

BEHAVE—An automated data analysis system for observed events

LAWRENCE J. FITZPATRICK

University of Minnesota, Minneapolis, Minnesota 55455

This paper first defines the context of automated data analysis systems and then mentions approaches to analysis of data by systems of this type. It then briefly describes the BEHAVE data structure and proceeds to a discussion of the BEHAVE data analysis package which provides the user with summary information concerning organization of data.

In recent years computers have assumed an ever increasing share of data analysis chores in the behavioral sciences. The primary component of these analyses involves utilization of available statistical packages to produce probability and clustering statements. Although computerized data collection and analysis accounts for only a small fraction of this use, it should grow rapidly as collection hardware becomes simpler and cheaper and as analysis packages become more sophisticated and readily available. Computerized data collection can be subdivided into two parts: (1) collection and analysis of various data in experimental situations using small general purpose computers, and (2) collection and subsequent processing of events from observational studies. This paper describes an event-processing system which serves the second of these functions.

There have been several approaches to analysis of multivariate behavioral data (Bobbitt, Gourevitch, Miller, & Jensen, 1969; Patterson & Cobb, 1972; Sackett, Stephenson, & Ruppenthal, 1973; Humphreys, Note 1; Rosenfeld, Note 2; Bakeman, Note 3, Note 4; Kaye, Note 5). Individual approaches to analysis vary widely in scope, in type of data read, in techniques employed to "break down" the data and in interpretability of output information. Serious comparison across analysis packages is in itself a major undertaking which could result in standardization of common procedures and reduction in the proliferation of data handling techniques. Some standardization in data structure is beginning to appear around the Datamyte input system (Electro/General Corporation, Note 6) in the work of Bakeman, Kaye, and Sackett.

Support for preparation of this manuscript and for development of the computer programs for data analysis reported herein was provided by Program Project Grant 1 PO HD 05027 from the National Institute of Child Health and Human Development, granted in part to Willard W. Hartup and William R. Charlesworth, Institute of Child Development, University of Minnesota. Computer time for program development was provided by the Human Learning Center, University of Minnesota, and by the University Computer Center, University of Minnesota. Requests for reprints should be sent to Lawrence J. Fitzpatrick, Institute of Child Development, 51 East River Road, University of Minnesota, Minneapolis, Minnesota 55455.

The most comprehensive of these approaches typically utilize large scale computers and take advantage of the extensive operating systems which these computers maintain to manipulate files, and in some cases to edit data and to utilize standard subroutines and statistical packages. Since many analytical approaches utilize different data structures, specific computerized analyses can only be implemented if the user's data already fit or are restructured to fit that specific analysis. It would be valuable at this point to collect the most rigorous and efficient of currently available computer analyses and build data conversion programs to permit more general access to these analyses.

Observable events that can each be on or off at any time are represented in two dimensions in Figure 1. This information can also be represented numerically. The data used by the BEHAVE system in the large scale computer lists the successive groups of simultaneously

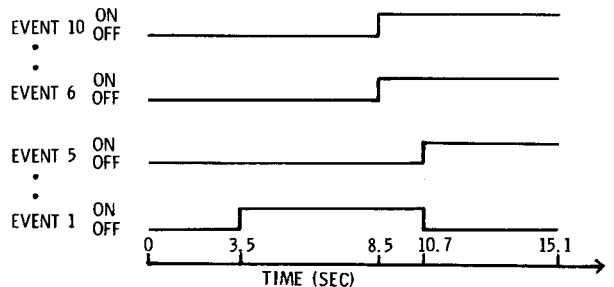


Figure 1. A sample data record represented graphically.

35	0				
50	1	0			
22	1	6	10	0	
44	5	6	10	0	
-1					

Figure 2. A sample data record represented numerically.

occurring events along with the period of occurrence for each group. A sample data record containing the Figure 1 information illustrates this data structure (see Figure 2). For each particular event combination, the time is listed in tenths of seconds immediately before the events. Each event combination is followed by a zero, and the complete data record is followed by a -1. The illustrated record contains 10 events. Reading from top to bottom, the reader can observe that there is a 3.5-sec period of no events followed by 5.0 sec of Event 1. Events 6 and 10 then occur along with Event 1 for 2.2 sec. At this point, Event 1 terminates and Event 5 occurs. Events 5, 6, and 10 continue for 4.4 sec to the end of the record.

