

A METHOD FOR THE STUDY OF INTERSTIMULUS SIMILARITY*

CLYDE H. COOMBS

UNIVERSITY OF MICHIGAN

A generalized method for collecting data on interstimulus similarity is presented and its special cases evaluated by use of information theory. A method of analyzing the data by the Unfolding Technique is presented which permits the study of the latent attribute structure underlying the similarity of stimuli for a single individual.

I. *Introduction*

For certain purposes it may be desirable to determine the relations among a set of stimuli with respect to their similarity. In certain instances, where the similarity and differences are ascribable to a single common latent attribute, the stimuli may be scaled in any of a number of different ways (1, 5, 6). The data utilized to construct such a scale, here called a stimulus scale, are either single judgments from each of a number of different judges or many experimentally independent replications of judgments from a single judge. The former are usually used when the stimuli are identifiable and there might thereby be experimental dependence in the replication of a single individual. The latter are necessary when the scale to be constructed is entirely a personal one and not common to other individuals. The Law of Comparative Judgment provides the best method now available for constructing a stimulus scale at the level of interval scale measurement but requires that the successive judgments be experimentally independent and that the stimuli be indiscriminable to some extent.

In certain instances the problem of constructing a stimulus scale for a single individual, when the stimuli are identifiable and 100 per cent discriminable, arises. This problem occurs in constructing scales of the utilities of objects and scales of psychological probability. In each such instance the stimuli may be clearly identifiable, and perfectly discriminable, and the scales may be unique to an individual. Constructing such scales is the problem for which the methods presented in this paper have been developed.

The methods developed here are an adaptation of the Unfolding Technique as it is applied to J scales (3, 4). In the case of a J scale, the different individuals are conceived of as standing on the J scale (the judges' respective "ideals") and evaluating each pair of stimuli as to which is "nearer" (pre-

*This study was carried out under Office of Naval Research Contract Nonr 374 (00) NR 041-011.

ferred). The judge's order of the stimuli from most preferred to least preferred is called an *I* scale and represents the *J* scale folded at the judge's ideal. The data, consisting of a set of *I* scales, may be unfolded to recover the *J* scale at the level of an ordered metric which may be interpreted as a common latent attribute underlying the preferences.

The problem in a stimulus scale is that the judges are not distributed on the continuum with the stimuli. In the case of a utility scale, for example, the locus of the stimuli on the scale corresponds to their "preferability" or their "desirability" to the individual.

To obtain metric relations between stimuli on the stimulus scale the method of collecting data may be modified to permit the use of the Unfolding Technique. What is necessary is a method of collecting data which will make the individual "stand" on the stimulus scale at a number of different places and evaluate the stimuli in order of increasing distance from each. For each position an *I* scale is obtained and the set of these may be unfolded to recover the stimulus scale with metric relations.

The general method of collecting data is called the Method of Cartwheels, primarily because it provides a ready terminology for the elements. All of the methods which are special cases of this general method differ only in the *amount* of information they provide; the *kind* of information is the same for all. The stimuli are always presented in sets of three or more in any of several arrangements and the basic judgment operation which the subject performs is to rank order the *distances* between pairs of stimuli. As far as I know, the first person to recognize and make use of such data was Richardson (8), followed by Klingberg (7) and more recently by Torgerson (10), all of whom applied the Law of Comparative Judgment to the analysis of such data. The method of analysis described here is quite different.

In the next section a special case of the Method of Cartwheels is presented, called the Method of Similarities, and the method of analysis is described in detail. The generalized method is then presented and some of its special cases pointed out. The analysis of data collected by any of the special cases is clear from the treatment of the Method of Similarities, as the various special cases differ only in the amount of information they provide.

II. *The Method of Similarities*

The stimuli are presented to the individual in triads, and he is asked to judge which two are most alike and which two are least alike in each triad. Then each triad is decomposed into three paired comparisons, each paired comparison becoming an element used in constructing a different *I* scale. The method is illustrated with a hypothetical example of five stimuli, *ABCDE*, presented in ten triads as shown in Figure 1.

The position of the stimuli in the triad and the serial order of the triads are randomized in an experiment. An individual is instructed to respond

to each of the ten triads indicating in each which pair of stimuli is most alike and which pair is least alike. This is the most general form of the instructions; variations are described below.

