DOCUMENT RESUME

ED 062 373 TM 001 302

AUTHOR	Friendly, Michael L.
TITLE	Proximity Analysis and the Structure of Organization in Free Recall.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C. Psychological Sciences Div.
REPORT NO	RB-72-3
PUB DATE	Feb 72
NOTE	179p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Cluster Analysis; *Individual Differences; Learning; *Learning Processes; Memory; Psychological Characteristics; Psychological Evaluation; *Psychological Patterns; Psychological Testing; Recognition; *Retention; *Thought Processes; Transfer of Training
IDENTIFIERS	Proximity Analysis

ABSTRACT

A method for assessing the structure of organization was developed on the basis of the ordinal separation, or proximity, between pairs ot items in recall protocols over a series of trials. The proximity measure is based on the assumption, common to all indices of organization, that items which are coded together in subjective memory units will consistently tend to be recalled contiguously in output. Methods of hierarchical cluster analysis are then employed to determine the structure of organization implied by the proximities between items. An experiment was designed to test the sensitivity of the method to differences in organizational structure. Existing data from several studies of part-whole transfer by Ornstein were reanalyzed to assess the explanatory power of the method of proximity analysis. It was concluded that the method of proximity analysis can be useful in attempts to elucidate the relationship between organization and memory. (Author/DG)

RB-72-3

PROXIMITY ANALYSIS AND THE STRUCTURE OF ORGANIZATION IN FREE RECALL

Michael L. Friendly
Princeton University

Prepared in connection with research sponsored by the Personnel and Training Research Programs Psychological Sciences Division, Office of Naval Research under Contract No. N00014-67-A-0151-0006, Contract Authority Identification Number, NR No. 150-302 and National Science Foundation Grant GB 8023X with Princeton University, Princeton, New Jersey Project on "Mathematical Techniques in Psychology" Principal Investigator: HAROLD GULLIKSEN

This work made use of computer facilities supported by National Science Foundation grants NSF-GJ-34 and NSF-GU-3157

Educational Testing Service
Princeton, New Jersey
February 1972

RB-72-3

PROXIMITY ANALYSIS AND THE STRUCTURE OF ORGANIZATION IN FREE RECALL

Michael L. Friendly
Princeton University

Prepared in connection with research sponsored by the Personnel and Training Research Programs Psychological Sciences Division, Office of Naval Research under Contract No. N00014-67-A-0151-0006, Contract Authority Identification Number, NR No. 150-302 and National Science Foundation Grant GB 8023X with Princeton University, Princeton, New Jersey Project on "Mathematical Techniques in Psychology" Principal Investigator: HAROLD GULLIKSEN

This work made use of computer facilities supported by National Science Foundation grants NSF-GJ-34 and NSF-GU-3157

Educational Testing Service
Princeton, New Jersey
February 1972

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Princeton University, Princeton, New Jersey (Psychology Department) Educational Testing Service, Princeton, New Jersey	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

3. REPORT TITLE

PROXIMITY ANALYSIS AND THE STRUCTURE OF ORGANIZATION IN FREE RECALL

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Michael L. Friendly

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February 1972	162	101

8a. CONTRACT OR GRANT NO. ONR Contract No. N00014-67-A-0151-0006	9a. ORIGINATOR'S REPORT NUMBER(S) RB-72-3
b. PROJECT NO. NR No. 150-302	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217

13. ABSTRACT

One of the central questions in the study of free recall learning concerns the role of organizational factors in retrieving information from memory. This work has been greatly facilitated by the development of procedures for measuring the amount of organization evidenced in recall. At a conceptual level, such measures may be thought of as indexing the formation of informationally-rich higher order memory units which serve as multiple access routes to the list items they subtend. Thus, a given list item may be retrieved either on its own merits or through prior retrieval of the subjective memory unit which includes it. There has been, however, no way to determine the actual manner of organization employed by individual subjects. Such a procedure would seem necessary in order to test directly hypotheses concerning the way in which organization influences performance and retention.

A method for assessing the structure of organization was developed on the basis of the ordinal separation, or proximity, between pairs of items in recall protocols over a series of trials. The proximity measure is based on the assumption, common to all indices of organization, that items which are coded together in subjective memory units will consistently tend to be recalled contiguously in output. Methods of hierarchical cluster analysis are then employed to determine the structure of organization implied by the proximities between items.

(See Continuation Sheet)

DD FORM 1473 (PAGE 1)
1 NOV 65

S/N 0101-807-6801

Security Classification

3ND PPSO 13152

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Free recall learning						
Memory						
Human memory						
Semantic memory						
Free recall memory						
Organization of memory						
Long term retention						
Part-whole transfer						
Subjective organization						
Hierarchical organization						
Individual differences						
Proximity analysis						
Cluster analysis						
Psychological scaling						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

Security Classification

DD Form 1473

13. Abstract (Continued)

An experiment was designed to test the sensitivity of the method to differences in organizational structure. All subjects learned a list consisting of words selected from hierarchically related taxonomic categories, and which could be organized in alternative ways. Three experimental groups were influenced to adopt the alternative organizations by using different blocked presentation orders of the items. Twelve acquisition trials were given and long-term retention was tested after either 1, 5, 10, or 20 days. All experimental groups receiving categorically blocked presentation recalled and retained more words than a random input-order control group. However, the experimental groups did not differ among themselves in recall during acquisition or retention. The proximity analyses produced results which were consistent with the predetermined patterns of organization and indicated that the different organizations of the list were maintained in the retention test.

Existing data from several studies of part-whole transfer by Ornstein were reanalyzed to assess the explanatory power of the method of proximity analysis. These studies had delineated some conditions under which prior learning of part of a list would facilitate or hinder subsequent learning of the whole list. One study demonstrated that random presentation of the whole list produced negative transfer, but that whole-list learning was facilitated by blocking the presentation order of the final list according to the "old" and "new" subsets of items. Applying proximity analysis to these data, it was found that the higher-order subjective units identified from the first-list protocols carried over to second list learning only for those subjects who had received blocked presentation of the final list. These results directly verified predictions which had been made from a theory of subjective organization (Tulving). It was concluded that the method of proximity analysis can be useful in attempts to elucidate the relationship between organization and memory.

PROXIMITY ANALYSIS AND THE STRUCTURE OF

ORGANIZATION IN FREE RECALL

by

Michael L. Friendly

A DISSERTATION

PRESENTED TO THE

FACULTY OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY THE

DEPARTMENT OF PSYCHOLOGY

October, 1971

PROXIMITY ANALYSIS AND THE STRUCTURE OF ORGANIZATION IN FREE RECALL

Michael L. Friendly

Abstract

One of the central questions in the study of free recall learning concerns the role of organizational factors in retrieving information from memory. This work has been greatly facilitated by the development of procedures for measuring the amount of organization evidenced in recall. At a conceptual level, such measures may be thought of as indexing the formation of informationally-rich higher order memory units which serve as multiple access routes to the list items they subtend. Thus, a given list item may be retrieved either on its own merits or through prior retrieval of the subjective memory unit which includes it. There has been, however, no way to determine the actual manner of organization employed by individual subjects. Such a procedure would seem necessary in order to test directly hypotheses concerning the way in which organization influences performance and retention.

A method for assessing the structure of organization was developed on the basis of the ordinal separation, or proximity, between pairs of items in recall protocols over a series of trials. The proximity measure is based on the assumption, common to all indices of organization, that items which are coded together in subjective memory units will consistently tend to be recalled contiguously in output. Methods of hierarchical cluster analysis are then employed to determine the structure of organization implied by the proximities between items.

An experiment was designed to test the sensitivity of the method to differences in organizational structure. All subjects learned a list consisting of words selected from hierarchically related taxonomic categories, and which could be organized in alternative ways. Three experimental groups were influenced to adopt the alternative organizations by using different blocked presentation orders of the items. Twelve acquisition trials were given and long-term retention was tested after either 1, 5, 10, or 20 days. All experimental groups receiving categorically blocked presentation recalled and retained more words than a random input-order control group. However, the experimental groups did not differ among themselves in recall during acquisition or retention. The proximity analyses produced results which were consistent with the predetermined patterns of organization and indicated that the different organizations of the list were maintained in the retention test.

Existing data from several studies of part-whole transfer by Ornstein (1970) were reanalyzed to assess the explanatory power of the method of proximity analysis. These studies had delineated some conditions under which prior learning of part of a list would facilitate or hinder subsequent learning of the whole list. One study demonstrated that random presentation of the whole list produced negative transfer, but that whole-list learning was facilitated by blocking the presentation order of the final list according to the "old" and "new" subsets of items. Applying proximity analysis to these data, it was found that the higher-order subjective units identified from the first-list protocols carried over to second list learning only for those subjects who had received blocked presentation of the final list. These results directly verified predictions which had been made from a theory of subjective organization (Tulving, 1966). It was concluded that the method of proximity analysis can be useful in attempts to elucidate the relationship between organization and memory.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the support and guidance received during all phases of this research. I am indebted most of all to the members of my thesis committee: Professor Harold Gulliksen, Professor Sam Glucksberg, and Professor Peter Ornstein. I am grateful to them collectively for helpful suggestions and comments at various stages and individually for their special roles in my training.

I am also grateful to Mr. Anthony Harris, Mr. Charles Liberty and Mr. Richard Murphy for their skillful assistance in the collection of data.

Many members of the Division of Psychological Studies at Educational Testing Service provided diligent and invaluable assistance. Mrs. Sara Matlack helped in ways too numerous to list but truly appreciated. Miss Henrietta Gallagher and her staff provided painstaking effort in much of the initial data tabulation and scoring. Mrs. Ann King supervised the preparation of the final manuscript. I also appreciate the splendid efforts of Miss Kathy Boyle, Mrs. Betty Clausen, Mrs. Laura Lenz and Mrs. Betty Springsteen for careful typing and Mrs. Marie Davis and Mrs. Frances Shaffer for proofreading the manuscript.

Much of this research was carried out as an Educational Testing Service Psychometric Fellow and later while I held a predoctoral fellowship from the National Science Foundation. The research was supported by ETS; by Office of Naval Research contract N000 14-67-A-0151-0006, Contract Authority Identification Number, NR-150-302, Princeton University; and by the National Science Foundation, Grant GB 8023X, administered through Princeton University, Project on "Mathematical Techniques in Psychology," Harold Gulliksen, Principal Investigator. Some equipment support was also received from National Institute of Mental Health, Grant NH-17620-01, Peter A. Ornstein, Principal Investigator, and from Public Health Service, Grant HD-01910, Sam Glucksberg, Principal Investigator. This work made considerable use of computer facilities supported in part by National Science Foundation, Grant NSF-GJ-34 and NSF-GU-3157.

PROXIMITY ANALYSIS AND THE STRUCTURE OF

ORGANIZATION IN FREE RECALL

Table of Contents

10

List of Tables

List of Figures

Figures		Page

CHAPTER 1

INTRODUCTION

The research reported here is broadly concerned with the role of organization in free recall learning (FRL). In the typical FRL experiment a list of items, usually familiar English words, is presented to the subject for study. He is then asked to reproduce from memory as many of the items as he can, in any order. Hence a subject's performance in this task may be considered as analogous to the operation of memory for verbal materials in natural situations, such as remembering the contents of a shopping list.

But more important than its similarity to real-life situations is the fact that the FRL experiment provides a vehicle for studying the role of structure or organization in memory. In allowing the subject to recall the items in any convenient order, the task imposes minimal restrictions on the possible strategies which may be used.

The experimental method of free recall has long been employed in psychology (Tulving, 1968). However, its significance for investigating organizational processes in memory was not fully realized until the appearance of Bousfield's (1953) classic paper on clustering in free recall. "If clustering can be quantified," Bousfield stated, "we are provided with a means for obtaining additional information on the nature of organization as it operates in the higher mental processes" (Bousfield, 1953, p. 229). Bousfield realized that the systematic discrepancies between input order and the output sequence of recalled responses, which would be regarded as errors in serial learning, provided important information about the operation of memory in free recall.

Experimental studies since then have shown that one basic phenomenon displayed by subjects in FRL is the grouping of items into recall units. That is, over a series of study-test trials with input order varied randomly, subjects (Ss) will tend to form increasingly consistent item groupings in their recall output.

1.1 Objective and Subjective Organization

The tendency to form stable recall groupings has been taken as a behavioral manifestation of organizational processes. Investigations in this area may be divided into two broad classes, distinguished by the nature of the to-be-remembered material.

The first class of studies, concerned with objective organization or clustering, has employed lists composed of two or more nonoverlapping subsets of items. In this paradigm, clustering is measured in terms of the observed tendency for items from the same subset to be recalled in immediately adjacent output positions. The subsets are defined by the experimenter in terms of membership in conceptual categories, as in the study by Bousfield (1953), or according to associative or other meaningful relations among the items.

Because the putative source of organization can be specified and manipulated by the experimenter (E), it has been possible to investigate the effects on clustering of a large number of stimulus variables and presentation conditions (see Shuell, 1969, for a recent review). The details of the measurement procedures for clustering will be taken up in a later section (1.2). However, it should be noted here that standard clustering measures will underestimate S's amount of organization to the extent that his grouping of the items diverges from that selected by the experimenter.

In the second class of free recall studies, the basis of organization is not predetermined by the experimenter. Here, the stimulus list is composed of "unrelated" words, chosen either without regard for inter-item relatedness or with a definite attempt to minimize such relatedness. This paradigm attempts to tap the development of organization based on the personal (but possibly shared) verbal dispositions with which the Ss enter the laboratory. Because the sources of this subjective organization (SO) are not imposed by E and may vary from one subject to the next, its measure must be sought in internal analyses of the consistency in output order over trials. Thus SO is determined by the degree to which S recalls the same sequences of words together on successive trials.

The concept and measurement of SO were developed by Tulving (1962a), who demonstrated that allegedly unrelated words were in fact organized in the course of FRL. Tulving also reported that both the degree of subjective organization and its communality across subjects increased over a series of trials.

The combined results obtained in these two paradigms point to organization as a central and pervasive factor in free recall learning. For example, if some readily apparent basis for grouping the items into cohesive subsets has been imposed on the materials by E, subjects will use this structure in their recall. As the salience of an E-defined organization decreases, apparent clustering will also decrease (Bousfield, Cohen, & Whitmarsh, 1958; Marshall, 1967). But this does not mean that Ss cease to organize their recall. In the limiting case, when "unrelated" words are to be learned, Ss nevertheless find common dimensions for relating item groups. The available evidence suggests, as Bower (1970) and Postman (1963) have indicated, that there is probably no such thing as a truly unrelated list of words: "with the adult's vast capabilities for searching out similarities and dissimilarities, almost any collection of 'unrelated' words can be partitioned into subsets within which items share a number of features" (Bower, 1970, p. 32). Therefore, it does not seem unreasonable at this time to assume the hypothesis that category clustering and subjective organization reflect the same basic processes.

1.2 Measurement of the Amount of Organization

The ability to quantify organization as a dependent variable in FRL has been onc reason for the interest in this paradigm and the analytical power of the theory it has generated.

The methods proposed to date for measuring the amount of organization fall into two distinct classes, corresponding to the two types of word lists which have generally been employed in FR experiments, viz., those based on E-defined groupings such as taxonomic categories or associative relations, and those based on "unrelated" lists. Only the general features of these methods will be considered here, since several recent reviews are available (Roenker, Thompson, & Brown, 1971; Shuell, 1969).

Measurement of categorical organization. When the to-be-remembered list can be partitioned a priori by E into mutually exclusive and exhaustive groups of items on semantic, associative, structural grounds, etc., it becomes interesting to determine the extent to which subjects actually use such groups in their recall. Items belonging to the same class are treated as indistinguishable and, following Bousfield and Bousfield (1966), clustering measures focus solely on the order of succession of the classes to which the items in any recall protocol belong (order properties). For example, of the two protocols,

(1a) horse, cow, dog, tie, gin, beer, socks, shirt

(1b) pen, hat, tip, field, teacher, herring, river, apple

the first may represent a subject's recall of a list composed of the taxonomic categories animals, beverages, and articles of clothing, while the classes in (1b) could be based on the structural property of word length--3, 5, or 7 letters. However, the information regarding the ordering of items from the various classes, and (by assumption) the organization reflected in the protocols, would be the same for (1a) and (1b) and can be represented by

(1c) A, A, A, B, C, C, B, B

that is, the first three items belong to the same class in (1a)--animals, as do the first three in (1b)--3-letter words, and so forth.

The fundamental assumption in all investigations and measures of organization in FR is that items which are stored/retrieved together should appear contiguously in the subject's output protocols. It is actually the converse of this assumption which is used to assess organization; that is, one assumes that contiguity in recall implies organization in storage or retrieval. In particular, for lists of items based on E-selected relationships, the goal in measuring organization, as implied above, is to determine the extent to which the grouping of items in Ss' protocols reflects the same grouping set up by the experimenter. That is, the categories or relations built into the list by E serve as the single standard against which the observed order of a subject's responses is compared to assess "how much" he is organizing.

From these assumptions, it has been natural and convenient to take as a unit of measurement for categorical clustering the category repetition, i.e., the occurrence in recall of an item from one class or category immediately following another item from the same class. Thus, (1c) above contains four repetitions, as indicated by the underscored items. Another possible measure is the number of runs in the series, where a run is defined as a maximal sequence of items of like class. Counting the number of runs in any series such as (1c) is equivalent to counting the nonunderscored items. Therefore, the number of runs (R) and number of category repetitions (C) give equivalent information about the occurrence of clustering, and are related as $C = n - R$, where n is the number of items recalled. In practice, measures of the degree of categorical organization are standardized so as to make the values obtained under varying conditions commensurable. For example, the observed number of repetitions in recall may be compared with what one would expect if the output order was determined by chance alone (Bousfield & Bousfield, 1966), or C may be divided by some maximum value (Bousfield, 1953; Bower, Lesgold, & Tieman, 1969), or both (Roenker et al., 1971).

Measurement of subjective organization. Estimation of the degree of organization appearing in the recall of "unrelated" lists again follows from the fundamental assumption of organization in FR, so that the tendency of S to recall the same items in contiguous groups over successive trials is taken as evidence for the existence of subject-imposed organization.

Because of the more open-ended nature of organization when its basis cannot be considered known, there has been less agreement on how it should be measured, even at a conceptual level. Unlike the case with categorized lists, where the grouping of items by E-defined relationships serves as an external standard, subjective organization must be estimated by a criterion of consistency internal to a set of free recall protocols.

Tulving (1962a) proposed that SO could be indexed by the degree of sequential redundancy, in information-theoretic terms, in the order of recall over a series of trials, relative to the maximum possible redundancy which would be observed if the S recalled the same items in a constant order on every trial. That is, SO measures the average degree to which a subject's i-th response can be predicted on a particular trial, given only the item recalled in the (i-1)-st position.

Subjective organization can also be assessed by letting each trial serve in turn as the standard for comparison with the order in which items were recalled on the immediately preceding trial (Bousfield & Bousfield, 1966), or by choosing the output order of one trial, for example the last, as the standard against which all other trials are compared (Ehrlich, 1965, 1970). The unit of subjective organization in Bousfield's measure is the intertrial repetition (ITR). An ITR is scored whenever an adjacent pair in the output of trial t also occurs contiguously in the same order on trial t+1. Ehrlich's measure, termed a coefficient of structuration, is essentially a correlation between the intraserial separation (i.e., number of other items intervening) between pairs of items on the final, criterion trial, and the separation of these pairs of items in output on each earlier trial.

Both Tulving's SO and ITR measure proposed by Bousfield reflect only the consistency in recall of immediately adjacent ordered pairs of items, and have been criticized for this reason. Several modifications have been proposed (e.g., Fagan, 1968), but by far the most ambitious revision of ITR to overcome this limitation has been worked out by Pellegrino (1971). Pellegrino has extended the definition and measurement of intertrial repetitions to include unordered item sequences of any specified size. That is, his procedure allows the examination of recall for output consistency in terms of groups of size 2, 3, 4, etc., and for any unit size, all possible orders are scored. This extension, therefore, provides for a more complete assessment of organization than is afforded by the ITR and SO measures.

1.3 Organization and Recall

The occurrence of clustering and subjective organization would be of slight interest, of course, if it were unrelated to the amount of recall or merely a by-product of practice. In his 1962 paper Tulving (1962a) demonstrated a strong correlation between SO and amount of recall. Subsequent experiments, showing that direct manipulations of organization produce predictable effects on recall, have supported the view of free recall memory as highly dependent on the development of stable organizational units.

Tulving (1962b) established that instructions to use an alphabetical organization in remembering unrelated words (which all had unique initial letters) produced a large and sustained facilitation of recall relative to control Ss instructed only to recall as best they could. An experiment by Mandler (1967a) further revealed that instructions to sort words into consistent subject-defined categories on the basis of meaning had the same facilitative effect on subsequent recall as instructions to remember the words. Ss given the subjective categorization task recalled as well whether or not they expected to be tested subsequently for recall. On the other hand, the recall performance of Ss who had sorted by rote, without regard to meaning, was only high when they had been explicitly instructed to remember. Experiments by Tulving (1966) have extended this latter result by demonstrating that rote repetition alone (without intent to recall) is insufficient to produce high recall when the same items are subsequently tested in multitrial FR.

Further, if trial-to-trial increments in recall are a direct consequence of the development of organizational groupings, then the rate of FRL should be retarded by inhibiting organization or inducing inappropriate grouping. The prediction of the effect of inhibiting organization was confirmed by Bower et al. (1969). They found that recall suffers when Ss are forced to change their groupings of unrelated words on every trial. Taken together these studies suggest that the formation of an appropriate organization may be both necessary and sufficient for efficient memorization to take place.

The theoretical significance of these observations stem from the fact that they allow a relatively parsimonious account of memory processes and the effects of repetition. The consistency of output order observed in recall tests has been regarded as evidence for the development of higher-order memory units, each composed of two or more list items. While the experimenter may conceive of the list in terms of L nominal units (E-units), the subject's organizational grouping may provide him with an effective list of less than L functional, higher-order units (Tulving, 1968). Since the actual higher-order units which develop are in general determined by S's own preexperimental verbal dispositions (regardless of whether the list is categorized or unrelated), the higher-order groups are termed subjective units, or S-units. The functional utility of S-units to the learner lies in the inherently limited capacity of human memory to store and retrieve information. If on any trial S can recall only a fixed number of subjective units, then increments in recall with practice must reflect the increased size and stability of these units.

In the original formulations of subjective organization theory (Tulving, 1962a, 1964), based on Miller's (1956) concept of chunking, organization was viewed as a process affecting the storage of material: "organizing processes . . . lead to an apparent increase in [storage] capacity by increasing the information load of individual units" (Tulving, 1962a, p. 344). In more recent expositions (Bower, Clark, Lesgold, & Winzenz, 1969; Slamecka, 1968; Tulving & Patterson, 1968) emphasis has shifted to the importance of organizational processes in retrieval, with S-units viewed as multiple routes by which access to stored traces may be achieved. At the present time, however, it is difficult to distinguish clearly between storage and retrieval effects, except in circumstances where one or the other can be isolated (e.g., in cuing studies, Tulving & Pearlstone, 1966).

1.4 The Present Research

As indicated above, much of the current research in FR is based on the notion that, in recalling the items from a particular list, S is not only telling the experimenter about the capacity of his memory, but also about the structure or arrangement of the items within his memory. Information about capacity is presumably contained in the number of items recalled while information about structure is usually derived from the order in which the items are recalled.

However, since the measurement procedures presently available for indexing organization are entirely concerned with the amount of organization rather than with its explicit structure, only indirect tests of organization theory have been possible. This methodological limitation has become more critical as mounting (and often conflicting) empirical observations have created an increased need for more clearly articulated theories. Recent statements by Mandler (1967a), Postman (1971), and Tulving (1968) have stressed the importance of focusing attention on the manner or pattern of subjective organization: "In order to evaluate fully the relation between type of subjective organization and recall, it is desirable to make the entire structure generated by the learner accessible to inspection" (Postman, 1971, p. 16).

The present investigation is concerned with the development and evaluation of one such method based on interitem proximities for determining how subjects are organizing lists of verbal items. This method subsumes the measurement of categorical clustering and subjective organization within a single unified framework in that it assumes no prior knowledge by the experimenter of the bases of organization. To the contrary, it offers an objective way to determine these bases and therefore provides a means of directly testing components of theories of memory which treat the subject as an active processor of mnemonic information.

The remainder of this report is divided into three major sections. The first section (Chapter 2) describes the method of proximity analysis and illustrates its use with sample data. Chapter 3 presents an experiment concerned with long-term retention of a hierarchically organized list, designed in part to test the validity of the proposed technique. In the final section (Chapter 4), available data from several studies of part-whole transfer are reanalyzed according to the present method to demonstrate the utility of assessing the structure of organization.

CHAPTER 2

PROXIMITY ANALYSIS

2.1 Limitations of Measuring the Amount of Organization

Many investigators recognize that present measures of the degree of organization have their limitations and that it is important to develop more adequate ones. It is worthwhile to consider some of these limitations before considering how the structure of organized recall may be assessed.

It should be noted, first of all, that the data collected in free recall experiments have many degrees of freedom. Recall protocols differ in complex aspects of the sequential patterning of the items recalled both within and across trials (cf. Tulving, 1964). In this connection, some comments by Cronbach (1955) concerning measurement in a different context may be applied to free recall. Whenever we describe the organization of recall data in a single, quantitative index, "we compress all the aspects of this variation into a single degree of freedom, and we must be careful that valuable information is not discarded or cancelled out" (Cronbach, 1955, p. 16).

Many theories of long-term memory make fairly explicit statements about the structural relations among units in the memory store. If we use only measures of the degree of categorical and subjective organization which compress all information about the structure of recall into a single index, there is no way to investigate these theories directly with free recall data.

Some examples may help to make this clear. Limited capacity theories hold that the memory system can store (Miller, 1956; Tulving, 1962a) or retrieve (Tulving, 1966, 1970) only a constant, limited number of memory units. Through repetition and rehearsal, it is supposed, Ss are able to pack greater numbers of nominally separate items (E-units) into each of the limited number of functional or subjective memory units (S-units). Current indices, however, provide no clear way of confronting such a statement with experimental data. In order to evaluate this theory directly, it would be necessary (a) to determine what the contents of the S-units are on each of a number of free recall trials and (b) to demonstrate that the "learning curve" of number of S-units recalled is a line with zero slope, while the corresponding function in terms of E-units is of the classical, negatively accelerated shape. Since researchers place great emphasis on models such as these in deriving predictions for experimental studies, it seems important to find ways of uncovering empirically the manner or structure of organization used by Ss in free recall tasks.

Another example, which will be taken up in detail in Chapter 4, concerns recent studies of transfer in FRL. Tulving (1966) showed that prior learning of part of a list of unrelated words produced negative transfer when the whole list was subsequently learned. Assuming the existence of an optimal organization of the whole list, interference would be predicted if part-list higher-order units persisted into the test stage, and Tulving explained his results on this basis. Although the expected consequences of this hypothesis have been confirmed in several recent studies (e.g., Bower & Lesgold, 1969; Ornstein, 1970), the persistence of inappropriate S-units has not been explicitly demonstrated. "In order to evaluate Tulving's position, we should have some documentation of just what the S-units are like at the end of (part-) list learning, and what they are like at various stages during (whole-) list learning" (Ornstein, 1968, p. 9).

A sole concern with measures of the amount of organization also creates problems for the interpretation of data. These are problems of a logical nature, concerning the validity and depth of inferences which may be drawn from these measures.

Two points need to be considered. First, it is not at all clear to what extent currently used measures actually index the development of S-units. If an S-unit consists of a network of interitem dependencies, then the number of different, organizationally equivalent orders in which the items may be recalled increases rapidly with the size of the unit. In fact, if an S-unit composed of N items were completely interconnected, the items could be recalled in N! different orders, all consistent with perfect organization in this sense. These sequences would, on the average, have relatively few repeated ordered pairs in common, yet the ITR and SO measures are typically restricted to such pairwise constancies.[1] What these sequences do have in common is that all members of an S-unit appear in close proximity. This theme will be developed in detail below.

The second interpretative difficulty is that strong inferences regarding the pattern of organization cannot, in most cases, be conclusively drawn from measures of the amount of organization even if infallible indices were available. For example, categorized lists are usually derived from norms collected from a large number of subjects, and thus reflect associative relationships common to the population from which these subjects were

---

[1]Pellegrino (1971) has recently presented a generalized ITR measure which counts all possible orders of a set of items and therefore overcomes the basis of this objection.

drawn. While it is true that such materials exploit the high communalities among subjects, it is impossible to determine if subjects are using bases other than those specified by E to organize their recall. In one type of study, different manipulations, instructions, etc., may be applied to experimental groups and the results examined in terms of curves showing average categorical organization over trials. When low levels of correspondence between E-categories and Ss' output orders are observed in studies of this sort, it is commonly concluded that subjects are not organizing, or that some variable designed to manipulate organization has been successful (unsuccessful) in inhibiting (facilitating) this process. In general, where strong clustering in terms of the experimenter's structuring of the list is not found, we do not know whether the items were difficult for the Ss to organize or whether the Ss were merely organizing in some manner that the experimenter had not considered. Alternatively, two groups of Ss may show the same numerical amount of sequential organization in their recall but may be performing qualitatively different operations on the input materials. Without an objective way to determine how subjects are organizing, the conclusions drawn from such data may be quite inappropriate. Mandler (1967a) and Postman (1971) have voiced similar cautions regarding the interpretation of degrees of E-defined organization when no independent checks are available.

2.2 A Method for Investigating the Structure of Organization

In general, functional memory units may be assumed to vary in strength as do single items. For example, instances of taxonomic categories with high normative frequency show greater clustering than do low frequency instances (Bousfield et al., 1958). It would be useful, therefore, if a method for identifying S-units were also to index the relative strengths of such units within a list. Also such a method should be applicable to data from individuals as well as to group data. As the strength of E-determined organization increases, idiosyncratic groupings and individual differences tend to decrease. Yet it is still important to determine whether substantial individual differences exist, or whether there are several homogeneous groups of Ss using disparate organizational strategies.

It is useful to proceed heuristically at first to develop the logic of the technique to be proposed. Following that, the crux of the method is presented formally (2.2.2) and then illustrated with sample data.

2.2.1 Rationale for Proximity Analysis

Consider a hypothetical subject presented with a categorized word list who recalls the following items on a given trial:

PANTS, SHIRT, SHOE, DOCTOR, SHRUB, BUSH, TREE, LAWYER, DENTIST

in that order. Counting the number of sequential repetitions of items from the same category, we find that there are five category repetitions in the above protocol.

This way of looking at contiguity in output as evidence for grouping in memory only considers pairs of items which are immediately adjacent. But if an S-unit consists of more than two items, all pairs of them cannot be immediately adjacent, and the degree of organization is probably underestimated. So, as a first step toward identifying the subjective units of recall, the rationale behind examining category repetitions can be generalized to allow for varying degrees of contiguity between items. Thus, PANTS and SHIRT, for example, are maximally close while PANTS and SHOE are less proximal, and so on for the other categories. The assumption made here is that S-unit "belongingness" is a graded property of groups of items, and that protocol separation beyond immediate adjacency also carries information about the relative strengths of S-units.

Going a step further, it is possible to look at the proximities between all pairs of words in the protocols, not just those within the given categories. For example, BUSH and TREE are more proximal than are SHOE and TREE, though the reverse could have occurred if the subject had thought of the compound noun, SHOETREE, and clustered on that basis. The actual outcome can be expressed quantitatively by giving the pair BUSH and TREE a higher proximity score for that trial than the pair SHOE and TREE, and so on for all pairs of items, basing the proximity score on their ordinal separation in the protocol. By combining proximity scores over blocks of trials, an item-by-item proximity matrix can be constructed with numerical entries representing the degree to which each pair tends to occur in contiguous output positions over the block of trials.

The modest step of considering the proximities between all pairs of items makes this way of looking at the subject's organization of a list independent of any knowledge of "best" or a priori categories. The use of the number of repetitions as an index of organization requires, by definition, a knowledge of which groups of items belong together. Through the use of proximities, however, it is possible to "discover" the grouping that the subject is in fact using, by defining the subjective units to be those groups of items that have mutually high interitem proximities.

Stated alternatively, we are asking what manner of grouping of the stimulus list into S-units would be most likely to result in the observed response protocols produced by individual Ss. In the analysis suggested here, the aspects of order information most relevant to the study of S-units may be represented by the proximities between all pairs of items. Questions concerning the organization of list items in memory can therefore be reduced to corresponding questions concerning the structure of proximities between items in recall. Thus, if items are organized into higher-order memory units which are recalled in contiguous groups, these S-units can be inferred by working backwards from the proximities. A by-product of the particular technique used for analyzing the proximities permits the assignment of relative strengths to the S-units so determined.

There is actually no logical necessity to invoke the notion of intraserial proximity in order to describe the contents of S-units.[2] The proximities are the middle men. They represent a construction--a device by which it is possible to bridge the gap between observed FR responses and a description of organization.

This discussion is not to imply a conception of S-units as fixed entities. Rather, it is hoped that this approach will yield a reasonably well-focused snapshot of organization as it develops over some block of trials.

2.2.2 Measure of Interitem Proximity

It remains to specify a way to quantify this notion of proximity, or its inverse, distance. One way to do this is to measure the distance between

---

[2]Discussion with John Hartigan has helped to clarify this and other points and is gratefully acknowledged here.

two items in terms of the number of other items which separate them in recall. Consider a list of N items presented to a group of S subjects for each of T trials under typical multitrial free recall (MFR) conditions. For a given subject, s, the data shall consist of T sequences of items, each of length r_{st}, where r_{st} is the number of words recalled by subject s on trial t. Where confusion is unlikely to arise, the subscript s is omitted in what follows. For ease of exposition, the simple (though unlikely) case where subjects recall only items from the list, and do not repeat responses, is considered initially. The problem of handling intrusions and repetitions is discussed in Appendix A.

