

Pythagorean distance and the judged similarity of schematic stimuli^{1,3}

WILLIAM C. RANKIN,³ ROBERT P. MARKLEY, AND SELBY H. EVANS
TEXAS CHRISTIAN UNIVERSITY

Independent groups of Ss rated the similarity of pairs of patterns belonging to the same class, either before or after a discrimination task of schematic concept formation (SCF). Mean similarity increased as a function of SCF pretraining. A linear relationship was found between a Pythagorean distance measure on the patterns and subjective similarity of pairs of stimuli ($r = .88$). No such relationship was obtained from an analysis of judgments from a group that viewed random, nonschematic stimuli. A secondary analysis of the discrimination judgments in the SCF task showed that the Pythagorean distance measure was predictive of judgments on pairs of stimuli from the same schema family, but the relationship did not hold for pairs representing two different schemata. Multidimensional scaling analyses indicated that mildly deviant schematic stimuli were perceived to be instances of a single family. The dimensions describing a schema cluster appeared to be specific to the sample of stimuli. A tendency for the Kruskal procedures to collapse certain types of stimulus clusters was observed and discussed.

Attneave and Arnoult (1956) were the first to state that populations of stimuli produced for the quantitative study of form should be defined in terms of the statistical parameters describing the populations. Contemporary research in form perception has been characterized by development of several stimulus generation methods meeting criteria established by Attneave and Arnoult, including those of Evans and Breckenridge (1968), Evans and Mueller (1966), Fitts and Leonard (1957), and Posner, Goldsmith, and Welton (1967).

Attneave and Arnoult also proposed that some statistical properties of stimuli should covary with the perceptual response, i.e., discriminability, similarity, recognition, etc. Consequently, there has been considerable research on measures of form that are associated with the perceptual response. The problems encountered in the quantitative study of form were reviewed by Michels and Zusne (1965). In many cases, the problem of stimulus definition has all but defied solution. Michels and Zusne stated that light energy may be

considered a first-order stimulus with only two basic dimensions, whereas form is a second-order stimulus, and the number of its dimensions is unknown and undoubtedly quite large. One successful attempt at quantifying the domain of random polygons has been reported by Brown and Owen (1967).

Still another approach comes from schema theory. A schema implies the existence of a set of stimulus attributes so that, in the multidimensional space representing these attributes, stimuli belonging to a single schema family will form a single cluster. Appropriately, stimulus definition in terms of belongingness to a particular schema population is both measurable and controllable. The work of Evans (Edmonds & Evans, 1966; Evans, 1964; Evans, 1967a; Evans & Arnoult, 1967) and his associates in schema theory has relied quite heavily on the histoform patterns produced by the VARGUS 7 (Evans, 1964, 1967b) and the VARGUS 9 (Evans & Mueller, 1966) systems. VARGUS stimuli offer the E the unique advantage of a priori knowledge and control of sources of variance and covariance in the stimuli. The VARGUS methodology permits the investigator to generate and to use samples of stimuli whose population parameters are specified beforehand.

VARGUS 9 stimuli are produced to satisfy the needs of schema research. They consist of populations of stimuli that have measurable deviations from a prototype. The stimuli can be located in a multidimensional physical space that could be considered conceptually similar to the "schema-with-correction" hypothesis of Woodworth (1938). Stimuli produced in this fashion have measures associated with them that may, according to schema theory, be relevant to judgments of similarity. Deviations are ratio-scale deviations from the prototype; thus, a pattern variance (PV) may be computed for each stimulus to indicate its degree of deviation from the schema.

Schematic concept formation (SCF) has been defined by Evans (1967b) as the development of the ability to assign objects to their corresponding schema families on the basis of the information derived from perceiving the objects. A schema family is a

population of objects that can be efficiently described in terms of the same prototype, e.g., dogs. A schema family may also be thought of as a statistical concept. The assignment of objects to their appropriate schema families is hypothesized to be possible without any aid from other sources of information (such as conventional knowledge of results) about the proper categorization. If only one schema is represented in a set of objects, Evans has theorized, learning of the schema occurs spontaneously when Ss have the opportunity to inspect several examples.

