SESSION IX

INNOVATIVE PROGRAMMING FOR PSYCHOLOGICAL RESEARCH

GEOFFREY LOFTUS, *University of Washington, President*

Interactive similarity ordering

FORREST W. YOUNG

University of North Carolina, Chapel Hill, North Carolina 27514

CYNTHIA H. NULL

College of William and Mary, Williamsburg, Virginia 23185

and

WARREN SARLE

University of North Carolina, Chapel Hill, North Carolina 27514

The interactive similarity ordering (ISO) system gathers similarity data that can be symmetric, asymmetric, or rectangular in form. Because the subject makes a very simple judgment and the number of judgments is interactively minimized, this procedure can be used for a relatively large number of stimuli. This system has been programmed on a PDP-11/45 and on a Motorola 6800-based microprocessor. An empirical investigation indicates that the ISO system compares favorably with other large stimulus set procedures.

When using a multidimensional scaling procedure, most researchers collect data by having subjects rate the similarity of all possible pairs of stimuli. When the number of stimuli increases to, say, 25, the number of pairs becomes very large, 300, and makes the similarity task unfeasible or, at least, undesirable. Since using a large number of stimuli is necessary if one expects a solution with three or more dimensions (Young, 1970), which is usually desirable, some way of limiting the size of the task is of great interest.

Many researchers avoid the large number of pairs required by a large number of stimuli by using a sorting or grouping task instead of interstimulus similarity judgments (e.g., Boorman & Arabie, 1972; Rosenberg & Sedlak, 1972). A similarity matrix is derived from the sorting information. One may question whether the derived matrix adequately approximates the subject's perception of the interstimulus distances. This method must be used with caution. Also, the method may be difficult to use in some situations. For example, it would be very difficult to sort auditory stimuli, since they appear in time and cannot be physically sorted into groups.

If the researcher collects data using ratings of pairs, the numbers of pairs must be reduced. Spence proposed using a cyclic design to choose a subset of the stimulus

pairs (Spence, 1977; Spence & Domoney, 1974; Spence & Janssen, Note 1), and demonstrated that gathering data for approximately 60% of the pairs yields results comparable to other procedures. He found that a solution space is best defined by its maximally distant pairs. This procedure does not guarantee the presentation of these pairs unless the experimenter has prior knowledge of the stimuli and chooses these pairs to appear in the cyclic subsets.

Young and Cliff (1972) developed an interactive procedure (ISIS) which chooses the subset of pairs to be presented on the basis of the subject's previous responses. The ISIS procedure determines the maximally distant stimuli so that prior knowledge is not necessary to obtain a well-defined space. The procedure requires the subject to make judgments on 25% to 45% of the pairs. This technique has been positively evaluated using both empirical and Monte Carlo procedures (Baker & Young, 1974; Girard & Cliff, 1976). However, the procedure is as time consuming as the Spence procedure (which uses more pairs) and assumes that the metric of the similarity judgments is euclidean.

The interactive similarity ordering (ISO) system proposed here allows the experimenter to obtain similarity data for as large a set of stimuli as do the procedures mentioned above. Additionally, this procedure

can be used to collect data that is symmetric, asymmetric, or rectangular. The ISO system has two main advantages: (1) the subject makes a very simple judgment and (2) the number of judgments is interactively minimized.

ISO SCALING PARADIGMS

Definitions

The data collected using the ISO system can be described as two-way data, where the number of ways corresponds to the number of experimental conditions. Two-way data can be either square or rectangular. The first involves the Cartesian product ($I \times I$) of a set of stimuli I with itself. For example, a researcher could ask a subject to rate the similarity of 25 comedians to each other. This data would be square two-way data and also symmetric. The distinguishing feature of a symmetric matrix O of observations is that the entries $o_{ij} = o_{ji}$ for all pairs of elements (see Figure 1). Square data can also be asymmetric. Asymmetric data is also called row conditional, that is, the observations in the i^{th} row are relative to the i^{th} stimulus. The experimenter asks the subject to rank order 24 of the comedians in relation to the 25th. The subject repeats the ranking task with each of the comedians as the standard. This data is square because the matrix is comedian by comedian, but is asymmetric since $o_{ij} \neq o_{ji}$ (see Figure 2).

Finally two-way data can be rectangular, which involves the Cartesian product ($I \times J$) of two sets, I (stimuli) and J (adjectives) (see Figure 3). The researcher can have a subject rate 20 musical timbres on each of eight adjective scales. For example, a subject

Figure 1. Schematic of symmetric data matrix.