Denote by ℓ_{it} the position of item i in the subject's output on trial t. Then the intraserial distance between two items, i and j, both recalled on a given trial will be $|\ell_{it} - \ell_{jt}|$. The absolute value of the difference is used since in most cases it seems sensible to consider the recall of items A,B equivalent to recall of B,A.

When both members of a pair of words are not recalled on a given trial, it is difficult to decide how a distance may be rationally assigned. A value could be assigned ad hoc, but it is probably better to assume that this event gives no information regarding the organization of that pair. It is necessary, therefore, to take varying degrees of item- and pair-recall into account.

Define a characteristic variable, ϕ, which shall be used to indicate the recall of particular items on given trials.

$$\phi_{it} = \begin{cases} 1, \text{ if word } i \text{ is recalled on trial } t \\ 0, \text{ otherwise} \end{cases} \quad (2.1)$$

$i = 1,\ldots,N$; $t = 1,\ldots,T$. Then the occurrence of pairs of items on particular trials may be expressed as

$$\phi_{ijt} = \phi_{it} \cdot \phi_{jt} = \begin{cases} 1, & \text{if words } i \text{ and } j \text{ are both recalled on trial } t \\ 0, & \text{otherwise .} \end{cases} \qquad (2.2)$$

That is, $\phi_{ijt} = 1$ if and only if both ϕ_{it} and ϕ_{jt} equal 1 . Since it is proximity rather than intraserial distance that is directly related to the tightness of organization, the positional difference measure can be "turned around" by subtracting it from a positive constant, so that large numbers represent more proximal items. The case where one or both members of the pair are not recalled is included by defining the proximity on trial t as

$$P^{*}_{ijt} = \phi_{ijt} \left[N - |\ell_{it} - \ell_{jt}| \right] ,$$

which is equal to zero when the pair is not recalled and when $i = j$. Considering all T trials (or only some block of them if we choose), an overall measure of proximity for items i and j is

$$P^{*}_{ij} = \sum_{t=1}^{T} P^{*}_{ijt} = \sum_{t=1}^{T} \phi_{ijt} \left[N - |\ell_{it} - \ell_{jt}| \right] , \qquad (2.3)$$

which will be termed the raw proximity between items i and j .

One problem with the P^* measure above is that it is not standardized with respect to the number of times that a pair is recalled. Consider the raw proximities for two pairs of items (W,X) and (Y,Z), recalled from an eight-item list (for which the maximum proximity value is 7) on a series of eight trials.

Trial

	1	2	3	4	5	6	7	8	Total = P^*
(W,X)	7	7	0	6	0	7	0	7	34
(Y,Z)	3	6	5	4	4	5	5	5	37

Thus, while (W,X) occur in immediately adjacent positions ($P^*_{WX} = 7$) on all but one of the trials on which they are both recalled, their raw proximity score for the eight trials is lower than (Y,Z) which are both recalled on all trials, but are never more proximal than (W,X). From this anomaly, it is seen that P^*_{ij} defined in Eq. (2.3) above is at least partially a measure of pair-recall, or performance. Since the proximities should not reflect recall performance per se, it is necessary to adjust for differences in recall frequencies among pairs of words. This may be done by dividing each P^*_{ij} by the number of trials, say n_{ij}, on which both members of the pair are recalled. Accordingly, define

$$P_{ij} = \frac{P^*_{ij}}{n_{ij}} = \frac{\sum_t \phi_{ijt}\left\{N - |\ell_{it} - \ell_{jt}|\right\}}{\sum_t \phi_{ijt}}$$

$$= N - \frac{\sum_t \phi_{ijt}\,|\ell_{it} - \ell_{jt}|}{\sum_t \phi_{ijt}} \quad . \qquad (2.4)$$

The proximity measure adopted is therefore the average proximity for the pair, over only those trials on which both members of the pair are recalled.[3] For the example above, this gives $P_{WX} = 34/5 = 6.8$, and $P_{YZ} = 37/8 = 4.5$, which agree more closely with intuition. Recognizing the second term on the right of Eq. (2.4) as the (average) intraserial distance, D_{ij}, gives

$$P_{ij} = N - D_{ij} \quad . \tag{2.5}$$

2.3 Illustrative Data

To make things more concrete, consider the data in Figure 1. Shown at the top of the figure are the protocols frm one subject on the last six trials of an eight-trial free recall session.[4] On each trial 12 unrelated words were presented visually in a different random order, and the subject's task was to recall as many words as possible.

Consider Trial 5. Items which are immediately adjacent, such as (HIGHWAY, STRUCTURE), and (INVENTOR, PROFESSOR), differ in ordinal position by one, so their proximity on that trial is N-1 or 11. On the other hand, words widely separated in the protocol have a lower proximity on that trial; for example, MAST and ASSAULT, which are 5 positions apart, have a proximity of 7.

---

[3]The decision to standardize the raw proximity values, so as to render the resultant measure independent of recall frequency (Eq. 2.4), appears to work quite well empirically, but creates an anomalous possibility. Thus, two items recalled concurrently only once, but in adjacent positions, would be considered as highly proximal as a pair recalled adjacently on all trials. One way to avoid this possibility is to set a threshold value, so that pairs recalled less often than this value are not considered, or have their proximity value reduced by some constant fraction.

[4]These data are from a study by Ornstein (1970, Exp. I), by whose kind permission they have been reanalyzed here.

The table on the lower left of the page shows the proximities of selected pairs of items for the six trials at the top of the page, and for each selected pair, the average proximity over all trials on which both members of the pair were recalled is also shown. Thus, QUARREL and ASSAULT were immediately adjacent on all six trials and have an average proximity of 11, the maximum possible for a list of 12 words. CAPTIVE and HIGHWAY, on the other hand, were consistently quite far apart, with an average proximity of 6.6. This means that, on the average, about five other words were interpolated between them in recall by this subject, and there would be little reason to believe that these two items belonged to the same functional memory unit for this subject.

Pairs of items also differ in the frequency with which both members of the pair are recalled. Thus CAPTIVE and ASSAULT were both recalled on all six trials. MAST and HIGHWAY, on the other hand, were both present in output on only three of the trials shown. When they were both recalled, however, they were quite proximal.

These proximities can be calculated for all pairs of words, and arranged in a square matrix as shown in the lower right of Figure 1. The matrix is necessarily symmetric by virtue of (2.4), so only the lower half is shown. The principal diagonal has also been omitted, since it conveys no information--$D_{ii} = 0$ for all items.

This matrix shows that there are several groups of words which have mutually high proximities within each group and relatively low proximities with items outside the group. INVENTOR and PROFESSOR, for example, seem to constitute a fairly distinct S-unit for this subject since their proximity to

Data

Trial 3:	urge	quarrel	assault	captive	decree	execution	mast	professor	inventor	north	
Trial 4:	urge	north	quarrel	assault	captive	execution	highway	professor	inventor		
Trial 5:	highway	structure	decree	mast	north	urge	execution	captive	assault	quarrel	inventor	professor
Trial 6:	quarrel	assault	decree	execution	captive	urge	north	structure	highway	inventor	professor	
Trial 7:	quarrel	assault	captive	execution	decree	north	structure	mast	highway	inventor	professor	urge
Trial 8:	quarrel	assault	captive	execution	decree	urge	structure	north	inventor	professor	highway	mast

Trial 5 groupings: highway–structure p = 11; decree–mast–north–urge p = 9; mast–north–urge–execution–captive–assault p = 7; execution–captive–assault p = 10; inventor–professor p = 11.

Proximities

Stimulus Pairs	Trial 3	4	5	6	7	8	Average Proximity
quarrel - assault	11	11	11	11	11	11	11.0
captive - assault	11	11	11	9	11	11	10.7
captive - highway	0	10	5	8	6	4	6.6
mast - highway	0	0	9	0	11	11	10.3
mast - quarrel	7	0	6	0	5	1	4.7
north - structure	0	0	9	11	11	11	10.5
north - urge	3	11	11	11	6	10	8.7
professor - structure	0	0	2	9	8	9	7.0
professor - inventor	11	11	11	11	11	11	11.0
decree - inventor	8	0	4	5	7	8	6.4
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Proximity Matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1 inventor												
2 professor	11.0											
3 highway	8.6	8.6										
4 mast	8.5	8.5	10.3									
5 north	8.3	7.7	8.7	9.5								
6 structure	8.0	7.0	10.0	9.3	10.5							
7 assault	6.0	5.7	5.2	5.7	7.3	6.2						
8 quarrel	5.3	5.0	4.2	4.7	6.7	5.2	11.0					
9 captive	7.0	6.7	6.6	6.8	8.3	7.8	10.7	9.7				
10 execution	7.5	7.2	7.2	8.0	8.8	8.2	9.8	8.8	10.8			
11 decree	6.4	5.8	7.5	8.8	9.0	9.5	9.0	8.0	10.4	10.4		
12 urge	7.0	7.0	7.6	7.5	8.7	9.0	7.7	7.3	8.3	8.2	8.4	

(Trials 3-8)

Fig. 1. Illustrative data for proximity analysis.

each other is 11, the maximum, while each of these words has relatively low proximities with all the other items (cols. 1 and 2 of the matrix). Similarly, the items ASSAULT, QUARREL, CAPTIVE, EXECUTION, and DEGREE are all highly proximal to one another in this subject's recall. A third highly organized group consists of HIGHWAY, MAST, NORTH, and STRUCTURE. The word URGE appears to be a singleton; it is recalled on all trials by this subject but it does not appear consistently near any other items. These four sets of words constitute a reasonable approximation to the subjective groups displayed in this subject's recall. Looking at the three groups of items whose proximities have been marked off in the triangular blocks, these S-units can be roughly ordered in terms of tightness of organization, from (INVENTOR, PROFESSOR) as the strongest down to (HIGHWAY through STRUCTURE) as the weakest unit.

Usually, however, the items will not be arranged in the proximity matrix so that their structure is so apparent. Indeed, in making up the table the rows and columns were reordered so that the groups of co-organized items would be together, giving rise to the triangular blocks of high proximities. In general the proximities will need to be subjected to further analytic scrutiny to reveal the underlying organization reflected in the order of recall. Several rather different methods are available for analyzing such data and a choice among them should depend on theoretical considerations.

2.4 Spatial Representations and Organization

Having determined a matrix of intraserial proximities, it is natural to think of some spatial or graphical representation of the items which in some sense summarizes the sequential output consistencies and depicts the contents of S-units. There are two basic spatial representations which occur repeatedly in psychological applications.

The first and most widely employed is the Euclidean representation embodied in multidimensional scaling (MDS) and factor analysis. According to such a conception, each item (word, test, stimulus) might be represented by a point in space, as in MDS, or by a vector as in factor analysis. The idea of representing words in Euclidean space is not foreign to verbal learning studies. The structure of verbal items has been explored by Deese (1965) in a factor analysis of word association data, by Friendly and Glucksberg (1970) using MDS to portray aspects of semantic change, and is inherent in the semantic differential technique (Osgood, Suci, & Tannenbaum, 1957). However, the attempt to locate items in Euclidean space implies that (a) a set of underlying dimensions exist such that each item has a value on every dimension, and (b) it is reasonable and useful to consider the relations among items in such terms.

The second class of graphical representations derives largely from biological taxonomy and consists of determining a taxonomic classification of the items, usually in the form of a tree diagram. Here the aim is to express the relationships among a set of items in terms of hierarchically arranged sets of optimally homogeneous subgroups. Methods which attempt this hierarchical classification are generally referred to as cluster analyses.[5]

[5]It is not appropriate to identify all methods of numerical classification or cluster analysis with a representation in terms of a tree diagram or hierarchy. "Cluster analysis" is a broad, generic term and many clustering techniques are designed to produce efficient classification by a minimum variance partition of Euclidean space. These include variants of discriminant analysis (Kendall, 1966) and principal components (Gower, 1966) and thus embody a Euclidean representation.

Compared with a Euclidean representation, hierarchical classifications can be considered to be based on the more limited assumption that each item has a value defined for only some of the components of the hierarchy (Miller, 1967).

The notion of a hierarchical system for the organization of items in memory finds support in the theory and data of free recall. Mandler (1967a) takes as his major theoretical argument the idea that "a hierarchical system recodes the input into chunks with a limited set of items per chunk and then goes on to the next level of organization, where the first order chunks are recoded into 'superchunks',..."(p. 332). Tulving's (1964) view of subjective organization focuses more on the retrieval side of memory but contains implicitly the idea that S-units may be nested into higher-order units. In recall the higher order units presumably provide access to the smaller units they contain which in turn facilitate the retrieval of individual list items.

The idea of hierarchical grouping is not a particularly new one. In 1550 the French philosopher Ramus wrote that "everything is formed of little units and the mind groups these." As an explanatory concept in the study of human memory, hierarchical organization became important with the publication of <u>Plans and the Structure of Behavior</u> in 1960 by Miller, Galanter, and Pribram. If memory is organized hierarchically, Miller et al. imply an adequate description of S-units must indicate their contents on all levels simultaneously. "We are trying to describe a process that is organized on several different levels, and the pattern of units at one level can be indicated only by giving the units at the next higher, or more molar, level of description" (Miller et al., 1960, p. 13).

The basis of the present technique also is not new in cognitive psychology. The idea of obtaining similarity values among a set of verbal items and applying cluster analysis to represent interitem relations was used by Miller (1969) to study semantic relationships in a word sorting task and by Martin (1970) in an investigation of subjective phrase structure. In all three cases (including the application discussed here) the use of a hierarchical representation is dictated by theoretical considerations.

Note at this point that the hierarchy is being used both as a theoretical model for organization in memory, and as a methodology for portraying the structure of items in FR protocols. In the context of some experiments, a hierarchical representation may not be reasonable. In such cases, the interitem dependencies may be analyzed by a nonhierarchical clustering procedure (e.g., Jardine & Sibson, 1968) instead of the algorithm discussed below.

2.5 Cluster Analysis of Proximities

On the basis of the view of organization as operating to form a nested system of S-units, it is appropriate to choose a method of analysis which will reveal any hierarchical structure underlying the proximity scores. The method adopted here is a hierarchical clustering procedure due to Johnson (1967). The discussion below is patterned after Johnson (1967) and Miller (1969). A *clustering* of a set of items is merely a partition of the set into mutually exclusive and exhaustive groups, or *clusters*. A hierarchical clustering scheme consists of a tree structure with numerical values at the branches representing the similarities among items. The tree structure describes a sequence of clusterings such that the first is composed of as many clusters as there are items, and each successive one in the series is formed by merging clusters from the immediately preceding clustering. The numerical levels can be chosen to represent the compactness of the clusters at each stage.

The method begins with the finest partition (the disjoint or "weak" clustering) in which all clusters consist of single items. The first non-trivial clustering is found by placing together those items which were consistently recalled most contiguously (the most proximal items). The merged items are then treated as a single element, and the proximities between this new cluster and all other items are entered in a new, smaller matrix. Again, the most similar items/clusters are joined, and so forth until all items have been merged into a single cluster (the conjoint or "strong" clustering).

The key to this process is the ability to merge items and replace them by a single element in the proximity matrix so that the distance between this cluster and other items or clusters can still be defined. Hence, identical operations can be performed on items and clusters; an item is merely a cluster of size one. Suppose that the two most proximal items are w_i and w_j which are separated by a distance of $D_{ij} = N - P_{ij}$ as in Eq. (2.5). These items are therefore merged to form the cluster (ij) and we are required to determine a reasonable distance to assign between the cluster (ij) and any other item, w_k. For example, in Figure 1, INVENTOR and PROFESSOR were recalled adjacently on all trials and have the highest possible proximity of 11. When these are joined to form a cluster, it is necessary to assign a proximity between this cluster and any other item, e.g., URGE, so that items and clusters can be treated alike. QUARREL and ASSAULT also merge at P = 11 (or D = 1), and so the same problem applies to this pair.

Clearly, this intercluster distance, $D_{(ij)k}$, will be some function of the distance from w_i to w_k and of the distance from w_j to w_k. In the simplest case, D_{ik} and D_{jk} would have equal values for any other item w_k, since this would make the choice unique. That is, if $D_{ik} = D_{jk}$ for all k, then when w_i and w_j are joined to form a cluster, it would be natural to assign to $D_{(ij)k}$ the common value of D_{ik} and D_{jk}. Since it is the closest items, w_i and w_j, which are clustered, the three distances in this simple case would be related as

$$D_{ij} \leq D_{ik} = D_{jk} \quad (2.6)$$

The above relation, when it holds for all triples of items (w_i, w_j, w_k), is called the <u>ultrametric inequality</u> (UMI). There are three distances between pairs of three items. Satisfaction of the ultrametric inequality means that either all three distances are equal, or if there is a smallest distance, the remaining two are equal. This can also be expressed as

$$D_{ij} \leq \max \left[D_{ik}, D_{jk} \right] . \quad (2.7)$$

The ultrametric inequality is more restrictive than the triangle inequality,

$$D_{ij} \leq D_{ik} + D_{jk} , \quad (2.8)$$

which must hold for any set of distances, since any distances satisfying Eq. (2.7) will satisfy Eq. (2.8) a fortiori, but not conversely.

The importance of this is that when the UMI holds for an empirical distance matrix, there is an exact equivalence between the distance matrix and a hierarchical clustering (Hartigan, 1967; Johnson, 1967; Miller, 1969). Information is neither added nor lost in going from one to the other.

In general, however, proximities computed from recall protocols will not satisfy the UMI, either because of "noise," or because the structure of the items does not conform to a hierarchy. In Figure 1, for example, with INVENTOR and PROFESSOR being merged, the UMI would require that the proximities in column 1 from ASSAULT down be equal to the corresponding column 2 entries. This is true for MAST and URGE; however, the proximity of (INVENTOR, PROFESSOR) to DECREE can range from 6.4 to 5.8.

The diameter and connectedness methods. Johnson has proposed two solutions, which in a sense provide upper and lower bounds on hierarchical clusterings which could be derived from the data. In one method, whenever a choice is necessary, as between P(INVENTOR, DECREE) = 6.4 and P(PROFESSOR, DECREE) = 5.8, the proximity of an item to a cluster is taken to be its proximity to the nearest item in the cluster (connectedness method). Alternatively in the second method, an item-cluster proximity is set equal to the proximity between the item and the farthest element in the cluster (diameter method). While other variants are possible[6] (Lance & Williams, 1967; Sokal &

[6] The maximum and minimum of cluster-object distances correspond to the boundary points of a one-parameter system of clustering strategies defined by

$$D_{(ij)k} = \min \left[D_{ik}, D_{jk} \right] + \eta \, |D_{ik} - D_{jk}| , \quad 0 \leq \eta \leq 1 .$$

In this family of clustering solutions, $\eta = 0$ gives the minimum method. $\eta = 1$ corresponds to the maximum method, while setting $\eta = 1/2$ will produce a mean-distance strategy. It is in this sense that the diameter and connectedness methods were referred to above as upper and lower bounds.

Sneath, 1963), such as the average, use of the minimum or maximum guarantees that the result of the clustering will be unaffected by any monotone transformation of the data.

Although these two proposals represent opposite extremes, the solutions they produce for any set of data will agree to the extent that the UMI is satisfied. Reversing the argument, the amount of agreement can be taken as an indication of how well the structure of the items can be represented as a hierarchy.

To illustrate how these methods work, they have been applied to the matrix for the 12 unrelated words in Figure 1.[7] The results are shown in Figure 2. Such a tree diagram, derived from free recall protocols can be called a memory diagram, or <u>M-gram</u>, for short. The first clusters formed contain those items which were recalled by this subject in immediately adjacent output positions on all trials and have the maximum proximity value, 11.0--(INVENTOR, PROFESSOR) and (ASSAULT, QUARREL). The next highest proximity is between CAPTIVE and EXECUTION, so these items are merged next, and so on until all items have been merged into one cluster.

In general, there is reasonably good agreement between the two methods. A measure of correlation computed between the two solutions (see Appendix B) has a value of .92. Both solutions indicate ASSAULT, QUARREL, CAPTIVE, EXECUTION, and DECREE as a higher-order S-unit, although they disagree on the order with which the smaller units (ASSAULT, QUARREL), (CAPTIVE, EXECUTION), and (DECREE) merged together. HIGHWAY, MAST, NORTH, and STRUCTURE

---

[7]The proximities shown in Fig. 1 were rounded to one decimal place for simplicity. The clustering in Fig. 2 represents the actual values.

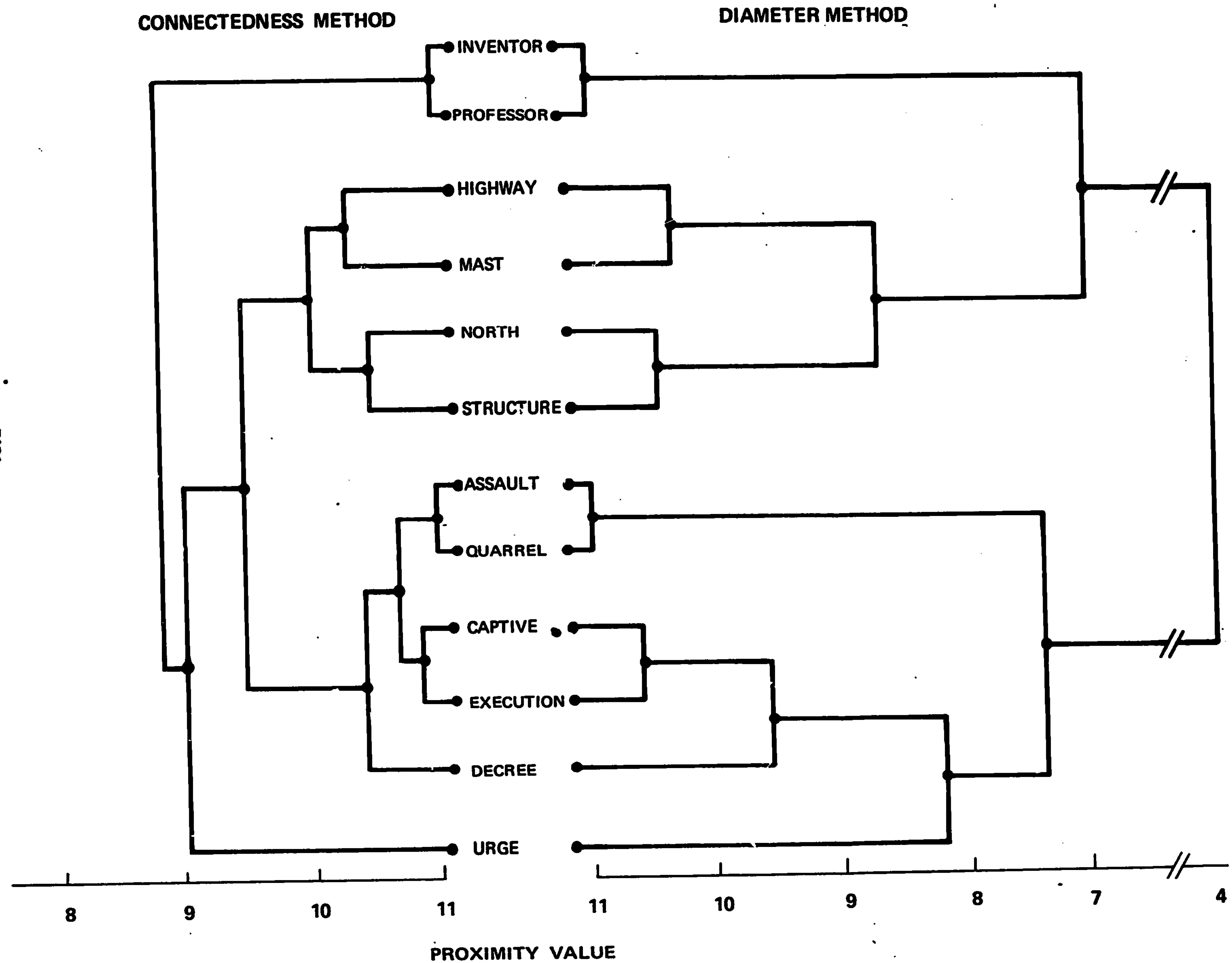


Fig. 2. Connectedness and diameter method clustering solutions for the data of Fig. 1.

are clustered by both methods, as are INVENTOR and PROFESSOR. The methods disagree most on the order in which these S-units and URGE (seemingly a loner) merge subsequently. Yet the clusters are not highly isolated at this stage, and it is probably unwise to interpret these final clusterings as superordinate S-units. Since the analysis will provide a hierarchical solution for any data, it seems safest to interpret only those clusterings which contain compact, isolated clusters.

This result is fairly typical of data from experiments using unrelated lists, at least in our experience. A moderate degree of subjective clustering is observed, but these clusters do not always appear to be tightly organized and sometimes no apparent structure above the level of pairs of items can be discerned. When subjects learn lists of related sets of items, on the other hand, subjective groupings of the items are more obvious, more consensual, and Ss' output orders reflect more highly constrained S-units (e.g., Cofer, 1965).

As an illustration of the organization of categorized lists, consider some data from another experiment by Ornstein (1970, Exp. II). Subjects in this experiment learned two categorized lists in succession. The first list for all Ss consisted of 24 items in six categories of four items each. For one group of Ss the categories used were Furniture, Gems, Professions, Parts of a home, Vegetables, and Vehicles. Subjects received visual presentation of the items for five alternate study-test trials. The diameter method M-gram for a typical S, with data pooled over all five trials, is shown as Figure 3.

The grouping of items into compact clusters, identical to the E-defined categories is striking. The smallest within-category proximity is 20.6

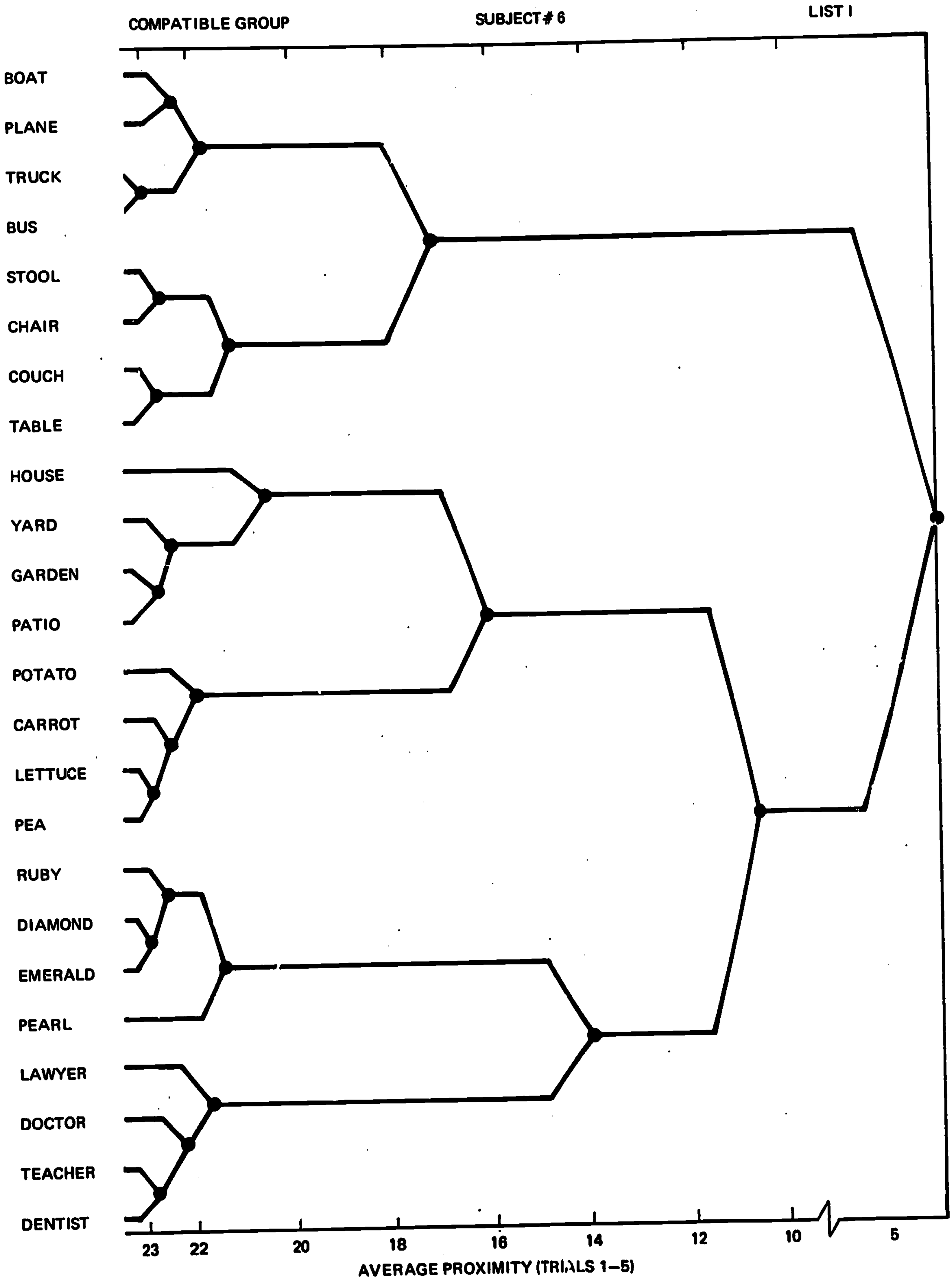


Fig. 3. M-gram for a categorized list.

between HOUSE and YARD. This value is 90% of the maximum value of 23.0 and corresponds to an average protocol separation of 2.4 items. The E-defined categories are highly isolated from each other; complete categories do not merge until relatively low levels of proximity are reached.

Interpretable subgroupings can also be identified within the categories. In the group HOUSE, GARDEN, YARD, and PATIO, the last three items are most similar semantically, and these items cluster before being joined with HOUSE. Similarly, among the Gems, RUBY, DIAMOND, and EMERALD are all stones, and they form the nucleus of this cluster. Without looking into the reliability and generality of these subgroupings, it is not wise to overstress them. We merely note an interesting (and possibly ephemeral) by-product, reminiscent of Bousfield and Sedgewick's (1944) finding of subgrouping in categorical associations. The major point to be noted is the strong grouping into S-units, and the identity of these units with the E-categories.

2.6 S-Units and Clusters

Whether or not the diameter and connectedness methods agree in practice, there are conceptual differences between them worthy of attention regarding the identification of S-units. In Johnson's connectedness method (Sokal and Sneath's "clustering by single-linkage" or nearest neighbor), choosing the minimum cluster-item distance ensures that a just-formed cluster will appear to move closer to some or all of the remaining objects/clusters and farther from none. Clustering methods which share this property are said to be space-contracting (Lance & Williams, 1967). This scheme will add an item to a cluster as soon as it is at a given distance from any item in the cluster, and the method tends to produce long chains, which are only locally compact.

By contrast, for a given criterial distance, the diameter method ("clustering by complete linkage" or farthest neighbor) does not admit an item to a cluster unless it is at least that close to all items in the cluster. This method therefore produces clusters which are globally connected. More explicitly, at any given stage in either method, a value for the clustering may be defined. In the diameter method, the largest distance within each cluster (the diameter) is found. The value of the clustering is then the maximum diameter of all clusters at that level. The merging of clusters at each stage in this method is performed so as to minimize the diameters of clusters.

Corresponding to the choice between these properties are two alternative conceptions of the nature of memory units. It is possible to think of S-units which form serial chains, so that each item is highly connected to its neighbors in the chain, but less so to more remote items. The cardinal compass points, North, South, East, and West, form such a series, as do mediated associative chains such as Billiards, Pool, and Water (Shapiro & Palermo, 1967). This type of "linear" grouping would also be expected if a list were organized alphabetically (Tulving, 1962b).

The connectedness method is well suited to revealing such sequences. Usually, however, an S-unit will be defined as a group of items with mutually high connectivity; recall of any one item in an S-unit should, with high probability, be accompanied by contiguous recall of the remaining items. The diameter method will tend to give a clearer picture of these highly compact groups.

Therefore, in applying proximity analysis to free recall data, greater emphasis will be given to the diameter method solutions for describing the contents of S-units.[8] Yet it is well to have some way of assessing the degree to which the connectedness method would give discrepant results. Stated in other terms, any hierarchical clustering scheme may be regarded as a method whereby the ultrametric inequality is imposed on a distance matrix. It would therefore be helpful to have some measure of this distortion. Some ways of achieving this are considered in Appendix B.

Since the cluster analysis provides a family of clusterings, rather than a single partition, we shall need some ways to talk about the strengths of S-units formed at different levels of proximity. In discussing Figures 2 and 3, two features of clusters were indicated which could serve to guide the interpretation of S-units--compactness and isolation. These notions may be defined precisely in terms of the cluster analysis.

For the maximum method, the cluster diameter (largest intracluster distance, or smallest proximity) provides a natural measure of compactness. The diameter of any cluster $(w_i, w_j, w_k, \ldots)$ may be defined as the node distance associated with the first clustering in which $w_i, w_j, w_k, \ldots$ are all in the same cluster. With proximity defined as in Eq. (2.5), the diameter of any cluster can be determined from the M-gram as $N - P(i,j,k,\ldots)$, where $P(i,j,k,\ldots)$ is the node proximity value of the cluster. In Figure 4, for instance, the diameter of the cluster (POTATO, CARROT, LETTUCE, PEA) is 24 - 22.0 or 2.0, while the diameter of (LETTUCE, PEA, RUBY, DIAMOND) is 13.3.

---

[8] This is not to imply that the diameter method is to be generally preferred, even in psychological applications. In any search for clusters or types, the investigator must begin with a substantive notion of a cluster, rather than with a statistical one.