The purpose of this study was to examine the judged similarity of pairs of stimuli within one schema family, before and after SCF training, and to ascertain the relationship, if any, between VARGUS 9 stimulus properties and judged similarity. The physical stimulus measure chosen was a Pythagorean distance computation suggested by Evans and Mueller (1966): "Another measure, analogous to PV, could be obtained for any pair of patterns by summing the squared column by column differences. Such a measure would be . . . a possible measure of similarity between pairs of patterns [p. 512]." Finally, the predictive utility of this type of measure was to be assessed in the context of the categorical judgments in the SCF task itself.

METHOD

Subjects

Thirty undergraduates enrolled in the introductory psychology course sections at Texas Christian University served as Ss in this investigation. They were assigned randomly to one of three treatment conditions as they came to the experiment.

Stimuli

Patterns were produced by a computer system, VARGUS 9 (Evans & Mueller, 1966). The VARGUS 9 system produces patterns of numbers randomly sampled from a defined population having specifiable information and redundancy characteristics. Redundancy and PV share an inverse relationship in that the higher the redundancy level in a population, the lower the average magnitude of PV. The sequences of numbers are mapped into

proportional column heights, resulting in patterns aptly described as histoforms. (Their appearance is similar to the histoforms of Fitts & Leonard, 1957.) It must be noted that the mapping of number sequences into proportional column heights or row widths is merely a convenience of computer output. The investigator is free to transform the number sequences by any number of mapping rules, including verbal material (Hollier & Evans, 1967).

In the present study, the VARGUS 9 histoform patterns were mapped into the shape of closed-frequency polygons. Examples of some of the polygons used in the study have been shown by Rankin and Evans (1968).

Three sets of patterns were required, designated as Schema I, Schema II, and Random. The random set consisted of 10 stimuli in which each element was determined by an independent random selection from the numbers 1 through 9 (the same range of column heights found in the VARGUS 9 system). Schemata I and II were to be discriminated in a SCF task. Both schemata were 70% redundant and consisted of a sample of 10 instances each. The samples comprising each schema were constrained to approximate the population distribution of PV for 70% redundant schemata. All stimuli were presented in booklets, with one pair per page.

Pythagorean distance between pairs of stimuli was computed by the following formula:

$$\left[\sum_{i=1}^n (X_i - Y_i)^2 \right]^{1/2},$$

where X_i is the i th column height or the

value of the i th attribute for a particular stimulus; Y_i is the corresponding column height for the comparison stimulus. For example, if X and Y had values (column heights) of 3, 7, 2, 9 and 4, 7, 4, 8, respectively, then the distance between X and Y would be:

$$[(3-4)^2 + (7-7)^2 + (2-4)^2 + (9-8)^2]^{1/2}.$$

Tasks and Procedure

Independent groups of 10 Ss each made similarity judgments on unmarked scales identified on the ends as "highly similar" or "highly dissimilar." Group PRE-SCF judged the similarity of four orders of all 45 possible pairs of Schema I stimuli before transferring to the SCF task. Group POST-SCF judged the same set of Schema I stimuli, but after they had the SCF task. The group that scaled random patterns for similarity had four orders of all 45 possible pairs of stimuli. Similarity judgments for all three groups were paced at 10 sec per pair of stimuli.

In the SCF task, Ss were told that the task would test their ability to learn to recognize different types of patterns, and that they should respond in terms of whether the pair of patterns on each page were examples of the same type of pattern or of different types of patterns. The SCF task required a discrimination among equal numbers of same-member or different-member pairings of patterns representing Schema I and Schema II. The 10 Schema I patterns used in the SCF task were a completely different sample from that used by the PRE-SCF and POST-SCF groups for scaling similarity. Out of the

190 possible pairs of stimuli from the two schemata samples, 96 were randomly sampled. This task was also paced at 10 sec per pair, and, since SCF does not require knowledge of results, none was provided throughout the entire 96 trials of the task. Performance on this particular SCF task has been reported elsewhere (Rankin & Evans, 1968).

