

Computer processing of free recall data: Program RECALL

MICHAEL L. FRIENDLY

York University, Toronto, Ontario M3J 1P3, Canada

A computer program for comprehensive analysis of data from free-recall experiments is described. The program scores recall protocols of each subject for over 30 measures of recall performance, subjective organization, and category clustering, any of which can be plotted for individuals or groups. In addition, the program calculates main-effect and interaction means for the experimental factors on any of the measures. A variety of other analyses, such as serial position curves and probabilities of item recall, are provided also.

This report describes a computer program designed to provide comprehensive and flexible analyses of data from multitrial free-recall (MFR) experiments. In such experiments, a list of items, usually common words, is presented to the subject for study, and the subject is required to recall as many list items as he can, in any convenient order. On subsequent trials, the same items are again presented, followed by attempted recall.

In analyzing the data from such experiments, the investigator is typically concerned with various aspects of recall performance, such as the number of items correctly recalled and measures of organization of recall. Manual scoring of the raw response protocols and calculation of various measures for each subject on each trial is extremely tedious and error prone. It is for this reason that the present program was written.

RECALL provides the facility for scoring the recall protocols for each subject for over 30 measures of recall performance, subjective organization, and category clustering, any of which can be plotted at the individual and group levels and written out onto an external device for further analysis. Probabilities of item recall can be calculated as a function of input serial position, nominal identity of individual items, and membership in experimenter-defined categories. In addition, the program can take into account the design structure of the experiment and calculate main-effect and interaction means for the experimental factors on any of the measures.

Input to the program consists of punched cards containing abbreviations (up to four characters) of the subject's responses, together with control information. If the items in a list have distinct initial letters, the data cards can be keypunched directly from the subject's written or spoken protocols without precoding or translation. This feature speeds the analysis of

experimental data considerably.

Related programs, described by Kowalski, Essene, Shapiro, & Jhangiani (1971) and Pellegrino (1972), perform some of the analyses described here. RECALL contains virtually all of the features of these programs, plus many others.

WHAT THE PROGRAM DOES

Each trial protocol for each subject is scored for a wide variety of measures of recall performance and organization. The measures are divided into three groups, as detailed below. At the user's option, each group of scores may be printed, written onto an output device (e.g., cards or disk) for subsequent analyses, or both. Some of the measures are plotted as a function of trials for each group. The plots identify each individual subject to provide some indication of individual variability and to detect deviant subjects.

Recall Performance Measures

The recalls for each subject are scored for the following measures of recall performance: (a) number of items correctly recalled, (b) number of intrusions—items "recalled" which did not appear on the stimulus list, (c) number of repetitions, (d) CC, CN, NC, NN.

The last set of four measures (CC, CN, NC, NN) are counts of items either recalled correctly (C) or not recalled (N) on two successive trials. Thus, NC is the number of items recalled on Trial *t* which the subject failed to recall on Trial *t-1*, while CC is the number of items recalled on both trials. CC may be regarded as an index of *intertrial* remembering, i.e., the portion of recall which is stable from trial to trial, while NC may be regarded as an index of *intratrial* remembering, the labile portion of recall. See Tulving (1964) for the description and theoretical account of these measures.

Subjective Organization

For an arbitrary set of list words, measures of subjective organization can be defined in terms of the

Research related to the development of this program was supported by the National Research Council of Canada, Grant No. A8615.

Requests for reprints should be sent to Michael Friendly, Department of Psychology, York University, Downsview, Ontario, Canada M3J 1P3.

extent to which subjects recall items in consistent groups from trial to trial. Tulving (1962) proposed that organization could be assessed by the amount of sequential redundancy (in information-theoretic terms) in the order of recall over a series of trials, relative to the maximum possible redundancy. The maximum redundancy occurs if the subject recalls the same items in a constant order on every trial. The program calculates Tulving's original SO measure (Tulving, 1962, Eq. 2), and some variations of SO, including SO₂, which uses a bidirectional scoring criterion (recall of items B, A is scored with recall of A, B). These measures can be computed over all trials, or over shorter blocks of successive trials, as determined by the user.

