

Information measurement of distinctiveness and similarity

D. C. DONDERI

McGill University, Montreal, Quebec, Canada

A method using information measurement is developed to derive euclidean interstimulus distances from stimulus-classification behavior. The method has two new properties: (1) It generates an origin for the resulting similarity space. The origin is the point of zero information or distinctiveness. (2) It generates a measure of informativeness or distinctiveness for each stimulus, which is the vector from the origin of the space to the point representing the stimulus in the space. It also generates an information measure that is equivalent to the interstimulus distance between every pair of stimuli. In addition, a new similarity judgment instruction is used—one that also generates a measure of distinctiveness for every stimulus and a measure of dissimilarity for every stimulus pair. The information measurement method and the similarity judgment instruction were tested on large independent samples of observers with a stimulus set consisting of 10 color photos. The two methods produced highly correlated measures of stimulus distinctiveness and stimulus-pair similarity, and also produced congruent stimulus configurations in euclidean similarity space.

Similarity is ubiquitous and intractable: ubiquitous because it enters into all relationships among stimuli and responses, and intractable because no theory of similarity (or its converse, dissimilarity) has yet explained its many ramifications. When describing how a portrait recalls the face and name of its subject, Alexander Bain wrote, "the consciousness of agreement or identity, the flash of a felt similarity, is real and distinct" (in Mill, 1869, p. 122). How can "the flash of a felt similarity" be described? How can the similarity of two objects be related to the distinctiveness of each? This article presents an answer based on a theory of psychological information measurement.

Uncertainty, Choice, and Information

Uncertainty, in the technical sense, measures the variability, over time or space, of a finite set of objects, which is independent of the particular objects in the set. Insofar as variability alone influences psychological judgments (including similarity), uncertainty measurement quantifies that influence. The uncertainty of a set is the number of binary choices (bits) required to select one object from a set, in which each object has an equal chance of being selected. The number of binary choices required to choose

a single object of a set equals the logarithm to the base 2 of the number of members of the set. This number is defined even when the number of members is not an integral power of 2. The larger the set, the larger the number of binary choices required to select one of its members.

Information quantifies the effect of choice on the uncertainty measured over a finite set of objects. For example, consider a set of eight identical cubes that can only be distinguished by color: they are black, red, orange, yellow, green, blue, violet, and white. You choose the black, red, orange, and white cubes. Information is a measure of what you have gained by your choice.

Information gained is uncertainty reduced. Choosing the black, red, orange, and white cubes reduces the set size from the original eight cubes, which required three binary choices (bits) to select a single cube, to a subset of four cubes, which requires only two bits to select a single cube. By reducing the uncertainty of the remaining set of cubes from three bits to two bits, choosing four cubes from a set of eight has resulted in a *gain* of one *bit* of information.

If you had chosen the green, yellow, and blue cubes instead, the uncertainty would have been reduced from three bits in the original set of eight cubes to 1.585 bits, which is the log (base 2) of three, the number of cubes in the smaller set. The information gained, which equals the uncertainty reduced, is therefore $3 - 1.585$, or 1.415 bits. For a constant original set size, the smaller the set of objects chosen, the larger the amount of information gained. And for a constant set of objects chosen, the larger the original set size, the larger the amount of information gained.

Information is the measure, on a logarithmic scale, of the effect of choice on uncertainty. In this paper, the term

The research was supported by a grant to D. C. Donderi and J. Sergeant from the *Fonds F.C.A.C. pour l'aide et le soutien à la recherche* of the Government of Quebec and a grant from Canada Manpower. The theoretical work reported here was completed while I was a visitor at the Institute for Perception Research, Eindhoven, The Netherlands, on sabbatical leave from McGill University. I thank Herman Bouma, D. G. Bouwhuis, and the entire staff for their warm hospitality. I also thank Kiran Amin, Olga Feider, J. Sergeant, A. S. Henderson, J. O. Ramsay, Yoshio Takane, A. A. J. Marley, and several anonymous reviewers. Correspondence may be addressed to me at Department of Psychology, McGill University, Montreal, PQ H3A 1B1, Canada.

information will be used with the meaning developed in the preceding paragraphs.

Stimulus Information Measurement

Donderi (1967) measured the *response information* of each stimulus in a set by calculating the response uncertainty reduced (response information gained) that depended on the dimensional attributes of the stimulus. For example, consider a set of four stimuli consisting of a large red square, a small red square, a large red circle, and a small green circle. Associate with each stimulus its own unique response (e.g., the letters A, B, C, D). The uncertainty of the response set is $\log_2(4) = 2$ bits. Suppose you choose the subset of stimuli with the attribute *green*. There is one stimulus in the set. The uncertainty of the response associated with the stimulus is $\log_2(1) = 0$ bits. The response uncertainty reduced (information gained) by choosing the attribute *green* is $2 - 0 = 2$ bits.

The response information in each stimulus was equated to the sum of the response information gained from each of the single attributes (*small*, *green*, *circle*) present in the stimulus, and also to the information gained from combinations of two or more attributes (*small and green*, *small and circle*, *green and circle*, *small and green and circle*). A multivariate analysis based on McGill's (1954) approach was used to separate the information gained through combinations of attributes from the information gained from single attributes. This measure depended on the specific S-R associations presented in the stimulus-response set and on the physical stimulus attributes of the set.