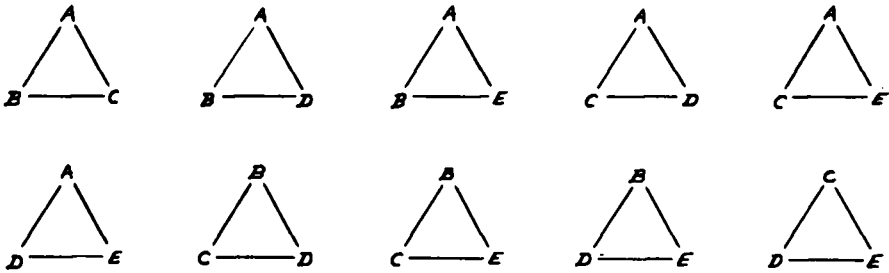


FIGURE 1
The Ten Triads of Five Stimuli

To illustrate the method of analyzing the data, assume the judgments obtained to be those presented in Table 1, and let the hypothesis be that the five stimuli lie on a common latent attribute. It is desired to test this hypothesis and, if sustained, to recover the latent attribute with the stimuli

TABLE 1
Sequence of Triads and the Response of the Judge*

**M* signifies the judgment of "most alike," and *L* signifies the judgment of "least alike."

ordered on it and with order relations on the distances between pairs of stimuli.

Analysis of the data begins by decomposing each triad into three paired comparisons and tabulating them as in Table 2. For example, the first triad is judged as shown in Figure 2. The decomposition of this triad into three paired comparisons proceeds as follows. Imagine that the individual were standing on the stimulus scale where stimulus *A* is and that he had been asked whether stimulus *B* or *E* were nearer *A*. Clearly, he has answered this

Thus, from these data one would have obtained a stimulus scale with the stimuli in the order *ABCDE* and with the metric relations given in Figure 3. This is as far as the data takes one. The problem of interpreting (identifying) the attribute, and of saying which end is which is an exercise for the experimenter's judgment. In the case of utility and of psychological probability, this is usually self-evident or may be obtained readily from the subject.

There are variations in the instructions that may be used with this method, appropriate for different purposes. It may be desirable to make the latent attribute explicit and ask which are most alike in brightness, in attitude toward the church, or whatever is relevant.

When the latent attribute is explicit, then it is only this subspace of the total stimulus space that is being investigated. One may study an individual's conception of such attributes as leadership, authority, or aesthetic merit, by the structural relations of the stimuli in such a subspace. When the latent attribute is left implicit, the relative similarity of stimuli and the attributes underlying similarity may be investigated, providing a new approach to the study of equivalence of stimuli.

In scaling utilities, we have used instructions like which two are hardest to choose between, and which two are easiest. In scaling psychological probabilities, we have instructed the subject to indicate which is most likely and which is least likely, and to place the intermediate statement nearer one of the two extremes.

Information theory is of assistance in evaluating various methods of collecting data. For example, in the Method of Similarities, each triad may be responded to in six different ways. The three "distances" in each triad are rank ordered by the two judgments "most" and "least" and, as the three distances may be ranked six different ways, there are six different response patterns. Assuming these to be equally likely, the number of bits of information contained in a single triad is $\log_2 6 = 2.58$. For n stimuli the number of triads is $\binom{n}{3}$ so the total number of bits is $2.58\binom{n}{3}$. In decomposing a triad into three paired comparisons, this is making three bits of information out of 2.58, and this increment represents the amount of information that has been added by imposing transitivity on paired comparison judgments within a triad. While there is experimental dependence among these three paired comparisons, the data are analyzed by associating paired comparisons from different triads which are experimentally independent.

A weakness of the method lies in the lack of replication and when there is error or vacillation in judgment, the I scales may not satisfy the necessary conditions for unfolding into a common latent attribute. The method is highly vulnerable and really constitutes a stringent criterion of whether a common latent attribute exists for an individual over these stimuli. As most human behavior seems to involve some "random" elements it was found desirable to generalize this method to provide replication and hence an

evaluation of inconsistency of judgments. The generalization of the Method of Similarities we call the Method of Cartwheels.

III. *The Method of Cartwheels*

The name Method of Cartwheels covers a great variety of methods which might be illustrated as follows. Of the n stimuli, p are presented at a time with not more than one in the center, or hub, and the others around the rim, as in Figure 4. The instructions are similar to those for the Method of

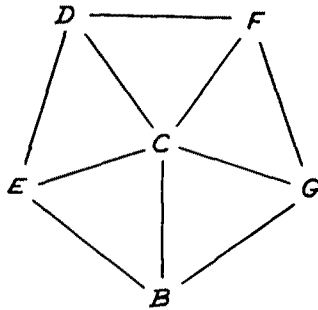


FIGURE 4
Illustrating the Method of Cartwheels

Similarities, except that the subject is asked to indicate which pair is most alike, which pair is next most alike, and so on, to which pair is least alike. A line in the figure, as a spoke, or on the rim, designates a pair of stimuli to be included in the ranking.

Clearly, with n stimuli, the number presented at a time, p , must satisfy $3 \leq p \leq n$. With $p = 3$, there are two variations. (See Figure 5.)

