \leqslant \frac{1}{2}$.*

If (A) and (B) hold, then for any fixed standard z, the scale values $w(x, z)$ are unique up to an additive constant. This is shown in the following theorem:

THEOREM 1. *Suppose that P, F, w are as above, with (A) and (B) satisfied, and that w' is another function such that when $P(x, y; z) \neq 0, 1$, then*

$$P(x, y; z) = F[w'(y, z) - w'(x, z)].$$

Then for each z there is a constant $k(z)$ such that for all x,

$$w'(x, z) = w(x, z) + k(z).$$

Proof. Let $k(z) = w'(z, z) - w(z, z)$. For arbitrary x we can use (B) to choose $x_0, ..., x_n$, with $x_0 = x$, $x_n = z$, and (x_{i-1}, x_i) confusable. Then

$$F[w'(x_i, z) - w'(x_{i-1}, z)] = F[w(x_i, z) - w(x_{i-1}, z)],$$

and since F is strictly increasing, it follows that

$$w'(x_i, z) - w(x_i, z) = w'(x_{i-1}, z) - w(x_{i-1}, z),$$

for $i = 1, ..., n$. Putting these equations together yields

$$w'(x, z) - w(x, z) = w'(z, z) - w(z, z) = k(z).$$

In Luce's analysis, Eq. 6 led to the triangle inequality. In the present analysis, we shall need a stronger assumption than $P(x, y; y) \leqslant P(x, y; z)$: namely, that, when the x to z distance is larger than the x to y distance, there is a chain $x_0, ..., x_n$, with $x_0 = x$, $x_n = y$, (x_{i-1}, x_i) confusable, and

$$P(x_{i-1}, x_i; y) \leqslant P(x_{i-1}, x_i; z).$$

[If all probabilities are nonzero, then this assumption will be implied by $P(x, y; y) \leqslant P(x, y; z)$, since the chain can consist simply of x and y.] In order to formulate the assumption rigorously, we introduce a formula for the "distance" from x to y:

$$d(x, y) = \tfrac{1}{2} [w(x, y) + w(y, x)] - \tfrac{1}{2} [w(x, x) + w(y, y)]. \tag{8}$$

Note that d is an absolute scale, invariant under the permissible transformations of the w-scale specified in Theorem 1. Note also its similarity to Shepard's distance measure as given by Eq. (1); if $w_{ij} = \log [1/P_{ij}^S]$, Shepard's formula is identical except for a multiplicative constant. If Eq. 8 is used to calculate d from w, then it is convenient to choose the zeros of the scales $w(\cdot, z)$ so that $w(z, z) = 0$ for all z.

The needed assumptions may be stated as follows:

(C) *Let d be defined by (8). If $d(x, y) \leqslant d(x, z)$ then there is a finite sequence $x_0, \ldots, x_n$ such that $x_0 = x$, $x_n = y$, and for $i = 1, \ldots, n$, (x_{i-1}, x_i) is confusable and $P(x_{i-1}, x_i; y) \leqslant P(x_{i-1}, x_i; z)$.*

(D) *For all x, y, $P(x, y; y) \leqslant \frac{1}{2}$.*

With these assumptions, we can establish the following theorem.

THEOREM 2. *Suppose that Assumptions (A)-(D) hold and that d is defined by Eq. 8. Then d is a pseudometric, i.e., for all x, y, z,*

(M1) $d(x, y) \geqslant 0$,

(M2) $d(x, x) = 0$,

(M3) $d(x, y) = d(y, x)$,

(M4) $d(x, y) + d(y, z) \geqslant d(x, z)$.

Note that d is a true *metric* if the converse to (M2) holds, i.e., $d(x, y) = 0$ only if $x = y$. In any pseudometric the relation $d(x, y) = 0$ is an equivalence relation, and if d is redefined so that its domain consists of the equivalence classes, then it becomes a true metric.

Proof. (M2) and (M3) follow from (8) without any assumptions.

(M1) follows easily from (D), using (A) and (B).