Figure 2. Schematic of asymmetric data matrix.

Figure 3. Schematic of rectangular data matrix.

can rate how "brilliant" each timbre is, how "hollow" each timbre is, etc.

The Task

When collecting symmetric data, the subject is presented with the following judgment situation: A list of

from 2 to r pairs of stimuli is presented, and he is asked to select the pair whose elements he judges most similar. The subject is repeatedly presented with such lists until the rank order of all the pairs is obtained. Figure 4 presents examples of two trials where the stimuli are American cities. The maximum length of the list presented, r , can vary from 2 through 7. From one presentation to the next, only one pair changes, so the task is quite simple. This method of collection is appropriate only when the number of stimuli is relatively small (say, less than 12) since it orders all pairs.

In the conditional rank-order situation (asymmetric data), one of the n stimuli is designated the "standard" stimulus and the remaining $n - 1$ the "comparison" stimuli. A list of from 2 to r comparison stimuli is presented to the subject, and he is asked to select the one he judges most like the standard. The subject is repeatedly presented with such lists until the rank order of the similarity of all $n - 1$ comparison stimuli to the standard is determined. From one list presentation to the next, it is often true that only one comparison stimulus changes, thus simplifying the task. The maximum length of the list of comparison stimuli presented to the subject is set by the experimenter and can vary from 2 through $n - 1$. In Figure 5 are three trials where the list length is 4 and the stimuli are again cities. For this set of trials Boston was chosen as the standard. The order of the standards is chosen by the experimenter. Also note that the list length, r , is the maximum number to be presented on any trial. Therefore, on any trial the subject may see as few as two comparison stimuli, as in the example. The length of the comparison list affects the total number of judgments a subject must make for each standard and the difficulty of each judgment.

After completing one conditional rank order, ISO selects a second standard and has the subject make selections from lists of comparison stimuli, thus constructing

WHICH OF THESE PAIRS IS CLOSEST?

- 2 MEMPHIS
SALT LAKE CITY
- 3 LOS ANGELES
SALT LAKE CITY
- 4 BOSTON
WASHINGTON

YOUR RESPONSE: 4

WHICH OF THESE PAIRS IS CLOSEST?

- 2 MEMPHIS
SALT LAKE CITY
- 3 LOS ANGELES
SALT LAKE CITY

YOUR RESPONSE: 3

Symmetric (Unconditional, one mode)

Figure 4. Example of subjects' judgment task using symmetric experimental paradigm.

WHICH OF THE FOLLOWING CITIES IS
CLOSEST TO: BOSTON

- 3 WASHINGTON
- 4 LOS ANGELES

YOUR RESPONSE: 3

WHICH OF THE FOLLOWING CITIES IS
CLOSEST TO: BOSTON

- 1 MEMPHIS
- 2 NEW ORLEANS
- 3 CHAPEL HILL
- 4 WASHINGTON

YOUR RESPONSE: 4

WHICH OF THE FOLLOWING CITIES IS
CLOSEST TO: BOSTON

- 1 MEMPHIS
- 2 NEW ORLEANS
- 3 CHAPEL HILL
- 4 LOS ANGELES

YOUR RESPONSE: 3

Asymmetric (Row conditional, one mode)

Figure 5. Example of subjects' judgment task using asymmetric experimental paradigm.

a second conditional rank order. This procedure is repeated until $c \leq n$ conditional rank orders have been completed, where c is also set by the experimenter and may vary from 1 to n . Note that the subject's judgment is to pick, from a short list of stimuli, the one most like the standard, a very simple judgment. This results in an asymmetric matrix of two-way similarity data.

This procedure is appropriate for a substantially large number of stimuli because the subject is rank ordering $n - 1$ items and not all pairs. Furthermore, not all standards must be rated and partial or incomplete orders of the comparison stimuli may be obtained without jeopardizing the solution. For a partial order the comparison stimuli are divided into sublists and the subjects determine a rank order for each of the sublists but not for the complete list. This reduces the number of judgments a subject must make for each standard. Both of these judgment-saving options are available in the ISO system and can be used at the experimenter's discretion to further shorten the task.

To obtain rating scale or rectangular data, the subject is presented with an adjective as the "standard" and all n stimuli as "comparisons." As in the conditional rank-order task, a list of from 2 to r comparison stimuli are presented to the subject and he is asked to choose the stimulus which is best described by the adjective or is most like the adjective, etc. In the example, Figure 6, the subject is asked to indicate which of the cities is most familiar to him. Lists of comparison stimuli are presented, one item changing per presentation, until a rank order for the stimuli with regard to the standard is obtained. This procedure is repeated for each adjective.