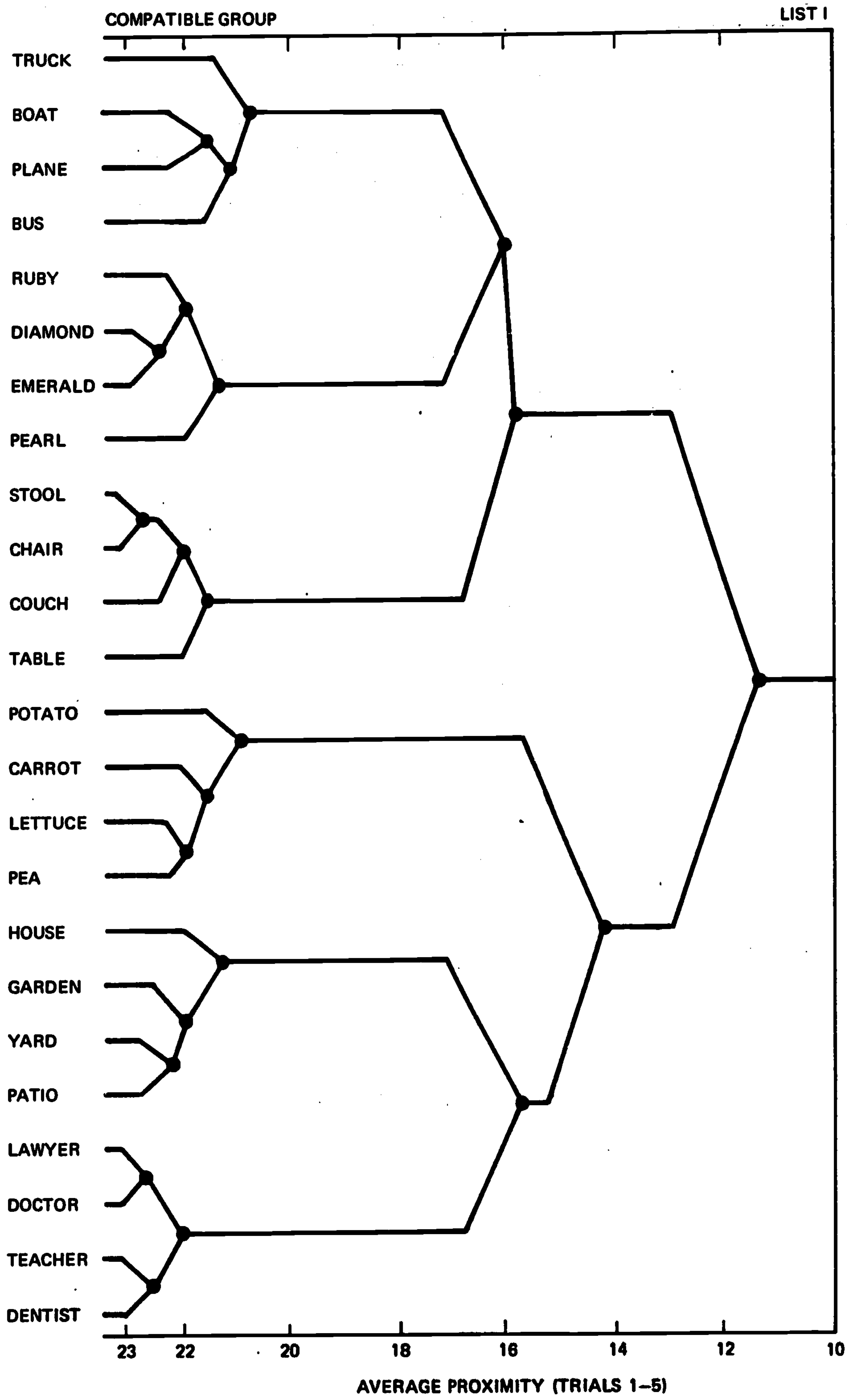


Fig. 4. Proximity analysis of pooled group data (S = 10).

While the cluster diameter gives an indication of the strength of an individual cluster, it says nothing about the relationship between clusters. The notion of cluster isolation can be used to distinguish among representations in terms of clusters at various levels in the hierarchy. The isolation of a cluster x expresses the diameter of x relative to the diameter of the first clustering in which x is merged with another cluster. In practice it will be convenient to take the difference between these two diameters as the measure of cluster isolation, although the ratio of the two could also be used. The isolation of a cluster can be thought of as a measure of the "empty space" around it, or the intercluster gap. In Figure 3 the diameter of the Professions category is 2.4; Professions next merge with Gems and this larger cluster has a diameter of 10.0. The isolation of the Profession category is therefore 10.0 - 2.4 or 7.6.

Up to this point the discussion of proximity analysis has been essentially concerned with the data from a single S in multitrial FR. The "modal" organization displayed by a group of Ss can be easily obtained by analyzing the average proximities for the group. Appendix A deals with this topic in more detail, and discusses several approaches to individual differences in organization. However, an example of organization determined from group data is useful at this point.

The high level of sequential organization usually found in the recall protocols from categorized lists was discussed in section 2.5 and illustrated in Figure 3 with the M-gram determined for a typical subject from one of Ornstein's groups. Figure 4 shows the M-gram derived from the pooled protocols of all seven Ss in that group. For the group data the six E-defined categories also emerge as compact, isolated clusters. Individual differences, if present, would appear as noise in the group analysis and tend to increase cluster diameters and reduce cluster isolation. The average cluster diameter for the group data, in terms of distance (Figure 4), is 1.68, which may be compared with the corresponding value of 1.55 for Figure 3. More precise comparisons do not seem warranted in the light of the strong similarity between the two figures. At the level of single categories, all Ss have utilized the same structure in their recall.

2.7 Related Work

Several other investigators have quite recently considered the problem of determining functional units in recall. Rather than using order of recall information directly as in the present approach, it is possible to attempt to identify S-units by obtaining supplementary information, independent of recall. Three workers have taken this approach in different ways. All three involve tasks designed to get S to reveal which sets of items go together in his memory.

Seibel (1964, 1965) introduced what he called the study-sheet technique, involving a modification of the typical input phase. With this procedure, S was given a sheet of paper with a large grid at the beginning of each trial. The subject was instructed to write each word as it was presented in any cell of the grid. This procedure allows S to establish a subjective categorization during input and to rehearse these categories as presentation proceeds. At the end of each presentation, S wrote the words he could remember on a new blank sheet of paper. This procedure differs from the usual method of presentation in that study time per item is uncontrolled and is probably cumulative over serial positions in input. Seibel found that items written together on the study sheet also appeared as output sequences in Ss' recall. A control group, instructed to write the items on the study sheet in the order of presentation, recalled less well than the group allowed to form subjective categories.

In a comprehensive series of experiments, Mandler (1967a, 1970; Mandler & Pearlstone, 1966) used a similar word-sorting task both to induce a stable, subject-determined organization and to make this organization directly observable by E. In these studies S was typically required to sort 50-100 words into anywhere from two to seven subjective groups using "any criterion, rule or category" (Mandler & Pearlstone, 1966, p. 127). Sorting trials were continued until S reached a criterion of 95% - 100% consistency in category assignments on two successive trials. This high criterion probably ensured a stable, well-learned categorization. After reaching criterion, FR memory for the items was tested, usually in a single trial. In these studies, Mandler was primarily concerned with the number of categories used in sorting as a predictor of subsequent recall performances and found a linear increase in recall as a function of this variable (up to approximately seven categories).

Assuming that the categorization established in either of the two procedures described above was the same as that utilized in subsequent recall, the categories generated by S could be considered to be the higher-order units. It would then be possible to investigate other characteristics of these subjective clusters. One potential problem is that the extent to which the sorting or study-sheet groupings and the functional units of recall actually overlap is not known, and little direct evidence on the point has been presented. Furthermore, it should be noted that in Mandler's procedure, all acquisition of the categorization scheme precedes the first (and typically only) test of memory for the words. Hence, this procedure provides little information about the acquisition of the organizational scheme itself. It would be relatively simple to remove both of these limitations by alternating sorting trials with FR test trials and using the technique of proximity analysis to investigate the correspondence between the two organizational structures.

The interitem dependencies in recall can also be dealt with in terms of the mathematical system of graph theory. This theme was developed extensively by Allen (1971). Allen argued that theories of organized memory could be coordinated with the formal language of directed graphs (digraphs) so that the analytic techniques of the latter could be usefully applied to studying organization. In applying graph theory to memory, Allen developed several methods for constructing empirical digraphs representing the structure of S-units for individual subjects. In a demonstration experiment, S learned a 20-item list comprised of high frequency unrelated nouns. After seven trials, Ss were given one of three "memory unit identification tasks." In two of these, S was given the list of words and required to write groups of list items which he felt went together in his memory in the cells of a matrix. In the third procedure, S was given a deck of 190 cards, each of which contained one of the possible pairs of list words. The task was to sort these cards into two piles, depending on whether S felt the members of a pair belonged to the same group in his memory. The instructions in all three cases stressed that the criterion for sorting should be whether S felt the words were together in his memory during the recall trials, not whether items merely seemed related. The information from these tasks was then used to generate a directed graph representing the subjective structure of memory items. Each point in the graph symbolized the trace of a list item; the lines connecting the points represented item pairs linked together in memory.

In operation, these procedures are quite similar to those of Mandler and Seibel. However, by imbedding these empirical tasks within the methods and concepts of graph theory, it is possible to investigate a large variety of important theoretical questions which cannot be studied by the use of these tasks alone. For example, Allen (1971) demonstrated that various aspects of recall were related to measures derivable from the graph representation of organization. Among these were the amount of organization (ITR), number correct, and the proximity between pairs of items in the protocols.

Allen's graph theory analysis is closely related to the present approach. The graph constructed from the subjective report task is equivalent to a square matrix (the adjacency matrix) containing 0 and 1 entries. The entry in row i and column j is unity if an S indicates that items i and j are together in his memory and is zero otherwise. The same matrix would result if a threshold value, c, were applied to the proximity matrix generated by the present approach such that any proximity greater than or equal to c were replaced by unity and any value less than c replaced by zero.

The proximity method thus includes Allen's adjacency matrix as a special case, where the relations among items in memory are considered to be all (1) or none (0) rather than of variable strength. The two techniques differ in one essential respect--the source from which information regarding interitem dependency is drawn. In the proximity method, pair relatedness is estimated directly from recall protocols while Allen introduces a supplementary task to obtain this information. They also appear to differ in a second respect, namely, the basis on which the interitem relatedness measures are further analyzed--hierarchical clustering versus graph theoretical procedures. However, these two methods are actually closely related. A number of methods of hierarchical cluster analysis are derived from graph theory (Bonner, 1964; Needham, 1961; Sokal & Sneath, 1963) and use a series of increasing threshold values as described above to produce a tree structure clustering.

Since the present research began, there have been two reports describing the application of Johnson's clustering procedure to FRL data. In attempting to provide evidence for a model of free recall based on semantic markers, Kintsch (1970) computed a measure of output adjacency in recall protocols. This measure can be derived from the adjacency matrix used in calculating Tulving's SO. The frequency, n_{ij}, with which item $\underline{j}$ immediately follows item $\underline{i}$ in recall output, is tabulated in this matrix. Kintsch's adjacency value, a_{ij}, for a pair of items is then calculated as

$$a_{ij} = \frac{n_{ij}}{n_i} + \frac{n_{ji}}{n_j} ,$$

where n_i, n_j are the marginal totals of the matrix, i.e., the number of times each item was recalled.

Thus, this measure takes into account only pairs of items recalled in immediately adjacent positions. Because it disregards information beyond this, more data are required to obtain reliable estimates of interitem dependency in recall, and the measure should probably be used only with group data. In fact, a rough calculation shows that for a list of N items Kintsch's method requires about N times as much data as a measure based on all pairs recalled.[9]

In spite of these deficiencies, Kintsch showed that this procedure allowed some information about the structure of organization to be extracted. Two 16-item lists were used in a demonstration experiment--a list composed of four equal-sized categories, and the unrelated list from Tulving's (1962a) original paper on SO. Two presentation orders were used for each list. The categorized words were arranged in either blocked or random order. The unrelated list appeared in orders from Tulving (1965) that either maximized or minimized normative sequential redundancy. Adjacency measures were calculated from group data on each of the three trials given.

Kintsch (1970) presented the hierarchical clustering for the first trial of the blocked presentation, categorized word protocols. As expected, the tree structure indicated that the list categories did appear as output units.

---

[9]This factor was determined as follows: If a subject recalls n_t items on trial t, there are $n_t(n_t - 1)$ pairs of items in his protocol, of which $(n_t - 1)$ are adjacent pairs. Since only the latter pairs are considered in Kintsch's measure, the protocol contributes $(n_t - 1)$ units of "proximity information" to the calculation, while all $n_t(n_t - 1)$ pairs contribute to the proximity measure in Eq. (2.4). The factor of relative efficiency of the present measure is actually closer to the average number of words recalled than to the number of words presented.

Kintsch reported that a reliably hierarchical structure (judged by the correspondence between the maximum and minimum method solutions) did not emerge in the random presentation-categorized data until Trial 3, and that no hierarchical organization could be found for the unrelated list with either presentation order. This latter finding is surprising in the light of (a) Tulving's (1962a) observation using the same words, that intersubject agreement in SO increased over trials, and (b) the fact that one of the presentation orders was chosen on the basis of maximum communality across subjects (cf. Tulving, 1965).

Koh, Vernon, and Bailey (1971) have applied Johnson's (1967) clustering technique to FRL data from deaf and hearing Ss of two age levels. In their experiment each S learned a categorized list and an unrelated list, both of 16 words, in multitrial free recall sessions. Their analysis is not explicitly described; however, they appear to have used, as a similarity measure, the proportion of times each pair was recalled adjacently on the last of 16 acquisition trials, collapsed over all Ss. The same reservations noted above apply to this measure also.

Koh et al. also report that better fit to a hierarchy was obtained for their categorized list than for unrelated words. In the clusterings derived for the unrelated words, the results for hearing Ss were more closely hierarchical than for deaf Ss; a small increase in hierarchical fit was also related to age.

Thus there have been a number of exciting and diverse attempts to deal with the structure of organized recall, most of them quite recent. As noted above, these approaches are not incompatible and can easily be applied in tandem. For example, it is quite feasible to combine an analysis based on clustering of interitem proximities with a subjective report or sorting task to specify more clearly the nature of S-units and provide more powerful ways of testing hypotheses about organized memory.

CHAPTER 3

ORGANIZATION AND LONG-TERM RETENTION OF HIERARCHICAL LISTS

3.1 Introduction

In Chapter 2, a procedure for investigating the structure of organized memory was described and illustrated with sample data. This chapter presents an experiment designed in part to provide empirical evidence regarding the validity and usefulness of this procedure. This methodological question was investigated in a situation where prevalent modes of list organization by Ss could be predicted in advance with some confidence, i.e., by making use of lists containing strong E-defined categories. In addition, data were obtained on the long-term retention of such lists.

In many studies concerned with the relation of organization and recall, organization is manipulated by constructing different lists which vary in characteristics relevant to the development and use of higher-order groupings, e.g., number and size of E-defined categories (Dallett, 1964), presence or absence of categorical retrieval cues (Tulving & Pearlstone, 1966), etc. In the present study, the specific material to be remembered was not manipulated. Instead, a list which could be categorized in alternate ways was constructed. It was hoped that, by manipulating the presentation order of the items, the experiment would induce different groups of Ss to employ the alternative modes of organization in recalling the list (cf. Wood, 1970).

The purpose of this manipulation was two-fold. The first intent was to assess the extent to which different presentation orders could produce variations in the manner in which subjects organize a single list. The second was to determine how well any such differences in organization could be detected by the proximity technique outlined earlier.

Taxonomic hierarchies with several levels provide one method for constructing a list which can be organized in more than one way. In such a list, the categories at the lower levels are nested within the categories of all levels superordinate to them. Figure 5 is an example of such a taxonomic hierarchy and contains the items used in this study.

The 42 items listed at the bottom of the figure can be regarded as belonging to three 14-item categories, or to six 7-item categories. Alternatively, the list may be conceptualized in terms of three systems of categories at different levels. At the most inclusive level, all of the items are EDIBLE SUBSTANCES, of which there are three broad classes at level 2; two subcategories at level 3 are nested within each level 2 group.

The acquisition of taxonomic hierarchies in free recall has been studied by Bower et al. (1969) and by Cohen and Bousfield (1956). The latter investigators used a dual-level list in which four major 10-item categories could each be divided into two 5-item subcategories. The major categories were independent rather than instances of some yet larger grouping. The occurrence of clustering in recall of this list was assessed on the basis of both four and eight categories, and the results were compared with those obtained in an earlier experiment (Bousfield & Cohen, 1956) with separate, single-level lists of four and eight categories. Recall of the dual-level list was greater than that of the earlier four category list but no different than that of the single-level, eight category list. Differences in clustering at either level of the dual list were negligibly small, though

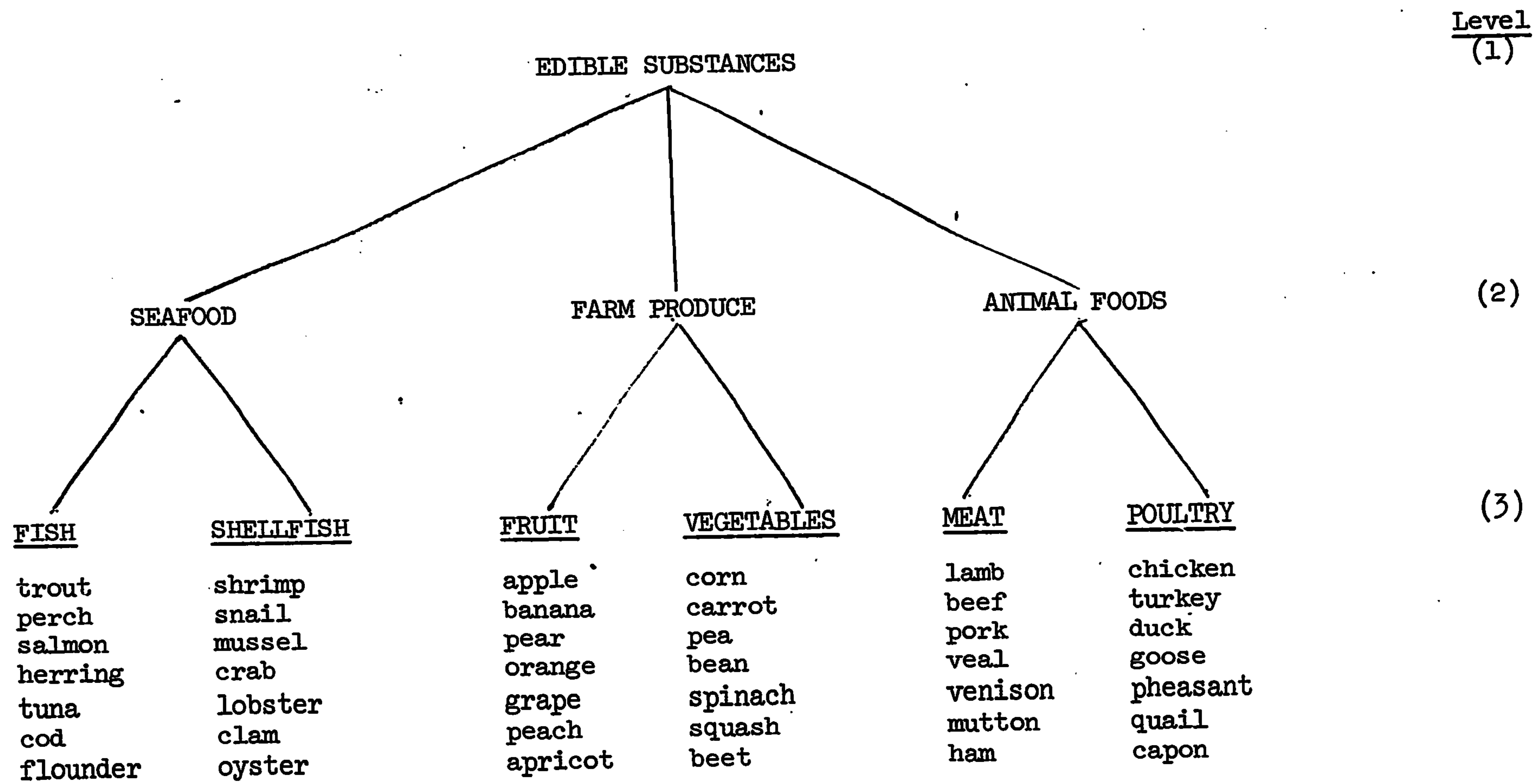


Fig. 5. Hierarchical structure of the words used in the experiment.

this is not surprising since only one presentation was given and input order was random.

In contrast to these small effects of several levels of organization is the dramatic facilitation of recall demonstrated by Bower et al. (1969) using hierarchical lists with varying methods of presentation. Words belonging to taxonomic hierarchies were learned by the method of complete presentation (i.e., all words presented simultaneously) with the words and category names arranged spatially in a vertical tree. The stimulus display thus appeared similar to Figure 5 here, without the connecting lines. For Ss in a Blocked group, the arrangement of items in the spatial tree corresponded to the hierarchical groupings in the list; for Ss in a Random group, the items were assigned randomly to the nodes of the spatial tree. Bower et al. found that blocking of the taxonomic hierarchies produced tremendous gains in recall. After two trials, the Blocked group recalled 95% of a 112-item list, while the random group recalled 35%.

The present study attempted to manipulate the type and mnemonic value of the information which S had about the structure of a hierarchical list by blocking the items according to its different levels. In blocked presentation, all members of an E-defined category are presented contiguously. If several input trials are given, the order of items within blocks is usually varied randomly from trial to trial, as is the order of the blocks themselves but the separate categories are not intermixed. Studies by Puff (1966) and by Dallett (1964) among others (cf. Shuell, 1969) have shown that blocked presentation facilitates recall and augments clustering according to the categories of the blocks.

Experimental groups included in the present study (see Table 1) differed according to whether input was blocked into three categories at level 2 of the hierarchy (Group B2), blocked according to six categories at level 3 (Group B3), or blocked according to both levels (Group B4). In recalling words from a categorized list, S must be able to retrieve items from within a given category and be able to move from one category to the next. It was expected that blocking at both levels of the hierarchy would provide information relevant to both these requirements and lead to the most efficient organization and acquisition of the list. Blocking at a single level (Groups B2 and B3) would not explicitly provide information about the relations among categories as readily, and was expected to lead to poorer performance.

Wood (1970) has also employed lists of words which can be categorized in more than one way. In Wood's list, the alternate classifications were incompatible, i.e., orthogonal to each other. In the hierarchical list used here, however, the alternative groupings were compatible in that they consisted of successively finer subdivisions of a single category. This arrangement essentially creates a stringent test for proximity analysis since the differences among alternative organizations of the hierarchical list would likely be fine grain ones.

In addition, it was decided to obtain data on long-term retention in the context of the manipulations described above. These data derive theoretical interest from the implication of organizational theory that long-term retention should depend on the stability and functional integrity of the higher-order groupings of a list of items developed during acquisition (Mandler, 1967a; Mandler, Pearlstone, & Koopmans, 1969; Postman, 1971). It has been demonstrated that recall performance during acquisition varies directly with the degree of organization in recall (Bousfield, Puff, & Cowan, 1964; Tulving, 1962a). However, the results concerning this relation beyond the time of original learning are scanty and conflicting (cf. Brand & Woods, 1958; Mandler, 1967a; Postman, 1970). This study was designed in part to shed some light on this problem. By comparing the organizational structures determined from acquisition with those derived from retention, the proximity analyses would indicate the extent to which organization remained intact after the retention period.

3.2 Method

Experimental Design

There were two phases of the experiment. In the original learning (OL) phase, all subjects were presented with the same list of 42 words on each of 12 trials. There were seven groups of Ss whose treatments differed in both the number and composition of blocks which were present in the input list.

Three experimental groups differed according to whether the items were blocked into major categories at level 2 of the hierarchy (Group B2), blocked according to minor categories at level 3 (Group B3) or blocked according to both level 2 and 3 categories (Group B4). For each experimental group, a control group (Groups R2, R3, and R4) learned the items with the same blocking structure, except that the items which consistently appeared together (blocked) were chosen randomly rather than according to conceptual relationships. These latter groups were used to evaluate the effects of blocking per se, i.e., to control for any facilitation which might occur only because a list was blocked, regardless of the contents of the blocks. An additional group (B1) received the items in a totally random fashion and served as a baseline for evaluating the effects of blocking alone and of blocking according to category membership.

Approximately equal numbers of subjects in each of the three experimental groups and Group B1 returned to the laboratory after 1, 5, 10, or 20 days for a retention test as the second phase of the experiment. In order to minimize rehearsal during the interval, subjects were told that the goal of the experiment was to investigate the relationship between list-learning performance and some paper-and-pencil tests of memory and cognitive ability and that they were to return to take these when they returned.

A major interest of the study concerned the effects on OL and retention of blocking according to different levels in a hierarchically structured list. Since Group B1 provided an overall control for blocking per se, the R conditions were only tested in retention at 1 and 5 days.

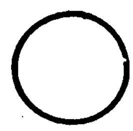
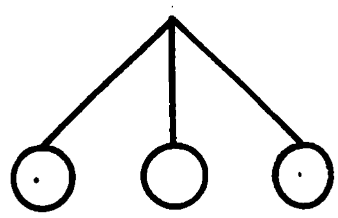
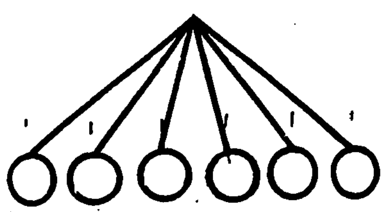
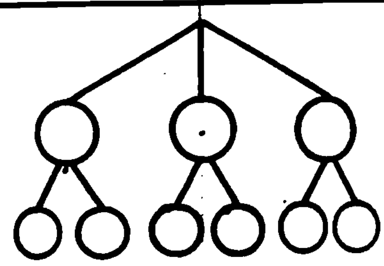
Subjects were run by four experimenters, counterbalanced over all groups and retention intervals. The design of the experiment, as well as the number of Ss per cell, is presented in Table 1. Additional subjects were run in the 20-day groups to protect against possible attrition after this long-time interval. The groups are described below.

Group B1.--The subjects in this group received a different random ordering of the stimulus list on each trial. For purposes of comparison with remaining groups, this condition can be considered as having the words blocked at level 1.

Group B2.--The blocks consisted of the categories at level 2 of the stimulus hierarchy, i.e., SEAFOOD, FARM PRODUCE, and MEAT. Thus, there were three blocks consisting of 14 words each, with the order of blocks and order of items within blocks randomized from trial to trial.

Table 1

Design of Experiment and Number of Subjects per Cell

Group	List Structure	Number of Subjects				
		OL[a]	Retention Interval (da.)			
			1	5	10	20
B1		42	8	8	9	11
B2		36	8	9	8	10
B3		37	8	9	8	9
B4		40	8	8	9	9
R2	same as B2	18	9	8		
R3	same as B3	19	8	9		
R4	same as B4	18	10	8		

[a]Numbers include those subjects not returning for Session II.

Group R2.--The structure of the blocking of items in this condition paralleled that in Group B2, that is, the list contained three blocks of 14 items each. However, items were assigned at random, rather than by category, to these blocks so that the influence of blocking according to conceptual categories (B2) could be evaluated against the effect of blocking alone (R2). Any difference in performance between Groups B2 and R2 could then be attributable to the presence of conceptual categories in the blocks for Group B2 rather than mere presence of consistently proximal input sets. Further, two different random partitions of the stimulus items into three blocks were generated and each presented to half of the R2 Ss to reduce the effect of any fortuitous groupings which might occur in assignment to blocks.

Group B3.--The items were arranged in blocks according to the partition at level 3 of the stimulus hierarchy. There were six blocks (e.g., FISH, SHELLFISH, FRUIT, etc.) composed of seven items each, with block order and within block order randomized over trials.

Group R3.--This group controls for the effect of blocking alone in Group B3 in the same way that Group R2 serves as a control for B2. Two of the 42-item list into six blocks of seven words each were generated and each used equally often over all subjects in this group.

Group B4.--The blocking of items in this condition was the most constrained and most congruent with the structure of the stimulus hierarchy (Figure 5). The items were first blocked into three major categories at level 2 in the hierarchy. Then, within each major category (e.g., FARM, PRODUCE) the 14 items were further divided into the two major categories, each consisting of seven words (e.g., FRUIT and VEGETABLES). On each trial, the three major categories were randomly ordered. Within each major category, the two minor categories were permuted and the order of individual items within minor categories also randomized. Since the blocking of items for this group gives the greatest amount of information regarding the list structure, performance and clustering for this group should be the greatest.

Group R4.--Subjects in this group had the words blocked in the same fashion as those in Group B4 except that the items which compared the blocks were chosen randomly from the stimulus lists. Again, two different random assignments of items to the blocks were used equally often.

Selection of stimulus materials. An initial pool of 61 items representing the categories of the list were chosen from the high frequency responses to categories in the Battig and Montague (1969) category norms. These norms were compiled by presenting a series of category names to subjects and asking for one or more instances of each category name. Hence, the (normalized) frequency of occurrence of a particular item *j* (say) as an instance of a category name can be thought of as a conditional probability--Prob(instance j|category name). However, studies of memory using categorized lists present the instances to the subject and assume that the set of instances will serve to generate the category name as an implicit response or cue. Because of this, it seems more appropriate to know the associative strengths in the direction opposite to that of the category norms, i.e., we should determine Prob(category name|instance $j_1 \ldots, j_n$) and use these to construct lists whose categories are balanced for the strength with which the items evoke the category name.[10]

[10] Such norms have recently been compiled by Loftus (personal communication), April 1971.

Rather than compiling instance-to-category norms, an item-sorting task (Friendly & Glucksberg, 1969; Miller, 1969) was used. Twenty Princeton undergraduates were individually presented with a deck of 61 cards, each card containing one of the items from the initial pool. These subjects were asked to sort the items into anywhere from 1 to 20 piles, putting in the same pile those items which they felt "belonged together." A "miscellaneous" category was allowed for items felt not to belong in any of the groups they had formed. After completing the sort, subjects were asked to provide a word or short phrase to describe each of the piles they had formed. From these data, an agreement matrix was constructed, giving for each pair of items the number of subjects who had put both members of that pair into the same pile. The agreement score can be thought of as an indicant of the extent to which a given pair of items tends to evoke a common concept or category name, while the number of times a given item was placed in the miscellaneous category is an index of that item's uniqueness in the conceptual environment provided by the remaining words.

The agreement matrix was used to select items for the stimulus list. First, any word placed in the miscellaneous category by three or more pilot subjects was eliminated from the pool. Then, hierarchical cluster analysis (Johnson, 1967) of the agreement matrix was used to select items which would give empirical categories of roughly equal strength (average interitem agreement score). The stimulus items chosen in this manner are shown in Figure 5.

Apparatus. The list items were typed in upper case letters on mimeograph stencils which were then mounted in 35 mm. slide frames. A Kodak Carousel projector was used to project the slides onto a translucent glass screen placed 1.5 feet from S. The projector was placed behind the screen at a distance required to produce a letter image one inch high on the screen. A small green light inside S's cubicle was used to indicate the start of the recall period and remained on for the duration of the recall interval. A SONY stereo tape recorder was used to record S's oral responses. The slide projector and recall light were controlled automatically by a timing circuit. An intercom was used to present instructions to S.

Subjects. A total of 191 Princeton University students of both sexes was run in both sessions of the experiment. An additional 19 Ss participated in the OL session, but failed to return for the retention tests, and six Ss were discarded during OL due either to equipment failure or E error. The Ss were volunteers and were paid $3.00 for participating. Assignment of Ss to treatment conditions was random with respect to groups, but was not completely random with respect to retention interval. Due to the complexities of scheduling, it was frequently necessary to assign a given S to a particular retention condition, rather than to a randomly determined one.

Procedure

Original learning. All Ss were tested individually in a darkened cubicle. Standard multitrial free recall instructions were read to S and indicated the nature of the task, the number of trials, that the words belonged to an unspecified number of conceptual categories, and that the items could be recalled in any order. To ensure attention during presentation, S was asked to read each word aloud as it appeared on the screen. The 42 items were presented at a 2.25 sec. rate (1.5 sec. on screen, with .75 sec. for slide change).

When recall immediately follows the presentation of the last item in a list, there is a strong tendency for Ss to begin recall with the last few words presented (recency), regardless of the characteristics of these items. Since our interest focused on the stable organization imposed by S, independent of such transient effects, an attempt was made to minimize the recency effect. Studies by Postman and Phillips (1965) and Glanzer and Cunitz (1966) have demonstrated that the recency effect is eliminated if recall is delayed for 10 to 30 sec. after presentation and S is occupied with a task designed to prevent rehearsal. Therefore, a 10-sec. delay was introduced following list presentation, during which S was required to count backwards from a number which appeared on the screen following the last stimulus word. At the end of the 10-sec. interval, the green recall light in the experimental cubicle was illuminated and S was given 80 sec. for oral recall. Subjects were given 12 alternating presentation-recall trials with this procedure.

Following the original learning trials, Ss were given a questionnaire designed to identify any strategies which they had used. The results were quite complex and will not be reported here.

Retention and relearning. Subjects returned to the laboratory after 1, 5, 10, or 20 days, ostensibly to complete a set of pencil-and-paper tests of memory and cognition. When S arrived for the second session he was first returned to the experimental cubicle and instructed to recall all the words he could remember from the first session. Approximately one minute elapsed between the time S was seated in the booth and the retention test. After the 80-sec. interval allowed for recall, S was instructed that four additional study-test trials would be given on the same set of items with a procedure identical to the original learning session.

Following the relearning trials, five short tests of memory and verbal abilities, selected from the Structure-of-Intellect series (Guilford, 1967), were administered to S. While these tests were found to have some relations to within-group differences in the free recall task, they proved unrelated to the major experimental variables of interest. Therefore, they will not be discussed further here.

In a brief post-experimental interview, Ss were asked whether they had expected to be asked to recall the stimulus list in the second session, and whether they had practiced the material during the retention interval. Because of the possibility of ingratiation in self-report, an attempt was made to phrase the questions so that S would not be reluctant to report rehearsal, and any possible bias introduced would tend to work against the experimental hypotheses.