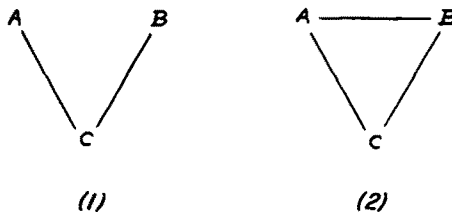


FIGURE 5
(1) The Method of Propellers
(2) The Method of Similarities

In (1), Figure 5, there is a stimulus at the hub and two spokes. The individual judges whether A or B is more like C . This is a direct paired comparison of the two distances. Every set of three stimuli is presented with

each stimulus at the hub in turn; this is spoken of as permuting hubs. Clearly, there is a full degree of freedom for each paired comparison in this method. The paired comparison from each such presentation is tabulated in the column corresponding to the hub, as was done in Table 2 with the Method of Similarities, where each stimulus in turn is regarded as the hub. Because of the number and variety of methods, we have taken to giving names to some of the more frequently used methods. This one is called the Method of Propellers.

In (2), Figure 5, the three stimuli are all on the rim and there is no hub. This is the Method of Similarities already discussed in detail. This method is slightly less powerful than the Method of Propellers when evaluated as follows. The 2.58 bits of information in a presentation in the Method of Similarities are distributed over the three paired comparisons obtained from the decomposition of the triad. Hence, on the average, each paired comparison conveys $(1/3)(2.58) = .86$ bits of information, and there is no replication of the paired comparisons in different triads. In the Method of Propellers each paired comparison conveys 1.00 bit of information. The latter method, however, takes longer to administer, as each set of three stimuli must be presented three times, permuting the hub, whereas in the Method of Similarities a set of three is presented only once. This is another illustration of the general principle that the value of data is measured by the effort required to collect it.

For all $p \geq 4$ there are further variations:

(1) The stimuli may be all on the rim with no hub (rims only); e.g., one presentation of four stimuli may be arranged as shown in Figure 6. Each set of four may be presented in three different arrangements with each

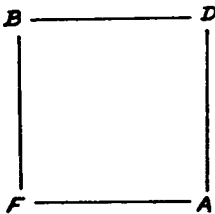


FIGURE 6

Another Variation of the
Method of Cartwheels

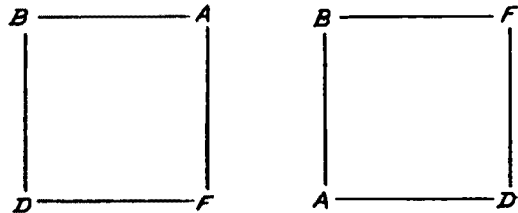


FIGURE 7

Other Arrangements of the Same
Four Stimuli

stimulus diagonally opposite each of the others in turn. For example, the four stimuli above may also be presented as shown in Figure 7.

The instructions again are to rank order the distances indicated, from smallest (most alike) to greatest (least alike). The four rank ordered distances may be decomposed into six paired comparisons, four of which are tabulated in the appropriate columns as in Table 2. The other two paired comparisons

are paired comparisons of disjoint distances and are tabulated separately. For example, in a presentation let the rank ordering from most alike to least alike be as illustrated by Figure 8. Imagine again that the individual stands

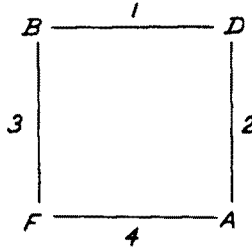


FIGURE 8
A Hypothetical Response Pattern

on each stimulus in turn and judges the relative similarity of the two adjacent stimuli. Then, in the manner of Table 2, these judgments would be tabulated as shown in Table 4. In addition the metric relations $AD < BF$, $AF > BD$

TABLE 4
Tabulation of Paired Comparisons
From Decomposition of the Response Pattern Given in Figure 8

A	B	...	D	...	F
D > F	D > B	...	B > A	...	B > A
.
.

are also implied in the rank ordering. The number of presentations of stimuli to which an individual responds in this method is $3\binom{n}{4}$, each set of four stimuli being presented in three arrangements. Each paired comparison, tabulated as in Table 2, is replicated $(n - 3)$ times.

A partial evaluation of this method may be made by applying information theory, as was done to the Method of Similarities. Here the amount of information in a presentation is $\log_2 24 = 4.57$. The judgments in a presentation are decomposed into six paired comparisons, so the 4.57 bits of information in a presentation may be regarded as being distributed over the six paired comparisons. On the average, then, each paired comparison conveys .76 bits of information. As this is cumulative over successive presentations if they are experimentally independent, the number of units of information gathered on any given paired comparison, essentially a measure of the amount of replication, is $.76(n - 3)$. This measure may be interpreted as an indication of the effectiveness of the method in controlling error or inconsistency.