A second operational definition of subjective organization is based on the *intertrial repetition* (ITR) analysis of Bousfield and Bousfield (1966). The basic ITR measure is simply the number of output pairs recalled adjacently on two successive trials. For example, suppose a given subject recalled the following items in the order shown on two successive trials: Trial t—*river, arm, life, name, golf, eye, year*; Trial t+1—*river, golf, name, eye, year, life*. For this pair of trials, there are two ITR pairs—*eye-year* and *golf-name*, if pairs are scored without regard to internal order, but only one (*eye-year*) if pairs are scored in a unidirectional fashion. The ITR concept has been extended to larger item groups by Pellegrino (1971). An unordered ITR unit of Size k is scored for each distinct set of k items which appear contiguously on two successive trials. In the above example, there are two (unordered) ITR units of Size 3, one unit of Size 4, and one unit of Size 5.

The observed number of ITR units, O_{ITR} , however, is at least partly confounded with number recalled: the more items recalled, the greater O_{ITR} will tend to be, even if recall order is determined at random. Therefore, it is common practice to standardize O_{ITR} so that its dependence on number recalled is eliminated. To achieve this, the program calculates the maximum value of O_{ITR} for units of Size k, and the expected value, E_{ITR} under the assumption of random ordering of the items recalled (Pellegrino, 1971). It then calculates two standardized measures of subjective organization based on ITR: a difference score, $O_{ITR} - E_{ITR}$ and a ratio score $(O_{ITR} - E_{ITR}) / (Max_{ITR} - E_{ITR})$. These measures are provided for each of Unit Sizes 2, 3, and 4.

Categorical Organization

When the to-be-remembered list can be partitioned into categories established by the experimenter, measures of stimulus-category clustering (SCR) can be used to determine the extent to which subjects use these categories in recall. SCR is simply a count of the number of times items from the same category are recalled in consecutive output positions in recall. For example, in learning a list composed from the taxonomic categories *animals, beverages, and articles of clothing*, a subject

might recall the following items in order: *horse, cow, dog, tie, gin, beer, socks, shirt*. In this protocol, the observed SCR count is 4, corresponding to the number of italicized items.

As with the subjective organization measures, SCR is usually standardized in some fashion to remove any necessary dependence of the category clustering measure on sheer recall. That is, as recall increases, more category repetitions would tend to appear even if the order of recall were random. The standardized measures attempt to partial out the number of words recalled so that the measures are independent of recall.

The RECALL program computes 10 measures for categorical word lists for each subject on every trial. These include the difference score suggested by Bousfield and Bousfield (1966), the Z-score measure of Hudson & Dunn (1969) and Frankel and Cole (1971), measures proposed by Dalrymple-Alford (1970), the "ratio of repetition" (Bower, Lesgold, & Tieman, 1969; Robinson, 1966), and the "adjusted ratio of clustering" of Roenker, Thompson, and Brown (1971). A complete list and description of all measures calculated by the program is available in a long write-up of the program, which may be obtained from the author.

Handling of Repetitions and Intrusions

In scoring free-recall protocols, some decision must be made regarding the handling of repetitions and intrusions. Whether such "irregular" responses are allowed to remain in the protocols, or are deleted, will affect the scoring of category clustering and subjective organization. If such responses are at all prevalent or if their frequency varies over treatment groups, the choice has some importance. To allow for some flexibility in this regard, RECALL performs all scoring on a separate (internal) copy of the recall data. In the standard version of the program, all repetitions (after the first appearance of an item) and intrusions are deleted before scoring for the measures of organization. It is a simple matter, however, to modify the program so that repetitions and intrusions are not edited out.

Item Recall Frequencies

In many studies, it is of interest to know something of the recall history of individual items within a list. For instance, in a mixed list design, the subject may be presented with a list of words that vary in word frequency, and it would be natural to look at the probability of recall of individual items as a function of their frequency of usage, or some other stimulus variable. Such information is also useful in retention studies in which the experimenter wishes to equate groups for degree of original learning (Underwood, 1964) or to assess recognition memory and relate recognition and recall.

To facilitate such analyses, RECALL tallies the frequency with which each item is recalled by each

subject across the set of trials. In the summary information provided for each subgroup, these frequencies are pooled over subjects and the proportion of times each item was recalled is printed out together with the item code. These subgroup proportions are often useful for spotting "unusual" words within a list, or determining why one list appears to be remembered better than another, though they had been matched on some stimulus variables.

Plots of Scores for Individual Subjects

After computing measures of performance and organization for a group, the program produces plots of these scores across trials for individual subjects. These plots may be requested by the user for any or all of the measures discussed earlier.

In a plot for a given measure, each subject's scores are identified by a different symbol. The program also plots the subgroup means for each trial, (M_t) and the values of $M_t \pm S_t$, $M_t \pm 2S_t$, etc., where S_t is the standard deviation on Trial t . These plots are quite useful in identifying subjects with unusual scores. Any subject whose scores are consistently outside the intervals $M \pm 2S$ should be regarded with curiosity—or suspicion—and his data should be carefully checked.