Donderi's (1967) approach defined an information measure for each stimulus separately, instead of defining a single information measure over the entire stimulus set. The purpose of defining information measures for each stimulus was to predict dependent variables for each stimulus (e.g., the number of trials required to learn each response to criterion). The partial success of this approach was limited by the same problem that has defeated other efforts to base useful information theory measures on the variation of physical attributes across stimuli. The necessary precondition that all physical stimulus attributes are psychologically equal is simply false. Not all physical attributes have the same psychological effect.

For example, consider the two kinds of stimulus dimensions defined by Garner (1974) as *separable* and *integral*. When stimuli composed of separable dimensions like color, size, or shape are scaled for similarity, and when each dimension is considered to be a dimension of the similarity space, the attributes of separable dimensions (e.g., *red*, *green*, *large*, *small*, *square*, *circle*) maintain perceptual independence throughout their combination with other attributes. The scaling produces a variety of noneuclidean results: city-block metrics, non-Minkowskian metrics, or dominance metrics in which either one of the dimensions is taken as the basis for similarity scaling and the other dimension is ignored (Shepard, 1964). On the other hand, such attributes of integral dimensions as *hue*

and *chroma* combine within the same stimulus to form an indivisible percept (e.g., *dark green*, *light blue*). The result of similarity scaling carried out on stimuli composed of integral dimensions is a multidimensional euclidean space whose coordinates can be identified with the original stimulus dimensions.

It has been exhaustively demonstrated that paired-associates learning (Donderi, 1967) and many other stimulus task outcomes (Garner, 1962; 1974) differ depending on whether the stimuli are composed of integral or separable dimensions. There also are dimensions whose properties are intermediate between integral and separable (Garner, 1978). Some people react to a particular physical variable as an integral dimension, while others react to it as a separable dimension (Chapman, 1981).

Psychological Information Measurement

Information measures based on physical stimulus measurements have been of limited value. An alternative approach is to derive information measures from the choice behavior of observers toward a set of stimuli. Although these psychological information measures take no direct account of the physical aspects of the stimuli, they have, in common with all information measures, a simple and useful relationship to distances measured in a euclidean similarity space, and they motivate an important extension to the concept of euclidean stimulus similarity. The similarity interpretation of psychological information measurement is correlated with the results of an appropriately extended method of direct similarity judgment.

Psychological information measurement of stimulus similarity has several advantages. First, it retains the advantages of the stimulus-based information measurement method developed by Donderi in 1967, because it generates information measures for each stimulus and each stimulus pair in the set. Second, it is based on stimulus classification behavior. It is easier to classify a large number of stimuli than it is to scale stimuli directly for similarity by using classical multidimensional scaling techniques. Classifying n stimuli into groups requires less work than judging $n(n-1)/2$ pairs of interstimulus distances. Third, a *distinctiveness* measure for each stimulus, which is a vector in the euclidean similarity space, is a natural by-product of the relationship between information measures and euclidean distances. The stimulus distinctiveness vector also defines an origin for the similarity space. The origin is the common point from which all of the distinctiveness vectors diverge. Fourth, as the theoretical analysis reported here was being developed, it became evident that *stimulus distinctiveness* might also have a place in conventional multidimensional scaling.

Information Measurement and Stimulus Classification

To generate psychological information measures, the observer is first shown a set of n stimuli and is allowed to inspect it freely. The instructions are for the observer to put the stimuli into groups; each group is defined as

having "some property that distinguishes it from the other groups." The only prohibited groupings are one group of all n stimuli or n groups of one stimulus each.

The results of the classification can be described by information measurement. The uncertainty of a single stimulus, I , relative to the original set of, for example, 10 stimuli is $\log_2(10) = 3.32$ bits. The uncertainty of the same stimulus relative to a subset or group of 3 stimuli is $\log_2(3) = 1.58$ bits. The information (uncertainty reduced), $H(I)$, associated with this stimulus as a result of its classification into the group of three stimuli is $3.32 - 1.58 = 1.74$ bits.

The joint information, $H(I,K)$, of every pair of stimuli I,K is also defined by stimulus classification. Joint information equals the sum of the information in one stimulus plus the information in the second stimulus that is independent of the information in the first. Independent information is defined as information gained by classifying two stimuli into separate groups, and nonindependent, or redundant, information is defined as information gained by classifying two stimuli into the same group.

If stimuli I and K are classified into separate groups, then the information in I is independent of the information in K , and $H(I,K) = H(I) + H(K)$. If stimuli I and K are classified into the same group, the information in I and K is not independent, and $H(I,K) = H(I) = H(K)$.

In both the independent (separate) and redundant (same) groups, the measure of information increases as the classification group size decreases. If stimuli I and K are classified into the the same group of three stimuli, the information in I , $H(I)$, equals the redundant information in K , $H(K) = 1.74$ bits, and the joint information of the stimulus pair $H(I,K)$ also equals 1.74 bits. If, however, stimuli I and K are classified into different groups of two stimuli each, the independent information in each stimulus is 2.32 bits [$\log_2(10) - \log_2(2)$], and the joint information of the stimulus pair $H(I,K)$ is the sum of the independent information in the two stimuli, or 4.64 bits.

Applying this principle to stimuli I and K , which are classified into separate groups of one stimulus each, from an original set of 10, the joint information of the stimulus pair, $H(I,K)$, is 6.64 bits, which equals the sum of the information in each stimulus: $\log_2(10)$ (3.32 bits) $- \log_2(1)$ (0 bits) = 3.32 bits. For two stimuli classified into the same group of eight, the information in K , $H(K)$, is redundant relative to the information in I , $H(I)$. The joint information $H(I,K)$ equals the information in either stimulus I or stimulus K , $H(I)$ or $H(K)$, which is $\log_2(10)$ (3.32 bits) $- \log_2(8)$ (3.0 bits) = 0.32 bits.