To prove (M4), let x, y, z be arbitrary. The required inequality is trivial (using (M1)) unless both $d(x, y) \leqslant d(x, z)$ and $d(y, z) \leqslant d(x, z)$. In the latter case, we can use (C) twice [once in conjunction with (M3)] to find finite sequences $x_0, \ldots, x_m$, $z_0, \ldots, z_n$ such that $x_0 = x$, $z_0 = z$, and $x_m = y = z_n$, and for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, (x_{i-1}, x_i) and (z_{j-1}, z_j) are confusable and

$$P(x_{i-1}, x_i; y) \leqslant P(x_{i-1}, x_i; z) \qquad \text{and} \qquad P(z_{j-1}, z_j; y) \leqslant P(z_{j-1}, z_j; x).$$

Then by (A), for $i = 1, \ldots, m$ and $j = 1, \ldots, n$,

$$w(x_i, y) - w(x_{i-1}, y) \leqslant w(x_i, z) - w(x_{i-1}, z)$$

$$w(z_j, y) - w(z_{j-1}, y) \leqslant w(z_j, x) - w(z_{j-1}, x).$$

Summing the left and right sides of all $n + m$ of these inequalities yields

$$w(y, y) - w(x, y) + w(y, y) - w(z, y) \leqslant w(y, z) - w(x, z) + w(y, x) - w(z, x).$$

Rearranging terms, subtracting $w(x, x) + w(z, z)$ from both sides, and applying (8) yields

$$d(x, y) + d(y, z) \geqslant d(x, z)$$

as required. This completes the proof.

Some justification is needed for using the function d in the formulation of Assumption (C) to capture the notion that the x to y distance is smaller than or equal to the x to z distance. This seems circular, since Assumption (C) is then used to prove that d has distance properties. There are two probability criteria that partially capture the same notion:

(i) $P(y, z; x) \geqslant \frac{1}{2}$,

(ii) $P(x, y; y) \geqslant P(x, z; z)$.

The first implies $w(z, x) \geqslant w(y, x)$, while the second one, for nonzero probability,s implies $w(y, y) - w(x, y) \geqslant w(z, z) - w(x, z)$. Averaging these two inequalities and rearranging yields $d(x, y) \leqslant d(x, z)$, by (8). Thus, d is a good compromise between the two obvious criteria, and our procedure is not really circular. Moreover, when the zeros of the $w(\cdot, z)$ scales can be chosen so that $w(x, y) = w(y, x)$ and $w(x, x) = 0$ for all x, y, then criteria (i) and (ii) are both equivalent to $d(x, y) \leqslant d(x, z)$ (see below), so the hypothesis of (C) could be changed to $P(y, z; x) \geqslant 1/2$.

It should be remarked that under some conditions, the zeros of the w-scales can be chosen so that $w(x, y) = w(y, x)$ for all x, y. This greatly reduces the number of free parameters in the scaling model. Under still further conditions, this symmetry of w is achieved by the assignment of zeros such that $w(x, x) = 0$ for all x. The distance formula (8) simplifies, in these cases, to one of the following forms:

$$d(x, y) = w(x, y) - \tfrac{1}{2} [w(x, x) + w(y, y)], \tag{9}$$

$$d(x, y) = w(x, y). \tag{10}$$

The exact conditions under which these simplifications are possible are developed in Krantz (1964).

There are two key assumptions under which the function d defined by (8), (9), or (10) is a distance function. These are the scaling assumption [Axiom (A)] and the generalization in (C) of the notion that confusion of comparison stimuli is least when the standard is equal to one of the comparison stimuli. The function d is thus a candidate for a *rational distance function*, since the pseudometric properties are related to meaningful psychological assumptions. The psychology involved in (C) is closely related to the assumptions implicit in Shepard's theory of generalization (Eq. 4). The scaling assumption (A) is a specialization of the more general equation of *simple scalability* (see Krantz, 1964):