3.3 Results

The Ss' response protocols were transcribed from tape and punched onto data cards for analysis. A general multitrial free recall program (Friendly, 1971) was used to score the protocols and to perform the proximity analyses.

Original Learning

Performance. Acquisition scores in terms of mean number of words correctly recalled are plotted in Figure 6. Since no reliable differences were apparent among the random-block conditions (R1, R2, and R3), they have been combined in Figure 6 (as well as in other graphs where they do not differ) and denoted collectively as Group R. A multivariate analysis of variance (Clyde, Cramer, & Sherin, 1966) was performed to test the hypothesis of equal mean learning

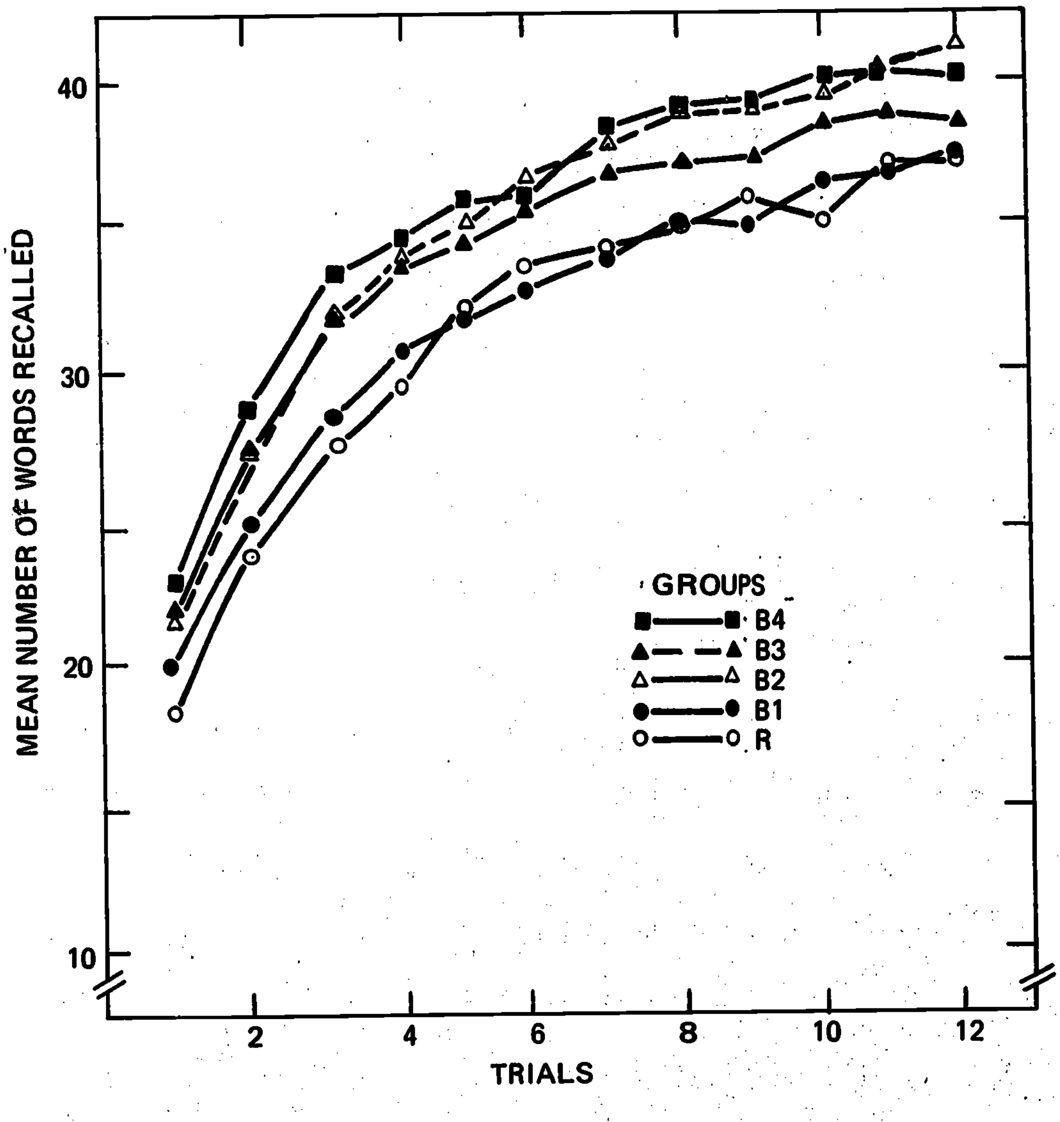


Fig. 6. Mean number of words recalled as a function of trials of original learning.

curves and to determine which trials contributed most to observed group differences. This analysis, as well as others reported below, included Groups (B1-B4, R2-R4), Experimenter, and Retention Interval as factors of classification and the trial-by-trial response measures as criteria. Overall tests based on Wilks' Λ criterion indicated that only differences due to Groups were reliable, $F(72,636) = 1.92$, $p < .01$.

To locate the source of group differences in acquisition, individual multivariate comparisons between groups were tested. In this analysis and others reported below, contrasts were chosen as orthogonal comparisons of Group B(I-1) minus the average of successive groups, i.e., B(I) to B(4). These comparisons are called Helmert Contrasts (Clyde et al., 1966). The essential result is that Groups B2, B3, and B4 differed from Groups B1 and R, $F(12,116) = 3.52$, $p < .001$, while neither the former set of three groups nor the latter set of two groups differed among themselves. The difference between experimental and control groups was highly significant on every trial by univariate tests, with F-ratios ranging between 10.0 and 28.8. Although differences among the experimental groups failed significance on the overall multivariate test, inspection of Figure 6 reveals that B2 and B4 recalled more words than B3 on all of the last 10 acquisition trials.

Total word recall was analyzed into two multiplicative components--number of categories recalled and number of items recalled per category (Cohen, 1966). A category was considered recalled if at least one member of the category was represented in output. The mean number of minor categories recalled did not differ across groups, the means ranging from 5.60 to 5.80 on Trial 1 and from 5.97 to 6.00 on Trial 12. The same results appeared when performance was scored in terms of the three superordinate categories.

Organization

Clustering. To what extent was blocking successful in differentially inducing Ss to organize at various levels of the stimulus hierarchy? This question may be answered in terms of measures of categorical organization (see section 1.2). The basic datum in these measures is the number of sequential repetitions, C, of items from the same category. The list used here can be thought of as comprising six 7-item categories or three 14-item categories. There are, therefore, two observed clustering scores, C_6 and C_3, for every subject-trial protocol. Since it was desired to make comparisons across groups for a given number of categories (6 or 3) and across categories for particular groups, the category repetition measures were standardized to a statistic,

$$\frac{C_k - \min(C_k)}{\max(C_k) - \min(C_k)} ,$$

suggested by Dalrymple-Alford (1970), which ranges from 0.0 (minimum clustering) to 1.0 (maximum clustering).[11] $\text{Min}(C_k)$ and $\max(C_k)$ are the

[11]The major virtue of this measure is that it allows comparison of clustering when the number of categories vary, since the values computed are always on the same scale. This is an attractive feature for graphical presentation, not shared by other measures of categorical clustering which the author nevertheless believes to be conceptually more sound. These are

$$z_1(C_k) = \frac{C_k - E(C_k | n_1, n_2, \ldots, n_k)}{\sigma(C_k | n_1, n_2, \ldots, n_k)} ,$$

and

$$z_2(C_k) = \frac{C_k - E(C_k | N)}{\sigma(C_k | N)} ,$$

minimum/maximum possible numbers of category repetitions which could be obtained by rearranging the items actually recalled on a given trial; *k* is the number of categories in the list.

The mean values of this category repetition statistic over the 12 trials of OL are plotted in Figure 7. Scored in terms of six categories (panel A) the graph indicates that groups receiving the items blocked at level 3 of the hierarchy (B3 and B4) cluster to a far greater extent than the other groups, multivariate $F(12,116) = 15.00$, $p < .001$. Only the Groups factor produced significant overall differences, $F(72,636) = 1.68$, $p < .001$. By Trial 12, Groups B1, B2, and R had not reached the same degree of clustering achieved by B3 and B4 at the second trial.

A similar pattern emerges when the data are scored in terms of the three superordinate categories (panel B). The major difference is that Group B2 in this analysis clusters to the same extent as B4. The contrast between experimental and control groups was highly significant, $F(12,116) = 5.84$, $p < .001$. In addition, B2 and B4 displayed more clustering on the last 10 trials than did B3, so that the relative ordering of B2 and B3 is opposite in the two analyses.

---

where N is the total number of words recalled, and n_i is the number of items recalled from category i, with $\Sigma n_i = N$. The expected values and standard errors are specified by the theoretical sampling distributions of C_k under two different null hypotheses of no clustering, in one case where the n_i are considered as fixed constants (z_1) and the other where they are considered to be random variables (z_2). The expected value under the null of z_1 was proposed by Bousfield and Bousfield (1966), while z_1 itself was first suggested by Hudson and Dunn (1969). Exactly what is meant by "chance clustering" is thus made perfectly explicit. Analyses parallel to those reported for the present data in terms of Dalrymple-Alford's 0-1 measure were carried out using z_1 and z_2. Essentially the same results were obtained with all three measures in the analyses reported here.

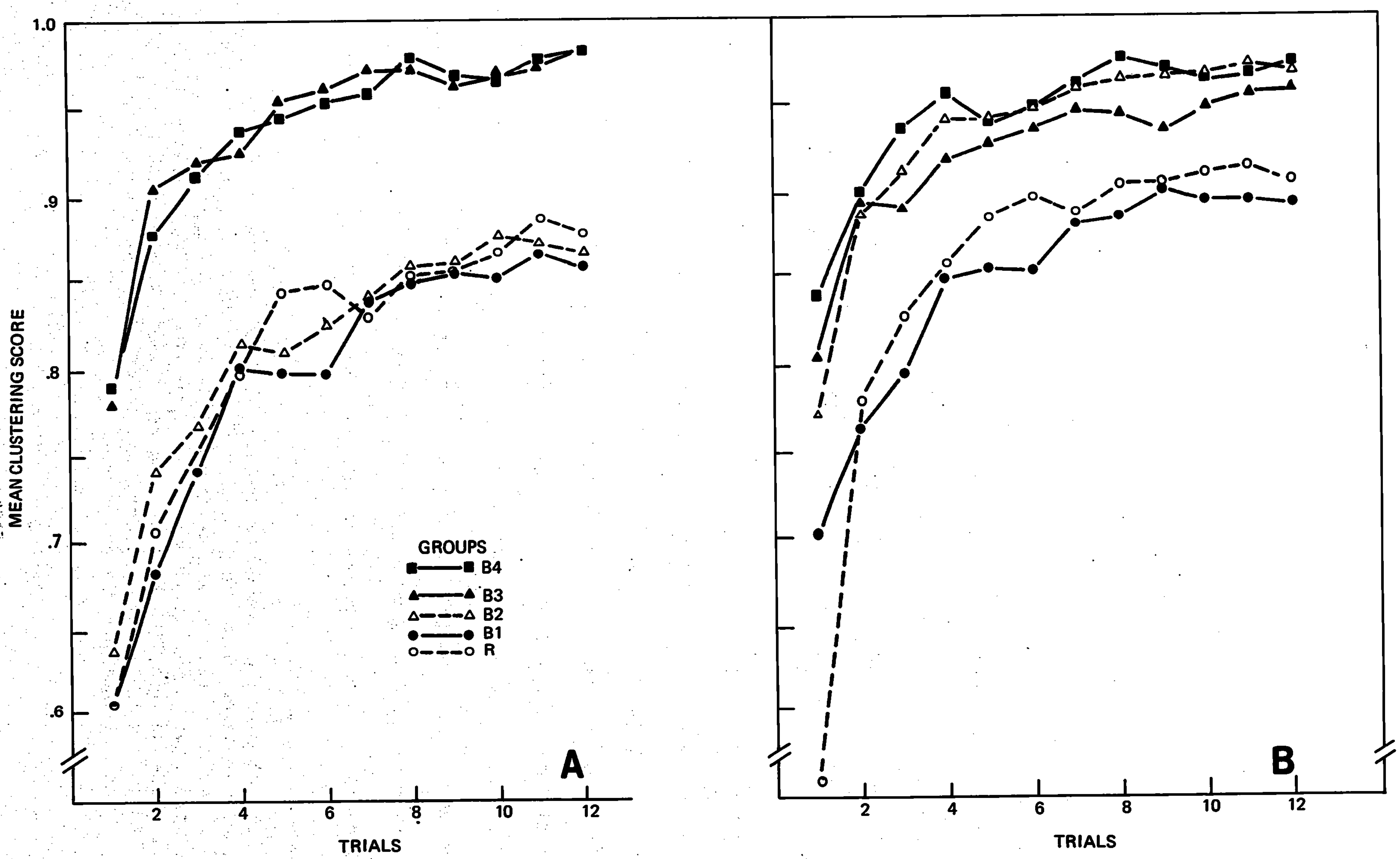


Fig. 7. Categorical organization scores, (C - min)/(max - min), as a function of trials of original learning. In panel A the data are scored by the six terminal categories in Fig. 16; panel B shows the result when the data are rescored in terms of the three superordinate categories.

This pattern of results is exactly what one would expect if all the B-groups organized their recall according to the blocked structure of presentation: B4 recalls according to six categories nested into three superordinates, so their clustering performance is high regardless of which way it is scored; B3 subjects group the items in terms of six independent categories, and their clustering drops somewhat when scored by the superordinate classes; B2 only clusters to a high degree in terms of three categories, while clustering in Groups B1 and R is uniformly low. These results do not depend on the particular clustering statistic used (see footnote 11). On the basis of these measures of sheer amount of organization, it appears that blocking of the list produced the desired effect of inducing Ss to group the items in alternative ways.

Proximity analyses. Average interitem proximities were computed for each B group over all acquisition trials, and analyzed by the hierarchical clustering procedure.[12] The diameter method solutions are shown in Figures 8-11. The filled circles indicate those clusters which emerged identically in the diameter and connectedness method solutions.

These analyses largely confirm the results obtained above with the measures of amount of organization but also reveal that the modal organization in Groups B2 and B3 was not restricted to a single level of the hierarchy (see Table 1) as the discussion above might imply. That is, Ss in Group B2

---

[12]The cluster analyses described in this chapter were performed using the Gruvaeus-Wainer (pers. comm.) algorithm. One deficiency of Johnson's (1967) program is that the clustering result is not invariant under permutation of the rows and columns of the proximity matrix. The Gruvaeus-Wainer program corrects this deficiency, but gives results otherwise identical to those obtained with Johnson's program. I am grateful to Gunnar Gruvaeus for making a copy of this program available.

(Figure 9) to some extent tended to subdivide the three input blocks into the level 3 categories. Also, B3 Ss (Figure 10) tended slightly to recall the six input blocks in pairs, according to the classification of the words at level 2. Thus the differences in organizational structure among these four groups reflected differing relative strengths of the category systems at level 2 and level 3 of the stimulus hierarchy.

Rather than examining the actual clusterings determined for these groups, the differences in organization can be better illustrated in terms of the measures of compactness and isolation which are derived from the clusterings (see section 2.6). The average diameters of clusters at both levels of the hierarchy were obtained from each group M-gram, and are displayed in Figure 12. For each group of Ss, the total height of the bar represents the mean diameter of the major categories. The shorter the bar, the more tightly-knit is the organization at this level. The average diameters of the minor categories are indicated by the filled portion of the bar, while the length of the unfilled portion indicates the isolation or separation of these two modes of organization.

It can be seen that the strength of organization in terms of the minor categories increases steadily (diameters decrease) from Group B1 to Group B4. A different picture is presented in terms of the diameters at level 2 and the degree of separation between the two organizational schemes. Subjects whose presentation was blocked at level 2 (B2) have the strongest organization at this level (shortest total height) and their clusters at level 3 are the least isolated. The reverse situation holds for Ss receiving independent blocks at level 3 (B3): they display the weakest

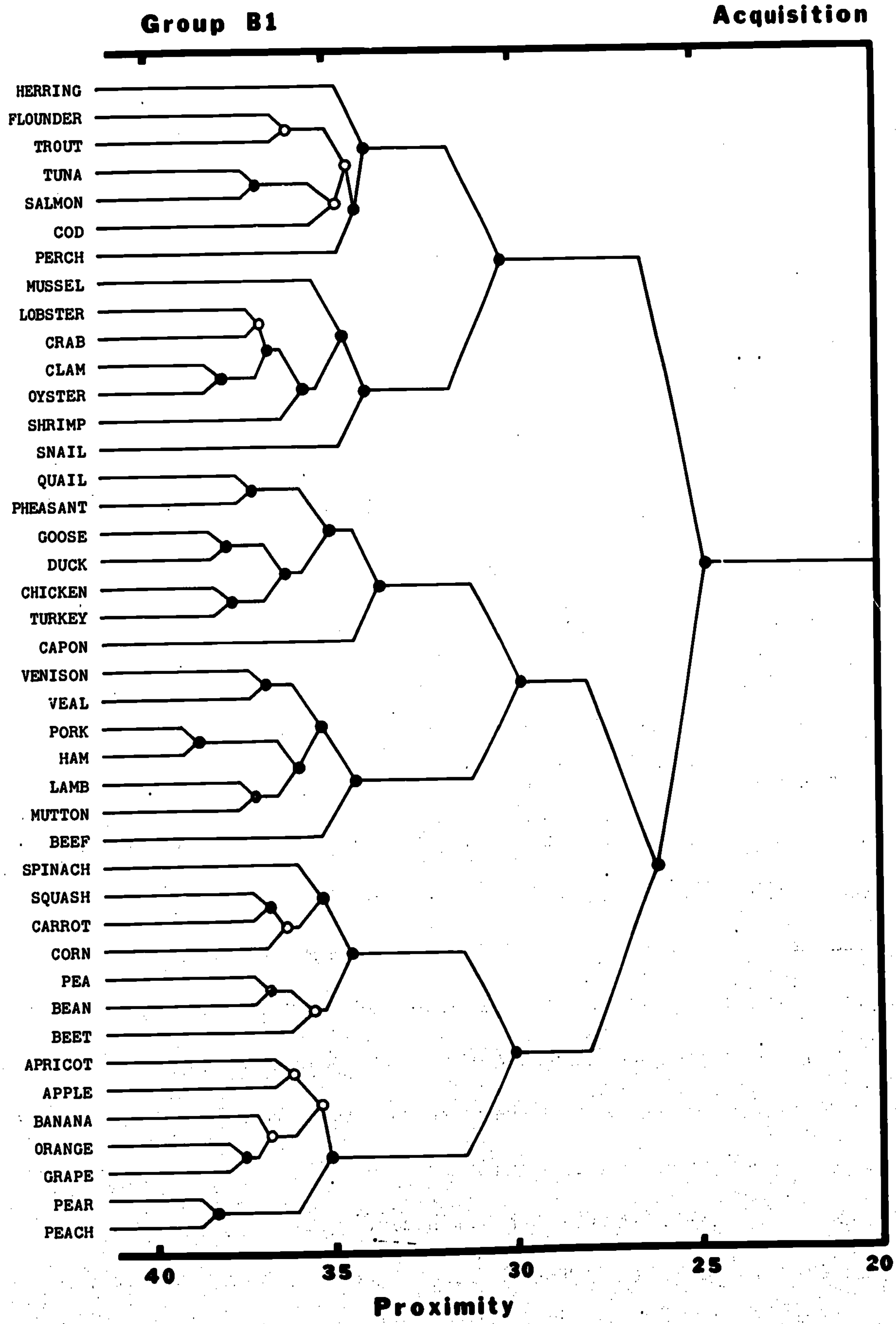


Fig. 8. Organizational structure for Group B1 in original learning. Data pooled over Ss and trials.

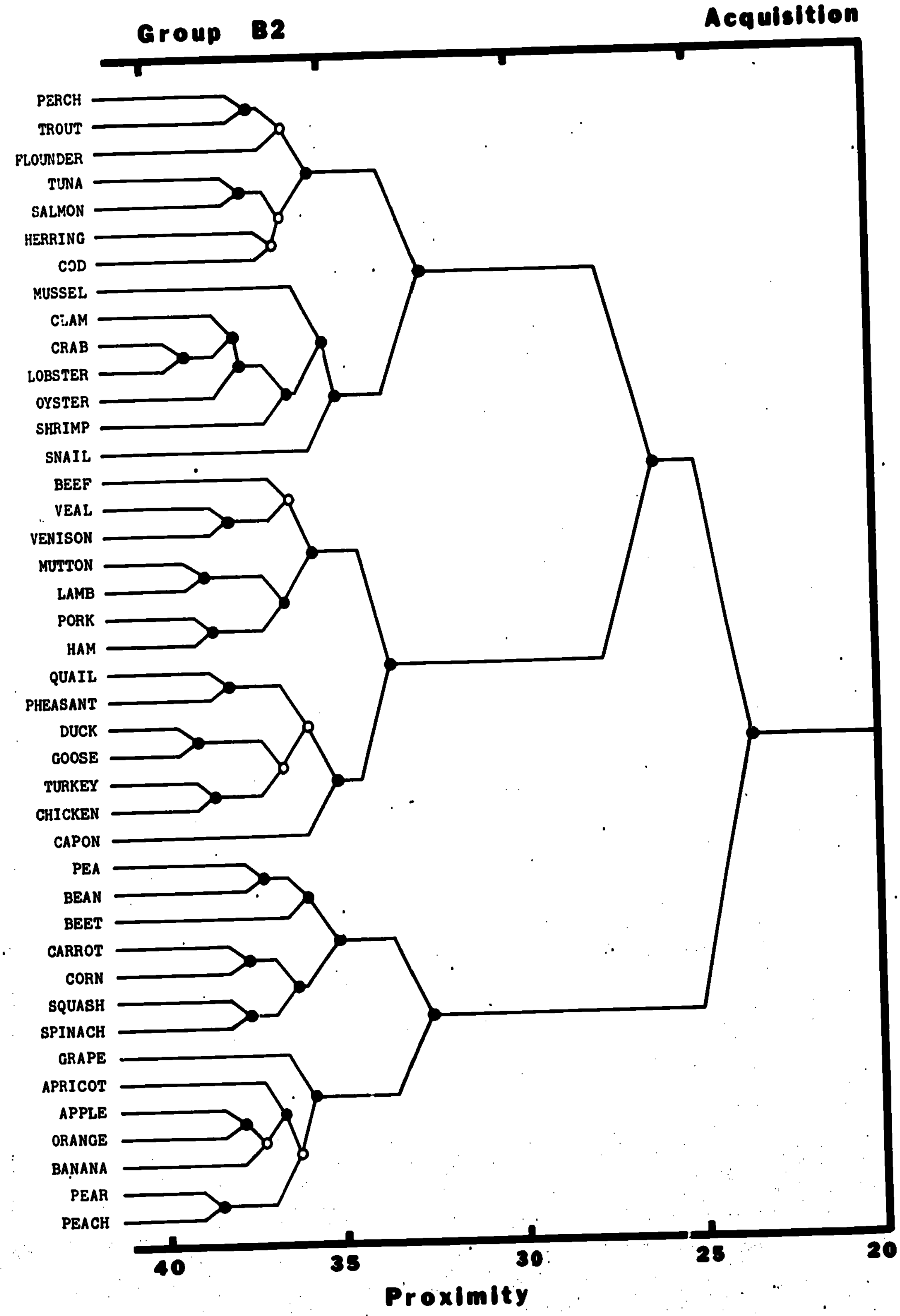


Fig. 9. Organizational structure for Group B2 in original learning. Data pooled over Ss and trials.

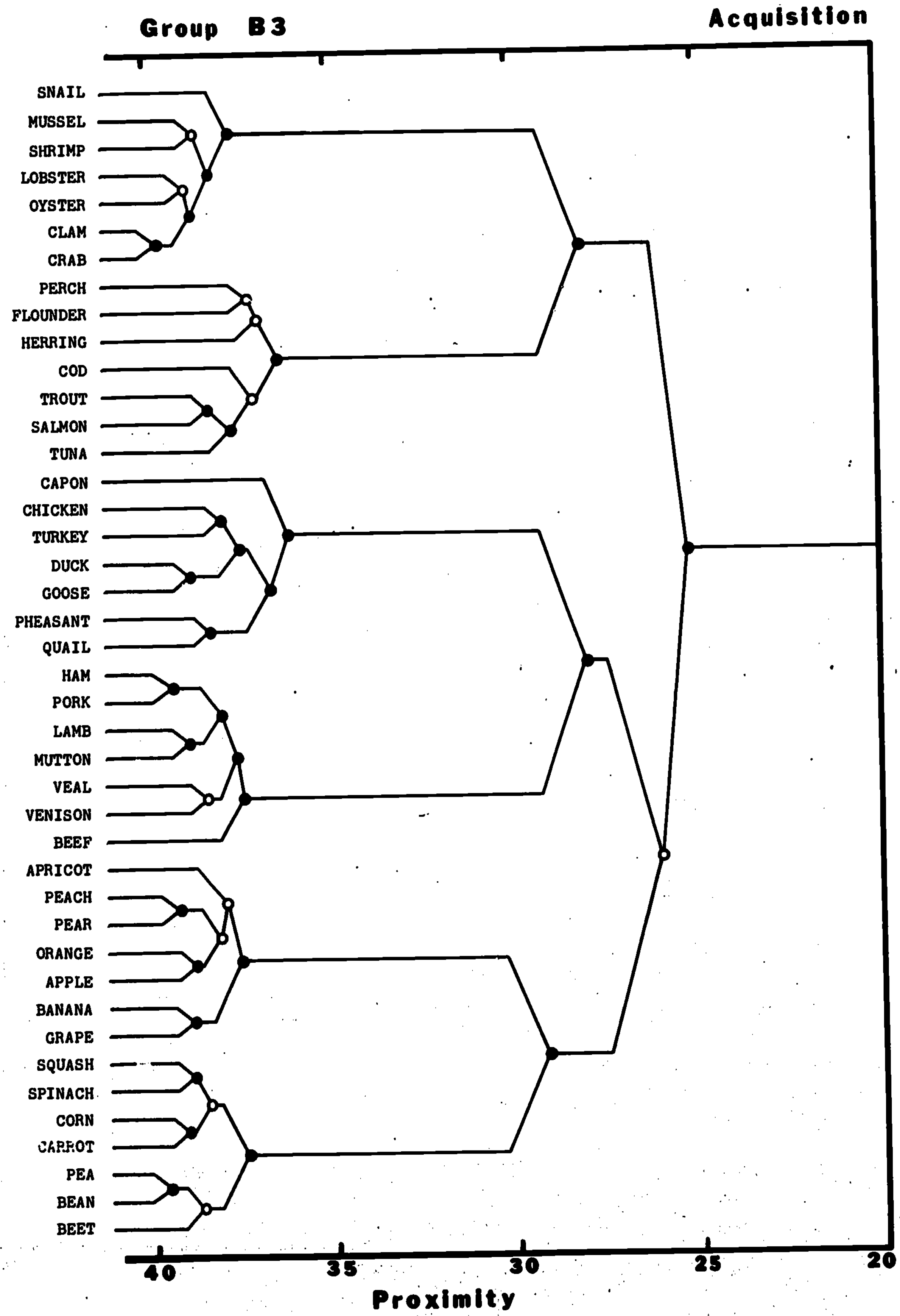


Fig. 10. Organizational structure for Group B3 in original learning. Data pooled over Ss and trials.

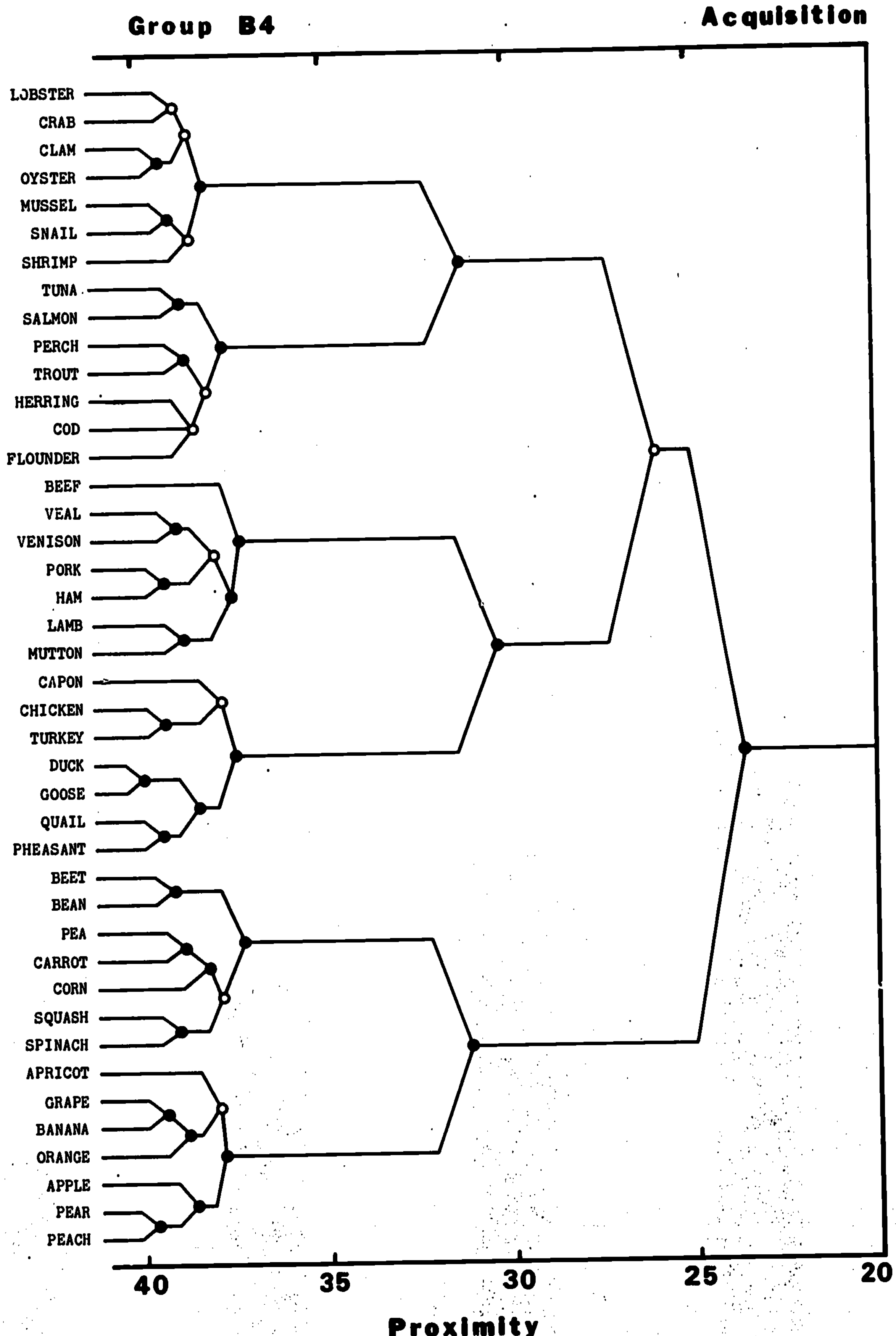


Fig. 11. Organizational structure for Group B4 in original learning. Data pooled over Ss and trials.

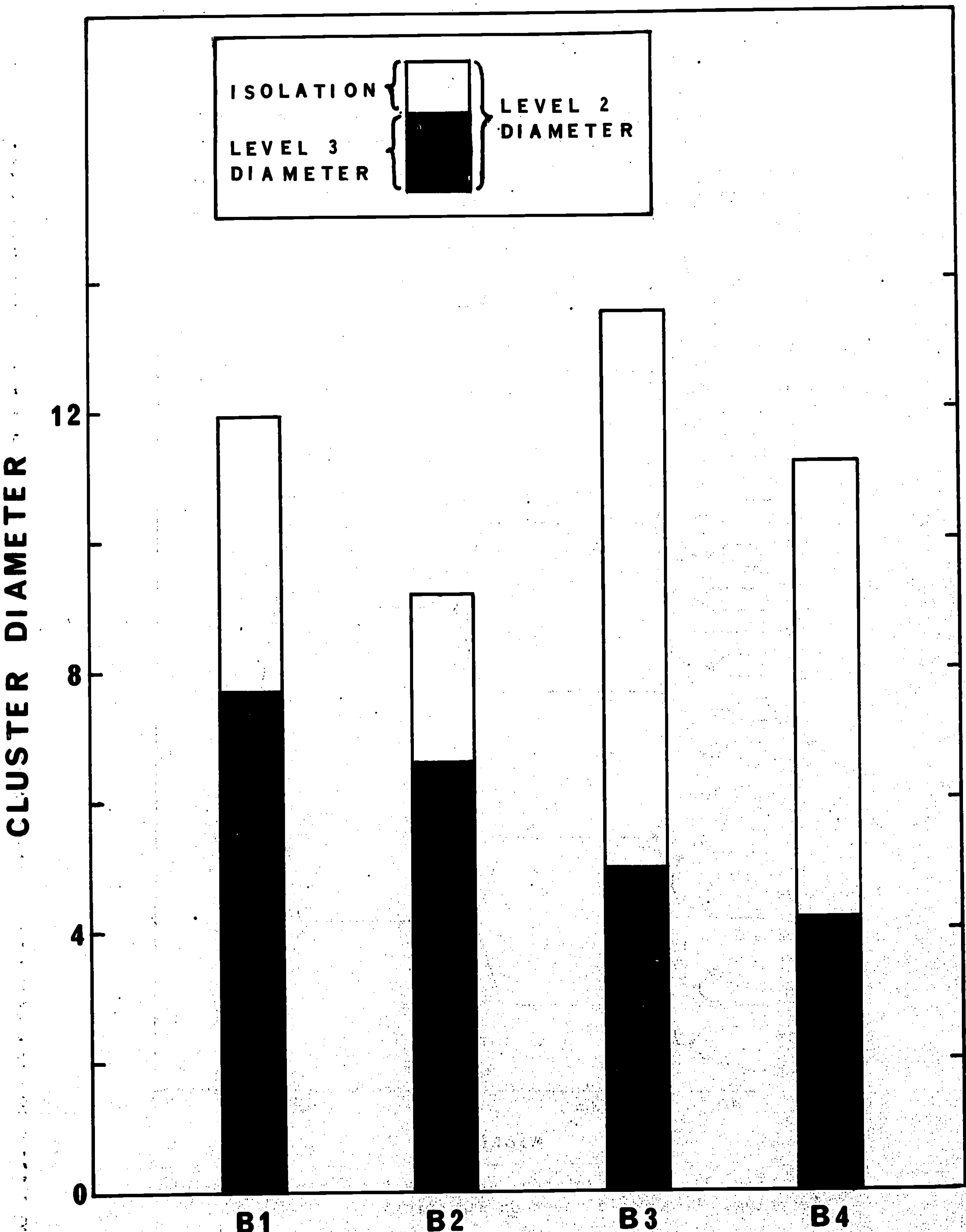


Fig. 12. Category diameters from cluster analyses of group proximities in acquisition, Trials 1-12.

organization at level 2 and the greatest isolation between the two category systems.