Additional information measures need to be defined before we can develop the relationship between psychological information and euclidean similarity. The first is *independent information*. The concept of independent information already has been used when deriving the joint information in the stimulus pair I,K from the information in stimuli I and K . The information in K is independent of the information in I when K and I have been classified into different groups. The independent information

in stimulus K , relative to the information in I , is symbolized as $H_i(K)$. It is defined as

$$H_i(K) = H(I,K) - H(I).$$

The final information measure relating stimuli I and K is called *transmitted information*, symbolized as $T(I:K)$. $T(I:K)$ is a measure of the information that stimuli I and K share in common. Relative to the total information in a stimulus, $H(K)$, $T(I:K)$ is the complement of the independent information defined above. In other words, $H(K) = H_i(K) + T(I:K)$. Alternatively, since $H(I,K) = H(I) + H_i(K)$, and

$$\begin{aligned} H(I) + H(K) - H(I,K) &= H(I) + H(K) - [H(I) + H_i(K)] \\ &= H(K) - H_i(K) = T(I:K), \end{aligned}$$

then

$$T(I:K) = H(I) + H(K) - H(I,K).$$

After a single classification of the stimulus set into groups, these information measures have a simple relationship. When I and K are classified into separate groups, $H_i(K) = H(K)$. This is because $H_i(K) = H(I,K) - H(I)$ by definition, $H(I,K) = H(I) + H(K)$ as defined above for separate groups, and therefore $H_i(K) = H(I) + H(K) - H(I) = H(K)$. Also, $T(I:K) = 0$, because $T(I:K) = H(I) + H(K) - H(I,K) = H(I) + H(K) - [H(I) + H(K)] = 0$.

When I and K are classified into the same group, $H_i(K) = 0$, because $H_i(K) = H(I,K) - H(I)$ by definition, $H(I,K) = H(I)$ as defined for classification into the same group, and $H_i(K) = H(I) - H(I) = 0$. Also, $T(I:K) = H(I) = H(K)$, because $T(I:K) = H(I) + H(K) - H(I,K)$, and $H(I,K) = H(I) = H(K)$, so $T(I:K) = H(I) + H(K) - H(K) = H(I)$, or $T(I:K) = H(I) + H(K) - H(I) = H(K)$.

These simple relationships occur because, following a single classification, every pair of stimuli I,K can only either be grouped together or not grouped together. This is not true when the information measures for stimulus pair I,K are averaged across many separate classifications by the same or different observers.

The following information measures have now been derived from stimulus classification behavior: $H(I)$ is the information in stimulus I ; $H_k(I)$ is the information in I that is independent of the information in stimulus K ; $H(I,K)$ is the joint information in the stimulus pair (I,K) ; and $T(I:K)$ is the information transmitted from stimulus I to stimulus K .

The following information measurement identities have been used in the previous paragraphs:

$$H_k(I) = H(I) - T(I:K), \quad (1)$$

$$H_i(K) = H(I,K) - H(I), \quad (2)$$

$$H_k(I) = H(I,K) - H(K), \quad (3)$$

and

$$T(I:K) = H(I) + H(K) - H(I,K). \quad (4)$$

The terms have been defined and the identities have been proven in many standard sources on discrete information measurement (Attneave, 1959; Garner, 1962; McGill, 1954; Shannon & Weaver, 1949).

The empirical definitions of the information quantities $H(I)$ and $H(I,K)$ for stimuli classified into the same or into different groups are consistent with the formal identities that comprise the structure of information theory.

Information Measurement and Euclidean Distance

Parke and Samson (1951) credit a lecture by Claude Shannon with the first mention of an informational distance measure, $d(IK)$, between sequences (messages, stimuli) I and K . They report that Shannon defined this measure as:

$$d(IK) = H_k(I) + H_i(K). \quad (5)$$

Parke and Samson then proved that $d(IK) = 0$ implies $I = K$, and that for all I, K , and J , $d(IK) \leq d(IJ) + d(KJ)$, thus establishing $d(IK)$ as a distance measure. As indicated by Equation 5, Shannon's measure $d(IK)$ of the distance between two stimuli is equal to the sum of the information in each stimulus that is independent of the information in the other stimulus.

In order to apply information measurement consistently and usefully to euclidean space, it was necessary to modify Shannon's informational definition of distance. The necessary modification is outlined in the following section. Consider the triangle IOK , in euclidean space, of Figure 1. Assign the points I and K to stimuli I and K , respectively. Let θ equal the angle IOK . Now make the following identities between information measures and euclidean distances $d(OI)$, $d(OK)$, and $d(IK)$:

$$\begin{aligned} H(I) &= d(OI)^2, & d(OI) &= H(I)^{1/2} \\ H(K) &= d(OK)^2, & d(OK) &= H(K)^{1/2} \\ H_k(I) + H_i(K) &= d(IK)^2, \\ d(IK) &= [H_k(I) + H_i(K)]^{1/2}. \end{aligned} \quad (6)$$

The trigonometric law of cosines for triangle IOK is:

$$d(IK)^2 = d(OI)^2 + d(OK)^2 - 2[d(OI)d(OK)]\cos \theta.$$

Substituting the information identities in this expression gives:

$$d(IK)^2 = H(I) + H(K) - 2[H(I)H(K)]^{1/2} \cos \theta. \quad (7)$$

Also, by Equation 6:

$$\begin{aligned} d(IK)^2 &= H_k(I) + H_i(K) \\ &= H(I) - T(I:K) + H(K) - T(I:K) \\ &= H(I) + H(K) - 2T(I:K). \end{aligned}$$

Therefore, by Equation 7,

$$T(I:K) = [H(I)H(K)]^{1/2} \cos \theta$$

and

$$\cos \theta = T(I:K)/[H(I)H(K)]^{1/2}. \quad (8)$$

Cosine θ is a normalized, or relative, measure of the information transmitted between stimuli I and K . It has a value of 1 for identical stimuli: $H(I) = H(K) = T(I:K)$, and a value of zero: $T(I:K) = 0$, for stimuli that do not transmit information to each other.