WHICH OF THE FOLLOWING CITIES IS
MOST FAMILIAR TO YOU:

- 2 WASHINGTON
- 4 BOSTON

YOUR RESPONSE: 2

WHICH OF THE FOLLOWING CITIES IS
MOST FAMILIAR TO YOU:

- 1 CHAPEL HILL
- 2 SALT LAKE CITY
- 3 DALLAS
- 4 LOS ANGELES

YOUR RESPONSE: 1

WHICH OF THE FOLLOWING CITIES IS
MOST FAMILIAR TO YOU:

- 1 DENVER
- 2 SALT LAKE CITY
- 3 DALLAS
- 4 LOS ANGELES

YOUR RESPONSE: 4

Rectangular (Row conditional, two mode)

Figure 6. Example of subjects' judgment task using rectangular experimental paradigm.

Partial and incomplete orders are also possible in this paradigm.

ALGORITHM

In the ISO system the composition of each list of comparison stimuli is determined from the subject's previous judgments by a MERGESORT (Knuth, 1973)

sorting routine in which the comparison operation has been replaced by the subject's judgment. MERGESORT is used because it presents a selection task and requires only $n * \log_2(n)$ comparisons (on the average), the least of any sorting procedure. In addition, the set of comparison stimuli changes by only one stimulus from comparison to comparison, which reduces the complexity of the task.

Table 1 shows a straight two-way merge sort with a comparison list length of 2 for the rectangular paradigm. The standard is familiarity and the stimuli are 16 American cities. This sorting method "burns the candle from both ends": The stimuli are presented to the subject two at a time from the top and from the bottom, working toward the middle. The first column illustrates a starting order for the stimuli. The subject is given Denver and Seattle as the first comparison list. To illustrate the process of selecting subsequent lists, consider the transformation from column 2 to column 3. (Note: The horizontal lines in the table indicate ordered subsets.) At the top are Denver and Seattle, with Denver rated as more familiar than Seattle and, at the bottom, with Chapel Hill as more familiar than Los Angeles. The subject is presented Denver and Chapel Hill to begin the merge. For this subject, Chapel Hill is more familiar, so the next presentation is Denver and Los Angeles; since Denver is more familiar than Los Angeles, the next pair is Los Angeles and Seattle. This gives the order Chapel Hill, Denver, Seattle, and Los Angeles, which is placed at the top of the next column. Then San Francisco, New Orleans, Miami, and Chicago are merged in the same manner and placed at the bottom of column 3, with the most familiar of the set, Chicago,

Table 1
Example of Straight Two-Way Sorting With List Length of 2 Using Familiarity to American Cities

Unranked Comparison Stimuli	2	3	4	Ranked Comparison Stimuli
<u>Denver</u>	Denver	Chapel Hill	Chicago	Chicago
<u>Chapel Hill</u>	Seattle	Denver	Chapel Hill	Chapel Hill
<u>San Francisco</u>	San Francisco	Seattle	Denver	Washington, D.C.
<u>Chicago</u>	New Orleans	Los Angeles	San Francisco	Boston
<u>Memphis</u>	Salt Lake City	Washington, D.C.	Miami	New York
<u>Boston</u>	Memphis	Boston	New Orleans	Philadelphia
<u>Dallas</u>	Philadelphia	Salt Lake City	Seattle	Denver
<u>New York</u>	Dallas	Memphis	Los Angeles	Salt Lake City
<u>Atlanta</u>	Atlanta	Dallas	Memphis	San Francisco
<u>Philadelphia</u>	New York	Atlanta	Dallas	Miami
<u>Washington, D.C.</u>	Boston	Philadelphia	Atlanta	Atlanta
<u>Salt Lake City</u>	Washington, D.C.	New York	Salt Lake City	New Orleans
<u>Miami</u>	Miami	New Orleans	Philadelphia	Seattle
<u>New Orleans</u>	Chicago	Miami	New York	Los Angeles
<u>Los Angeles</u>	Los Angeles	San Francisco	Boston	Dallas
<u>Seattle</u>	Chapel Hill	Chicago	Washington, D.C.	Memphis

at the bottom. The process proceeds by merging Salt Lake City, Memphis, Boston, and Washington, D.C., with the result placed after the first merge with the most familiar first. The results from the final merge of Philadelphia, Dallas, Atlanta, and New York are merged and placed in the remaining spaces in column 3, with the most familiar toward the bottom. Column 2 was transformed from column 1 in this manner, as was 4 from 3 and 5 from 4. This procedure can be generalized for the situation where r , list length, is greater than 2.