RESULTS

Arithmetic mean similarity ratings were computed over Ss for each pair of patterns judged by each group of Ss. The relationship between mean judged similarity and Pythagorean distance is shown in the three panels of Fig. 1. A strong linear relationship was found for the PRE-SCF and POST-SCF groups ($r = +.88$). No such regularity was observed for the stimuli judged by the random group. Circled points are those pairs of stimuli in which the prototype appears. The groups of points enclosed by a balloon on Fig. 1 represent stimulus pairs containing the most deviant instance of Schema I.

The influence of SCF is evident in that the mean similarity judgment was higher after SCF (see the POST-SCF panel in Fig. 1). The overall similarity means for PRE- and POST-SCF groups were 4.5 and 5.3, respectively, which was significant ($t = 10.1$, $df = 44$, $p < .001$). The increased similarity is shown dramatically in Fig. 2, where only those stimulus pairs including the Schema I prototype are plotted. Figure 2 shows judged similarity as a function of the rank order of deviation (PV) from the prototype for stimulus pairs containing the prototype. With one exception, the POST-SCF group judged the pairs of stimuli to be more similar than did

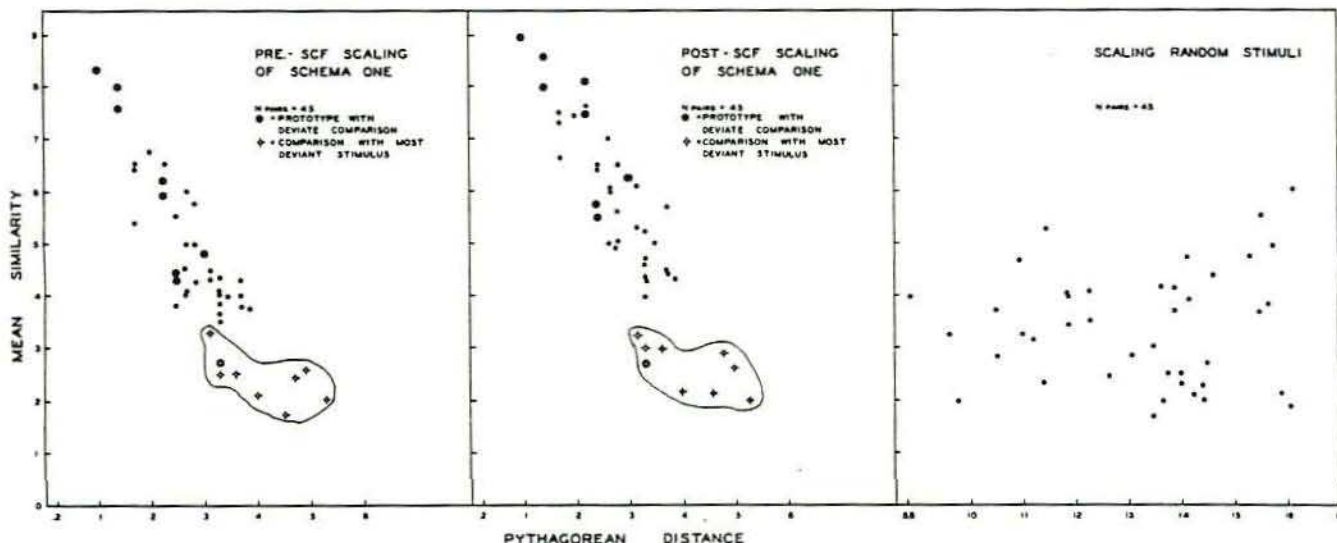


Fig. 1. Mean similarity as a function of Pythagorean distance between pairs of stimuli for three groups of Ss.