$$P(x, y; z) = F[w(x, z), w(y, z)], \tag{11}$$

where F is a function of two variables. Simple scalability states that the influence of the pair (x, z) on the probability of choice is entirely described by one number $w(x, z)$, the scale value of the pair. Thus, if (x, z) and (x', z) have the same scale value, they will be discriminated equally well from any alternative (y, z). Unfortunately, this

assumption can easily fail in practice: two subjective differences (x, z) and (x', z) may be approximately equal, but may differ very greatly in the "*difficulty of choice*" when confronted with (y, z). If x lies in the "same direction" from the standard z as does y, and x' lies in the "opposite direction," then (x, z) and (y, z) will be easily discriminated, leading to choice probabilities far from $\frac{1}{2}$, while (x', z) and (y, z) may be hard to discriminate. "Direction" is an "irrelevant" dimension (see Coombs, 1964, Ch. 23 for a discussion of "laterality") which affects difficulty of choice. Similarly, two tones which have about the same loudness, but differ greatly in the "irrelevant" dimension of pitch, may differ greatly also in ease of discriminability in loudness from a third tone which is more similar to one of them in pitch. In fact, the above phenomenon does occur in data using similarity comparisons: see Section V below.

One main difficulty, then, with the function d defined above is that the psychological assumptions underlying the basic scaling equation overlook a possibly important determinant of choice behavior.

V. EVALUATION OF THE MODEL

The scaling model described in the previous section was applied in the domain of color similarities. These experiments are described elsewhere in more detail (Krantz, 1964; 1967) and only a few essential features will be presented here.

It should be noted, however, that models for color differences have a long history, dating to Helmholtz, which is rather independent of the development of multidimensional scaling in other areas. The experiments below were undertaken in hope of contributing to the color-difference problem, and were considered to be an *application* of the above scaling model, rather than a *test* of it. Thus their design is highly unsuitable, in a number of ways, for testing the model. That their results do actually test the model is a consequence of the fact that both major assumptions, (A) and (C), are so utterly incorrect that their falsity was made evident by crude data.

A. EXPERIMENTAL METHOD

Subjects were presented triads consisting of a standard stimulus, z, and two comparison stimuli, x and y, and asked which color difference was smaller, that between x and z or that between y and z. The stimuli were monochromatic lights adjusted for each individual subject to a constant brightness level, slightly above the brightness of the surround-field. The surround looked nearly neutral. Figure 1 shows a schematic drawing of the stimulus display. Essential stimulus parameters are summarized in Table 1. The top stimulus was designated as standard, the two bottom one as comparison stimuli. Each triad was presented equally often with the comparison stimuli in the two spatial orders.
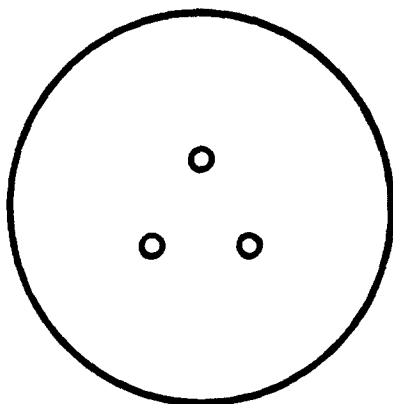
Fɪɢ. 1. Schematic drawing of the stimulus display. The three monochromatic stimuli appeared in three holes arranged at the vertices of a triangle, with the standard at the top. The rest of the large circle appeared nearly neutral; the area outside the large circle was black. See Table 1 for detailed parameters.

TABLE 1

Essential Stimulus Parameters

Test-field visual angle	44′
Separation of test-field centers	3°42′
Surround-field visual angle	23°
Viewing distance of test-field apertures	25.5 in.
Bandwidth of monochromatic fields	50 % of the light passed through exit slits at $\lambda \pm 3.3m\mu$, for nominal wavelength setting of λ.
Test-field luminance	1.55 equiv. ft.-candles
Surround-field color Munsell paper 5 PB 2/6	
Illuminant color temperature	somewhat below 2800° K
Approximate surround-field color temperature	somewhat below 6500° K
Surround field luminance	1.15 equiv. ft.-candles

A total of five subjects was used. Three of them, JPE, KR, and SI, were paid for observing. They had no previous experience in psychophysical experiments. The other two, RMS and HC, were experienced observers, but were naive with respect to the purpose of the experiment. All subjects were color-normal, as indicated by midpoint and range settings for the Rayleigh equation on a Nagel anomaloscope.