PROGRAMS

The ISO system has been programmed on both a mini- and microcomputer. One program is written in FORTRAN IV PLUS for a PDP-11/45 with an RSX-11D operating system and a scope operating at 9,600 baud. The program collects data for any of the paradigms mentioned above. It also collects data where the task is to insert into a list of comparison stimuli (using the QUICKSORT algorithm) as opposed to picking from a comparison list as described here. The program, in addition to collecting similarity data, counts the number of judgments per order and the amount of time a subject takes to make each judgment. The subject may interrupt a data collection session between standards and return at a later time to complete the task. This allows the subject to stop when tired and return when fresh. This facility improves judgment reliability and validity. It also enables collection of data over several sessions, thus further increasing the potential number of stimuli.

For the Sphere microprocessor, a Motorola 6800-based system, the ISO system has two programs. The programs are written in Pittman's TINY BASIC and numerical machine code. The first of the two programs collects asymmetric and rectangular data for the alphanumeric type of stimuli as described above. The second program presents discrete graphical patterns to the subject for only the asymmetric paradigm. An example from this program is presented in Figure 7. Note that the length of the comparison list can be only 2 or 3 in the second program due to the limited size of the CRT screen. Both programs record the number of judgments necessary for each order. The interruption facility is available in this set of programs, but subject judgment time and the insertion instead of pick option are not available at this time.

EMPIRICAL EVALUATION

As a preliminary investigation of the effectiveness of the ISO system, data were collected in the asymmetric and rectangular paradigms. The effect of comparison list length was assessed on two dependent variables: the number of judgments the subject made per standard and the length of time for each judgment. Comparison lists were 2, 8, or 16 stimuli long. The use of partial

Figure 7. Example of picture judgment task using asymmetric experimental paradigm.

orders was also investigated. Instead of ranking the complete list for each standard, some subjects in the list length conditions of 2 and 8 rank ordered two sublists. Thus, there was complete ranking of the stimuli for presentation list lengths of 2, 8, and 16 and partial (or separately ranking each half of the stimuli) for list lengths of 2 and 8.

Method

Subjects. Forty-five undergraduate introductory psychology students at the University of North Carolina at Chapel Hill served as subjects. They received course credit for their participation. The subjects were randomly assigned to one of the five experimental conditions.

Stimuli. Sixteen American cities were used as stimuli: Atlanta, Boston, Chapel Hill, Chicago, Dallas, Denver, Los Angeles, Memphis, Miami, New Orleans, New York, Philadelphia, Salt Lake City, San Francisco, Seattle, and Washington, D.C. City familiarity was used as the standard in collecting the rectangular data.

Apparatus. Stimuli were presented via the ISO system on a PDP-11/45 with an RSX-11D operating system and a scope operating at 9,600 baud.

Procedure. The subject sat in front of the computer scope and, after brief instructions about using the keyboard, the experiment began. All instructions appeared on the scope. The subject began by rank ordering the cities on familiarity. The comparison set of stimuli was displayed as shown in Figure 6. The subject chose the stimulus from the list that was most familiar to him. The subject continued to make these judgments until he had rank ordered the list. After familiarity ranking was complete the subject rank ordered, by distances, 15 cities to a 16th city chosen as the standard, as described above and illustrated in Figure 5. The subjects repeated the ordering task for approximately eight standards.

Results and Discussion

A summary of the results are contained in Table 2. As expected, judgment time increased as length of comparison list increased. In the complete ordering conditions, the number of judgments decreased dramatically as list length increased. However, the average amount of time subjects spent ranking for each standard did not

Table 2
Evaluation of the ISO System for Asymmetric and Rectangular Experimental Paradigms

	List Length				
	2		8		16
	Complete Ordering	Partial Ordering	Complete Ordering	Partial Ordering	Complete Ordering
Average time per judgment	4.707	5.519	9.169	9.983	12.348
Standard deviation (seconds)	2.076	1.813	2.805	6.091	4.572
Average number of judgments per order	41.284	14.589	21.21	6.566	14.571
Standard deviation	2.972	2.022	.0721	.496	.495
Average number of minutes per standard	3.239	2.684	3.241	2.185	2.999
Number of standards per hour	18	22	18	27	20
Number of orders	67	180	62	129	56
Number of subjects	9	10	8	10	8

vary for list length in the complete condition. For the list length of 2 the subject took about 3.2 min/standard and for the list length of 16, about 3 min/standard. That is, the subject made a few long judgments in the long-list-length condition and many quick judgments on

the short list. There are many fewer judgments to be made in the partial list conditions since the subject is completely ranking two sublists. The average time per standard is decreased by this work reduction. From these results it is estimated that subjects can rank order

Figure 8. Multidimensional scaling solution for asymmetric data from the complete order condition.