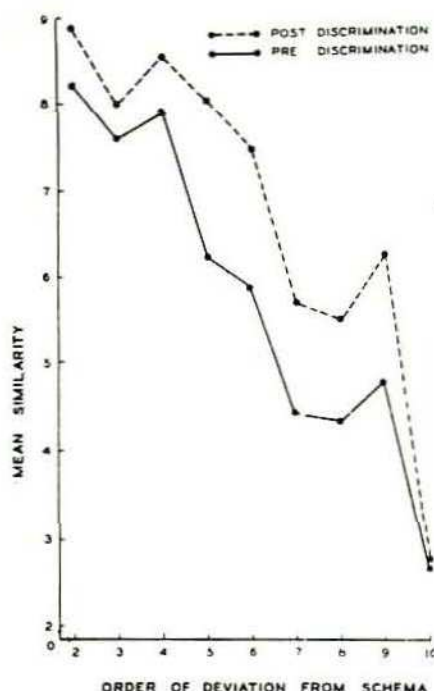


Fig. 2. Plots of mean similarity between stimuli and the schema prototype as a function of degree of deviation from the prototype.

the PRE-SCF group ($p < .01$, sign test).

In the SCF task, Ss made a binary "same" or "different" response to each pair of patterns. The proportion of Ss who responded "different" to a given stimulus pair should roughly reflect the apparent similarity of the pair. Figure 3 shows the proportion of Ss in each group who responded "different" to pairs of stimuli, plotted as a function of the Pythagorean

distance between the stimuli. In Fig. 3, judgments of pairs containing patterns from the same schema family (within schema) are shown separately from judgments of pairs with patterns from two schemata (between schemata). The relationship between Pythagorean distance and proportion of different responses was essentially linear for within-schema judgments ($r = .51$ to $.66$). Proportions for between-schema pairs were uniformly higher than within-schema proportions but not regularly related to distance ($r = .02$ to $.10$).

A standard psychophysical approach would include the assumption that the subjective distance between pairs of stimuli remains relatively constant over time. Schema theory predicts that learning, manifested as a decrease in distance between members of within-schema pairs, will occur. In the data reported above, the occurrence of SCF in the same-different task is shown by the increased similarity of within-schema pairs evidenced by the POST-SCF group. SCF did not alter the form of the relationship between Pythagorean distance and judged similarity for within-schema judgments.

MULTIDIMENSIONAL SCALING ANALYSES

While POST-SCF similarity judgments were generally higher than PRE-SCF judgments, the rank order of the similarity of pairs was essentially the same ($\rho = +.93$). Nonmetric multidimensional scaling analyses (Kruskal, 1964), using the rank order of the group mean similarity judgments, were carried out to obtain psychological spatial configurations for the stimuli. The Euclidean distance metric was used for all

configurations to be reported here. Several starting configurations were used, and solutions were found in 6, 5, 4, 3, 2, and 1 dimensions. The computer program iterated each configuration until: (1) a goodness-of-fit measure, stress, fell below .005; (2) 100 iterations were carried out; or (3) a minimum stress was obtained (i.e., the program could find no way to change the configuration to reduce stress). Tied ranked similarity judgments were treated in the primary manner, as described by Kruskal (1964).

The stress statistic is a measure of the absence of a monotonic relationship between the rank orders of the distances in the derived configuration and in the similarity measures. The stress criterion of .005 was deliberately chosen to be smaller than previously published (and normally programmed) satisfactory stress values (Kruskal, 1964) in order to allow the program full opportunity to find a minimum stress solution (cf. McGee's 1966 comments with regard to satisfactory stress within a given stimulus domain).

PRE-SCF and POST-SCF Groups

All solutions, regardless of dimensionality, for the 10 Schema I stimuli collapsed the stimuli into two clusters, with nine stimuli in one group and Stimulus 10 (the pattern most deviant in terms of PV) standing alone. Distances within the nine-stimuli clusters were essentially of zero magnitude. The typical J curve, relating stress to dimensionality, was not found. No increase in stress was observed as dimensionality decreased. The one-dimensional configurations with low stress were apparently sufficient descriptions of these data. Additional dimensions were redundant.