Intragroup differences. The analyses of organizational structure described above were based on the average proximities for each group, and therefore reflect the aspects of organization common to each group as a whole. To determine the extent to which individuals within a group differed in their patterns of organization, the proximity procedure was applied to the protocols of each S in the B groups. Inspection of the individual M-grams revealed variation across Ss in several aspects of their hierarchies.

It proved difficult, however, to extract any meaningful generalities, or to gauge the degree of intersubject variation with precision. Therefore a procedure developed by Gruvaeus and Wainer (see Appendix B) was used to obtain correlations between the tree structure clustering solutions for all pairs of Ss in a group. In general, the correlations were quite high; the median intersubject rank correlations for Groups B1 to B4 were .65, .74, .80, and .83, respectively. Thus, although all groups learned the same set of words, as the degree of structure present in the input order increased, so too did the agreement among subjects in the structure of their organization.

The inter-S correlations are measures of the similarity of their organization. It is possible, therefore, to apply the clustering procedure to the Ss themselves to reveal the presence of subgroups sharing a common pattern of organization. The average proximity matrix for each group was included in this analysis as a point of reference. The results of this analysis showed that within a given group, rather than forming homogeneous subgroups, Ss tended to vary in the degree to which their organization resembled the modal organization for the group.

Each group of Ss was divided into roughly equal halves--those whose organization was most like ("central") and least like ("remote") the average for the group. Pooling the proximities within each subgroup separately, it was found that the remote Ss differed mainly in that their organization was less cohesive (compact) at the level of the minor categories of the hierarchy (see Figure 13). However, some qualitative differences between remote and central Ss in the pattern of organization were apparent. For example, most of the Ss classified as remote in Group B2 organized the items according to some or all of the three major categories with little subgrouping according to the minor categories. Many of the remote B3 subjects also organized primarily at one level--that of the minor categories.

The category diameters determined for these subgroups appear in Figure 13 which also shows performance in recall, averaged over trials for each subgroup. Comparison of the shaded portions of the two panels shows that recall varies directly with the cohesiveness (inversely with diameters) of the level-3 categories. The recall results are quite surprising. They indicate that the difference in recall between subgroups determined empirically within a given experimental group is approximately as large as the range of mean recall scores across all groups in this experiment (cf. Figure 6). Since all Ss within a given experimental group are treated identically, and since the use of categorized words tends usually to reduce inter-subject variability (Marshall, 1967), it may be that the magnitude of individual differences in free recall has been vastly underestimated.

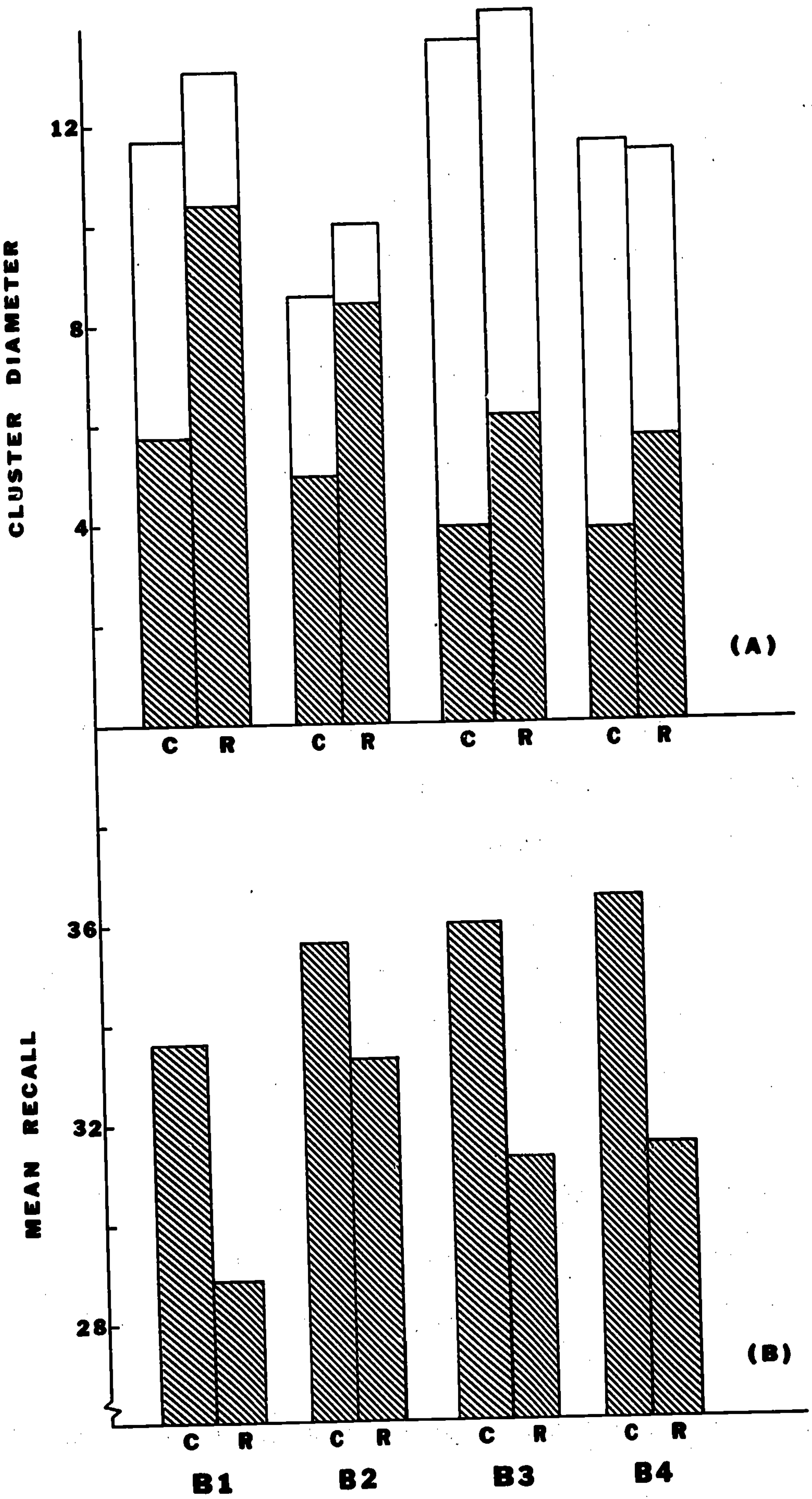


Fig. 13. Category diameters and mean recall for empirically isolated subgroups. C = Central Subgroup; R = Remote Subgroup. Panel A shows mean diameters of minor categories (shaded portion) and major categories (total height). Panel B shows average recall over all trials of OL for the subgroups.

Retention and Relearning

Recall. Mean retention (Trial 1 of Session II) for the B-groups--expressed as a percentage of recall on the final trial of Session I--is plotted in Figure 14. (The groups receiving randomly blocked presentation are not shown but retained amounts intermediate between B2 and B1 at one- and five-day intervals.) Both Retention Interval and Groups were between-S factors so that each point represents the mean of a different set of eight or more Ss. In general, retention is at a relatively high level throughout with a grand mean of 82% over all groups and retention intervals. An analysis of variance performed on the number of words recalled on the retention trial is summarized in Table 2. Groups differed reliably on the retention trial, $F(6,107) = 3.22$, $p < .01$. These differences were largely accounted for by a comparison between B1 and Groups B2, B3, and B4. The greatest source of variation was that associated with retention interval, $F(3,107) = 17.78$, $p < .005$. As is evident from Figure 14, the decrease in amount retained over time is for the most part linear, with a first-degree orthogonal polynomial accounting for 88% of the sums of squares due to RI. Although the retention of B4 Ss appears to decline at a slower rate over the long retention intervals, the interaction of B-groups with retention interval was not large enough to cause rejection of parallelism.

Studies of retention are frequently prone to methodological difficulties which affect interpretation. In the present instance, different groups of Ss learned the items under presentation conditions which differentially facilitated the performance of the experimental groups; these groups also recalled the greatest amount on the first trial of Session II. Underwood

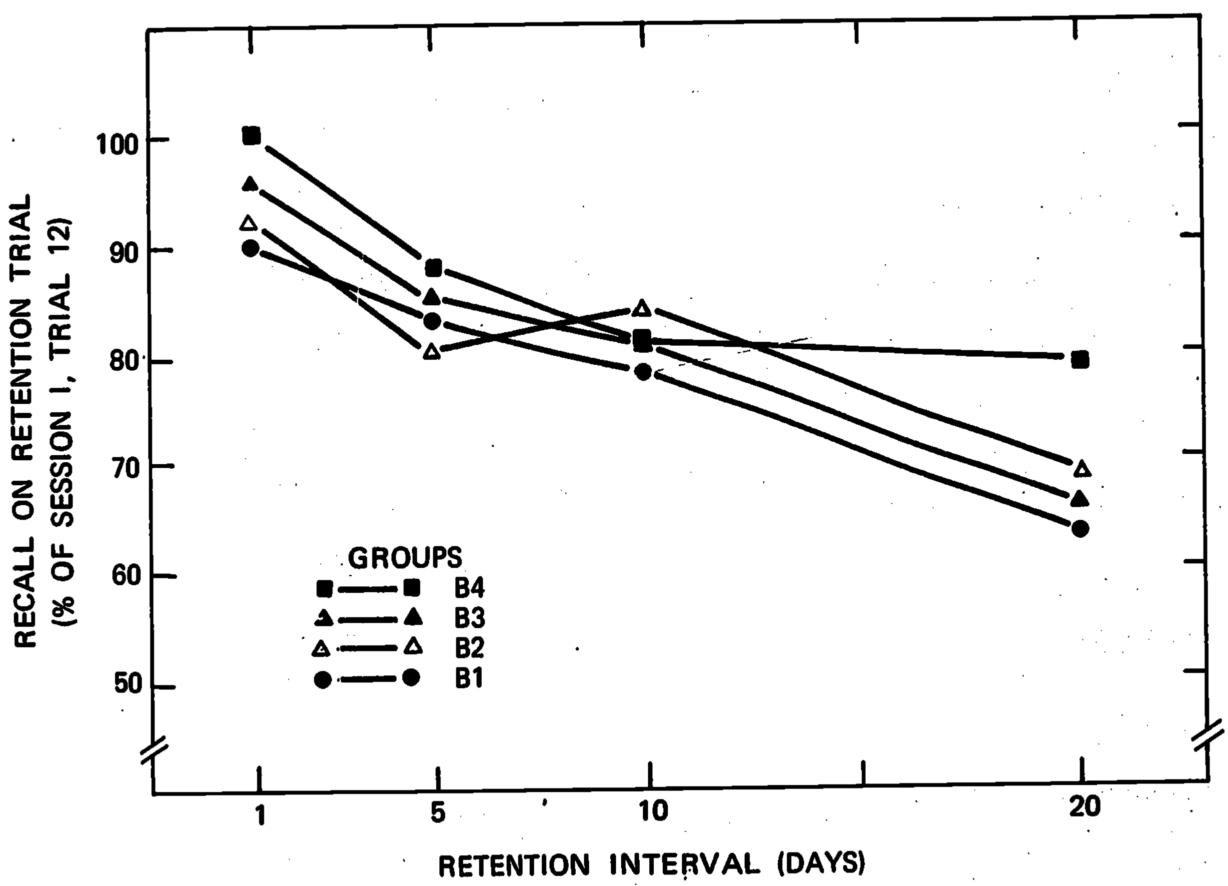


Fig. 14. Recall on the retention trial as a percentage of recall at the end of OL.

Table 2

Summary of Analysis of Variance Performed on Number of Words Recalled on the Retention Test (Trial 1) of Session II

Source	df	MS	F
Groups (G)	6	(121.571)	3.22**
G1: B4 vs. B3	1	110.471	2.92
G2: (B3, B4) vs. B2	1	2.355	0.06
G3: (B2, B3, B4) vs. B1	1	389.959	10.32***
G4: Among R-groups	2	65.695	1.74
G5: Remainder	1	95.271	2.52
Retention Interval (R)	3	(671.915)	17.78***
R1: Linear	1	1775.890	46.98****
R2: Quadratic	1	67.662	1.79
R3: Cubic	1	172.189	4.56*
Experimenters (E)	3	32.720	0.87
B-groups x R (G1R + G2R + G3R)[a]	9	28.002	0.74
G x E	18	44.869	1.19
R x E	9	27.472	0.73
Residual	31	39.878	1.05
Within Cells	107	37.798	--

*$p < .05$

**$p < .01$

***$p < .005$

****$p < .001$

[a]The design of the study precluded extraction of the complete interaction of G x R.

(1964) has argued that such differences in rate or final level of acquisition make it difficult to regard subsequent differences in recall as reflecting greater retention per se rather than just the degree to which the materials were learned initially. The present study is further complicated by the fact that, in the post-experimental interview, some Ss did report rehearsing the words during the retention interval.

A limited solution to these difficulties may be obtained by an analysis of covariance. Reported rehearsal may be regarded as random with respect to the treatment conditions, since it showed no relation to groups, $\chi^2(6) = 4.51$, $p < .50$, or to RI, $\chi^2(3) = 6.31$, $p < .10$. Covariance analysis using the acquisition scores as concommitant variables is appropriate for determining whether, apart from any differences in OL, differences in retention also exist according to the conditions of training (Cochran & Cox, 1964). That is, are the effects of the organization of materials on long-term retention simply a reflection of their effects on performance during OL, or is there something more?

The relevant data appear in Table 3. Taken together, these covariables are strongly related to the amount retained, as indicated by F - value for regression, $F(3,104) = 24.48$, $p < .001$. Two additional analyses were then performed to determine which of these covariables were related to retention. In one, rehearsal alone was covaried and yielded an F - value for regression less than 1.0 while Groups remained significant, $F(6,106) = 3.71$, $p < .005$. In the second, only the OL recall scores were covaried and both regression and Groups were significant. Thus, only the recall scores were significant predictors of retention. Because Groups remain significant, however,

Table 3

Summary of Analysis of Covariance Performed on Number of Words Recalled on the Retention Test (Trial 1) of Session II (Covariates: Number correct on last two trials of OL and Reported Rehearsal)

Source	df	MS	F
Regression	3	558.025	24.48****
Groups (G)	6	(86.944)	3.82***
G1: B4 vs. B3	1	80.782	3.54
G2: (B3, B4) vs. B2	1	47.390	2.08
G3: (B2, B3, B4) vs. B1	1	188.733	8.28***
G4: Among R-groups	2	85.786	3.76*
G5: Remainder	1	33.176	1.46
Retention Interval (R)	3	(629.484)	27.81****
R1: Linear	1	1762.876	77.35****
R2: Quadratic	1	84.081	3.69
R3: Cubic	1	41.476	1.82
Experimenters	3	4.958	0.22
Residual	67	26.287	1.16
Within Cells	104	22.791	--
<u>Raw Regression Weights</u>			
Rehearsal	1.608		
OL, Trial 12	.758		
OL, Trial 11	.165		

*$p < .05$; **$p < .01$; ***$p < .005$; ****$p < .001$

$F(6,104) = 3.82$, $p < .005$ (Table 3), when final acquisition differences are removed, differences among the groups in retention are not attributable merely to the differences during Session I. (On the other hand, the cubic component of trend in retention can be attributed to OL differences since this effect fails significance in the analysis of covariance.) It may be concluded that variation in the amount retained is more than a simple reflection of the residual effects of inequalities in the degree of original learning.

Relearning. Performance in relearning is shown in Figure 15 (Trials 2-5) for each treatment combination of presentation condition and retention interval. (Trial 1 is the retention test.) A repeated-measures analysis of variance was performed on the data for the B-groups[13] (Table 4) and revealed significant effects due to both Groups, $F(3,112) = 7.38$, $p < .001$, and retention interval, $F(3,112) = 3.41$, $p < .02$.

To provide more detailed information on the course of relearning, an orthogonal polynomial trend analysis, summarized in Table 4, was also performed on these data. The overall interaction of RI and Trials was highly significant, $F(12,448) = 12.06$, $p < .001$. This interaction can be seen more clearly in Figure 16, in which the B-group curves from Figure 15 have been pooled at each retention interval. As is evident from Figure 16, the RI groups differed significantly in the slopes (linear trend) of their relearning curves, $F(1,112) = 15.39$, $p < .001$, as well as in curvatures, $F(9,336) = 8.74$,

[13]The cell ns were equated for this analysis. By reference to a table of random digits, a total of 11 out of 139 Ss were deleted from the 16 B-group-RI cells.

Table 4

Summary of Analysis of Variance and Trend Analysis Performed on the Number of Words Recalled in Session II[a]

Source	df	MS	F	p
Between Ss				
Groups (G)	3	518.49	7.38	<.001
Retention Interval (R)	3	239.37	3.41	<.020
G x R	9	31.80	0.45	
Ss (G x R)	112	70.22		
Within Ss[b]				
Overall				
Trials (T)	4	1418.05	181.39	<.001
G x T	12	7.67	0.99	
R x T	12	93.65	12.06	<.001
G x R x T	36	6.86	0.88	
Ss (G x R) x T	448	7.77		
Linear				
T	1	4443.93	299.23	<.001
G x T	3	19.19	1.29	
R x T	3	228.51	15.39	<.001
G x R x T	9	11.73	0.79	
Ss (G x R) x T	112	14.85		
Curvature				
T	3	409.42	75.33	<.001
G x T	9	3.83	0.71	
R x T	9	47.59	8.74	<.001
G x R x T	27	5.24	0.96	
Ss (G x R) x T	336	5.44		
Total	639			

[a]Analysis of B-groups only, with number of Ss per cell equated.

[b]The same significance levels result from a conservative test (Greenhouse & Geisser, 1959) applied to within-S effect.

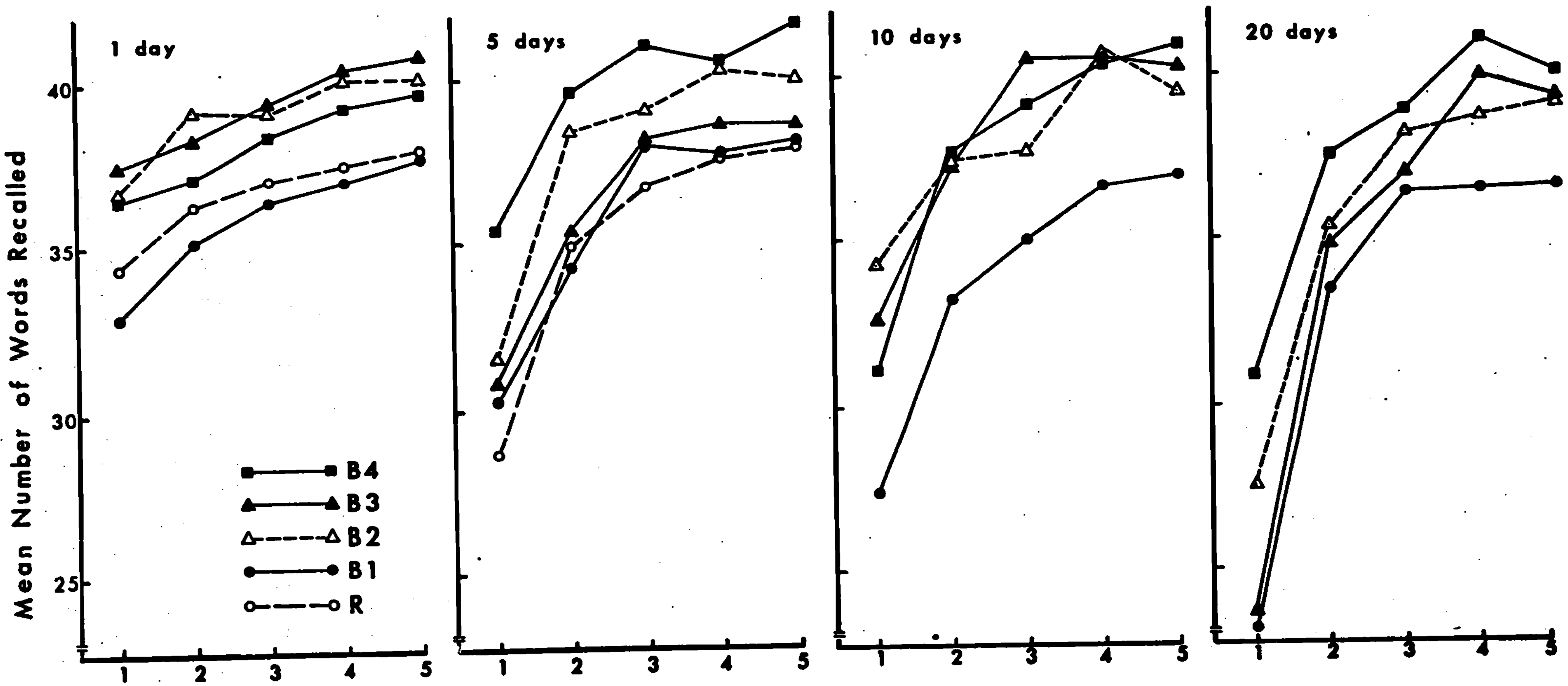


Fig. 15. Mean number of words recalled by Groups and by Retention Interval as a function of trials of Session II.

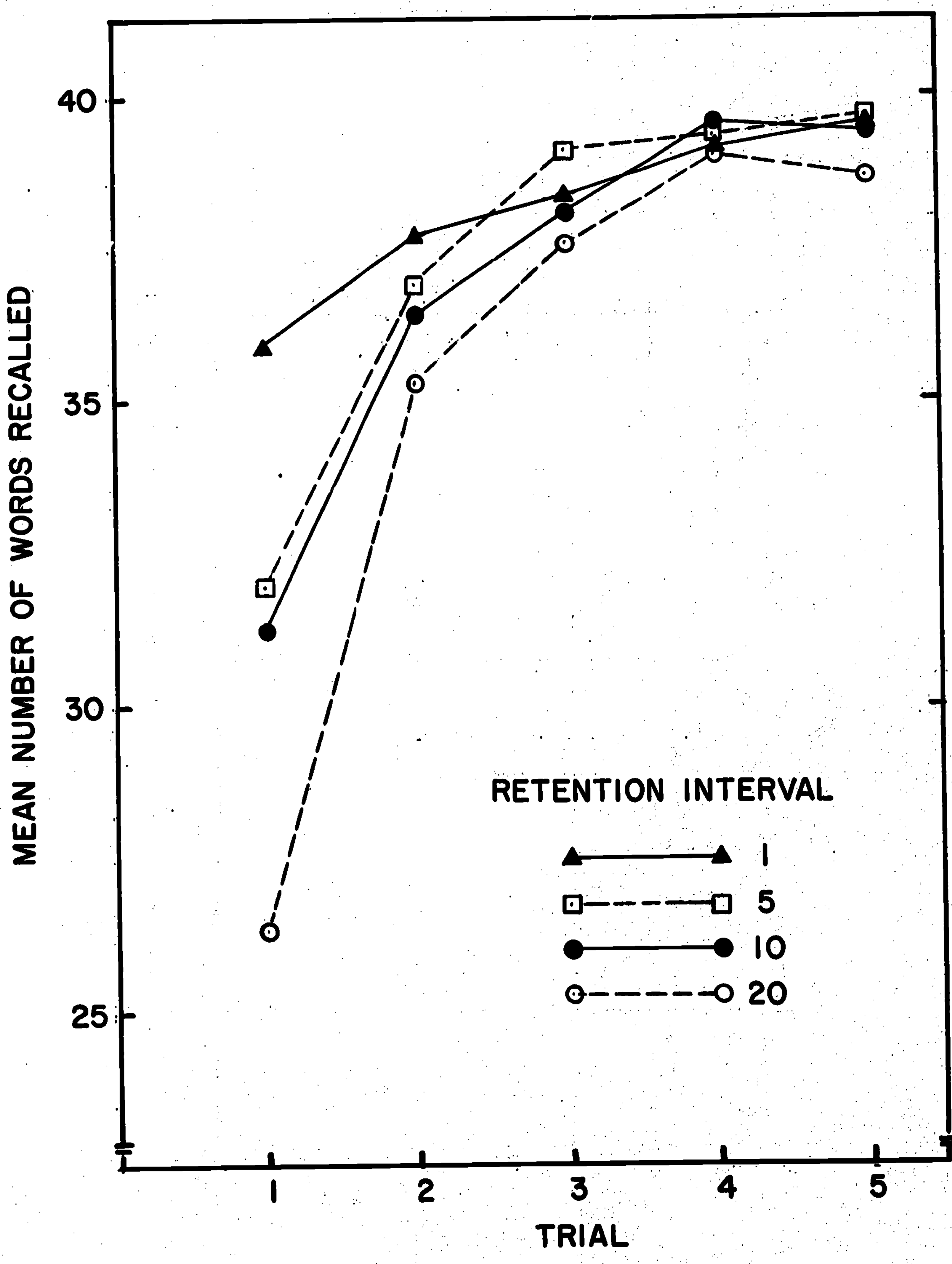


Fig. 16. Mean recall as a function of retention and relearning trials for groups varying in retention interval.

$p < .001$. None of the interactions with Groups proved significant, indicating that at any given retention period all training groups relearned the items at roughly constant rates. Thus, the effect of retention period dissipates over relearning trials (Figure 16) but group differences, in general, remain.

An analysis of categories represented in recall and the number of words recalled per category indicated that all of the above effects in retention and relearning reflected differences in within-category recall. Category recall was virtually perfect, even on the retention trial after the longest intervals.

Organization

Measures of categorical clustering were computed for the relearning data in a similar fashion as for original learning. Mean clustering by Groups in terms of six categories using the standardized $z_1(C_k)$ measure (see footnote 11) is shown in Figure 17. The same data are replotted in Figure 8 with RI as a parameter. A multivariate analysis of variance performed on these data revealed significant overall effects due to Groups, $F(30,414) = 1.54$, $p < .05$, and Retention Interval, $F(15,284) = 4.69$, $p < .001$. Testing particular contrasts in the group main effect indicated that the following differences among groups contributed to the effect: Group B1 relearning showed significantly less clustering by six categories than Groups B2, B3, and B4, $F(5,155) = 3.77$, $p < .004$; Group B2 in turn clustered less than B3 and B4, $F(5,155) = 3.01$, $p < .02$, while B3 and B4 did not differ $F(5,155) = 0.422$, $p > .10$. In the first two comparisons the groups differed in clustering on every trial by univariate tests, each $p < .005$, while B3 and B4 were not reliably different on any trial. Groups varying in retention period differed substantially in categorical organization on the first trial of Session II, $F(3,107) = 8.21$ $p < .001$, but did not differ thereafter (Figure 18). When the data were rescored for clustering according to the three superordinate categories, the same results obtained as in OL: Groups B2, B3, and B4 clustered more than B1, but did not differ among themselves.

Proximity analyses. The correspondence between amounts of organization and of retention exhibited in these data provide some confirmation for the idea that retention depends upon the maintenance of a stable category system. A clearer view of the organization which persists over the retention period can be provided by the proximity technique.

Proximity analysis was applied to the pooled group data from Trial 1 of the second session. The resulting diameter method cluster analyses for Groups B1 and B4 are shown in Figures 19 and 20. Again, filled circles indicate those clusters common to the diameter method and connectedness method solutions. In general, the proximities for all four groups conform reasonably to the ultrametric inequality and therefore may be adequately represented by a tree structure. The measure of badness-of-fit to a hierarchy, suggested in Appendix B, gives values of 5.0%, 3.2%, 3.6%, and 1.8% for Groups B1, B2, B3, and B4, respectively.

Comparison of the group clustering solutions indicates that the four groups do not differ in the overall structure of organization on the retention trial. In all four M-grams the items are clustered "appropriately" into the six minor categories, which in turn are nested into the three

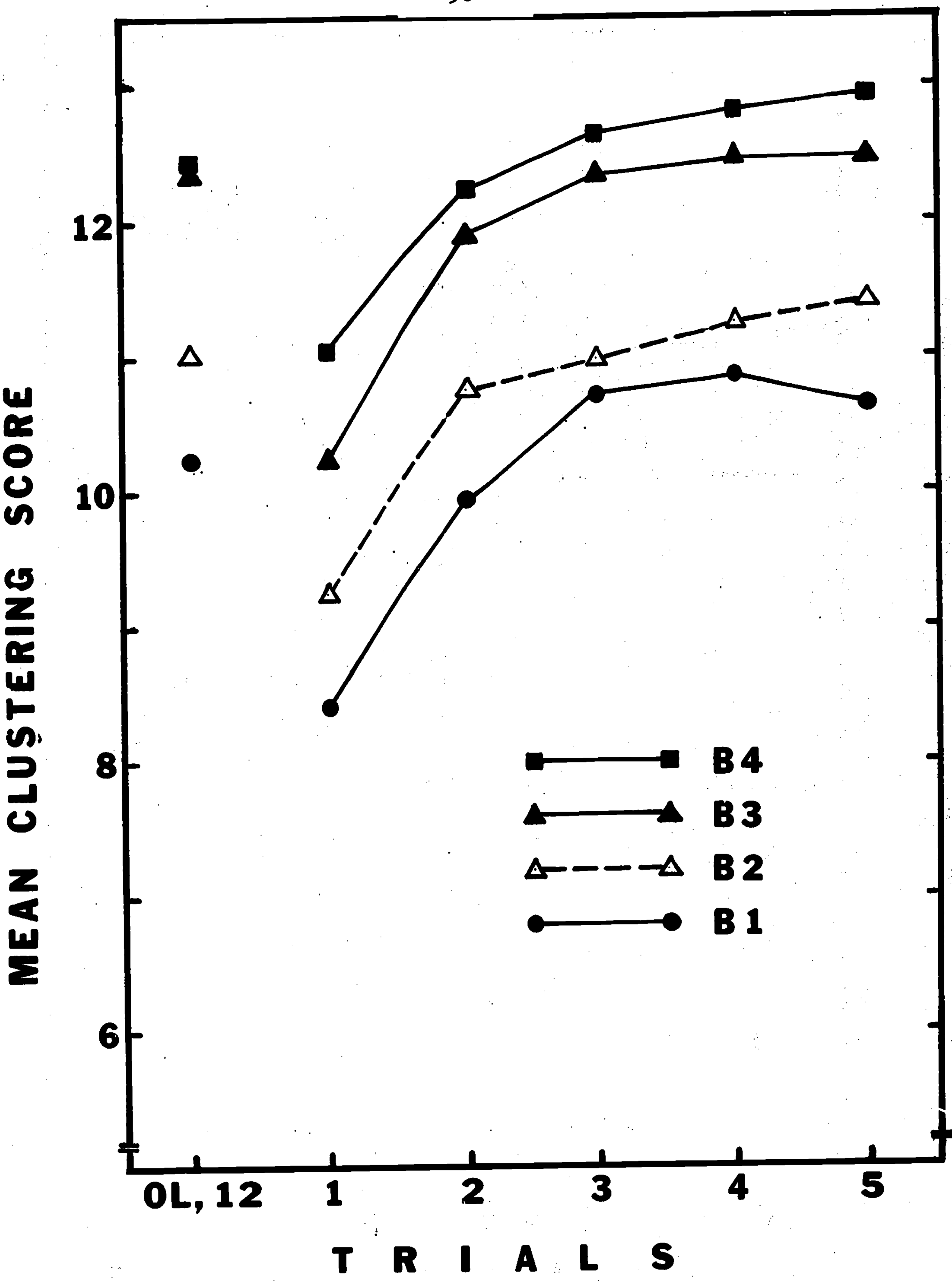


Fig. 17. Mean categorical clustering, $[C_6 - E(C_6)]/\sigma(C_6)$, in retention and relearning by Groups, pooled over all retention intervals. The unjoined points represent the clustering scores on the last trial of original learning.