The correspondence between information measures and euclidean distances has now been defined for points corresponding to all stimuli I, K and a third point, O . The spatial interpretation of this identification of information with euclidean distance is clear from Figure 1. The distance $d(IK)$ is the conventional interstimulus distance between two stimulus objects identified as points in a euclidean stimulus space. The distances $d(OI)$, $d(OK)$ are the distances between a single common point, O , and the points representing the stimulus objects I, K in euclidean space. These distances correspond to the informativeness, or distinctiveness, of each stimulus. The point O , since it is zero distance from itself, has an information measure $H(O) = [H(O)]^{1/2} = d(O,O) = 0$. Since its information measure is zero, it is the point of noninformativeness, or nondistinctiveness, in the space, and can be considered a natural origin of the similarity space generated by the vectors radiating from it.

In summary, distance from the origin, which is the point of zero informativeness or distinctiveness, is associated with information. *Informative stimuli are distinctive*. Increased distance between two stimuli is associated with increased independence of information. *Informationally independent stimuli are dissimilar*. Stimulus pairs whose vector cosine is close to 1 transmit a high proportion of the geometric mean of their total information. *Informationally redundant stimuli are similar*.

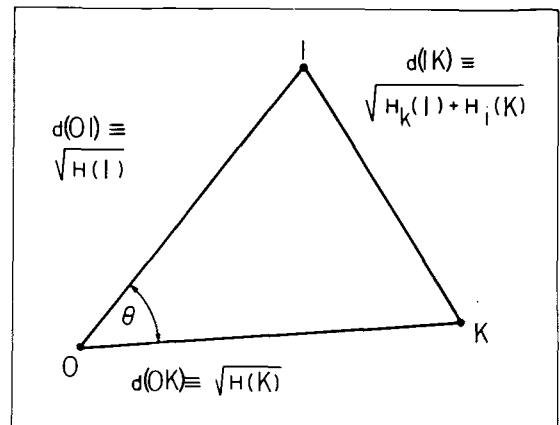


Figure 1. Triangle IOK in euclidean space, illustrating the isomorphism between euclidean distances $d(OI)$, $d(OK)$ and $d(IK)$, and the information measures $H(I)^{1/2}$, $H(K)^{1/2}$, and $[H_k(I) + H_i(K)]^{1/2}$.

The derivation of information measures from stimulus classification behavior, and the identification of information measures with euclidean similarity space, is now complete.

DEMONSTRATION

The equivalence between information measures and distance in euclidean similarity space suggested that measures in the information and similarity scaling domains should be correlated. This correlation was demonstrated by carrying out two tasks. The first was stimulus classification, and the second was similarity scaling. Both tasks employed a set of 10 photographs selected from *National Geographic Magazine*.

Stimulus Classification

The purpose was to generate empirical information measures for each stimulus and stimulus pair using the stimulus classification technique that was explained previously.

Method

Stimuli. The stimuli were 10 color photographs mounted on 14×14 cm cards. Each stimulus was labeled on the back with a code letter, which precedes its description in the following list: (A) arctic twilight across a settlement, (B) a woman and donkey in a field, (C) a chimpanzee sitting on the ground, (D) an old woman in a misty forest, (E) overturned boats used as fishing shelters on the seacoast, (F) Inca ruins in a grassy field, (G) Roman ruins in the desert, (H) the face of an aborigine, (I) an aerial cityscape with apartments and canals, (J) people walking on a winter street.

Observers. Forty high school and college students, recruited through an employment agency and paid \$10 each, participated in this experiment.

Procedure. The observers were instructed to sort the stimuli into groups so that "the pictures in each group are alike but each group is different from the other groups in the way that is most clear to you." Prohibited groupings were one group of 10 photos, or 10 groups of one photo. Observers were asked to sort the photos five times, "a different way each time, using the clearest way the first time, down through to the least important way the fifth time."

At the beginning of each of the five sorts, the stimuli were placed face up on a table in front of the observer in haphazard order. The observer selected the stimuli, either stacking them or placing them into compact separate groups. There was no time limit. After the observer completed each sort, the experimenter recorded the stimulus subsets produced by the observer and also recorded the observer's explanation of the basis for sorting on each trial.

Results

Each observer made five successive classifications, and the information measures obtained from each classification were averaged within each observer. Then the results were averaged across all observers.

The first classification per observer gave better results (higher correlations with scaled distances and paired-associates confusions) than the average of five classifications. However, since all five classifications were collected, and since the average result confirms the hypothesis, the average results are reported. Future research will

explore the optimum number of classifications per observer and possible second-order effects due to differences in experience with the stimuli.

Average information measures were generated for each single stimulus and every stimulus pair, using the techniques outlined earlier. For every stimulus pair IK , the information measures $H(I)$, $H(K)$, $H_i(K)$ and $H_k(I)$ are sufficient to generate all the other information measures of interest. The information measures calculated for each of the 10 stimuli, A ... J, are presented in Table 1.