Figure 9. Multidimensional scaling solution for asymmetric data from the partial order condition.

16 comparison stimuli to about 18 standards in 1 h in the complete order condition. In the partial order condition the subject can make judgments on 22 standards in 1 h when the list length is 2, and about 27 when the list length is 8. These times are faster than the usual times for making similarity judgments on all possible pairs of 16 stimuli. However, 16 stimuli cannot be considered a large set. If the number of stimuli is 32, we estimate the subject can complete partial orders for about 16 standards in 75 min. This compares favorably with both the ISIS procedure and Spence's incomplete design.

The multidimensional scaling solutions for the complete and partial order conditions appear in Figures 8 and 9. Solutions for each list length are not separately presented since list length did not affect solution quality. The solution for the complete order condition is topologically equivalent to a map of the United States, with the exception of the location of San Francisco and the reversing in the east-west direction of Chapel Hill and Atlanta. These discrepancies are due to the limited number of standards judged by each subject and the fact that the standards are the same for all

subjects. Previous anecdotal results indicated that only topological equivalence to a United States map should be expected from subjective distance judgments. It can be argued that the subjects can more accurately judge distances in the part of the country where they reside and that distant locations are lumped together. Thus, the distances between western cities should be decreased relative to true distances for residents of Chapel Hill.

The MDS solution for the partial ordering asymmetric data (Figure 9) is very different from a map of the United States. The cities that served as standards are fairly well located (New Orleans, Chapel Hill, Denver, Los Angeles, Miami, Seattle, Boston, and Washington, D.C.), and the remaining cities are grossly misplaced. These discrepancies were caused by a simple design error. The comparison stimuli were divided into two sublists. The same division was used to pick the standard to be used; that is, all of one sublist served as standards and none of the other. Therefore, the two subsets were completely disconnected: There were no data to determine the distances between the subsets. This problem can be eliminated by having each subject use a different randomly selected set of standards, where the sublists

for each standard are either different by a set preselection or different for each subject by random selection.

In conclusion, the ISO system interactively collects symmetric, asymmetric, and rectangular data. The results presented here indicate that it is feasible to use this system to collect asymmetric data for a large set of stimuli if one uses partial orders and limits the number of standards. Caution must be used when choosing the standards and when choosing the sublists so that all stimuli are connected.

REFERENCE NOTE

1. I. Spence & H. Janssen. *Cyclic incomplete designs as an alternative to ISIS models in nonmetric scaling*. Paper presented at Psychometric Society meetings, Stanford University, 1974.

REFERENCES

- BAKER, R. F., & YOUNG, F. W. *Interactive scaling with individual subjects: II. An improved system* (L. L. Thurstone Psychometric Laboratory Report 132). Chapel Hill: University of North Carolina, 1974.
- BOORMAN, S. A., & ARABIE, P. Structural measures and the method of sorting. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1). New York: Seminar Press, 1972. Pp. 225-249.
- GIRARD, R. A., & CLIFF, N. A Monte Carlo evaluation of interactive multidimensional scaling. *Psychometrika*, 1976, **41**, 43-64.
- KNUTH, D. E. Sorting and searching. *The art of computer programming* (Vol. 3). Reading, Mass: Addison-Wesley, 1973.
- ROSENBERG, S., & SEDLAK, A. Structural representations of perceived personality trait relationships. In A. K. Romney, R. N. Shepard, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 2). New York: Seminar Press, 1972. Pp. 133-162.
- SPENCE, I. Incomplete experimental designs for multidimensional scaling. In R. Colledge & J. N. Rayer (Eds.), *Multidimensional data analysis of large data sets*. Columbus, Ohio: Ohio State University Press, 1977.
- SPENCE, I., & DOMONEY, D. W. Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika*, 1974, **39**, 469-490.
- YOUNG, F. W. Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, 1970, **35**, 455-473.
- YOUNG, F. W., & CLIFF, N. Interactive scaling with individual subjects. *Psychometrika*, 1972, **37**, 385-415.