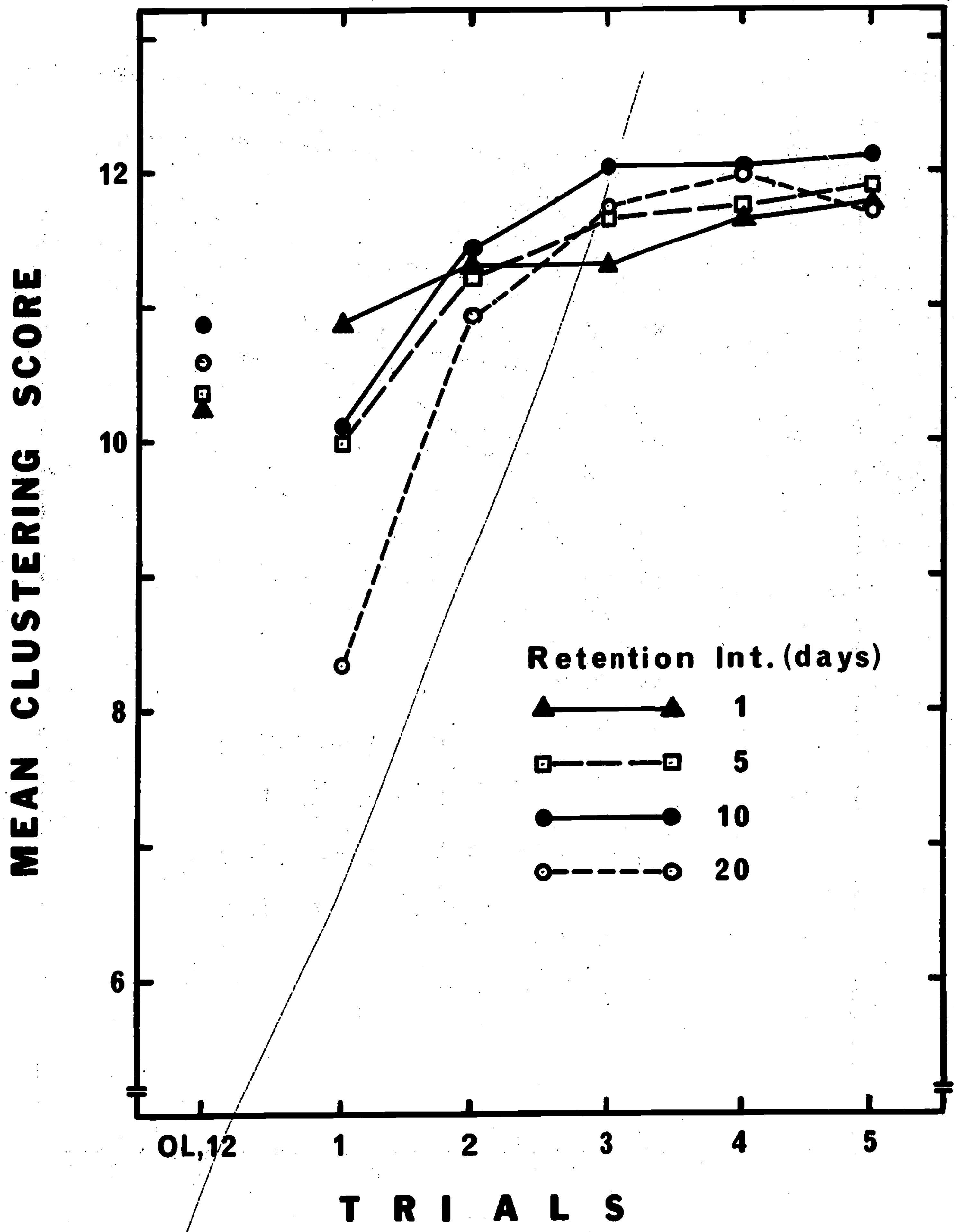


Fig. 18. Mean categorical clustering, $[C_6 - E(C_6)]/\sigma(C_6)$, in retention and relearning by Retention Interval, collapsed over B-groups. The unjoined points represent the clustering scores on the last trial of original learning.

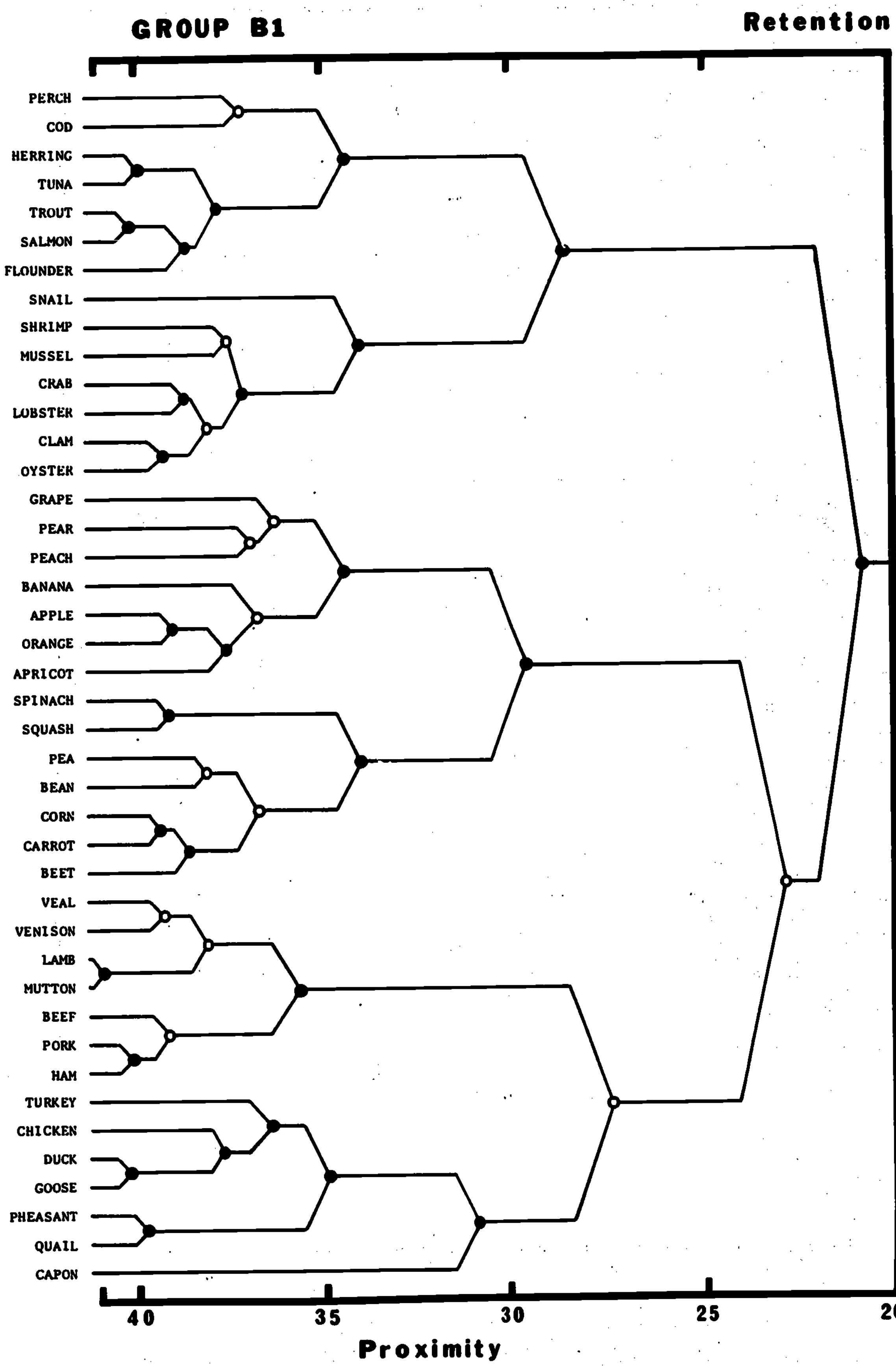


Fig. 19. Organization of the list for Group B1 on Trial 1 of Session II.

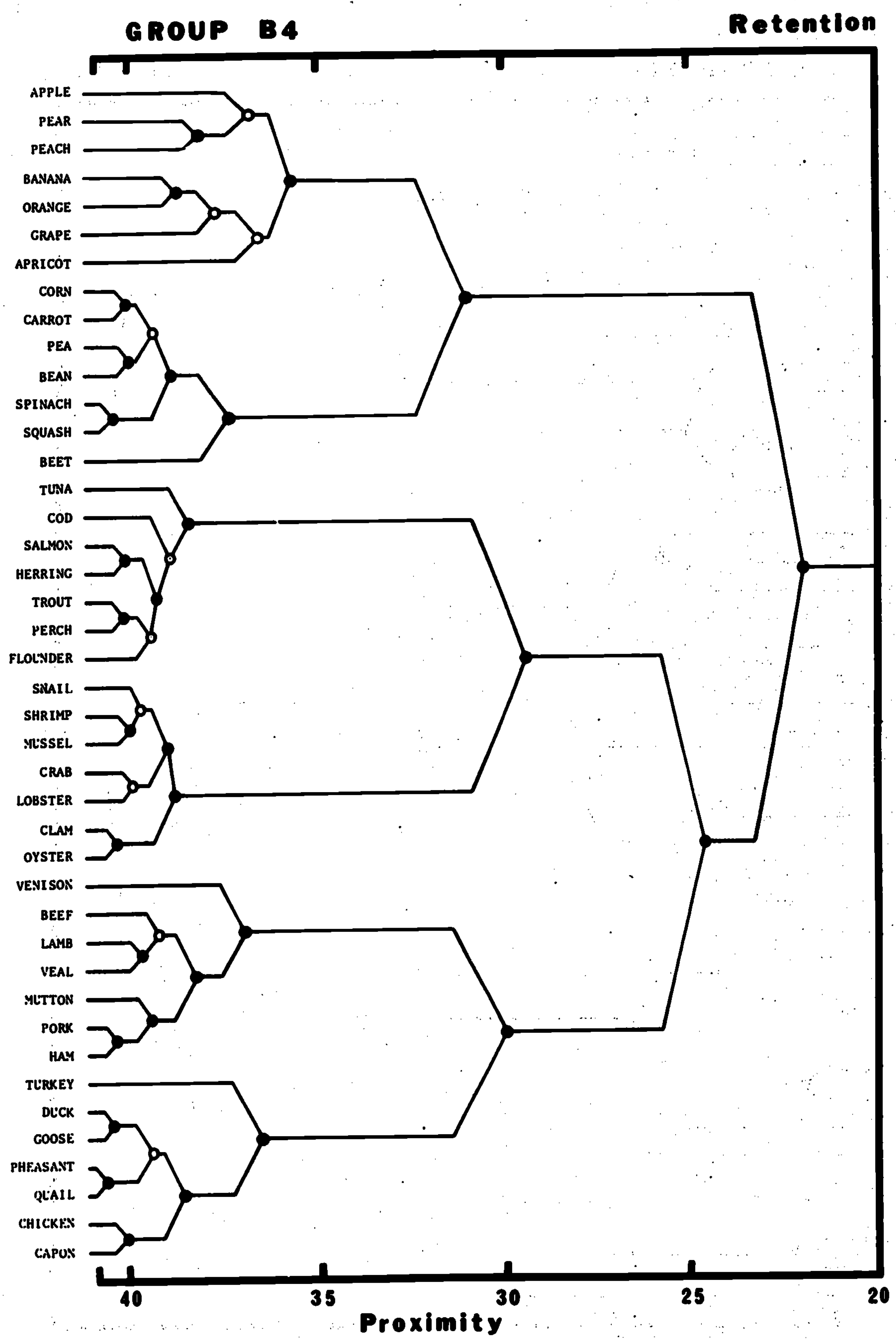


Fig. 20. Organization of the list for Group B4 on Trial 1 of Session II.

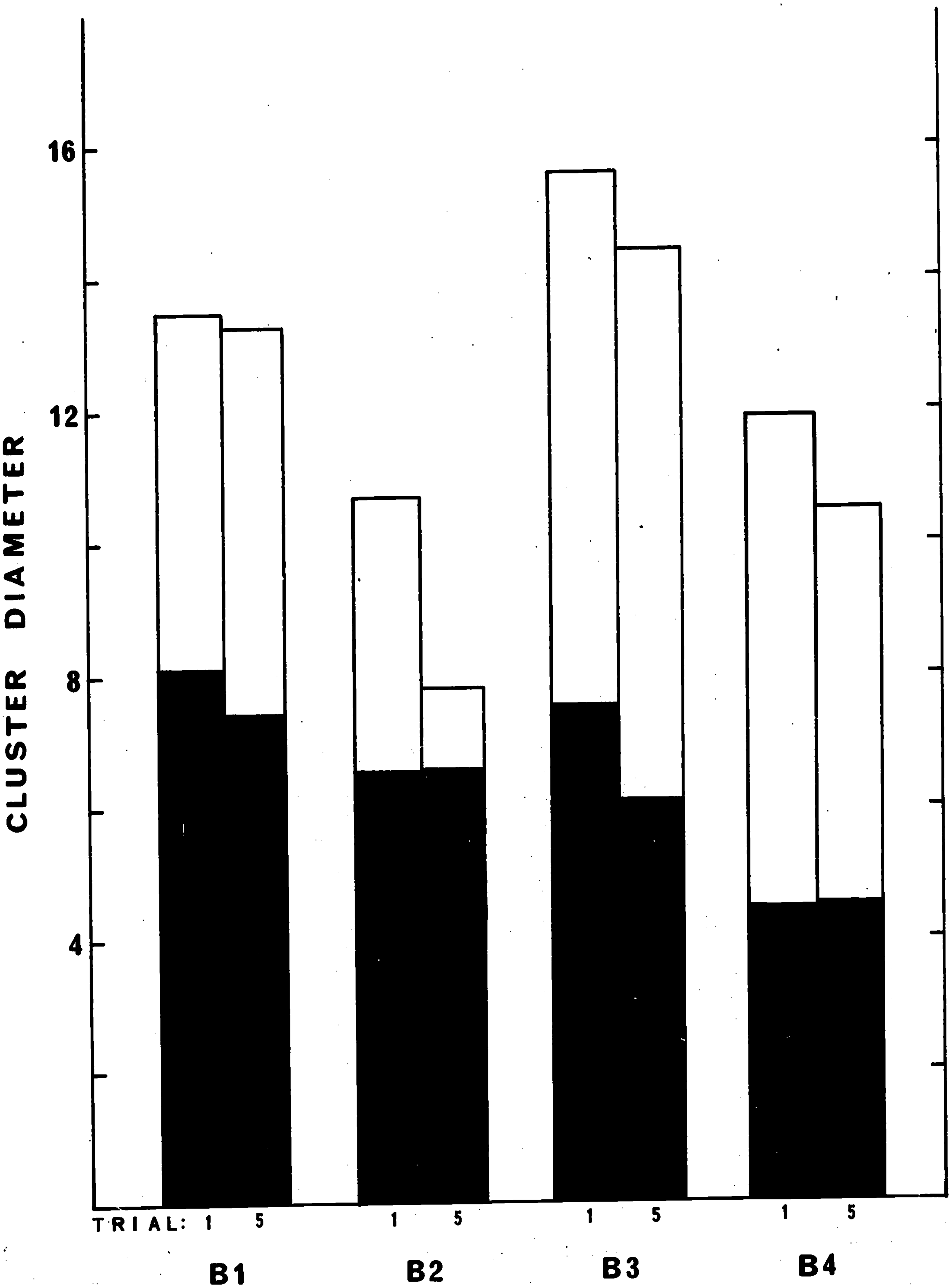


Fig. 21. Category diameters from cluster analyses of group proximities in Session II, Trials 1 and 5.

superordinate categories. As with the OL data, however, it is more instructive to consider the diameters of the clusters at both ordinate levels. Figure 21 displays the diameter values for the retention trial (Trial 1) and for the last relearning trial (Trial 5). On Trial 1, the level-3 organization of Group B4 is most compact, that of B1 is most diffuse, while B2 and B3 are intermediate. Further, on this trial the organization of Group B3 shows the greatest isolation between the two levels, while B2 shows the least. These results are quite similar to those obtained for original learning (Figure 12). By Trial 5, all four groups organize somewhat more compactly at level 2, with little change at level 3.

3.4 Discussion

By manipulating the presentation order of a hierarchically categorized list, this experiment attempted to lead groups of Ss to organize this list in several different ways. The experiment was performed to provide evidence regarding the utility and sensitivity of proximity analysis in a situation where fine discriminations among alternative patterns of organization would be required. Additionally, it was hoped to obtain data on the acquisition and retention of words which conform to a taxonomic hierarchy.

Since the list was constructed to consist of E-defined groupings, it was possible to assess the occurrence of category clustering at both levels of the hierarchy. The results obtained using these measures of the amount of organization were consistent with the view that each of Groups B2, B3, and B4 organized the list according to the different structures imposed on presentation order. Substantially the same overall interpretation was derived from cluster analyses performed in the interitem proximities in the recall protocols. The cohesiveness of item clusters determined in these analyses was found to vary in accordance with the predetermined modes of organization.

There were, however, some discrepancies between these two summaries of organization. For instance, the substantial gap separating Groups B3 and B4 from the remaining groups in the six-category analysis of the amount of organization (Figure 7A) did not appear in the cluster diameters derived from the proximity analysis. However, in view of the basic differences between these two procedures in purpose (amount vs. structure of organization) and in detail (trial-by-trial vs. overall summary) the correspondence of the results seems reasonably good.

The proximity also indicated that subjects receiving less than completely structured input discovered, to some extent, the additional taxonomic levels on their own. Thus, in the clustering of the average proximities for Group B2, each of the three major categories contained the appropriate minor categories as subclusters. In the B3 analysis, the six minor categories merged to form the appropriate superordinate clusters. To interpret this result it should be noted that the group proximities represent only the organizational tendencies common to a group and that some evidence was found of individual differences in organization within the groups. In general the differences among the experimental groups in the structure of organization appear mainly in the diameters of the clusters at the two levels. The groups are not aligned along a single dimension of amount of organization, since the cohesiveness of clusters at both levels must be considered simultaneously. Since the clusterings for the experimental groups differed in these terms in accordance with the predetermined patterns of organization, it may be concluded that the method of proximity analysis performs as it is supposed to do.

At a substantive level, the present experiment confirmed previous findings (Cofer, 1967; Dallett, 1964; Puff, 1966) that free recall learning of a categorized list is facilitated by blocked presentation of the category members. All groups receiving categorically blocked input learned more rapidly and retained more words than groups receiving either randomized input or randomly chosen blocks. That the R groups performed no better than Group B1 suggests that blocking of a list, of itself, does not facilitate memorization.

A differentiation among the groups receiving blocked presentation had also been predicted. Although Groups B2 and B4 recalled more words than Group B3, these differences were small and nonsignificant. Thus the predicted differentiation among these experimental groups was not obtained. A likely explanation of this result is that providing categorical cues at even one level of the stimulus hierarchy made it sufficiently easy for Ss to discover and utilize the additional level of categories. If this was the case, as suggested by the proximity analyses, then the lack of substantial facilitation of Group B4 relative to B2 and B3 is understandable.

In addition, categorically blocked presentation produced sizable increments in retention over the entire range of intervals studied. Again, essentially no differential effect appeared according to the level at which items were blocked. The overall blocked vs. random difference is consistent with the view that knowledge of the list structure provided by blocking influences not only the formation but also the temporal stability of higher-order memory units. This interpretation is strengthened by the result that differences in the amount recalled on the retention test were matched rather well by differences in the amount of organization, across both OL groups and retention periods. Correlations between recall and organization have been found within a single learning session (Tulving, 1962a, 1964). This result has been interpreted to reflect a causal dependence of recall upon the formation of higher-order units. On the retention trial in the present study the correlation (within cells) of recall with organization was $\underline{r}$ = .876. This result demonstrates that recall continues to covary with the degree of organization in retention.

Further evidence relating to the stability of organization was derived from the proximity analyses. It was observed that the organizational structures determined from the retention trial protocols were quite similar to those determined in OL, and that group differences in the cluster diameters also remained relatively constant over retention intervals.

These results, of course, provide only indirect support for the claim that retention is dependent on the maintenance of higher-order units. It might be possible to test this hypothesis in a more convincing fashion by making within-subject comparisons of recall and organizational units. In the present study, for example, it was observed that individual Ss formed highly cohesive groupings of some items, while other words were less tightly organized. Subjects also consistently remembered some items and rarely remembered others. The hypothesized relation between the stability of organization and retention would be considerably strengthened if it could be shown that the best-remembered items were in fact those which have been most tightly organized.

The finding that no group differences appeared in the number of categories recalled during acquisition or retention deserves comment. This indicates that all differences in recall and retention in the present study may be attributed to differential access to items within the categories rather than to variation in the number of accessible categories. This result is in sharp contrast to other findings with categorized lists (Cohen, 1963, 1966; Tulving & Pearlstone, 1966) where word recall per category remained constant while the number of categories recalled varied as a function of experimental conditions. The studies cited above typically employed more categories than used here and the number of categories also exceeded the number of items per category.

These conflicting results point to a trade-off relation between item recall and category recall which varies with the composition of the list. They also suggest that a single mechanism may be responsible for the retrieval of categories and of items within categories, with limited capacity at both levels. A list composed of many independent categories places a greater strain on category recall in such a system, and experimental manipulations which facilitate recall overall should benefit recall of categories most. On the other hand, if relatively few categories are to be recalled and a higher level scheme for grouping the categories exists, as in the present experiment, then experimental conditions should mainly affect the recall of items per category.

3.5 Artificial Experiment

One test of a proposed technique for studying mnemonic organization is that it should perform appropriately when a prevalent grouping of the items may be confidently predicted. In the present experiment, it was shown that the effects of the different blocking conditions did appear in the cluster analyses in terms of the diameters of clusters at both levels of the stimulus hierarchy.

While this is a necessary test for any technique to satisfy, it is also important to study the behavior of proximity analysis in the null case, i.e., when no organization is present. To do this, statistical subjects were generated in an artificial experiment. Statistical Ss were yoked to real subjects under two possible models of random organization. Under an independent trace (IT) model, a statistical subject was matched to each real S in terms of number of items recalled only, the specific items "recalled" by the statistical S and their sequential order was chosen at random with uniform probability. According to a dependent trace (DT) model, a yoked subject was matched item-for-item to a real S, with only recall order left to chance. Repetitions and intrusions were eliminated from the protocols in both cases.

Essentially, these two models consider the information contained in a real S's protocol as consisting of three parts: (a) the number of items recalled; (b) all conditional probabilities of recall, $P(i|j)$, $P(i|j,k)$,... $P(i|j,k,...,l)$; and (c) the sequential order in which the items are recalled. Artificial subjects generated under the independent trace model are equated with real Ss in the first component only. If the proximity method is indeed independent of recall performance per se, no semblance to the real Ss' organization should appear in the IT data. Any differences between real and IT organization should depend only on recall order and the probabilities that some items are recalled, given that other items appear in output.

On the other hand, artificial Ss generated under the dependent trace model match real Ss in all but the last of the three components. A comparison of the proximity results of real Ss with their DT yoked counterparts should depend only on the order of recall. The notion of "item properties," which Bousfield and Bousfield (1966) felt should be excluded from measures of organization, encompasses both total recall and conditional recall probabilities. Their measures of category clustering (SCR) and subjective organization (ITR) are therefore based upon a comparison of observed values with chance expectation under the dependent trace model.

Finally, the extent to which the mere co-occurrence of particular sets of items in recall influences the proximity results can be judged by comparing the results for IT and DT statistical Ss, since they differ only in that the conditional probabilities of item recall are included in the latter. The concept of a higher-order memory unit implies that recall of a single item from such a unit should increase the probability that other items from that unit are also recalled. Therefore, the conditional probabilities might be expected to provide some information regarding organization.

Interitem proximities were computed from the protocols of IT and DT statistical Ss in an analysis parallel to that described for the original learning data. To summarize these results, two measures of organization were derived from the proximity matrices. To the extent that subjects consistently organize groups of items, some proximities will be high and others will be low. Thus the range of proximity values is one indicant of the degree of subjective organization. Also, if subjects organize according to some predetermined set of categories, the average value of proximities for pairs belonging to the same category should exceed the average value for pairs belonging to different categories. The difference between these two average values can be taken as a simple index of categorical organization.

The results in terms of these statistics were quite simple. Artificial Ss generated under both models displayed no semblance of organization in the proximities among items. Table 5 presents the summary statistics from the analyses carried out for Ss yoked to Groups B1 and B4. The difference of within-category and between-category proximities determined from real data exceeded the corresponding values for both types of statistical Ss by several orders of magnitude. Similarly, the range of proximity scores for real Ss was about four times that of statistical Ss. However, Table 5 also shows small differences between the DT and IT models. The dependent trace Ss, matched in terms of the actual items recalled by real Ss, display slightly more organization by these measures than their independent trace counterparts.

It may be concluded that proximity analysis is (a) dependent almost entirely on the order in which items are recalled, (b) is influenced to a slight extent by the conditional probabilities among items in recall, but (c) is virtually independent of the overall level of recall. One further

Table 5

Partial Summary of Proximity Analyses for

Real and Statistical Subjects

Group	Data	Subjective Organization (Range of Proximities)	Category Clustering Within	Category Clustering Between	Category Clustering Difference	Badness-of-Fit to Hierarchy (%)
	Real	13.88	35.741	28.051	7.690	1.40
B1	IT	3.12	30.119	30.242	-0.123	1.45
	DT	3.99	30.186	30.177	0.009	1.52
	Real	16.17	38.305	27.295	11.010	1.12
B4	IT	3.40	29.325	29.354	-0.029	1.36
	DT	4.60	29.362	29.343	0.020	1.57

point should be noted regarding the use of the badness-of-fit measure to evaluate cluster analysis results (see Table 5, last column). While the tree structures determined for statistical Ss were not meaningful in any sense, they did fit a hierarchical clustering scheme as well as the solutions derived from real data. Thus, although a good fit to a hierarchy is a necessary condition for interpreting an organizational structure, it is by no means a sufficient condition for useful results.

CHAPTER 4

INAPPROPRIATE S-UNITS IN PART-WHOLE TRANSFER: REANALYSIS OF ORNSTEIN'S DATA

4.1 Introduction

The application of proximity analysis to the hierarchical list experiment produced reasonable results and indicated some aspects of organization which could not be readily determined on the basis of the amount of organization alone. On the whole, however, this technique did not contribute greatly to the interpretation of the results: the conclusions drawn therein can be based with equal force on measures previously available. This chapter attempts to demonstrate the utility of proximity analysis in a situation where the amount of organization alone provides insufficient evidence for strong conclusions.

The application described concerns the effects of organization on transfer in free recall learning. In transfer studies, S learns one list for several study-test trials and then learns a second list under similar conditions. Typically the lists are related in some fashion. For example, the items on the first list may be a subset of those to be learned on the final list (Tulving, 1966); or they may be instances of the same taxonomic categories which make up the second list (Birnbaum, 1968; De Rosa, Doane, & Russel, 1970).

Transfer tests are typically used to assess the effects of one learning experience on another. In the case of free recall, the transfer paradigm provides a means for determining the functional significance in a subsequent task of higher-order units which have been developed in prior learning.

That is, if higher-order units are more than a momentary product of learning, the relation between units formed in the two tasks should be an important determinant of performance in the second task. As an illustration of the use of proximity analysis, transfer studies are of particular interest, therefore, since their interpretation is based upon a comparison of the organizational patterns developed in the two learning experiences.

The data from two experiments concerned with part-whole transfer by Ornstein (1970) have been made available to the author. They are discussed and reanalyzed below by the techniques proposed in Chapter 2. The use of available data for illustrative purposes also has the virtue of evaluating a new method by experiments whose results are known.

4.2 Part-Whole Transfer

Prior learning of part of a list retards subsequent learning of the whole list. This somewhat counterintuitive result, first demonstrated in a free recall task by Tulving (1966), suggests (a) that practice or repetition of material in free recall is not always sufficient to produce efficient memorization and (b) that a satisfactory explanation of the (ordinarily facilitative) effects of practice must include more theoretical machinery than just the notion of independent strengthening of individual item-traces.

In one of Tulving's experiments Ss first learned a list of 18 unrelated words for eight trials and then learned a 36-word list on which eight presentation-recall trials were also given. Two groups of Ss learned different initial lists, but transferred to a common second list. For a part-whole (PW) group, all of the List 1 words appeared on the second list. A control group (C) first learned 18 words which did not reappear on List 2.

The surprising finding concerned the comparison of the two groups in their performance on the second list. Group PW, which has already learned one-half of the second list, did no better than Group C, which had learned 18 irrelevant items. In fact, the control Ss appeared to learn List 2 at a faster rate, evidenced by a slope difference in their mean performance curves. In interpreting this result, Tulving argued that the subjective organization imposed on the part list by experimental (PW) Ss was not appropriate for learning the whole list. If it is assumed that the number of S-units which can be retrieved on a given trial is limited, then learning the final list would require the PW Ss to reorganize or modify the S-units formed in learning List 1 in order to accommodate the new items; the necessity to restructure resulted in a performance decrement relative to control Ss for whom no reorganization was necessary.

Tulving's account is quite plausible and derivations from the SO theory have been confirmed in a number of other transfer studies (Birnbaum, 1968; Bower & Lesgold, 1969; Novinski, 1969; Ornstein, 1970). Tulving's (1966) data, however, do not compel an explanation based on inappropriate S-units. In fact, there is another explanation which is equally compatible with the data.

It is possible that PW Ss employ an input strategy of selectively attending and rehearsing the new items in List 2 at the expense of old items. This is related to the fact that newly learned items tend to be recalled earlier in output than old items, both in single-list free recall (Battig, Allen, & Jensen, 1965) and in part-whole transfer (Roberts, 1969). Such a strategy would make new items less susceptible to intratrial forgetting

(Tulving, 1964) during the recall period. However, the combined effects of input strategy (selectively attend to new items) and recall strategy (recall new items first) would cause old, previously learned items to undergo interference. That is, the learning of old and new items would conform to a retroactive interference paradigm on the old words--Learn A (old), Learn B (new), Test B, Test A. Essentially the recall of old items would be attempted after greater intervening time and interpolated recall. This explanation of negative transfer has also been suggested by Postman (1971) and is supported by the finding that prior part-list learning produces a greater negative effect on the recall of old words than of new words (Bower & Lesgold, 1969).

The effects of RI--an inability to recall previously learned material as a consequence of learning some other material--are well established in free recall (Postman & Keppel, 1967; Shuell, 1968; Tulving & Psotka, 1971), and it is also known that RI increases with interlist similarity (Shuell, 1968; Wood, 1970). Hence, this selective attention-RI explanation would predict that the PW group, having already learned a randomly selected portion of the final list, would experience interference in List 2 learning, to which control Ss would not be susceptible. In this view, negative transfer is ascribed to changes in the nature of stored traces as a result of subsequent input, rather than to the inability of the retrieval mechanism to provide access to more than a limited number of units of intact units, as implied by the organizational interpretation.

On the basis of Tulving's transfer studies (Tulving, 1966; Tulving & Osler, 1967), nothing more can be said to decide between these two explanations.

However, the SO account can be tested directly by using the method of proximity analysis to determine the contents of S-units at the end of List 1 learning and their composition at various stages in List 2 learning. Presumably, any changes in organization which occur in learning the final list should go in the direction of producing S-units which are more optimal for the whole list. However, it is difficult to substantiate the organizational explanation by testing it in this form, since "optimal" groupings may vary from one subject to the next, and hence any interlist modification of S-units might be taken as supportive evidence for Tulving's position.

A considerably stronger test would result from an experiment in which List 1 S-units remained appropriate for final list learning for some Ss, while other Ss were forced to reorganize. In such a situation, Tulving's position would require that (a) the former Ss should show positive transfer while the latter Ss should not, relative to the control group, (b) the organization of old items embedded in List 2 for Ss with appropriate transfer should be consonant with their own organization of these same items when first learned, and (c) List 2 M-grams for Ss with inappropriate transfer should indicate that prior-list groupings have been abandoned or modified in final list learning.

4.3 Ornstein's Experiment I

Several studies by Ornstein (1970) have employed this logic of manipulating prior-list organization to test prediction (a) above. While verification of (a) requires only inspection of the group performance curves, (b) and (c) depend on the availability of a method for indicating the contents of memory groupings.

One of Ornstein's experiments (1970, Exp. I) attempted to maintain prior-list subjective organization by presenting List 2 in blocks of old and new items, in contrast to the Tulving study in which the two sets of words were randomly intermixed on the final list. Blocked presentation should serve to facilitate discrimination of old and new subsets and allow Ss to develop a separate parallel organization for the new items, while preserving List 1 groupings of the old words. In addition to groups replicating Tulving (Groups Part Whole-Random and Control, with List 2 randomly arranged), Ornstein's design included two groups which received the final list in a blocked fashion. One of these saw all the old words first, followed by all new words on each trial of final-list learning (Group PW-O/N). In the other group, old and new items were each divided into two equal subsets and presented in alternating blocks (Group PW-O/N/O/N). Transfer was from a 12-word list to a 24-word list, all unrelated words, and eight trials were given on both lists.

The test of proposition (a) involves the comparison of group recall performance on the final recall task. As in Tulving's study, Group PW-R, which had received the List 2 items randomly arranged, did no better than the group which had had no prior relevant learning (Group C). Group PW-O/N/O/N recalled more items than control Ss on Trial 1 of the second list, but this superiority disappeared on subsequent trials. Group PW-O/N, which had the greatest advantage of blocked presentation, showed large, positive transfer.

This result is consistent with the organizational interpretation, but we can provide the strong, direct test of (b) and (c) by analyzing the proximities among items in List 2 recall for subjects in the various groups. In order for Tulving's hypothesis to be supported, Group PW-O/N Ss should maintain the organizational pattern developed in List 1, while Ss in the random presentation, part-whole group should show structures for which the organization of old items is fragmented with respect to List 1 organization.[14] It is not clear what to expect in Group PW-O/N/O/N, since the partition of old items into two subsets might tend to conflict with prior-list organization to an unknown degree. Thus, the output order of old words in List 2 learning would probably represent the combined influence of prior groupings and List 2 input order.

The data first used to illustrate the method of proximity analysis (Figure 1) were taken from the List 1 recall protocols of one of the PW-O/N Ss. The cluster analysis performed on the proximities from this S's last six trials of List 1 (Figure 2) indicated a hierarchical organization which could be described by three S-units. Figure 22 presents the organizational structure (diameter method) for this S derived from the List 2 protocols (Trials 1-8). The corresponding List 1 M-gram for old words has been redrawn at the left of Figure 22 for ease of comparison. The most striking feature of the List 2 organization is the separation of the tree structure into "old" and "new" components. The separation is not perfect--LABORATORY and SEAT merge with the old rather than new items--but these two words are

[14] It should be noted that it is not appropriate to take high category clustering scores (e.g., SCR) in terms of old vs. new items as evidence that List 1 organization has been preserved (but see Birnbaum, 1968; Bower & Lesgold, 1969). Marked old/new clustering indicates that Ss are organizing old and new items separately, but does not necessarily "reflect the maintenance of part-list organization" (Birnbaum, 1968, p. 1041), nor does it give any information about the degree of sequential consistency within these subsets.