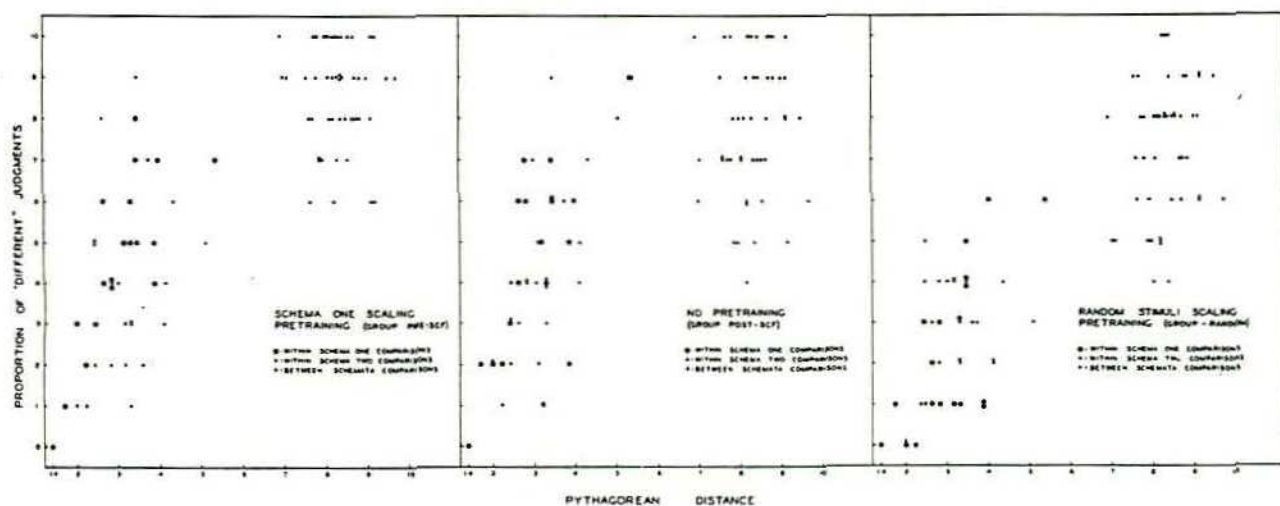


Fig. 3. Frequency data from discrimination judgments in the SCF task as a function of Pythagorean distance between stimulus pairs for three groups of Ss.

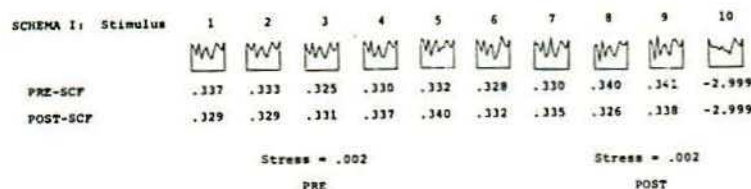


Figure 4 shows the stimuli and typical one-dimensional configurations for the PRE-SCF and POST-SCF groups. It appears that, perceptually, the 10 stimuli consisted of nine highly similar patterns clustered together, while one pattern (Pattern 10 in Fig. 4) differed greatly from the cluster.

It seems reasonable to believe Ss were making more than categorical schema vs not-schema responses in the similarity task. Thus, the stimulus pairs containing 10 were therefore excluded and the nine clustered patterns of Schema I reanalyzed with the Kruskal procedure. Varimax rotations were applied to all solutions. Stress as a function of dimensionality is shown in Fig. 5.

Choice of proper dimensionality for stimuli analyzed by the Kruskal procedure is, in the absence of a prior model, at best a subjective attempt to balance accuracy and parsimony. For these data, the one- and two-dimensional solutions have stress values that are too large. The five- and six-dimensional configurations have too many dimensions. The authors favor a four-dimensional configuration for the PRE-SCF data and the three-dimensional configuration for the POST-SCF group. These configurations are given in Table 1.