Serial Position Analyses

In the analyses of recall, performance is scored as a function of the nominal identity of item types. It is often worthwhile to check the manner in which recall varies with the position of items in the presentation sequence, rather than with the nominal identity of items. The results of such an analysis in terms of presentation order or serial position are typically presented in the form of a serial-position curve (SPC) showing the proportion of items recalled as a function of input serial position. Many of the effects of experimental variables in free recall on the SPC have been reviewed by Glanzer (1972).

The program calculates a separate serial-position curve for each trial of the experiment. The proportion of items recalled at each serial position is displayed both numerically and in plotted form. In addition, each SPC is smoothed by fitting at each serial position (excepting the endpoints) the least-squares polynomial of first degree (linear), relevant to three successive points (Hildebrand, 1956). The smoothed curve is also displayed on the plots.

Analysis of Experimental Factors

The analysis of recall protocols by the program can be divided into two phases. In the first phase, the protocols for subjects in each subgroup of the design are scored for the measures selected by the user. The program prints out the result for each subject as well as means for the subgroup and the plots described above. Each subgroup is analyzed in turn, without regard to any other subjects in the experiment.

The second phase of the RECALL analysis is concerned with the experimental factors in the design. If the experimental design includes two or more crossed or nested factors, it is useful to obtain main effect and interaction means for some or all of the measures discussed above. At the user's option, the program will compute any or all of the main-effect or interaction means for the experimental factors.

LIMITATIONS

The following restrictions have been placed on the current version of the program: (1) No more than five between-subject factors are included in the experimental design. (2) There is no limit of the number of levels for any single factor, but the number of groups or cells in the design may not exceed 40. (3) No more than 60 stimulus items are used. (4) Trials are limited to no more than 20. (5) No more than 25 categories are in the stimulus list. (6) No more than 10 different orders of presentation are involved if serial position analyses are desired.

LANGUAGE AND COMPUTER

The program is written in standard FORTRAN IV for IBM 360/370 series machines. It will compile and execute under both the G- and H-level compilers. Using the OVERLAY feature, the program will run in under 120 K bytes of main core.

AVAILABILITY

A long write-up containing detailed information necessary to set up the data deck is available from the author. Source copies of the program, together with control cards and sample data are available on nine-track tape for the cost of processing. The cost is \$20.00 if you supply a tape reel (600 ft or larger); \$50.00 if you do not supply the tape reel. Address requests for information to Michael L. Friendly, Department of Psychology, York University, Downsview, Ontario M3J 1P3, Canada.

REFERENCES

- Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and sequential constancies in repeated free recall. *Psychological Reports*, 1966, 19, 935-942.
- Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 481-493.
- Dalrymple-Alford, E. C. The measurement of clustering in free recall. *Psychological Bulletin*, 1970, 74, 32-34.
- Frankel, F., & Cole, M. Measures of category clustering. *Psychological Bulletin*, 1971, 76, 39-44.
- Glanzer, M. Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5). New York: Academic Press, 1972.
- Hildebrand, F. B. *Introduction to numerical analysis*. New York: McGraw-Hill, 1956.
- Hudson, R. L., & Dunn, J. E. A major modification of the Bousfield (1966) measure of category clustering. *Behavior Research Methods & Instrumentation*, 1969, 1, 110-111.
- Kowalski, C. S., Essene, K. A., Shapiro, S. I., & Jhangiani, A. K.

- Computer analysis of free recall, intertrial organization, commonality of organization and clustering. *Psychonomic Monograph Supplements*, 1971, 4, 33-56.
- Pellegrino, J. W. A general measure of organization in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods & Instrumentation*, 1971, 3, 241-246.
- Pellegrino, J. W. A FORTRAN IV program for analyzing higher-order subjective organization units in free recall learning. *Behavior Research Methods & Instrumentation*, 1972, 4, 215-217.
- Robinson, J. A. Category clustering in free recall. *Journal of Psychology*, 1966, 62, 279-285.
- Roenker, D. L., Thompson, C. P., & Brown, S. C. Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 1971, 76, 45-48.
- Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, 69, 344-354.
- Tulving, E. Intratrial and intertrial retention; notes toward a theory of free recall verbal learning. *Psychological Review*, 1964, 71, 219-237.
- Underwood, B. J. Degree of learning and the measurement of retention. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 112-129.

(Received for publication October 23, 1974;
accepted November 2, 1974.)