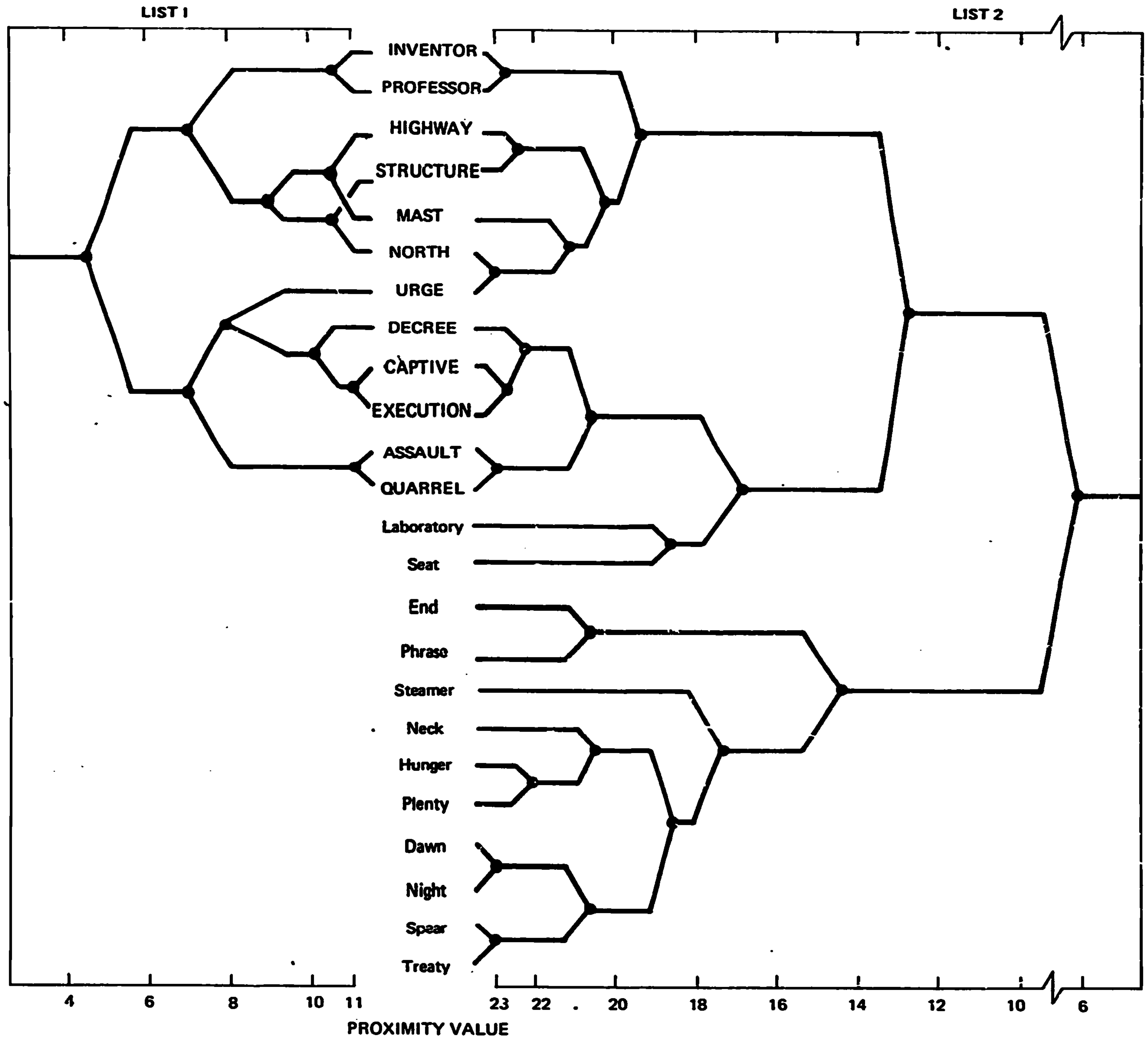


Fig. 22. Comparison of List 1 and List 2 organization for a subject receiving blocked presentation (O/N) of List 2. Data from Ornstein (1970, Exp. I).

only weakly associated with the old items. The groupings of the new items (shown in lower case) also seem to make sense semantically--(END, PHRASE), (HUNGER, PLENTY), (DAWN, NIGHT) and (SPEAR, TREATY). Also, comparing the organization of the old items with this S's structure of these items in prior-list learning, it can be seen that the major subjective units uncovered earlier have remained intact--(INVENTOR, PROFESSOR), (HIGHWAY, STRUCTURE, MAST, NORTH), and (DECREE, CAPTIVE, EXECUTION, ASSAULT, QUARREL).[15]

If the cluster analysis is believed to give a relatively accurate portrayal of the fine-grain structure, then it would be of interest to interpret the organization with S-units. Comparison of the first- and second-list solutions indicates that local, intra-unit differences do appear. However, the most tightly-knit groupings (which we shall call primary S-units) from List 1 learning--(INVENTOR, PROFESSOR), (ASSAULT, QUARREL), (CAPTIVE, EXECUTION)--do remain perfectly intact in transfer to List 2 and are also among the most tightly-knit units in that list.

Although this analysis was in terms of a single S, the most general results, i.e., segregation into old and new components, and maintenance of primary S-units and higher-order units of old items from List 1, also obtain at the group level. Figure 23 shows the clustering results for the pooled data of all Ss in this group (S = 7). The group analysis also

---

[15]Perhaps the greatest difference between the two M-grams is in the position of URGE. This word did not appear consistently near any other word during List 1 learning, but is merged with NORTH at the highest proximity level in List 2. The reason for this is not entirely clear, but the proximity of NORTH and URGE may have been underestimated in List 1. The lower left panel of Figure 1 shows the trial-to-trial proximities of these two items. On trial 3 these items appeared at opposite ends of the protocol (p = 3), but on four of the five remaining trials, they were recalled in adjacent or penadjacent positions.

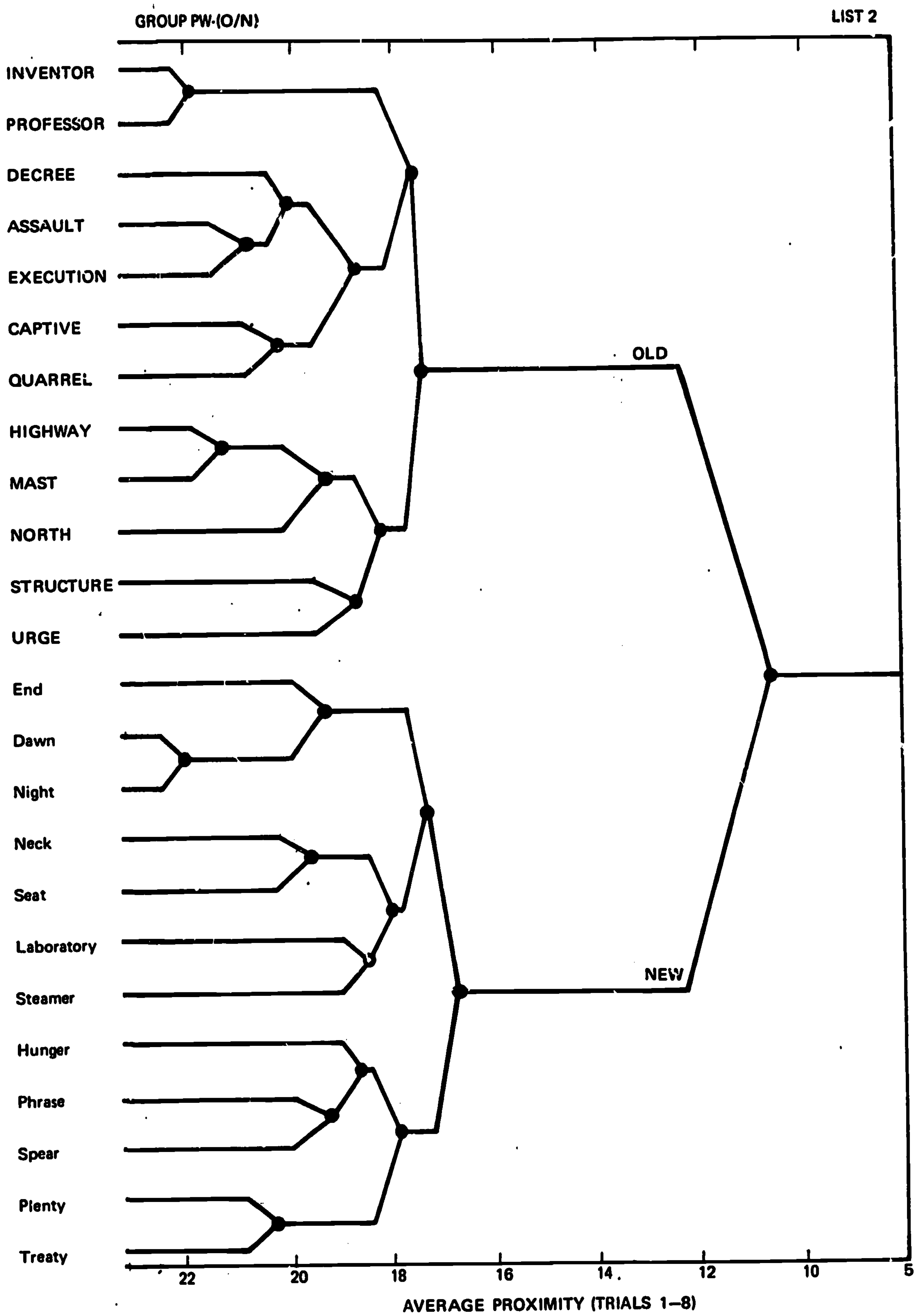


Fig. 23. List 2 organization for Group PW-(O/N).

indicates the major branching into old and new items. Individual differences, which would be treated as noise, cannot be large at this level. In the organization within the two subsets, items first begin to cluster at lower proximity values than in M-grams for individual Ss, indicating some individual variation in S-units within the subsets.

In the random presentation, part-whole group (PW-R), no positive transfer occurred. Subjects in this group learned the same lists as those in Group PW-O/N, differing only in the random presentation order of List 2. What light can proximity analysis shed on their decreased performance? The structure of List 2 organization for a fairly typical subject from this group is shown in Figure 24, with old items typed in upper case, new items in lower case. The old and new items in the organizational pattern of this S are completely mingled. The groupings extracted from the first and second list protocols differed so markedly that the two hierarchies could not be drawn juxtaposed in the same figure without considerable crossing of lines. This mixing of items from the two subsets in the organization of List 2 occurred for every S in Group PW-R.

By comparing this S's List 2 S-units with those which emerge in prior-list learning, it is possible to see what, if anything, he was able to maintain in transfer to the longer list. First-list organization for this S appears in Figure 25. Unlike the situation in Group PW-O/N, Ss in the random group seem to have either lost or discarded the higher level S-units in whole list learning. Comparison of the two M-grams indicates, however, that several of the highly proximal pairs of old items carry over when the whole list is learned--(DECREE, EXECUTION), (HIGHWAY, STRUCTURE), and

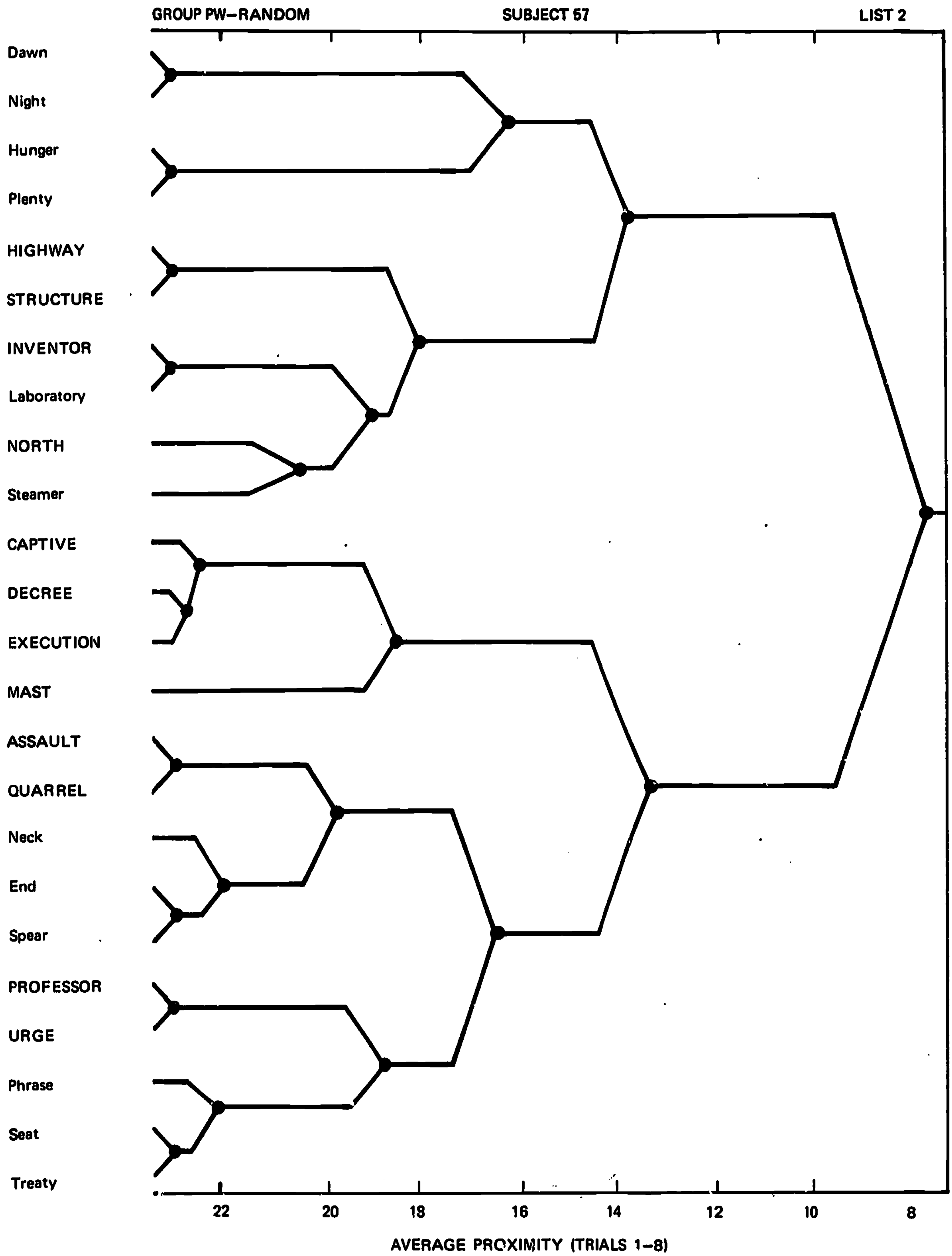


Fig. 24. List 2 organization for a subject from group PW-R. Data from Ornstein (1970).

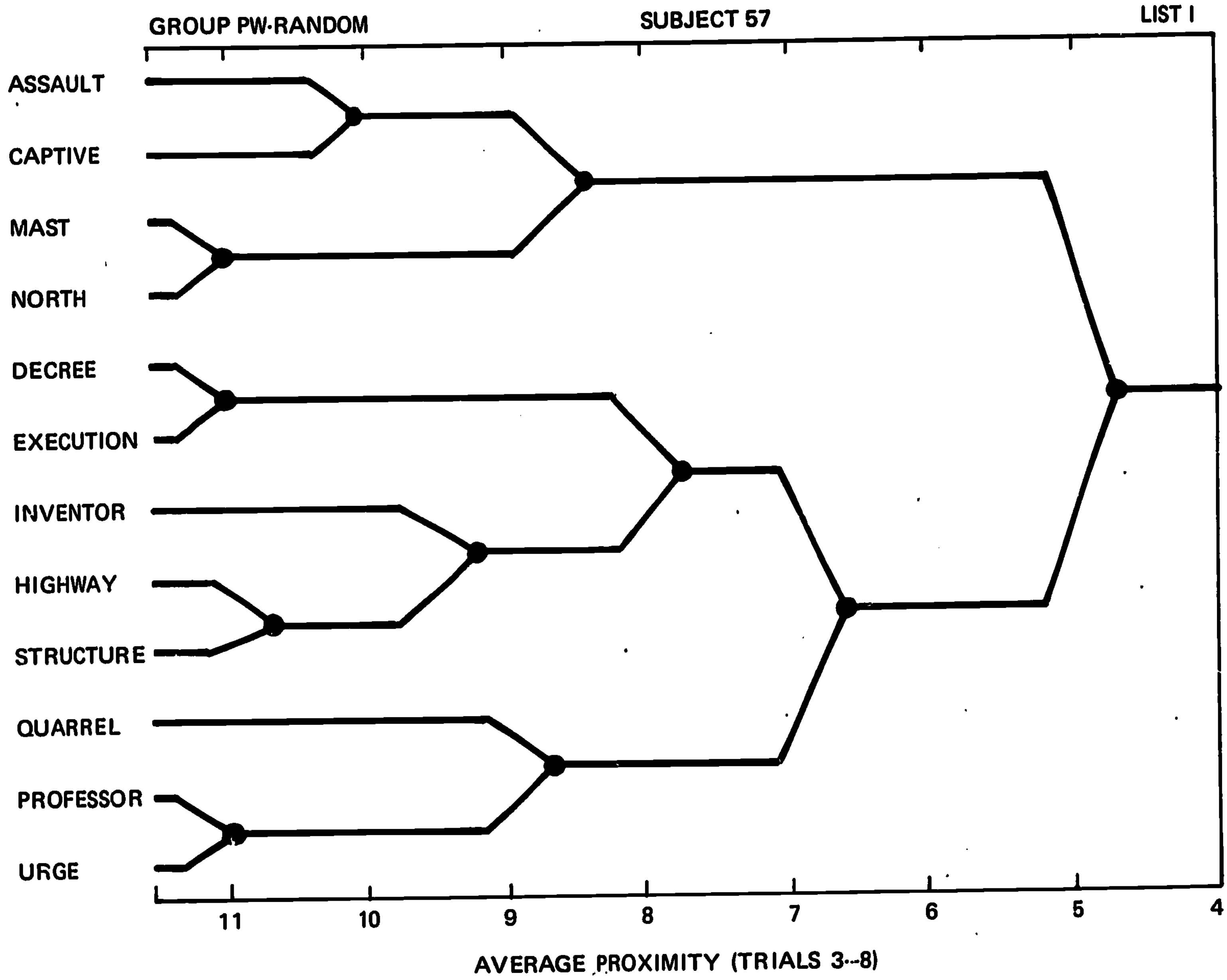


Fig. 25. Organization of "old" words in List 1 learning for the subject whose List 2 organization is shown in Fig. 24. Data from Ornstein (1970).

(PROFESSOR, URGE). Again, this general pattern of intermixing of old and new in List 2 structure, with maintenance of only the strongest primary S-units, appears for almost all subjects in this group.

The effect on List 2 organization of dividing each of the old and new subsets into equal halves and presenting the items in four alternating blocks can be seen in the M-gram determined from the pooled data of Group PW-O/N/O/N (Figure 26). In this analysis, recall protocols were aggregated over all trials of final-list learning as well as over Ss. The membership of items in the various blocks is indicated in the legend. As in Group PW-O/N, these Ss develop a separate organization for the new items. In addition, Group PW-O/N/O/N structures the new items exactly according to the arrangement in blocks. The old items, on the other hand, do not display any grouping according to the contents of the blocks. Although the diameter and connectedness methods agree in the features noted above, they show little agreement in the organization within the old items. This indicates noise or individual differences, and hence interpretation of the groupings within these old items is not warranted.

As a result of these analyses, what can be said about the lack of positive transfer for the random presentation group, and how does blocking of the whole list facilitate the performance of Group PW-O/N? It seems that for both groups, the highly organized, primary S-units acquired in learning the part list are maintained and used by the subjects in recalling the whole list. What differentiates the groups is the degree to which they use the higher order units of List 1 to aid recall of List 2. Higher order units can be thought of as access routes which guide the retrieval system from one

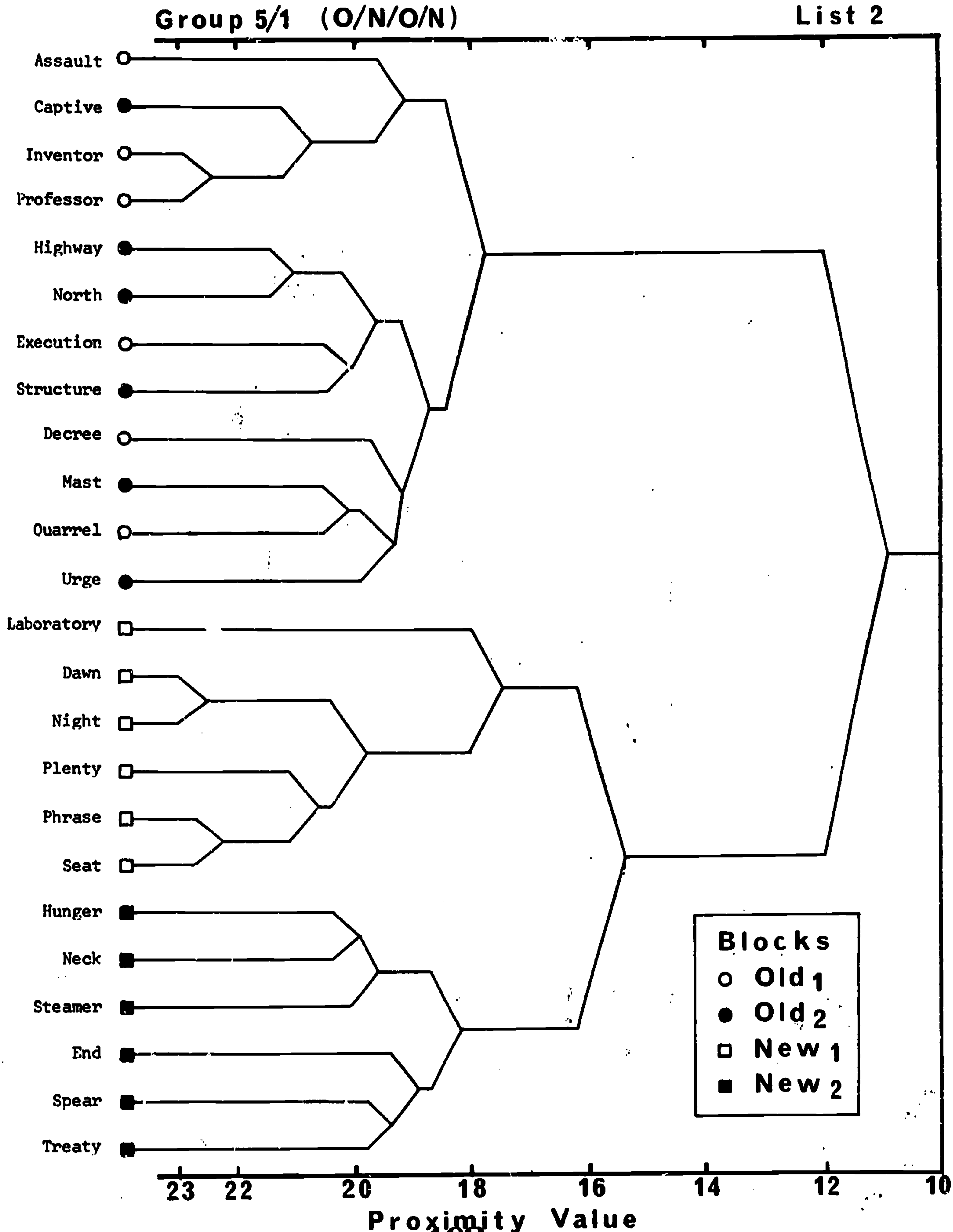


Fig. 26. List 2 organization for Group PW-(O/N/O/N) of Ornstein's (1970) experiment.

primary S-unit to the next. Several theorists have argued that the basic limitation in free recall is utilization of information in the memory store, rather than how much information can be packed into it (Mandler, 1967b; Tulving, 1966, 1970). That is, information is often available, but not accessible (Tulving & Pearlstone, 1966). If this is so, then the higher order S-units would be important since each one presumably serves as a retrieval aid for a large number of items. It follows that anything which interferes with these informationally rich units should have a disruptive effect on the overall success of recall. This appears to be precisely what has occurred in Group PW-R. Subjects receiving blocked input on the whole list, however, maintain the higher-order units of List 1. For the most part they develop a separate and parallel organization for the new items.

4.4 Ornstein's Experiment II

In a second experiment, Ornstein (1970) attempted to manipulate the appropriateness of prior-list organization in part-whole transfer. This experiment employed lists containing polysemous words which could be categorized in two different ways. For example, under one reading, the word yard could be categorized with patio, garden and house while by a second meaning it would go with foot, meter and rod. Three groups of Ss learned a common final list of 56 words containing 14 equal-sized categories after having learned different initial lists of 24 items, grouped into six categories of four words each. Five trials were given on each list. For a Compatible group, three four-word groups from List 1 were carried over to List 2 and were categorized identically on both lists. For a Conflicting group, 12 prior-list items also appeared on List 2, but were organized in 12 distinct categories on the basis of the alternate meaning of each word. Thus, one of the first-list categories learned by this group was yard, foot, meter and rod. On the final list, foot was grouped with nose, eye and arm; meter with dial, gauge and scale, etc. A Neutral group learned an initial list which contained neither items nor categories from the final list. Presentation was blocked according to nominal categories on List 1 and the first trial of List 2 for all groups; the remaining trials of the final list were presented randomly.

Since on the prior list the Compatible group alone had the opportunity to learn categories appropriate for List 2, it was predicted that this group would perform better on the final list than the Conflicting and Neutral groups. The prediction received partial confirmation in that positive transfer was obtained for the Compatible group on Trial 1 of second list learning, though the effect did not persist thereafter.

The recall protocols for Ss in the Compatible and Conflicting groups were subjected to proximity analysis. An analysis of List 1 learning indicated that each group had utilized the intended categorization of the items in their recall. Figures 3 and 4 presented the organization of List 1 learning for a typical S, and the pooled data, respectively, from the Compatible group. An analysis was performed for half of the Ss in each group on the final list, pooling over Ss (S = 10) and trials (T = 5). The tree structure of organization for the Compatible and Conflicting groups are shown in Figures 27 and 28. The 12 items which transferred from the first list for each group are typed in upper case. It can be seen that first-list

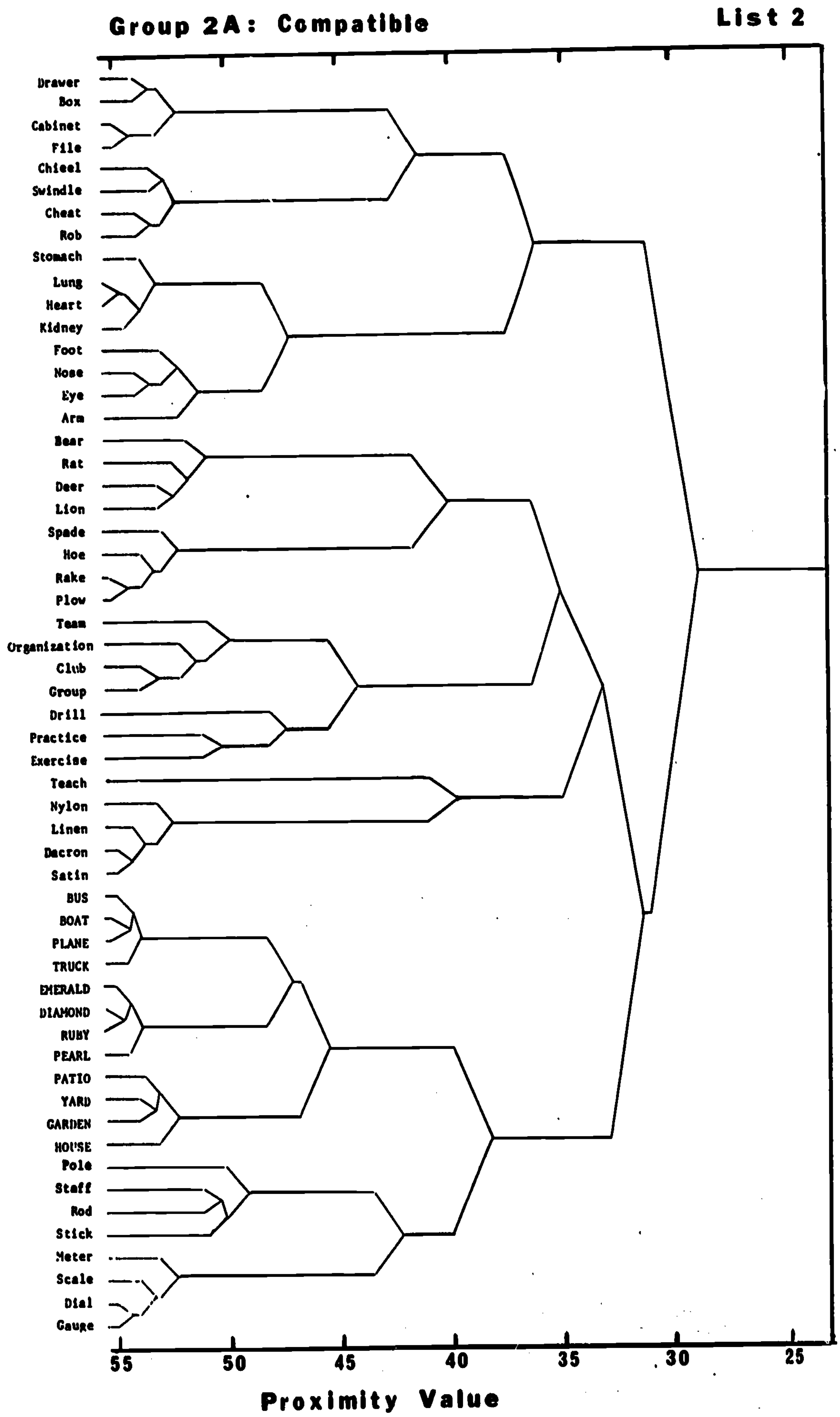


Fig. 27. List 2 organization for Compatible Group.

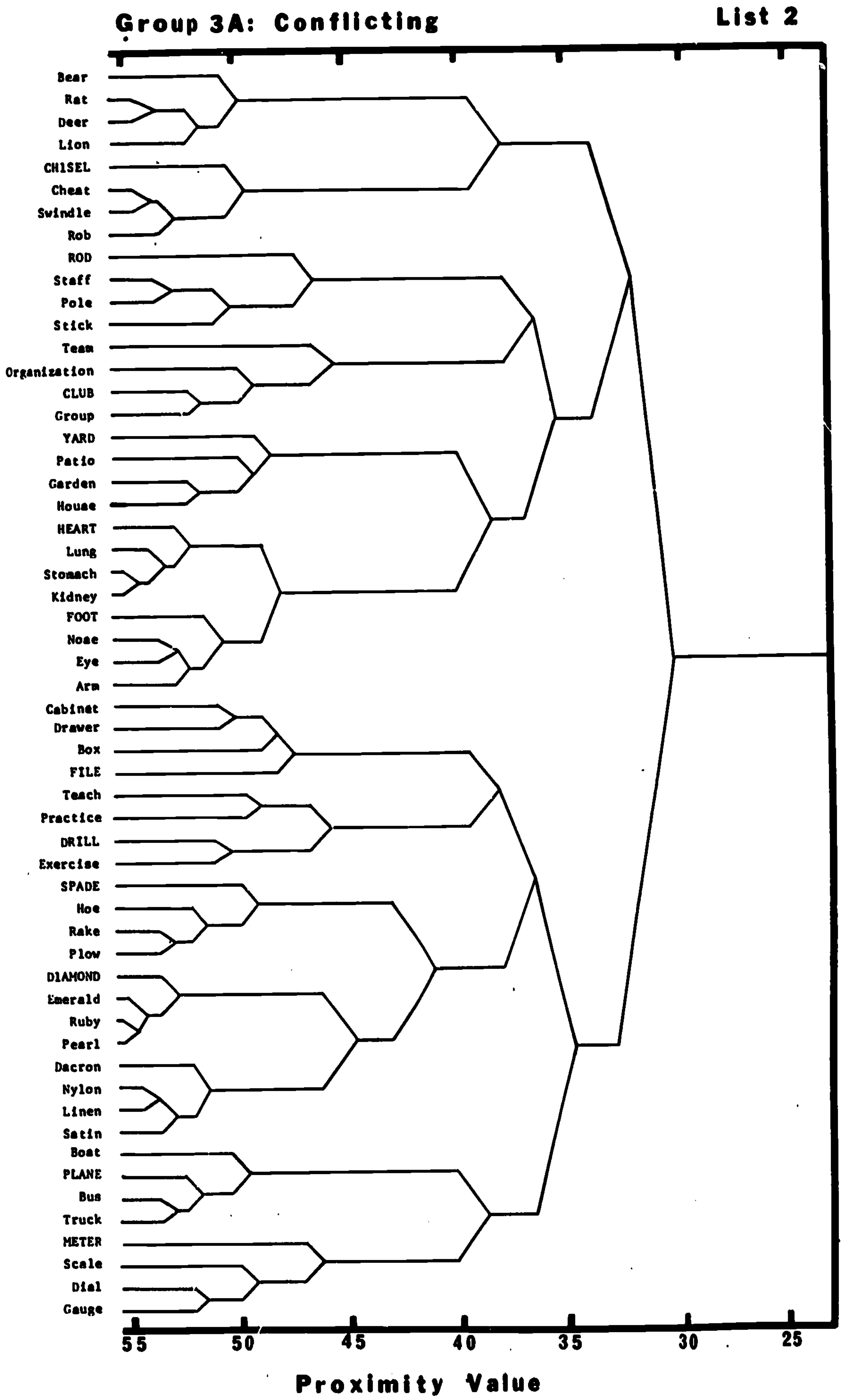


Fig. 28. List 2 organization for Conflicting Group.

items form the most highly organized groups in List 2 learning for the Compatible group. Of the 14 final-list categories, the diameters of the old categories rank 1, 2, and 6.[16] Over the five trials of List 2, the Conflicting group reveals substantially the same organizational pattern. The grouping of the old items from List 1 has evidently been discarded in favor of the appropriate final-list categories. The residual effects of first-list organization may, however, be discerned in the order with which the old items merge in the List 2 categories, shown in Figure 28. Of the 12 old items, nine are the last to join, i.e., least integrated members of their respective categories. A result this extreme or more has a chance probability of 0.006 on the null hypothesis of random orderings within categories.