Each subject judged repeated presentations of between 90 and 225 different triads. These triads were chosen so as to yield estimates of similarity-scale parameters for a series of interlocked small and large differences spanning a portion of the visible spectrum. During any experimental session, judgments were obtained from one subject for 13-19 different triads, each repeated 8-12 times. Since the total amount of data was limited by the complicated manual control of the apparatus, the sequence of trials in any one session was designed to minimize the number of apparatus changes and to maximize the number of observations. Most triads were replicated in from 3 to 6 sessions per subject, although occasional ones were replicated less often.

B. Tests of the Basic Scaling Assumption

Two types of test of Assumption (A), the basic scaling assumption of Section IV, were carried out. One type of test involved estimation (by maximum likelihood) of the similarity-scale parameters for an application of Luce's choice model to a limited set of triads. In each instance, a chi-square test of fit was performed. The parameters were estimated and the test performed using data from a single subject for the triads in question. Each test of this first type used all possible triads composed of either three or four comparison stimuli with a fixed standard. For example, RMS and JPE were each tested with $600m\mu$ as standard wavelength and 590, 591, 606, and $608m\mu$ as comparison stimuli; a total of six different triads could be formed. Three independent parameters could be estimated for the four color differences, for the model equations

$$P(x, y; z) = \frac{v(x, z)}{v(x, z) + v(y, z)} = F[w(y, z) - w(x, z)],$$

where $F(t) = (1 + e^{-t})^{-1}$, $w(x, z) = \log [1/v(x, z)]$. (There are only three parameters because one of the v-scale values may be chosen arbitrarily.) The prediction of six independent observed frequencies from a three-parameter model led to a chi-square test with three degrees of freedom. Similarly, for a test of the same type, using three comparison stimuli, there were three triads, two independent parameters, and one degree of freedom. For each test, the maximum-likelihood technique led to a set of simultaneous nonlinear equations, which were solved approximately by an iterative method.

From one to three tests of the above type were carried out on each of four subjects, for a total of eight tests. These tests are summarized in the first eight entries in Table 2. The last column of the table refers to the direction of the deviation of observed from predicted frequency; this will be explained below. The number of observations per triad for these tests averaged 46.

The statistical significance of the deviations of observed from predicted frequencies for these tests is beyond question. Moreover, the choice model is rejected even though

TABLE 2

CHI-SQUARE TESTS OF THE SCALING MODEL

Subject	Pairs	df	χ^2	p	Direction
RMS	(560, 567)	1	2.08	.10–.20	extreme
	(560, 552)				
	(560, 540)				
	(600, 608)	3	6.07	.10 +	extreme
	(600, 606)				
	(600, 591)				
	(600, 590)				
JPE	(572, 540)	1	.00	.90–.95	—
	(572, 581)				
	(572, 591)				
	(591, 608)	1	3.14	.05–.10	less extreme
	(591, 581)				
	(591, 572)				
	(600, 608)	3	21.48	.00 +	extreme
	(600, 606)				
	(600, 591)				
	(600, 590)				
SI	(608, 600)	1	8.94	.001	extreme
	(608, 620)				
	(608, 628)				
	(600, 610)	3	4.23	.20–.30	extreme
	(600, 608)				
	(600, 591)				
	(600, 590)				
KR	(608, 600)	1	1.12	.20–.30	extreme
	(608, 620)				
	(608, 628)				
	(600, 608)	–	—	—	extreme
	(600, 606)				
	(600, 592)				
	(600, 591)				
	(600, 590)				
Total		14	47.06	.00 +	7/9 extreme
Total excluding JPE's 21.48		11	25.58	.01	6/8 extreme

the tests involved were biased toward accepting it, in two ways. For one thing, the parameter estimates were not required to be consistent with estimates of the same parameters obtained in the context of the entire experiment. (Some of the color differences involved were used in other triads.) For another thing, a choice was available as to what data should be used for any given test. This choice arose from the circumstance that in each case, there were previous data for some triads collected before the decision was made to include additional triads and test the model. In each case, the older data were included only if a better fit to the model was obtained by including them.