Identifiable characteristic contours were found for most of the patterns appearing at the extremes of the dimensions shown in Table 1. For example, a dimension found in both groups (PRE-SCF II and POST-SCF I in Table 1) contrasted Patterns 8 and 9 with Patterns 4 and 7. Patterns 8 and 9 (shown in Fig. 4) were the only patterns with valleys that ascended from left to right. Pattern 7 has the only central peak that is symmetric and rises above all other peaks. Both Pattern 4 and

Pattern 7 have central peaks with concave left walls. All other patterns are convex at that point.

It is not clear that this effort to identify cues used by Ss will prove to be of general use. There is no way to insure that the characteristics identified here will persist through other samples of patterns from this or any other schema. These speculations, however, may be partially indicative of the kind of strategies that Ss develop for the ordering of pattern stimuli. Ss were clearly able to find characteristic regularities (subschemata) within a pattern and to use these to perform consistently in the tasks imposed by E.

Of interest also are the differences in dimensions used by the two groups. The PRE-SCF group's dimensions, I and IV, defined by Patterns 9, 5 and 8, 6, respectively, appeared to have coalesced into a single dimension (II) for the POST-SCF group. A PRE-SCF dimension (III) correlated with PV ($\rho = +.93$). This dimension appeared as a bipolar dimension, defined by Patterns 9, 8 and 2, 3, for the POST-SCF group.

SCF Data

The frequency data from the SCF task reported above were also analyzed by the Kruskal procedure. The results for both PRE- and POST-SCF groups supported the hypothesis that the Ss were able to form perceptual groupings on the basis of schematic characteristics of the stimuli. The best two-dimensional solution is shown in Fig. 6. There are two distinct clusters of points, each cluster corresponding to one of the schemata.

It has been pointed out, in connection

Fig. 4. Typical one-dimensional solutions for PRE- and POST-SCF groups.

with the procedures developed by Shepard (1962), that a configuration containing two clusters with large between-cluster and small within-cluster distances will collapse into one dimension with zero between-cluster distance and a large within-cluster distance (Klemmer & Shrimpton, 1963). The results shown in Fig. 4 indicate that the same phenomenon occurred in the first Kruskal analyses. The Kruskal procedure is derived from Shepard's and, evidently, has the same characteristic. This result should not necessarily be viewed as a fault in the method. If a similarity judgment is a measure of perceptual equivalence or substitutability within the context of the current judgmental situation, then the message provided by the first Kruskal analyses, i.e., that Stimulus 10 is psychologically quite different within the realm of Schema I, is both valid and meaningful.

DISCUSSION

The striking relationship between the objective measure of distance between stimuli from the same schema family and psychological distance has several implications. First, human Ss seemed to be able to estimate well distances between the stimuli in m -dimensional space (for this study, $m = 14$). Second, the data suggest a successful attempt at quantifying the domain of stimuli produced by the VARGUS 9 methodology. To the extent

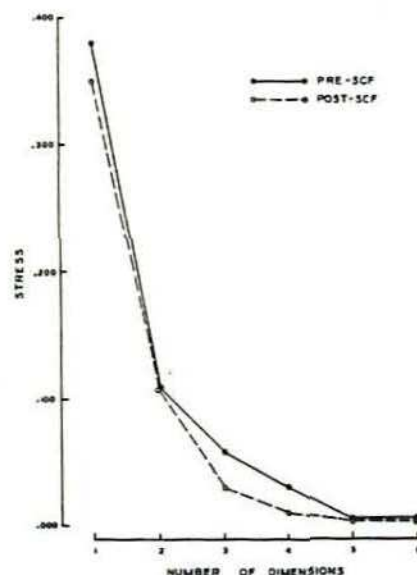


Fig. 5. Stress as a function of dimensionality.