Similarity Scaling

The purpose of this task was to generate an interstimulus distance matrix by similarity scaling. As described earlier, the development of an informational distinctiveness measure in euclidean space suggested that the standard similarity scaling technique might be modified to estimate the informativeness, or distinctiveness, distances $d(OI)$ and $d(OK)$, of Figure 1, and the interstimulus distance $d(IK)$. This was accomplished in the present task by asking the observer to judge the distinctiveness of each stimulus and stimulus pair. When a single stimulus was presented, its distinctiveness was judged with respect to all the other stimuli in the set. When a stimulus pair was presented, the distinctiveness of the two stimuli was judged by comparing each to the other.

Method

Stimuli. The stimuli were the set of color photographs used in the previous task.

Observers. The observers were a new sample of 28 high school and college students recruited through the same government employment agency used to recruit the 40 observers for the previous task. These new observers were also paid \$10 for participating in this experiment. These observers had not participated in the classification task.

Procedure. Each photo was identified on its back with one of the letters A through J. The observer sat across from the experimenter at the same desk, in the same room, and under the same lighting conditions as in the classification task.

Each observer received a response booklet consisting of 11 22×28 cm ($8\frac{1}{2} \times 11$ in) horizontally oriented pages. Each page contained five horizontally oriented 20-cm long lines. The phrase "Completely identical or nondistinctive" was written at the bot-

Table 1
 $H(I)$, $H(K)$, $H_i(K)$ and $H_k(I)$ Calculated for Each Stimulus and Stimulus Pair from the Data of the Classification Task

$H_i(K)$, where i is the row index and K , the column index*										
	A	B	C	D	E	F	G	H	I	J
A	7.16	5.28	7.58	6.82	3.92	3.70	4.42	6.92	4.87	8.05
B	5.20	7.24	5.26	4.95	5.64	5.00	4.61	6.17	7.22	7.58
C	6.61	4.37	8.11	4.20	5.32	6.09	6.26	5.32	7.11	7.11
D	6.36	4.57	4.70	7.62	4.57	5.76	6.20	5.86	7.05	6.93
E	4.05	5.84	6.41	5.16	7.05	4.58	5.59	6.92	5.14	8.18
F	3.90	5.28	7.26	6.42	4.66	6.95	3.49	7.26	5.25	8.14
G	4.41	4.67	7.21	6.65	5.45	3.27	7.17	6.94	5.27	7.84
H	6.09	5.41	5.45	5.48	5.96	6.21	6.12	8.00	7.23	4.45
I	4.25	6.68	7.46	6.89	4.40	4.23	4.66	7.44	7.78	5.64
J	6.44	6.05	6.46	5.79	6.44	6.33	6.24	3.58	4.65	8.77

*Entries in the main diagonal, where $i = k$, are $H(K)$.

tom of the page under the left end of the five lines. The phrase "Greatest imaginable difference or distinctiveness" was written under the right end of the five lines.

Each line was labeled just beyond its left end with either a single letter or a two-letter combination, corresponding to the photo or photo pair to be evaluated on that line. Instructions were for the observer to place a vertical mark across the line at the point corresponding to the judgment for the single photo (from "nondistinctive" to "greatest distinctiveness") or the photo pair (from "completely identical" to "greatest imaginable difference").

The order of evaluation of the complete set of 45 [$N(N-1)/2$] photo pairs plus 10 individual photos was separately randomized for each observer. At the end of the task, the marked distances were measured from the left edge to the nearest centimeter, and the distances for all observers were averaged for each stimulus and for each stimulus pair.

Results

The average scaled distance between each photo pair and the average scaled distinctiveness of each photo are recorded in Table 2. The symmetrical photo-pair distances appear in the upper triangle of the matrix, and the average distinctiveness for each photo is recorded on the main diagonal.

ANALYSIS

The most direct test of the claim that information measures establish a euclidean similarity space with a theoretically meaningful origin is to compare the information measures and the similarity judgment measures that were obtained for the same set of stimuli. The two sets of measures were obtained from independent groups of observers, $n = 40$ for the information measures and $n = 28$ for the similarity judgments.

Three figures illustrate the comparison. Figure 2 plots the mean judged distinctiveness of each of the 10 stimuli, as measured by the similarity scaling task on the ordinate, against the information measure $H(I)^{1/2}$, which was derived from the stimulus classification task, on the abscissa [Pearson correlation coefficient $r(9) = .91, p < .001$].

Figure 3 plots the mean judged interstimulus distance for each stimulus pair on the ordinate against its informational equivalent in euclidean space, $[H_i(K) +$

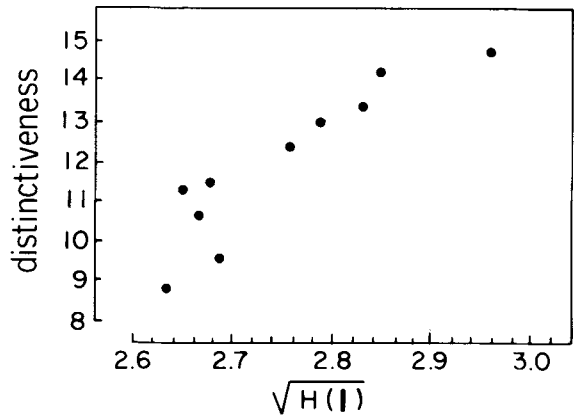


Figure 2. The average judged distinctiveness of each of the 10 stimuli (D , maximum = 20) plotted against its informational equivalent, the average square root of the information gained in each stimulus $H(I)^{1/2}$ [Pearson $r(9) = .91, p < .001$].