The BEHAVE system contains an extensive analysis package (Fitzpatrick, Note 7; Fitzpatrick & Laliberte, Note 8) which is not yet reported in the behavioral literature and will be outlined here. Specific computer programs which make up this package are available from the author upon request. Copyright procedures were used to establish authorship of BEHAVE algorithms and not to restrict access or use of the procedures. Since computer listings provide only the technical language necessary to implement the analysis package on large scale installations, requests should be limited to those who wish to implement the specific algorithms. Formal data analysis is carried out on a large computer using data as described above and computer programs from the BEHAVE analysis package. The package can be divided into three functional groups: a sorting group which allows the user to combine and select events and to select portions of data records to be studied; a counting and summing group which allows the user to make frequency counts of selected events and to compute durations, rates and optional duration histograms of selected events; and an organizational description group which lets the user (1) compute temporal relationships and optional histograms of those relationships between pairs of selected events, (2) perform tests to detect non-random simultaneous occurrence of selected events, and (3) perform tests to detect nonrandom order of selected event onsets. Each group will now be described in more detail.

SORTING GROUP

The sorting group allows the user to specify periods within data records upon which he wishes to focus and to select and combine events within records. These sorting functions permit the user to focus on subranges of time and events which he deems valuable to study in separate units. It is advantageous to minimize the number of events considered at a time since smaller groups of variables require less computer time to analyze and, if chosen wisely, provide simpler and better organized structures among output variables. The time periods and events produced by the sorting group are

used in the remaining analyses. Periods of time can either be directly specified, as "from 0 to 30 sec in Record A," or can refer to recorded events such as "while the mother holds the infant," "for 30 sec before the infant becomes distressed," "from the time when the mother leaves the room until the infant becomes distressed," et cetera. This allows maximum flexibility in specification of time and also uses a form of specification closely related to the way we employ time. Periods can also be combined using the standard logical functions "and," "or (inclusive)," and "not," standard parenthetical notation, and standard rules of precedence for logical operators. This allows the user to specify "while the mother held the infant AND the infant cried," for instance. Events can be combined using the same logical functions and rules that apply to temporal states. A user may wish to combine "wave" and "vocal greet" into the generic event "greet," for instance, or he might wish to study "infant looking AND NOT vocalizing" separately from "infant looking AND vocalizing."

COUNTING AND SUMMING GROUP

Analyses which provide information concerning frequency and duration of behavioral events are available in the BEHAVE analysis package. This type of information has always been important to experimental and observational assessment, and the ability to gather this information in larger quantities using less labor should increase its utility in various assessment applications. The counting and summary information available for each event within the time period specified include its frequency, rate, duration, density of duration, average duration, and standard deviation of duration. Optional histograms of event durations are also available.

ORGANIZATIONAL DESCRIPTION GROUP

The organizational description programs deal with the organizational network which connects specified events in data records. They include a contingency analysis program, a simultaneous event program, and a sequential event analysis program.

The contingency analysis program produces a summary of temporal relationships which exist between specified pairs of events. To perform this summary, the second member of the pair is treated as contingent or dependent upon the first or target member. The summary information is in the form of a table, suitable for categorical analysis, of the contingent event onset and offset frequency and the contingent duration and density during specified periods before, during, and after the target member. Optional histograms of the temporal relationships between contingent and target events are also available.

The analysis of simultaneous event occurrence uses discrete multivariate techniques and for a specified set of

events determines statistically the members of that set which occur together more (or less) frequently than chance and within that interactive subset determines the individual interactive components. It lists the significant interactions, including the likelihood ratio chi square for that term, and the actual and expected tables of cell counts which summarize that interaction test.

The sequential analysis program utilizes the sequential order of event onsets and also uses discrete multivariate techniques and a specified set of events. The analysis first produces a list of the order of occurrence of the specified event onsets. It then uses these onset chains to determine the existing nonrandom order effects and the specific event combinations which contribute most strongly to those order effects. It lists these event combinations along with their expected vs. actual cell counts and a Freeman-Tukey chi square to indicate their contribution to the order effect.