Nevertheless, this first type of test is not entirely convincing with respect to the rejection of Assumption (A). One reason is that it relies on a specific assumption (Luce's choice model) regarding the shape of the scaling function F. Perhaps some alternative scaling function would fit the data better. In order to reject simple scalability (Eq. 11) and, *a fortiori*, to reject (A), we show that the deviations from the choice model are in a direction that agrees qualitatively with an alternative laterality model, which is incompatible with simple scalability.

As was pointed out in Section IV, difficulty of choice should be much less, and probabilities should be more extreme, for "unilateral" triads where both comparison stimuli lie on one side of the standard, than for "bilateral" triads. In particular, this is precisely the prediction of a Thurstonian model, where the discriminal process for a pair of stimuli is the difference random variable of the discriminal processes of the two stimuli in the pair; the resulting correlation of two unilateral pairs is positive, corresponding to an effective reduction of discriminal dispersions (Coombs, 1964, Ch. 23).

Looking at the last column in Table 2 we see that for the eight tests mentioned earlier, the observed frequencies are more extreme than the predicted ones in eight out of eleven unilateral triads, less extreme in two cases, equal in one case. (The three cases of less-extreme-or-equal observed frequencies all occur in JPE's data; however, the large value of chi-square for JPE is due to the one more-extreme-than-predicted unilateral triad.) The ninth test in Table 2, for KR, could not be put in the form of a chi-square, due to an experimental error in which the wrong unilateral triad was presented; however, the results for this unilateral triad are very extreme, as would be predicted from the above analysis.

Some caution must be exercised in interpreting these results statistically, since the expected proportion of deviations of observed frequencies in the more extreme direction will often be slightly greater than 50%, due to the skewness of the binomial distribution. Nevertheless, the above results for unilateral triads seem to support the idea that deviations from simple scalability occur in the anticipated direction.

The second type of test of Assumption (A) involved predicting the data for unilateral triads on the basis of parameter estimates based solely on bilateral triads. These parameter estimates were obtained, for each subject, from all the bilateral triads

presented to that subject. The model used was a modification of the simple Luce choice model, which included a bias mechanism that introduced one additional bias parameter in each wavelength region studied. The bias model is described in detail in Krantz (1964). It still involves a form of simple scalability, and, as discussed in the reference cited, the bias parameter does not essentially affect the predictions regarding unilateral triads. The parameter estimates from bilateral triads could thus be used to obtain predicted choice frequencies for the unilateral triads.

The results of the comparisons of observed with predicted choice frequencies for one class of unilateral triad are summarized in Table 3. Deviations are marked "$+$"

TABLE 3

SQUARED DEVIATIONS OF OBSERVED FROM PREDICTED CHOICE PROBABILITIES FOR
UNILATERAL TRIADS

Subject	$(+)$	$\chi^2_{(+)}$	$(-)$	$\chi^2_{(-)}$
RMS	13	35.18	4	13.89
JPE	8	52.59	3	11.24
KR	11	53.83	1	0.36
SI	6	26.26	4	5.10
HC	7	13.54	1	1.92
Total	45	181.40	13	32.51

if the observed frequency is more extreme than the one predicted one, "$-$" otherwise. For each subject, separate chi-square statistics (which, of course, do *not* have the chi-square distribution given the null hypothesis) were calculated based on $+$ and $-$ deviations. A few additional such tests, involving triads with much larger color differences, were inconclusive. These results, although obtained in the context of a specific scaling model, suggest strongly that there is a qualitative failure of simple scalability in the direction anticipated and qualitative agreement with models such as the Thurstonian one described above, which are incompatible with simple scalability. Model-free tests of simple scalability have been devised (Krantz, 1964) but unfortunately they were not available when the experiments reported here were run. Nevertheless, the present data seem to yield fairly strong evidence against the validity of Assumption (A).