$H_k(I)]^{1/2}$, on the abscissa [Pearson $r(44) = .79, p < .001$].

Figure 4 displays two-dimensional euclidean spatial representations for two ensembles of 11 points each. Each ensemble includes points representing the 10 stimuli and an 11th point corresponding to the origin of the space. The Category ensemble plots the 11 points in a space defined by the information measures, while the Scaling ensemble plots the points in a space defined by the similarity judgments. Each set of data was produced by a different group of observers. Each configuration was generated independently, using the parametric multidimensional scaling program MULTISCALE (Ramsay, 1982). The origins of each space (O) were placed in coincidence on the figure. Then the longest distances in the two configurations (the distance from the origin O to the stimulus D) were made equal, and the points D were put into coincidence. The other distances were scaled proportionally without changing their relative positions. The program constraints that produced these results were identical for both sets of data. They were the default conditions for operation of the MULTISCALE program, which involve the fewest restrictions on the configurations.

The two-dimensional structure of interstimulus distances reveals clusters related by pictorial content. Closest to the origin, and least distinctive, is a group of three photos (B, F, and G) that are rhythmic studies of nature: one of alternative tufts of yellow grass and green bushes on an ancient hillside, crossed by a figure leading a donkey; another, the ancient rhythmically serrated walls of an Inca fortress in the Andes; the third, a sandstone mausoleum carved out of desert rock in Syria. Another cluster of photos (I and J) consists of modern urban scenes: people hurrying down a busy street in winter, and an aerial cityscape of apartments, trees, and water. The two most distinctive photos (A and D) are scenes of human distance and isolation in nature: one of a woman in a forest, and the other of an Arctic coast settlement under a low sun.

Table 2
Scaled Average Stimulus Distinctiveness and Interstimulus Distance for Scaling Task

	Stimulus									
	A	B	C	D	E	F	G	H	I	J
A	10.6	10.2	17.2	14.1	12.2	8.1	8.9	13.0	13.9	15.7
B		9.6	12.3	10.9	10.7	7.7	13.4	11.1	15.9	15.3
C			14.3	8.3	13.6	15.2	14.8	11.3	18.6	13.0
D				12.4	9.2	13.0	14.0	13.8	17.1	15.4
E					11.3	10.5	13.8	11.3	14.1	17.0
F						8.9	8.4	11.8	11.2	16.3
G							11.4	14.4	10.5	16.1
H								13.4	16.5	10.5
I									13.1	11.9
J										14.7

Note—Entries on the main diagonal are measures of stimulus distinctiveness.

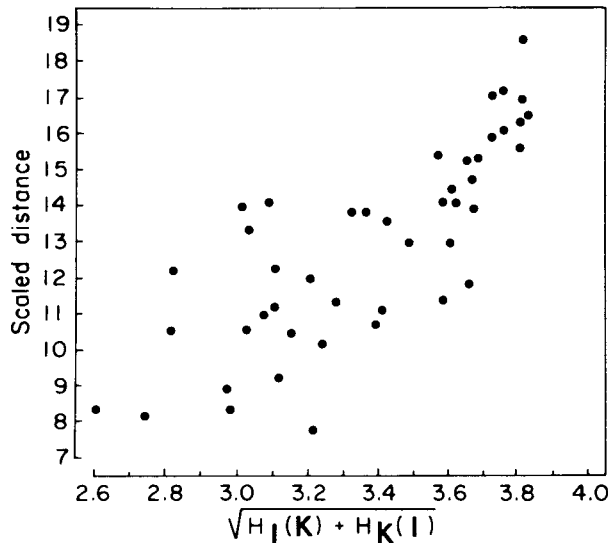


Figure 3. The average judged interstimulus distance for each stimulus pair (Scaled distance, maximum = 20) on the ordinate, plotted against its informational equivalent, the average square root of the sum of the conditional information in each stimulus $[H_i(K) + H_i(I)]^{1/2}$, on the abscissa [Pearson $r(44) = .79, p < .001$].

The cluster of photos C and H consists of the only two animate closeups in the set: one a chimpanzee, the other an aborigine. The ambiguously placed photo E, of overturned boats used as shelters, was placed close to the rhythmic nature group (B, F, and G) by the similarity judgments, and closer to the urban scenes of people and buildings by the information measures.

The high correlations between the distance and the information measures (Figure 3), and the similarity of their spatial configurations as represented in two-dimensional euclidean space (Figure 4), argue strongly for the equivalence

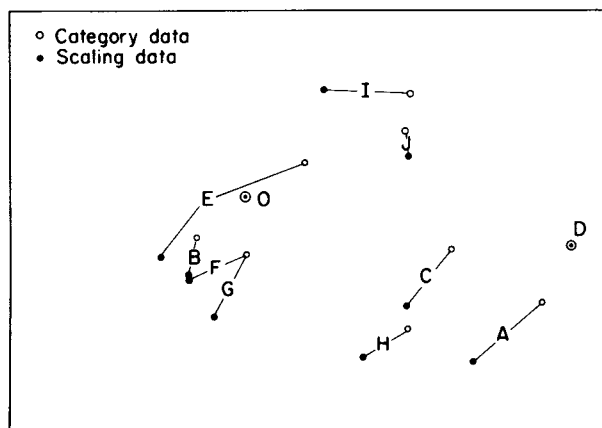


Figure 4. Two-dimensional euclidean spatial configurations for the ensemble of 11 points defined by the 10 stimuli and the origin of the space. One configuration (Category data) is based on the stimulus classification task. The other configuration (Scaling data) is based on the similarity scaling task. Each configuration was generated independently using the parametric multidimensional scaling program MULTISCALE (Ramsay, 1982).

of the psychological quantities being measured in the two experiments. Therefore, the information measures derived from stimulus classification are empirically, and theoretically, equivalent to the distance measures generated by direct estimation from individual stimuli and stimulus pairs.