Taken together, this group of three tests can provide a comprehensive view of the organization existing in the user's data records. To illustrate, we will discuss analysis of some hypothetical mother-infant interaction data which might profit from this approach. Mother-infant interactions can be viewed as the organization which exists between members of a set of events containing a subset performed by mother and a subset performed by infant. If we consider "mother look at infant," "mother hold infant," and "infant look at mother" as three separate events, we might use the simultaneous event analysis to find that when mother holds infant she looks at him and he also looks at her longer than expected, or that, when mother is not holding the infant, she looks at him more but he looks at her less than expected. Both of these results would be important and would deal only with overlaps in time of the three selected events. Additional information would appear if we performed a sequential analysis using these same three events. If we found that "mother look at infant" was followed by "mother hold infant" and then by "infant look at mother," we could assume a mother strongly soliciting her infant's attention; if "mother look" was followed by "infant look" and then by "mother hold," we would assume more active infant participation in the process. Both of these outcomes would be compatible with the simultaneous findings and would produce a different overall interpretation of the data.

If we then chose to focus upon "mother hold" and "infant look" using contingency analysis, we might find that infant looks started mostly from 10 to 15 sec before mother began holding and ended during mother hold. This would imply a stronger contingency between "infant look" and "mother hold" than a finding that infant looks started mostly before mother began holding, but that the range of onsets was more variable and more offsets of "infant look" occurred also before "mother hold" began. Thus, with the contingency test we can combine information concerning the occurrence

of one event before another (sequential information) and the occurrence of one event during the occurrence of another (simultaneous information). This allows us to gain some detail about that pair of events which was unavailable in either the sequential or simultaneous tests.

In general, the test of simultaneous event occurrence deals only with overlap in occurrence among events, looking across events at each successive point in time (see Figure 1) without concern for the temporal location of the successive points or the order of occurrence of specific events. Conversely, the test of sequential order of event onsets deals with the temporal ordering of events, taking into account only the temporal location of event onsets and eliminating from consideration any event overlaps which may occur. In contrast, the contingency analysis deals with both sequential and simultaneous effects, merely describing for pairs of events where, in time, one event occurs with respect to the other. The sequential and simultaneous tests can be used to study multiple interactions in clusters of events and to provide separate assessment of structure across and within time, while the contingency analysis can focus upon specific event pairs and summarize in detail the location of one event with respect to the other. A more detailed discussion of all aspects of the BEHAVE system can be found in Fitzpatrick (Note 9).

In conclusion, data analysis systems are in their infancy and researchers need to agree on common data structures in order to encourage commonality and power in data analysis programs. If this can be accomplished, analysis systems of which this is an early example should continue to become simpler, more flexible and wider in scope, and will provide the behavioral researcher with an increasingly powerful tool to help study important relationships which exist between events in the behavior stream.

REFERENCE NOTES

1. Humphreys, P. PRIMATE users manual for Version 4. *Monograph 74/1*. Brunel University, Uxbridge, 1974.
2. Rosenfeld, H. M. *Time series analysis of mother-infant interaction*. Paper presented at the meeting of The Society for Research in Child Development, Philadelphia, 1973.
3. Bakeman, R. *Data editing procedures [technical report 1]*. Atlanta: Georgia State University, 1975.
4. Bakeman, R. *Data analyzing procedures [technical report 2]*. Atlanta: Georgia State University, 1975.
5. Kaye, K. *The CRESCAT software system* (Reference in DATAMYTE Newsletter No. 1, 1976).
6. Electro-General Corporation. *Datamyte Newsletter, Nos. 1 & 2*. Minnetonka, Minnesota, 1976.
7. Fitzpatrick, L. J. Algorithms for extraction of frequency and temporal information in multivariate data records. *U. S. Copyright*, 1976.
8. Fitzpatrick, L. J., & Laliberte, D. M. Algorithms for the extraction of sequential and simultaneous order existing in multivariate sequential data records. *U. S. Copyright*, 1976.
9. Fitzpatrick, L. J. *An approach to complex behavioral analysis*. Submitted for publication, 1977.

REFERENCES

- BOBBITT, R. A., GOUREVITCH, V. P., MILLER, L. E., & JENSEN, E. Dynamics of social interactive behavior: A computerized procedure for analyzing trends, patterns, and sequences. *Psychological Bulletin*, 1969, **71**, 110-121.
- PATTERSON, G. R., & COBB, J. A. Stimulus control for classes of behaviors. In J. F. Knutson (Ed.), *The control of aggression: Implications from basic research*. New York: Aldine, 1972.
- SACKETT, G. P., STEPHENSON, E., & RUPPENTHAL, G. D. Digital data acquisition systems for observing behavior in laboratory and field settings. *Behavior Research Methods & Instrumentation*, 1973, **5**, 344-348.

(Received for publication February 16, 1977;
revision accepted April 21, 1977.)