Table 1
Rotated Kruskal Configurations for Pre- and Post-SCF Groups Nine Stimuli Analyses

Stimuli	Pre-SCF Dimension				Post-SCF Dimension		
	I	II	III	IV	I	II	III
1	0.048	-0.103	-0.497	-0.155	0.019	-0.224	-0.201
2	-0.243	0.148	-0.734	0.050	0.023	-0.439	-0.825
3	-0.588	-0.027	-0.256	-0.067	-0.291	0.158	-0.707
4	-0.177	-0.562	-0.014	-0.141	0.439	0.000	0.121
5	0.173	0.174	0.406	1.164	0.102	1.323	-0.117
6	1.158	0.093	-0.111	-0.018	0.119	-0.975	0.046
7	0.182	-1.113	0.035	0.195	1.079	0.092	-0.072
8	-0.710	0.657	0.721	-0.101	-1.012	0.431	0.589
9	0.157	0.733	0.451	-0.915	-0.481	-0.367	1.167
	Stress = .029				Stress = .030		

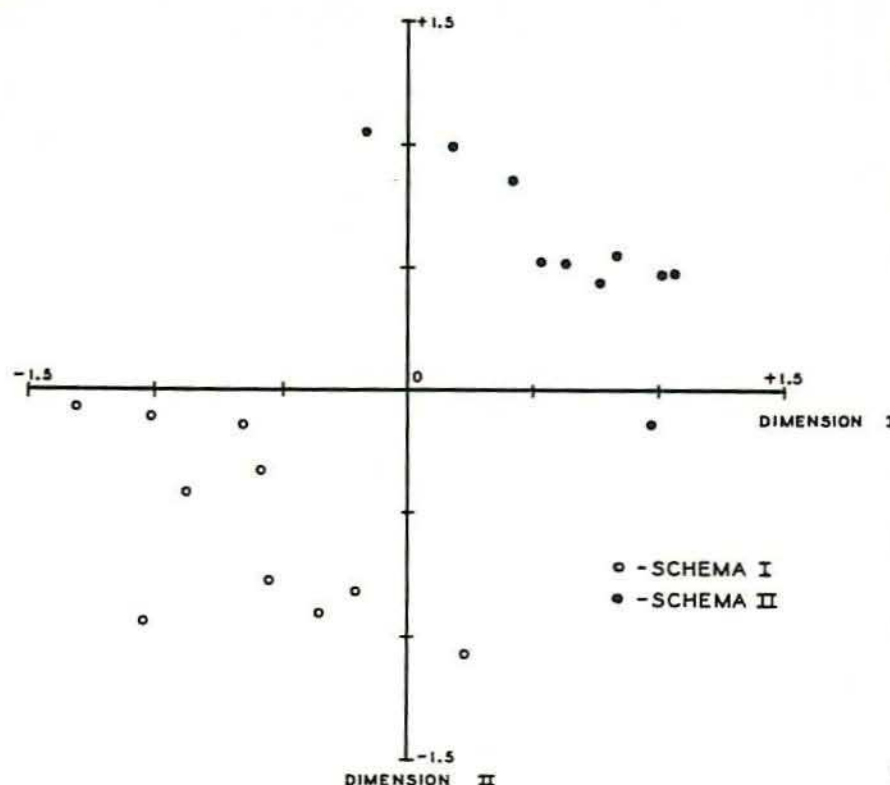


Fig. 6. Typical two-dimensional solution for stimuli in the SCF task for PRE-SCF group only, stress = .072.

that patterns or forms in the real environment may be represented by the characteristics of stimuli produced via this generation procedure; a valuable contribution to the quantitative study of form perception has been made.

Failure to establish the formal relationship between objective distance and similarity judgments on between-schemata pairs of stimuli poses a problem for describing variables that may underlie the process of SCF. The Pythagorean distance measure was suggested on purely rational grounds by Evans and Mueller (1966), based upon their method of statistical control of the variability associated with the attributes of the stimuli. Why this measure does not predict between-schemata judgments is not known; further research will be necessary to determine a function for between-schemata similarity judgments. Such possibilities as a higher-order metric for the Pythagorean distance measure or even a discriminant function could be attempted. It would be desirable, however, to arrive at the particular function on the basis of some rational model.