The evidence against Assumption (A) can be reinforced by examination of previous color-triads data, which were analyzed using a different instance of simple scalability, Thurstone's Case V. These data were reported by Torgerson (1958). They consist of group proportions (based on one judgment per subject for each triad) for similarity judgments for triads made up of nine Munsell colors of hue 5R. The Case V param-

eter estimates were based on all the data, from unilateral, bilateral, and in-between triads. For reanalysis I selected a subset of 36 triads that, with respect to the original Munsell space, seemed to be "unilateral." The observed frequency was more extreme than the one predicted from the parameters in 32 of these 36 cases (see Krantz, 1964).

C. Tests of the "Least Confusion" Assumption

The second critical assumption made in Section IV, Assumption (C), embodies psychological content which corresponds to the geometric property of the triangle inequality for the transformation $d(x, y) = w(x, y) = \log [1/v(x, y)]$. In particular, (C) implies that $P(x, y; y) \leqslant P(x, y; z)$. This was tested in 64 cases where $(x, y; z)$ was a "unilateral" triad, with the (x, y) and (y, z) differences quite small. In 59 of 64 cases, the observed relative frequency of choosing x as nearer than y to the standard was greater when y was the standard than when z was the standard. This clearly contradicts (C).

D. Tests of the Triangle Inequality

To what extent is the failure of (C), shown above, a consequence of the failure of simple scalability? That is, when the standard is equal to one comparison stimulus, the triad may be more nearly "bilateral," and the "difficulty of choice" may elevate $P(x, y; y)$. One possible way to test this idea is to see whether the function $d(x, y) = \log [1/v(x, y)]$ does act as a distance measure, when the v-scale parameters are estimated using only bilateral triads. Hopefully, the difficulty-of-choice factor will be approximately constant for all bilateral triads, and the relative values of $v(x, y)$ thus obtained will be close to the "true" values. One could then ascribe the fact that $P(x, y; z)$ is underestimated from these "true" values, for unilateral triads, to "ease of choice," or high correlation of discriminal processes.

This possibility was tested by 101 comparisons of $\log [1/v(x, z)]$ with the sum

$$\log \frac{1}{v(x, y)} + \log \frac{1}{v(y, z)}$$

where y is between x and z. In 59 instances, the (x, y) and (y, z) differences were very small; in 41 of these 59 cases, $\log [1/v(x, z)]$ was *larger* than the sum of the two smaller "distances", while in only 18 cases was the triangle inequality satisfied. Of the 42 instances involving large (x, y) and (y, z) differences, the triangle inequality held in 39 cases; usually, it was "oversatisfied" in that $\log [1/v(x, z)]$ was much smaller than the sum of the two smaller "distances."

Thus, even when scale values are estimated from bilateral triads, hopefully eliminating most of the difficulty-of-choice factor (in fact a moderately good fit of the scaling model was obtained for bilateral triads), the triangle inequality fails in a highly system-

atic way for the function $\log [1/v(x, y)]$. This result seems to indicate that Assumption (C), plausible as it may be, is very much on the wrong track. Because of the qualitative similarity of (C) to Eq. 4 above, some doubt is cast also on the appropriateness of Shepard's theory as summarized in Section II. A much better approximation to the triangle inequality was obtained from the function

$$d(x, y) = \left[1 - \frac{1}{v(x, y)} \right]^{1/5}.$$

Psychological distances obtained from this formula agreed well with previous color-difference measurements or with predictions from color theory (Krantz, 1964; 1967).