DISCUSSION

Other Classification Similarity Measures

Other approaches to the analysis of stimulus-sorting data begin by defining a co-occurrence matrix, C_{ij} , in which the datum at the intersection of row i and column j is the number of times, summed over observers, that the stimulus pair i, j was classified into the same group. Miller (1969) showed that the transformation

$$d(xy) = N - n(ij)$$

is a distance measure. $n(ij)$ is the entry in cell i, j of the co-occurrence matrix, and N equals the number of rows (or columns) in the matrix. Miller used the distribution of co-occurrences to study the lexical characteristics of a set of 50 common nouns. Miller's analysis used only the necessarily symmetrical off-diagonal entries.

Burton (1972) calculated a derived measure of similarity that accounted for the informativeness of co-occurrences. Two stimuli were considered more similar if they were grouped together into a smaller co-occurrence group. The weighting factor for co-occurrence in group i was $-\log_2 p(i, a)$, where $p(i, a)$ is the chance probability that any pair of stimuli could be found in group i , of size a . This measure, like Miller's simpler one, applies only to the co-occurrence of pairs of stimuli in a group, and does not measure the distinctiveness of single stimuli.

Information measurement has also been used previously to generate interstimulus distance measures. Nishisato (1975, pp. 23-26) discussed the use of the Rajski distance

$$d(xy) = 1 - T(X:Y)/H(X, Y)$$

as an index of similarity for two nominal variables with discrete distributions. Nishisato calculated this distance from probabilities of joint co-occurrence collected over the nominal response categories of such variables as the occurrence of above- or below-median scores on any of six Rorschach response variables in a population of different diagnostic types. The Rajski distance is almost the complement of

$$\cos \theta = T(X:Y)/[H(X) H(Y)]^{1/2}$$

developed in this paper. Like the complement of $\cos \theta$, it ranges from 0 to 1, reaches a minimum when X is equivalent to Y , and reaches a maximum when X and Y are informationally independent. Its value as a distance measure is limited because it is normalized. Our interstimulus distance measure

$$[H_i(Y) + H_i(X)]^{1/2}$$

is not normalized, and so has a greater potential range.

Takane (1984) developed a method for generating stimulus clusters in multidimensional space directly from stimulus sorting data. His technique does not measure the distinctiveness of individual stimuli, nor does it provide an origin for the similarity space. However, Takane's very flexible notation is adapted, in the Appendix, to describe the information analysis developed in this paper.

CONCLUSION

The distinctiveness and similarity measures derived from information measurement of stimulus-classification behavior are almost the same as those derived from similarity judgments using the new similarity judgment instruction. Strong evidence for their equivalence is the close agreement between the independently generated two-dimensional euclidean spatial configurations shown in Figure 4. The position of only one stimulus is substantially different in the two configurations. One of the configurations is based on all 55 information measures, and the other configuration is based on all 55 similarity measures. Each configuration was generated by an independent group of observers.

The euclidean spatial configuration generated from the new stimulus-classification measures of informational distance, and the configuration generated by the new scaling judgment instruction, include a point that is the origin of the euclidean similarity space. It is the point of zero stimulus information, or zero stimulus distinctiveness. The advantage of identifying such a point is illustrated in Figure 2, where the judged distinctiveness of each shape is shown to be correlated with the information measure of the distance between the origin and the point representing the stimulus. It is also illustrated in Figure 3, where the spatial configuration of stimuli around the point defined as the origin is seen to be consistent between both measurement techniques.

The origin of a psychologically generated euclidean similarity space has been given two parallel definitions. The information measurement definition of the origin is: the locus of all of the stimulus vectors OI, OK , measured by $H(I)^{1/2}, H(K)^{1/2}, \dots$, as derived from the information analysis of stimulus classification behavior. The similarity judgment definition is the result of the new scaling instruction, "scale for distinctiveness," which measures the distance between each stimulus and the origin of the space in a similarity scaling task. Both of these definitions are new. The "scale for distinctiveness" instruction grew from the realization, while the information measure theory was being developed, that there should be a behavioral equivalent to distinctiveness, or informativeness, in conventional multidimensional scaling judgments and in stimulus classification behavior.

Information measurement derived from stimulus classification and the distinctiveness similarity scaling instruction are two new tools for measuring stimulus similarity and stimulus distinctiveness in a common euclidean space.

The tools differ in their use. Extensive applied research has shown that repeated similarity scaling is very tiring, and that the stimulus set size cannot grow much beyond 8 or 10 before the task becomes very difficult to complete (Lyne et al., 1983; Lyne, Whiteman, & Donderi, 1984).

On the other hand, classifying stimuli is easier, because the entire stimulus set need be handled once only, and more freely, by the observer. We intend to use the classification task with sets as large as 20 or 30 stimuli, and an effort will be made to test the practicality and reliability of the technique with stimulus sets as large as 100.

Validating the usefulness of these new tools has begun, but much remains to be done. Analysis of the stimulus-classification data suggests that it may be possible to simplify the task without reducing the precision of the resulting information measures. One classification sort per observer, averaged over 40 observers, may be just as valid and reliable as five sorts per observer averaged over 40 observers.