(2) One of the stimuli may be at the hub and all others on the rim, and this has two variations: (a) spokes only, permuting hubs; and (b) both rims and spokes, permuting hubs; e.g., as shown in Figure 9, the instructions



FIGURE 9

Two More Variations for Presenting Four Stimuli at a Time

always being to rank order the pairs or distances indicated from most alike to least alike.

Evaluation of these methods by information theory may proceed as follows. For spokes only, permuting hubs, each presentation yields $\log_2 6 = 2.58$ bits of information, assuming each alternative response pattern is equally likely. These 2.58 units are distributed over 3 paired comparisons, so each receives on the average .86 bits from a single presentation. A given paired comparison is replicated $n - 3$ times in this method, so each paired comparison accumulates a total of .86 ($n - 3$) bits of information. From the point of view of controlling inconsistency, this method appears more powerful than the preceding.

For the method $p = 4$, both spokes and rims, permuting hubs, the number of alternative response patterns to one presentation is $6! = 720$. So the amount of information in a response pattern is $\log_2 720 = 9.97$. Distributing these over the 15 paired comparisons which a rank ordering of 6 elements decomposes into, each paired comparison may be valued at .66 bits. But in this method each paired comparison is replicated $2(n - 3)$ times, so each paired comparison accumulates 1.32 ($n - 3$) bits of information. This method is clearly the most powerful in this respect of any of those presented here. In fact, we have found that this method taxes the subject too greatly, and increasingly so with increasing complexity of stimuli.

These examples of $p = 3$ and $p = 4$ illustrate some of the variations possible in presenting p stimuli at a time. The major variations may be classified as spokes only, rims only, or both, as illustrated above. The variations differ in their theoretical effectiveness for controlling inconsistency of judgment. With sufficient replications on each paired comparison, inconsistency of paired comparison judgments may be controlled and the dominant preference in each pair determined. The analysis of these paired comparisons into rank ordered I scales and their unfolding proceeds as with

the Method of Similarities if the necessary conditions are satisfied. The result, then, is a scale with the stimuli ordered and with some of the distances between them ordered—an ordered metric.

IV. *General Comments*

The context in which these methods have been presented is the construction of a unidimensional scale for the stimuli. The data may sometimes require a multidimensional structure. Criteria for the number of dimensions necessary to account for the data have been developed by Bennett (2), but procedures for recovering a J space are as yet incompletely developed.

It is interesting to note that data collected by these methods may be analyzed also by Torgerson's multidimensional psychophysical scaling method involving an extension of the Law of Comparative Judgment (10).

It would be interesting to collect data from a single judge and analyze them by both methods. Application of the Law of Comparative Judgment requires many replications of each paired comparison to permit a reasonably reliable normal curve transformation on the percentages. One of the more powerful versions of the Method of Cartwheels would yield a sufficient number of replications itself, or a less powerful method may be used with replication of the method. We have found the latter to be more feasible because the more powerful versions of the Method of Cartwheels are beyond the judges' capacity to execute and hence generate noise. The scale obtained from Torgerson's method would be an interval scale, the scale from the method of analysis described here would be an ordered metric; but they could be compared.

The methods presented here are most appropriate for the development of stimulus scales which may be peculiar to individuals, and for testing whether different individuals have the same latent structure for a given set of stimuli. They may also prove useful in the study of equivalence of stimuli, generalization, and transfer. By virtue of providing an independent source of metric information, these methods will also be useful in the study of various attributes of response as possible psychological distance functions, e.g., amplitude, latency, and consistency.

REFERENCES

1. Attneave, Fred. A method of graded dichotomies for the scaling of judgments. *Psychological Review*, 1949, **56**, 334-340.
2. Bennett, J. F. A method for determining the dimensionality of a set of rank-orders. Unpublished doctor's dissertation, University of Michigan, 1951.
3. Coombs, C. H. Psychological scaling without a unit of measurement. *Psychological Review*, 1950, **57**, 145-158.
4. Coombs, C. H. A theory of psychological scaling. *Engineering Research Bulletin*, (University of Michigan Press), 1952, No. 34.

5. Edwards, A. L. and Thurstone, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1952, **17**, 169-180.
6. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
7. Klingberg, F. L. Studies in measurement of the relations among sovereign states. *Psychometrika*, 1941, **6**, 335-352.
8. Richardson, M. W. Multidimensional psychophysics. *Psychological Bulletin*, 1938, **35**, 659-660.
9. Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, **34**, 273-286.
10. Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17**, 401-419.

Manuscript received 8/3/53

Revised manuscript received 10/26/53