VI. SIMPLE SCALABILITY AND THE METHOD OF TRIADS

At first blush, the program of developing rational distance functions seems attractive, since the metric space representation of differences would be justified in terms of psychological theory, as well as mathematical convenience. Furthermore, the method of triads is a "natural" one, in that it uses "direct" comparison of subjective differences and permits one to subsume discrimination experiments as part of the similarity comparisons (by comparing "small" differences with a zero-difference). However, the obstacles seem severe: there is neither an adequate theory of choice probabilities for the method of triads, not adequate data that might suggest new theories. One major class of theories, those involving simple scalability, are apparently ruled out by existing data. The first problem, if the hope of obtaining rational distance functions based on comparisons of subjective differences is not to be abandoned, is to find an adequate substitute for simple scalability. A Thurstonian theory is one possibility, with non-constant correlations assumed between the discriminal processes corresponding to the choice elements. Highly similar pairs of differences have highly correlated discriminal processes, so that a given difference in "mean similarity" corresponds to a more extreme choice probability. The trouble with this sort of theory is that the number of free parameters grows as fast as the number of choice probabilities. Some restriction governing the correlations must therefore be devised. One possibility might be to assume uncorrelated discriminal processes of known form for individual stimuli and to generate from these the forms of the discriminal processes for stimulus differences as in the example treated by Coombs (1964, Ch. 23). As an alternative to use of correlations, one might try to account for the effect of "over-all similarity" of two differences in terms of the entire configuration of differences determined by the endpoints of the pairs. New parameters are not introduced, but a question arises concerning the form of dependence of the choice probabilities (or of the correlations of discriminal processes) on the "irrelevant" similarity parameters.

These difficulties are sufficiently formidable that one might well abandon the attempt to extract more than the ordering of pairs, and turn to the "ordinal" methods of multidimensional scaling (see Section I). For the method of triads, one would use only the information as to whether a choice probability was greater or less than $\frac{1}{2}$, rather than the actual estimate of the probability, which is affected by "irrelevant" factors. However, the study of factors that affect choice probabilities is in itself a fairly interesting problem. Development of a satisfactory theoretical treatment of this problem seems to be a prerequisite for reopening the question of rational distance functions based on similarity comparisons.

REFERENCES

BEALS, R. W., KRANTZ, D. H., AND TVERSKY, A. The foundations of multidimensional scaling. *Michigan Mathematical Psychology Technical Report* MMPP 67-2, 1967.

BENNETT, J. F., AND HAYS, W. L. Multidimensional unfolding: determining the dimensionality of ranked preference data. *Psychometrika*, 1960, **25**, 27-43.

COOMBS, C. H. A theory of psychological scaling. *Engineering Research Institute Bulletin*, No. 23. Ann Arbor: Univer. of Michigan Press, 1952.

COOMBS, C. H. *A theory of data.* New York: Wiley, 1964.

INDOW, T., AND UCHIZONO, T. Multidimensional mapping of Munsell colors varying in hue and chroma. *Journal of Experimental Psychology*, 1960, **59**, 321-329.

KRANTZ, D. H. *The scaling of small and large color differences.* Unpublished doctoral dissertation, University of Pennsylvania, 1964.

KRANTZ, D. H. Small-step and large-step color differences for monochromatic stimuli of constant brightness. *Journal of the Optical Society of America* (in press, 1967).

LUCE, R. D. *Individual choice behavior.* New York: Wiley, 1959.

LUCE, R. D. A choice theory analysis of similarity judgments. *Psychometrika*, 1961, **26**, 151-163.

LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. I. New York: Wiley, 1963.

MESSICK, S. J. An empirical evaluation of multidimensional successive intervals. *Psychometrika*, 1956, **21**, 1-17.

SHEPARD, R. N. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325-345.

SHEPARD, R. N. Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psychological Review*, 1958, **65**, 242-256. (a)

SHEPARD, R. N. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, 1958, **55**, 509-523. (b)

SHEPARD, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, **27**, 125-140.
TORGERSON, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17**, 401-420.
TORGERSON, W. S. *Theory and method of scaling*. New York: Wiley, 1958.
TVERSKY, A. The dimensional representation and the metric structure of similarity data. Mimeographed. Center for Cognitive Studies, Harvard University, 1966.