Both the information classification and the similarity scaling techniques can be used with other stimuli, such as words, objects, or sounds. The usefulness of the new similarity techniques will be tested by verifying their consistent extension to other sets of stimuli.

Information theory measures of interstimulus distance can be generated from classifications made by a single observer and from classifications aggregated over groups. It remains to be seen whether these measures will predict the data generated by single observers as well as they predicted the group data presented in this article.

REFERENCES

- ATTNEAVE, F. (1959). *Applications of information theory to psychology*. New York: Holt, Rinehart & Winston.
- BURTON, M. (1972). Semantic dimensions of occupation names. In A. K. Romney, R. N. Shepard, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 2, pp. 55-71). New York: Seminar Press.
- CHAPMAN, M. (1981). Dimensional separability or flexibility of attention? Age trends in perceiving configural stimuli. *Journal of Experimental Child Psychology*, 31, 332-349.
- DONDERI, D. C. (1967). Information measurement of single multidimensional stimuli. *Canadian Journal of Psychology*, 21, 93-110.
- GARNER, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GARNER, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology, General*, 107, 287-308.
- LYNE, M. B., PARUSH, A., RICHER, F., JORDAN, B. D., DONDERI, D. C., & RAMSAY, J. O. (1983). A multidimensional analysis of paper-related factors in print quality. In *The role of fundamental research in paper-making: Transactions of the Seventh British Paper and Board Industry Federation, Cambridge, 1981* (pp. 655-683). London: Mechanical Engineering Publications.
- LYNE, M. B., WHITEMAN, A., & DONDERI, D. C. (1984). Multidimensional scaling of tissue quality. *Pulp & Paper Canada*, 85, 43-50.
- MCGILL, W. J. (1954). Multivariate transmission of information. *Psychometrika*, 19, 97-116.

- MILL, J. S. (1869). *Analysis of the phenomena of the human mind* (Vol. 1). London: Longmans.
- MILLER, G. A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 6, 169-191.
- NISHISATO, S. (1975). *Elements of applied scaling, Part I*. Toronto: Ontario Institute for Studies in Education.
- PARKE, N. G., & SAMSON, E. W. (1951). *Distance and equivalence in sequence space* (Laboratory Note E4080). Cambridge, MA: Air Force Cambridge Research Laboratories.
- RAMSAY, J. O. (1982). *Multiscale II Manual*. Unpublished manuscript, McGill University Department of Psychology, Montreal.
- SHANNON, C. E., & WEAVER, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, 54-87.
- TAKANE, Y. (1984). Multidimensional scaling of sorting data. In Y. P. Chaubey & T. D. Dwivedi (Eds.), *Topics in applied statistics*. Montreal: Concordia University Press.

APPENDIX

Takane's Notation for Classification Distances

Takane (1984) developed a method to calculate multidimensional distances between the centroids of stimulus clusters, and among the stimuli within each cluster, by using the off-diagonal elements of a classification co-occurrence matrix (the number of times stimuli i and j are placed in the same sorting classification). While Takane's method does not generate a measure of stimulus distinctiveness, it does use a notation that is excellent for describing many scaling methods based on co-occurrence matrices.

Let $a_m(k)$ = one group, m , of stimuli that are classified together by observer k . Let i and j index the stimuli that are being classified. $i \in a_m(k)$ means that stimulus i is classified into the group $a_m(k)$. $j \notin a_m(k)$ means that stimulus j is not classified into the group $a_m(k)$.

The term $g_m^i(k)$, which is associated with stimulus i , group m and observer k , describes the membership status of stimulus i in each of the m groups for each of the k observers. Let $g_m^i(k) = 1$ if $i \in a_m(k)$, and $g_m^i(k) = 0$ if $i \notin a_m(k)$.

Let $na_m(k)$ = the total number of stimuli sorted into group a_m by observer k . Let N = the total number of stimuli. Let $pa_m(k)$

= the probability that a stimulus was sorted into group a_m by observer k . By definition, $pa_m(k) = na_m(k)/N$.

Let $wa_m(k)$ be any theoretical parameter that is unique to group $a_m(k)$. Using this notation, both Takane's distance measure and the measure developed in this article can be expressed in common form. For Takane's measure, $wa_m(k) = 1/pa_m(k)$. For the measure developed in this paper, $wa_m(k) = -\log_2 pa_m(k)$. Making the appropriate substitution for the theoretical parameter in the following formula, both the Takane interstimulus distance and the interstimulus distance measure developed in this paper can be expressed as

$$d(ij) = \frac{1}{k} \sum_k \left\{ \sum_{a_m(k)}^M wa_m(k) [g_m^i(k) - g_m^j(k)]^2 \right\}^{1/2}.$$

Takane's notation can be extended to include the informational measure of distinctiveness developed in this paper. Let

$$d(Oi) = \frac{1}{k} \sum_k \left\{ \sum_{a_m(k)}^M wa_m(k) [g_m^i(k) - (0)]^2 \right\}^{1/2}.$$

The distances $d(Oi)$ originate at a point O that is the common locus of the distance vectors $d(Oi)$ over i . The point O has zero distance from itself [$d(OO) = 0$], and is defined as the origin of the space. The length of the vectors $d(Oi)$ from this point measure the distinctiveness of each stimulus. The distinctiveness is measured in the same units as the interstimulus distances $d(ij)$.

The complete set of distances $d(ij)$ and $d(Oi)$ over stimuli i and j , and origin O , define the multidimensional similarity space generated by the information analysis of classification data described in this paper.

(Manuscript received November 21, 1987;
revision accepted for publication June 13, 1988.)