The fact that the Kruskal procedure shows stimuli clustering, rather than being spaced along dimensions, seemed somewhat puzzling. However, with all of the stimuli coming from a single schema

family, it is reasonable that they should be close together in psychological space. It is also interesting to note that the Kruskal procedure was sensitive to SCF. This was evidenced by the bipolar, "schema-like" dimension observed in the analysis of SCF data. Apparently, this multidimensional scaling procedure has a potential as a clustering program. As such, it may make additional contributions to schema theory.

This research supports the strategy of generating stimuli defined in terms of belonging to a particular schema population whose parameters are both measurable and manipulable. Statistical modeling of stimulus populations thus provide us with improved grasp of the problem of stimulus definition when researching form and pattern perception in a probabilistic environment.

REFERENCES

- ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychological Bulletin*, 1956, 53, 452-471.
- BROWN, D. R., & OWEN, D. H. The metrics of visual form: Methodological dyspepsia. *Psychological Bulletin*, 1967, 68, 243-249.
- EDMONDS, E. M., & EVANS, S. H. Schema learning without a prototype. *Psychonomic Science*, 1966, 5, 247-248.
- EVANS, S. H. A model for perceptual category formation. Unpublished doctoral dissertation, Texas Christian University, 1964.
- EVANS, S. H. VARGUS 7: Computed patterns

- from Markov processes. *Behavioral Science*, 1967a, 12, 323-328.
- EVANS, S. H. A brief statement of schema theory. *Psychonomic Science*, 1967b, 8, 87-88.
- EVANS, S. H., & ARNOULT, M. D. Schematic concept formation: Demonstration in a free sorting task. *Psychonomic Science*, 1967, 9, 221-222.
- EVANS, S. H., & BRECKENRIDGE, R. L. Generation of patterns for: Schema research, ecological validity, and control of other variables relevant to pattern perception. Paper presented at the meeting of the Psychonomic Society, St. Louis, October 1968.
- EVANS, S. H., & MUELLER, M. R. VARGUS 9: Computed stimuli for schema research. *Psychonomic Science*, 1966, 12, 511-512.
- FITTS, P. M., & LEONARD, J. A. *Stimulus correlates of visual pattern recognition: A probability approach*. Columbus: Ohio State University Press, 1957.
- HOLLIER, J., & EVANS, S. H. Schematic concept formation with linguiform patterns. *Psychonomic Science*, 1967, 9, 89-90.
- KLEMMER, E. T., & SHRIMPSON, M. W. Preference scaling via a modification of Shepard's proximity analysis method. *Human Factors*, 1963, 5, 163-168.
- KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, 29, 115-129.
- McGEE, V. E. The multidimensional analysis of "elastic" distances. *The British Journal of Mathematical & Statistical Psychology*, 1966, 19, 181-196.
- MICHELS, K. M., & ZUSNE, L. Metrics of visual form. *Psychological Bulletin*, 1965, 63, 74-86.
- POSNER, M. I., GOLDSMITH, R., & WELTON, K. E. Classification of distorted patterns. *Journal of Experimental Psychology*, 1967, 73, 28-38.
- RANKIN, W. C., & EVANS, S. H. Facilitation of schematic concept formation as a function of two within-schema pretraining modes. *Psychonomic Science*, 1968, 13, 325-326.
- SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, 27, 125-140.
- WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

NOTES

1. This research was supported by the Department of Defense, Project THEMIS Contract (DAAD05-68-C-0176), under the Department of the Army to the Institute for the Study of Cognitive Systems through the TCU Research Foundation. Further reproduction is authorized to satisfy needs of the U.S. government.
2. The authors gratefully acknowledge manuscript review and comments of Professor Richard M. Fenker.
3. Address: Institute for the Study of Cognitive Systems, TCU Research Foundation, Fort Worth, Texas 76129.

(Accepted for publication April 17, 1969.)