It appears from the group M-grams, then, that the Ss in the Conflicting group failed to show a sustained deficit in List 2 performance, relative to the Compatible group, because they were able to discard easily their prior-list S-units. Hence their old organization did not interfere with the development of an appropriate strategy for List 2 as much as had been intended. To test this explanation, the following analysis was performed. The proximities between all pairs of items were computed on each trial of List 2 learning separately for all Ss in the Conflicting group and pooled over Ss. A group matrix for each trial was thereby produced giving the

---

[16] One of the final categories--drill, practice, exercise, teach--does not appear to have been consistently recalled as a unit by these Ss. In particular, the word teach does not function in recall as a member of the category. In this connection, it should be indicated that proximity analysis can be used to evaluate the success with which categorized materials were chosen. Category norms (e.g., Battig & Montague, 1969) do not provide a measure of the strength of a set of instances to the category label, but give instead the strength of the reverse association, category name to instance.

recall relatedness of item pairs. From each matrix, all pairs belonging to the three 4-item categories carried over from List 1 (e.g., foot, yard, rod, meter) were extracted and their proximities were averaged. This provided an index of the strength of first-list organization during the acquisition of a conflicting second list. In a similar fashion, average proximities were obtained within the 12 "appropriate" List 2 categories[17] (e.g., foot, arm, eye, nose). By design, the Compatible Ss learned only "appropriate" categories on List 2, and the within-category proximities were computed on each trial for this group. The results appear in Figure 29, where the within-category values are plotted as a function of trials of List 2 learning. Despite the fact that presentation on Trial 1 was blocked according to the new (appropriate) categories, the inappropriate categories of the Conflicting group still maintained considerable strength on this trial. The graph shows a progressive disbanding of the old categories thereafter. From Trial 2 on, all Ss received the words in random order, yet the upper curves indicate that the appropriate categorization had been readily picked up by both groups. Compatible Ss, having had prior practice recalling three of these categories, recall them in slightly tighter-knit groupings than Conflicting Ss.

Although Ss in the Conflicting group seem to adopt readily the new stimulus categories, it is possible that some residual effects of their prior learning experience remain during second-list learning. Assuming (for the present) that a categorized list would normally be organized hierarchically, then to the extent that two competing modes of organization contributed to the

---

[17]Two categories on List 2 were completely new for all Ss (e.g., dacron, nylon, linen, satin). These items were excluded from the analysis.

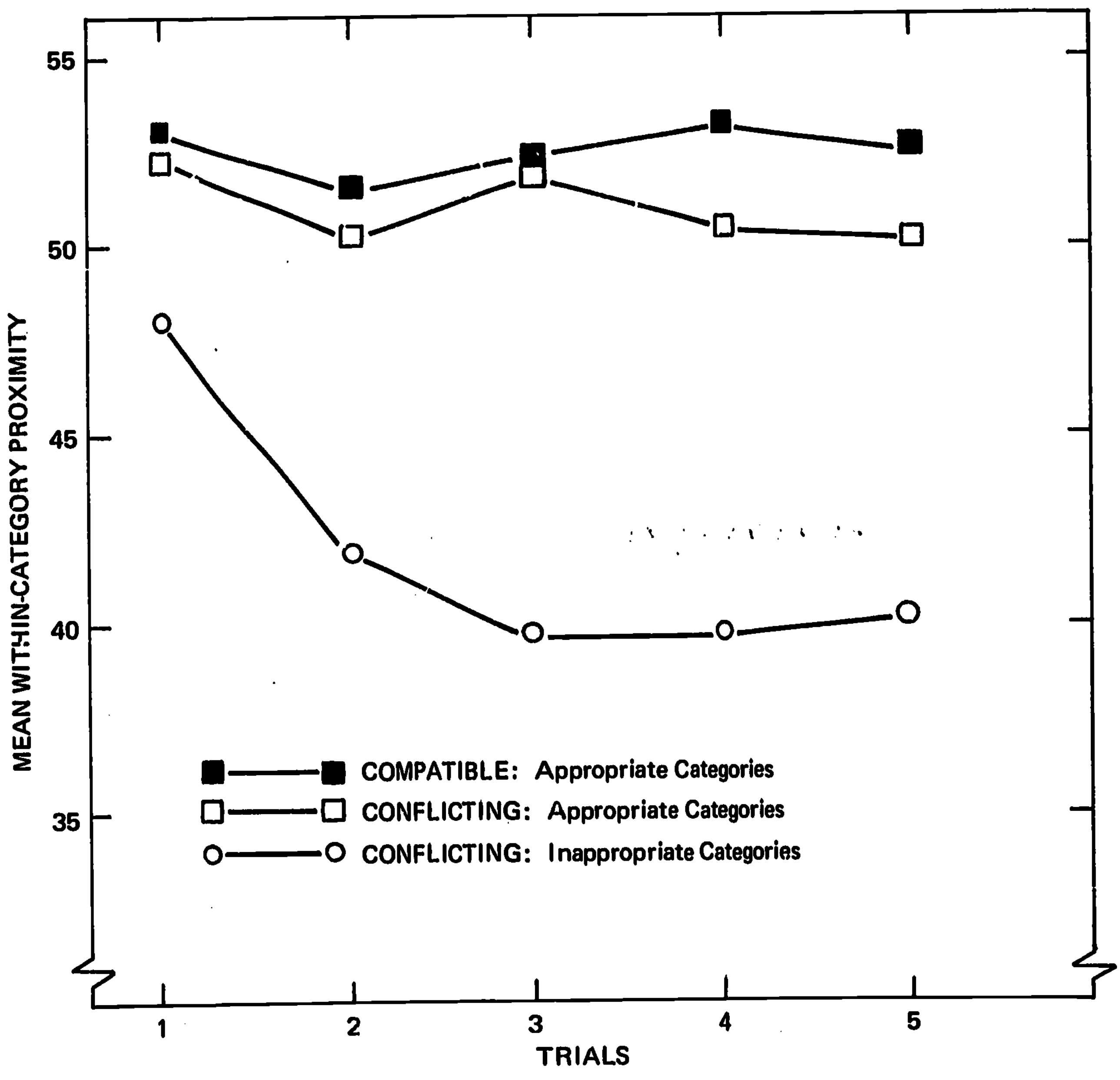


Fig. 29. Average proximities within appropriate and inappropriate categories as a function of trials of List 2 learning.

order of recall of these Ss, one would expect the Conflicting group proximities to diverge from a true hierarchy. Application of Johnson's clustering algorithm always produces a hierarchical representation. However, the numerical index of fit to a hierarchy, described in Appendix B, allows the examination of possible residual effects by comparing the fit values of the Conflicting and Compatible groups at various stages of List 2 learning. The group proximity matrices described above were clustered for each trial by the diameter method, and the measure of badness-of-fit was computed for each solution. These values, plotted in Figure 30, show a progressive decrease over trials. That is, for both groups, the modal organization becomes increasingly hierarchical as acquisition of the second list proceeds. On Trial 1 the hierarchical fit for both groups is quite poor, but the Conflicting group fits least well, suggesting some carry-over effects of prior organization. Beyond the first trial, however, the two curves do not differ.

It seems relatively clear, then, both from the group M-gram (Figure 28) and from the analyses just discussed, that the Conflicting group was but briefly hindered by their old organization and readily abandoned it. The blocked presentation according to the new categories on Trial 1 was evidently sufficient to produce a stable realignment of mnemonic units for the remainder of List 2 learning. It would be interesting to know whether Ss could as easily discard an inappropriate prior organization without the additional cues provided by a blocked input order.

In its strong form, Tulving's original explanation of the negative transfer effect in free recall (Tulving, 1966) implies that mnemonic units

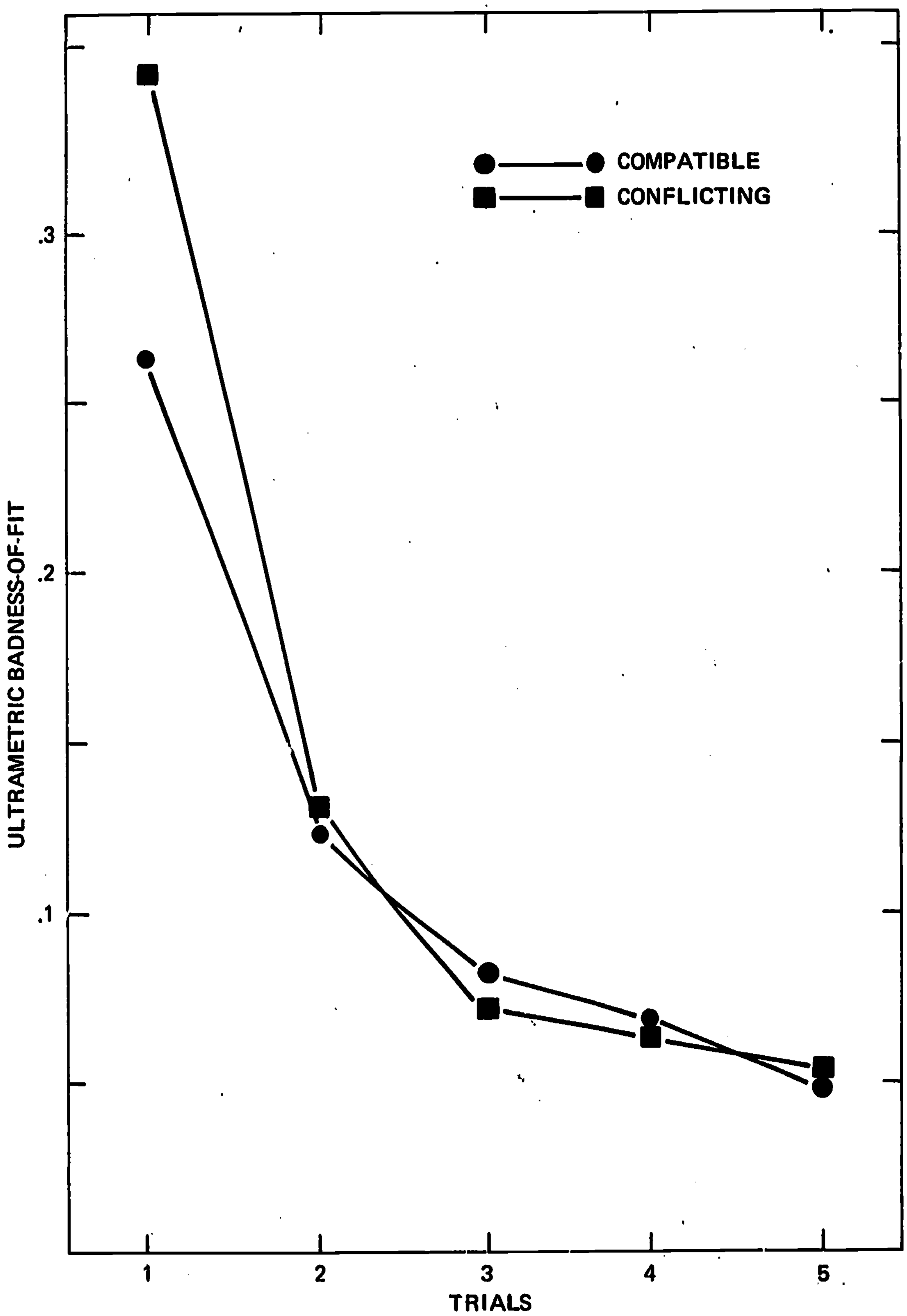


Fig. 30. Badness-of-fit to a hierarchical clustering scheme as a function of trials for Conflicting and Compatible groups.

remain more or less intact when the items they contain reappear in altered context on a second list. The reanalysis of Ornstein's experiment clearly demonstrates one counterexample to the strong interpretation of Tulving's explanation. Tulving's argument can also be interpreted in a weaker sense. According to this alternative interpretation the original mnemonic units do not necessarily persist in the transfer stage, but may be actively modified or abandoned. The present results are in agreement with this account and suggest the need for more detailed study of the conditions under which prior-list organization will transfer and to what extent its maintenance is under S's control.

4.5 Summary

In summary, data from two experiments concerned with part-whole transfer in free recall have been reanalyzed and discussed in terms of the method of proximity analysis. Both experiments attempted to test implications of the organizational explanation offered by Tulving (1966) for the finding of negative transfer in this paradigm.

In the first experiment, the extent to which Ss could make use of prior-list S-units in learning the final list was manipulated by blocked presentation of the final list. Proximity analysis of the final list protocols revealed that the major difference between Ss who had received random presentation and those who had had the advantage of blocked input lay in the greater ability of the latter Ss to make use of higher-order units from the prior list.

In the second experiment, Ss whose prior-list categories were appropriate for learning the final list showed a slight facilitation with respect to a group whose initial list contained categories which conflicted with those on the final list. By analyzing the manner of organization for both groups, it was possible to come to a clearer understanding of these results. The Compatible group M-gram indicated highly cohesive groupings of the items which had appeared on the prior list. The analysis for the Conflicting group showed, however, that these Ss did not have a great deal of difficulty in discarding the S-units developed earlier in favor of the more appropriate groupings for List 2.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The present research has been concerned with the development of a technique for studying the structure of organization in free recall learning. The existence of higher-order units in recall is typically inferred from consistency in the order of recall over a series of trials. Starting with this observation, it was proposed that the degree to which pairs of items shared common membership in a higher-order unit could be indexed by the ordinal separation or proximity between pairs in the recall protocols. By applying methods of cluster analysis to the interitem proximities, it was shown that a description of the pattern of organization and the contents of higher-order units could be determined.

An experiment was performed involving acquisition and retention of a hierarchically categorized list. This experiment led to the following conclusions regarding the method of proximity analysis: (1) The cluster analyses produced results which were consistent with E-determined patterns of organization. (2) Measures of the amount of organization derived from the proximities produced results essentially equivalent to those obtained with previous measures. (3) The patterns of organization developed during acquisition were maintained in the retention test. (4) A simulation experiment with artificial Ss matched in recall performance to real Ss demonstrated that the method performs appropriately when no organization is present.

Data from two studies of part-whole transfer (Ornstein, 1970) were reanalyzed by the proximity method. For these experiments, the analyses confirmed the hypothesis that the amount and direction of transfer in part-whole learning depends on the congruence between the subjective units developed in the two tasks. Reanalysis of the first experiment provided direct evidence that negative transfer in such situations is accompanied by a failure to maintain the prior organizational units. In the second experiment the direction of transfer had been predicted as a function of the appropriateness of part-list organization for learning the whole list. Sustained negative transfer was not obtained when the two lists conflicted in organization. The proximity method indicated, however, that the conflicting part-list organization did not persist into the test stage. It is concluded that the method of proximity analysis can be useful in testing theories concerned with the structural relations among items in memory.

References

Allen, M. Graph theory, organization and memory. Paper presented at the meetings of Eastern Psychological Association, New York, April 1971.

Battig, W. F., Allen, M., & Jensen, A. R. Priority of free recall of newly learned items. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 175-179.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication of the Connecticut Category norms. Journal of Experimental Psychology, 1969, 80, (No. 3, part 2), 1-46.

Birnbaum, I. M. Free-recall learning as a function of prior-list organization. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 1037-1042.

Bonner, R. E. On some clustering techniques. IBM Journal of Research and Development, 1964, 8, 22-32.

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 1953, 49, 229-240.

Bousfield, W. A., & Cohen, B. H. Clustering in recall as a function of the number of word categories in stimulus word lists. Journal of General Psychology, 1956, 54, 95-106.

Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. Psychological Reports, 1958, 4, 39-44.

Bousfield, W. A., Puff, C. R., & Cowan, T. M. The development of constancies in sequential organization during repeated free recall. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 449-459.

Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. Journal of General Psychology, 1944, 30, 149-165.

Bower, G. H. Organizational factors in memory. Cognitive Psychology, 1970, 1, 18-46.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 323-343.

Bower, G. H., & Lesgold, A. M. Organization as a determinant of part-to-whole transfer in free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 501-606.

Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 481-493.

Brand, H., & Woods, P. J. The organization of the retention of verbal material. Journal of General Psychology, 1958, 58, 55-68.

Clyde, D. J., Cramer, E. M., & Sherin, R. J. Multivariate statistical programs. Coral Gables, Fla.: Biometric Laboratory, University of Miami, 1966.

Cochran, W. G., & Cox, G. M. Experimental designs. New York: Wiley, 1964.

Cofer, C. N. On some factors in the organizational characteristics of free recall. American Psychologist, 1965, 20, 261-272.

Cofer, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmutz (Ed.), Concepts and the structure of memory. New York: Wiley, 1967.

Cohen, B. H. An investigation of recoding in free recall. Journal of Experimental Psychology, 1963, 65, 368-376.

Cohen, B. H. Some-or-none characteristics of coding behavior. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 182-187.

Cohen, B. H., & Bousfield, W. A. The effect of a dual-level stimulus-word list on the occurrence of clustering in recall. Journal of General Psychology, 1956, 55, 51-58.

Craik, F. I. M. Modality effects in short-term storage. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 658-664.

Craik, F. I. M. The fate of primary memory items in free recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 143-148.

Cronbach, L. J. On the nonrational use of information measures in psychology. In H. Quastler (Ed.), Information theory in psychology. Glencoe, Ill.: Free Press, 1955. Pp. 14-30.

Dallett, K. M. Number of categories and category information in free recall. Journal of Experimental Psychology, 1964, 68, 1-12.

Dalrymple-Alford, E. C. The measurement of clustering in free recall. Psychological Bulletin, 1970, 74, 32-34.

Deese, J. On the prediction of occurrence of particular verbal intrusions in immediate free recall. Journal of Experimental Psychology, 1959, 58, 17-22.

Deese, J. The structure of associations in language and thought. Baltimore, Md.: The Johns Hopkins Press, 1965.

De Rosa, D. V., Doane, D. S., & Russel, B. The influence of first-list organization upon second-list free recall learning. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 269-273.

Ehrlich, S. Le role de la structuration dans l'apprentissage verbal. Psychologie Francaise, 1965, 10, 119-146.

Ehrlich, S. Structuration and destructuration of responses in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 282-286.

Fagan, J. F., III. Measuring verbal recall II: The ITR score expressed as a ratio. Psychonomic Science, 1968, 11, 205.

Friendly, M. L. RECALL: A computer program for analysis of multi-trial free recall protocols, 1971, in preparation.

Friendly, M. L., & Glucksberg, S. On the description of subcultural lexicons: A multidimensional approach. Journal of Personality and Social Psychology, 1970, 14, 55-65.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 351-360.

Gower, J. C. Some distance properties of latent root and vector methods in multivariate analysis. Biometrika, 1966, 53, 325-328.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. Psychometrika, 1959, 24, 95-112.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Hartigan, J. A. Representation of similarity matrices by trees. Journal of the American Statistical Association, 1967, 62, 1140-1158.

Hudson, R. L., & Dunn, J. E. A major modification of the Bousfield (1966) measure of category clustering. Behavioral Research Methods and Instrumentation, 1969, 1, 110-111.

Jardine, N., & Sibson, R. The construction of hierarchic and nonhierarchic classifications. The Computer Journal, 1968, 11, 177-184.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Kendall, M. G. Discrimination and classification. In P. R. Krishnaiah (Ed.), Multivariate analysis. New York: Academic Press, 1966.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), Models of human memory. New York: Wiley, 1970.

Koh, S. D., Vernon, M., & Bailey, W. Free-recall learning of word lists by prelingual deaf subjects. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 542-547.

Lance, G. N., & Williams, W. T. A general theory of classificatory sorting strategies I. Hierarchical systems. The Computer Journal, 1967, 9, 373-380.

Mandler, G. Organization and memory. In K. W. Spense & J. T. Spense (Eds.), The psychology of learning and motivation, 1967, 1. New York: Academic Press. (a)

Mandler, G. Verbal learning. In New directions in psychology, III. New York: Holt, 1967. (b)

Mandler, G. Words, lists and categories: An experimental view of organized memory. In J. L. Cowan (Ed.), Thought and language. Tucson: University of Arizona Press, 1970.

Mandler, G., & Pearlstone, Z. Free and constrained concept learning and subsequent recall. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 126-131.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. Effects of organization and semantic similarity on recall and recognition. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 410-423.

Marshall, G. R. The organization of verbal material in free recall: The effects of patterns of associative overlap on clustering. Unpublished doctoral dissertation, New York University, 1963. (Cited in Cofer, 1965.)

Marshall, G. R. Stimulus characteristics contributing to organization in free recall. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 364-374.

Marshall, G. R., & Cofer, C. N. Associative indices as measures of word relatedness: A summary and comparison of ten methods. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 408-421.

Martin, E. Toward an analysis of subjective phrase structure. Psychological Bulletin, 1970, 74, 153-166.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-96.

Miller, G. A. Psycholinguistic approaches to the study of communication. In D. L. Arm (Ed.), Journeys in science: Small steps--great strides. Albuquerque: University of New Mexico Press, 1967.

Miller, G. A. A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 1969, 6, 169-191.

Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. Psychological Review, 1949, 56, 311-324.

Miller, G. A., Galanter, E., & Pribram, K. Plans and the structure of behavior. New York: Holt, Rinehart, & Winston, 1960.

Murray, D. J. Vocalization-at-presentation and immediate recall with varying presentation rates. Quarterly Journal of Experimental Psychology, 1965, 17, 47-56.

Needham, R. M. The theory of clumps, II. Report M. L. 139, Cambridge Language Research Unit, Cambridge, England, 1961.

Novinski, L. Part-whole and whole-part free recall learning. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 152-154.

Ornstein, P. A. The role of prior-list organization in a free-recall transfer task. Unpublished doctoral dissertation, University of Wisconsin, 1968.

Ornstein, P. A. The role of prior-list organization in a free-recall transfer task. Journal of Experimental Psychology, 1970, 86, 32-37.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1967.

Pellegrino, J. W. A general measure of organization in free recall for variable unit size and internal sequence consistency. Behavioral Research Methods and Instrumentation, 1971, 3, 241-246.

Pollio, H. R., Kasschau, R. A., & DeNise, H. E. Associative structure and the temporal characteristics of free recall. Journal of Experimental Psychology, 1968, 76, 190-197.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Postman, L. Does interference theory predict too much forgetting? Journal of Verbal Learning and Verbal Behavior, 1963, 2, 40-48.

Postman, L. Effects of word frequency on acquisition and retention under conditions of free-recall learning. _Quarterly Journal of Experimental Psychology_, 1970, _22_, 185-195.

Postman, L. A pragmatic view of organization theory. Paper presented at the Pittsburgh Conference on Organization and Memory, March 1971.

Postman, L., & Keppel, G. A. Retroactive inhibition in free recall. _Journal of Experimental Psychology_, 1967, _74_, 203-211.

Postman, L., & Phillips, L. W. Short-term temporal changes in free recall. _Quarterly Journal of Experimental Psychology_, 1965, _17_, 132-138.

Puff, C. R. Clustering as a function of the sequential organization of stimulus word lists. _Journal of Verbal Learning and Verbal Behavior_, 1966, _5_, 503-506.

Roberts, W. A. The priority of recall of new items in transfer from part-list to whole-list learning. _Journal of Verbal Learning and Verbal Behavior_, 1969, _8_, 645-652.

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.

Seibel, R. An experimental paradigm for studying the organization and strategies utilized by individual _Ss_ in human learning and an experimental evaluation of it. Paper presented at meetings of The Psychonomic Society, Niagara Falls, Ont., October 1964. (Cited in Mandler, 1967a.)

Seibel, R. Organization in human learning: The study sheet paradigm and experiments one and two. Paper presented at meetings of The Psychonomic Society, Chicago, October 1965.

Shapiro, S. I., & Palermo, D. S. Mediated clustering in free recall. Journal of Experimental Psychology, 1967, 75, 365-371.

Shuell, T. J. Retroactive inhibition in free-recall learning of categorized lists. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 797-805.

Shuell, T. J. Clustering and organization in free recall. Psychological Bulletin, 1969, 72, 353-374.

Slamecka, N. J. An examination of trace storage in free recall. Journal of Experimental Psychology, 1968, 81, 557-560.

Sokal, R. R., & Sneath, P. H. A. Principles of numerical taxonomy. San Francisco: W. H. Freeman & Co., 1963.

Tucker, L. R, & Messick, S. J. An individual differences model for multidimensional scaling. Psychometrika, 1963, 28, 333-367.

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychological Review, 1962, 69, 344-354. (a)

Tulving, E. The effect of alphabetical organization on memorizing unrelated words. Canadian Journal of Psychology, 1962, 16, 185-191. (b)

Tulving, E. Intratrial and intertrial retention: Notes toward a theory of free recall verbal learning. Psychological Review, 1964, 71, 219-237.

Tulving, E. The effect of order of presentation on learning of "unrelated" words. Psychonomic Science, 1965, 3, 337-338.

Tulving, E. Subjective organization and effects of repetition in multi-trial free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 193-197.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), Verbal behavior and general behavior theory. Englewood Cliffs, N. J.: Prentice Hall, 1968.

Tulving, E. Short and long term memory: Different retrieval mechanisms. In K. H. Pribram & D. E. Broadbent (Eds.), Biology of memory. New York: Academic Press, 1970.

Tulving, E., & Osler, S. Transfer effects in whole/part free recall learning. Canadian Journal of Psychology, 1967, 21, 253-262.

Tulving, E., & Patterson, R. D. Functional units and retrieval processes in free recall. Journal of Experimental Psychology, 1968, 77, 239-248.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 381-391.

Tulving, E., & Psotka, J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. Journal of Experimental Psychology, 1971, 87, 1-8.

Underwood, B. J. Degree of learning and the measurement of retention. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 112-129.

Waugh, N. C., & Norman, D. A. Primary memory. Psychological Review, 1965, 72, 89-104.

Wood, G. Organization, retroactive inhibition, and free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 327-333.

APPENDIX A

FURTHER DEVELOPMENTS OF PROXIMITY ANALYSIS

This appendix describes several extensions and elaborations of the proximity technique presented in Chapter 2. The first section deals with the problem of handling repetitions and intrusions. The second section considers the problem of determining the organization of the "average" subject in a group, and the assumptions this entails. Following, some ways to explore individual differences are discussed. A third section considers the use of interword response times to index item relatedness when oral recall is obtained.

A.1 Repetitions and Intrusions

In free recall studies, subjects typically produce words in output which did not appear on the input lists, and produce the same word more than once on a given trial. Without strong reasons for excluding these "errors," intrusions and repetitions should be considered as integral to the data as responses scored "correct." A complete discussion of the measurement of organization in FRL, therefore, should make explicit the treatment of such responses.

To discuss the approach taken here, consider a list of 10 items denoted by the letters A through J, and suppose that a subject on a given trial has produced the sequence

C B G A X J B D .

The subject thus has one intrusion (the item X) and one repetition (item B).

Repetitions have most often been dealt with by arbitrarily ignoring all but one occurrence of a list item, usually retaining the first. Since there is no reason to favor the first or second occurrence of B in the list above, we shall consider both as potentially informative. If item B is in fact organized for this subject along with items C and G, rather than items J and D, this should be indicated by the contiguous occurrence of items B, C, and G on other trials, thus giving larger proximity scores over a block of trials for B with C and G than with J.

There is another case in which the argument for retaining all instances of an item is more compelling. It is plausible to think of an organized schema as building up in stages, with S-units growing in size, and perhaps breaking up, reorganizing, and merging with increased practice. Now, if we analyzed the trials prior to the one shown above and, separately, the trials following it, finding that B clustered with items C and G in the former block of trials but with J and D in the latter, then the occurrence of a repetition on the given trial would be quite significant. With reasonable confidence we could infer this trial to be the locus of the reorganization of memorial units.

Since extra-list intrusions are usually highly idiosyncratic (except, as Deese, 1959, has shown, where all members of a list are free associates of a given word), adding them as additional words in the analysis will probably not be overly revealing. The position in which an intrusion occurs, however, is important. Thus while we do not want to count X as an additional item in the sequence given earlier, we still want to say that A occurred two positions away from J, rather than one. When it is important to consider specific intrusions, as is the case in Deese (1959), these words may be added to the proximity matrix as additional rows and columns and included in the cluster analysis.

A.2 Group Data and Individual Differences

One asset of the proximity method is that it allows a determination of the organization displayed by each subject. There are practical reasons, however, for which an investigator will want to combine individual data and determine if there are any components of organization common to a group as a whole, or whether there are empirically identifiable subgroups of subjects organizing in different ways. Although the investigator would like to know that statements he makes about a group hold as well for individual subjects, there is a danger of being buried by an avalanche of data. If at the outset he performs a separate cluster analysis for each subject, he may lose sight of the forest for the trees. It is often wise to begin simply and look at the "modal" or "typical" organization for a group. If the method of proximity analysis is to be generally useful, it is desirable that it be sufficiently flexible so that the level of detail can be chosen to suit the needs of inquiry.

Estimates of proximity from group data. There are several alternative assumptions concerning the nature and importance of individual differences in organization which might motivate an analysis of group organizational structure. First, one may assume that all subjects organize in essentially the same way and that any differences between subjects represent minor, random variations from the organizational strategy which is believed to characterize the recall behavior of the group. In the light of Marshall's (1963; 1967, Exp. II) studies of idiosyncratic clustering, this assumption seems appropriate to the extent that there is a strong or transparent organization inherent in the list itself, as for example in categorized lists. In this study pairs of items were selected at varying levels of associative relatedness (Marshall & Cofer's, 1963, MR measure). As MR level increased (items more "objectively" related), subjects indicated successively fewer idiosyncratically related item-pairs and these accounted for a decreasing proportion of their total clustering scores. Additionally, Tulving's (1962a) work indicates that there is some degree of communality of subjective organization across subjects learning unrelated words, and that this overlap increases with practice in FRL.

Alternatively, an investigator may decide that the only interesting aspects of organization are those which are common to the majority of subjects. Recall strategies which are shared only within small subgroups are felt to be unimportant.

The data from individual subjects may be combined to give group estimates of proximity in a variety of ways. One way to do this is to consider all of the data for a group as if it were S replications of a single subject and take the average proximity over all subject-trials on which both members of the pair (i,j) were recalled. Formally, the group estimate, $P_{ij.}$, of the proximity of items i and j over trials t_1 to t_2, may be expressed as

$$P_{ij.} = \frac{\sum_{s=1}^{S} \sum_{t=t_1}^{t=t_2} P^*_{ijst}}{\sum_{s=1}^{S} \sum_{t=t_1}^{t=t_2} n_{ijst}}$$

Equivalently, this may be described as pooling the protocols for all subjects over trials t_1 to t_2 and considering these as the recall of one subject for $S \cdot (t_2 - t_1 + 1)$ trials.

It should be noted that it must be possible to consider subjects strictly as replications in order for averaging to give meaningful results. For, if subjects organize very differently, the average proximities may represent the organization of none of them.

To illustrate the problem of averaging when subjects are organizing in radically different ways, suppose three subjects, I, II, and III, learn a list composed of four items, A, B, C, D. Each of our hypothetical subjects forms two S-units of two items each, but to be perfectly diabolical, each subject chooses a different one of the three ways in which this can be done, viz.,

I: [A, B] , [C, D]

II: [A, C] , [B, D]

III: [A, D] , [B, C] .

Assume that on each of eight trials, each subject recalls all four items in a different one of the eight possible orders consistent with his organization. For instance, subject I could recall A,B,C,D; A,B,D,C; C,D,A,B; ..., but not, say, A,C,B,D. The upper section of Figure A1 shows the proximities which would be derived from such protocols. For subject I, the pairs (A,B) and (C,D) both have an average proximity of 3.0, the maximum for a four-item list. All other pairs have a proximity of 2.0. Subjects II and III each have the same distribution of proximity scores, but arranged according to the composition of their S-units. The clusterings at the bottom of

Figure A1 show the organization of each subject clearly. When we average these proximities over subjects, however, all information concerning organization is lost; the group proximities are uniformly equal to 2.33, and the cluster analysis portrays an undifferentiated four-item group.

When there are relatively strong groupings built into the list by the experimenter, it is to be expected that most subjects will display them in recall, and averaging will not cause undue concern. More caution is required with unrelated lists and lists whose items have been drawn from weak levels of some scheme of relatedness, e.g., associative frequency, taxonomic strength, concept dominance, etc. In any case, if the cluster analysis of group data reveals item groupings which cluster at relatively high levels of proximity, that is to say, the clusters are highly compact, then univocal organization may be inferred. Figure A1 indicates that the effect of aggregating discordant organizational groupings is to contract the proximities to a middle range, i.e., to reduce their variance. High average proximities can only obtain for groups of items which cluster in recall for most of the subjects in a group.

Individual differences in organizational strategy. In the last section it was shown that the averaging of proximities over all subjects in a group is only appropriate to the extent that subjects are all organizing in the same manner. When this nomothetic analysis is ruled out, because it is not valid, the idiographic alternative of separate analyses for each individual may be equally unattractive because it is unwieldy.

Individual differences in organization can be thought of as arising from a combination of three components: (a) completely idiosyncratic differences reflecting personal verbal predispositions, (b) systematic

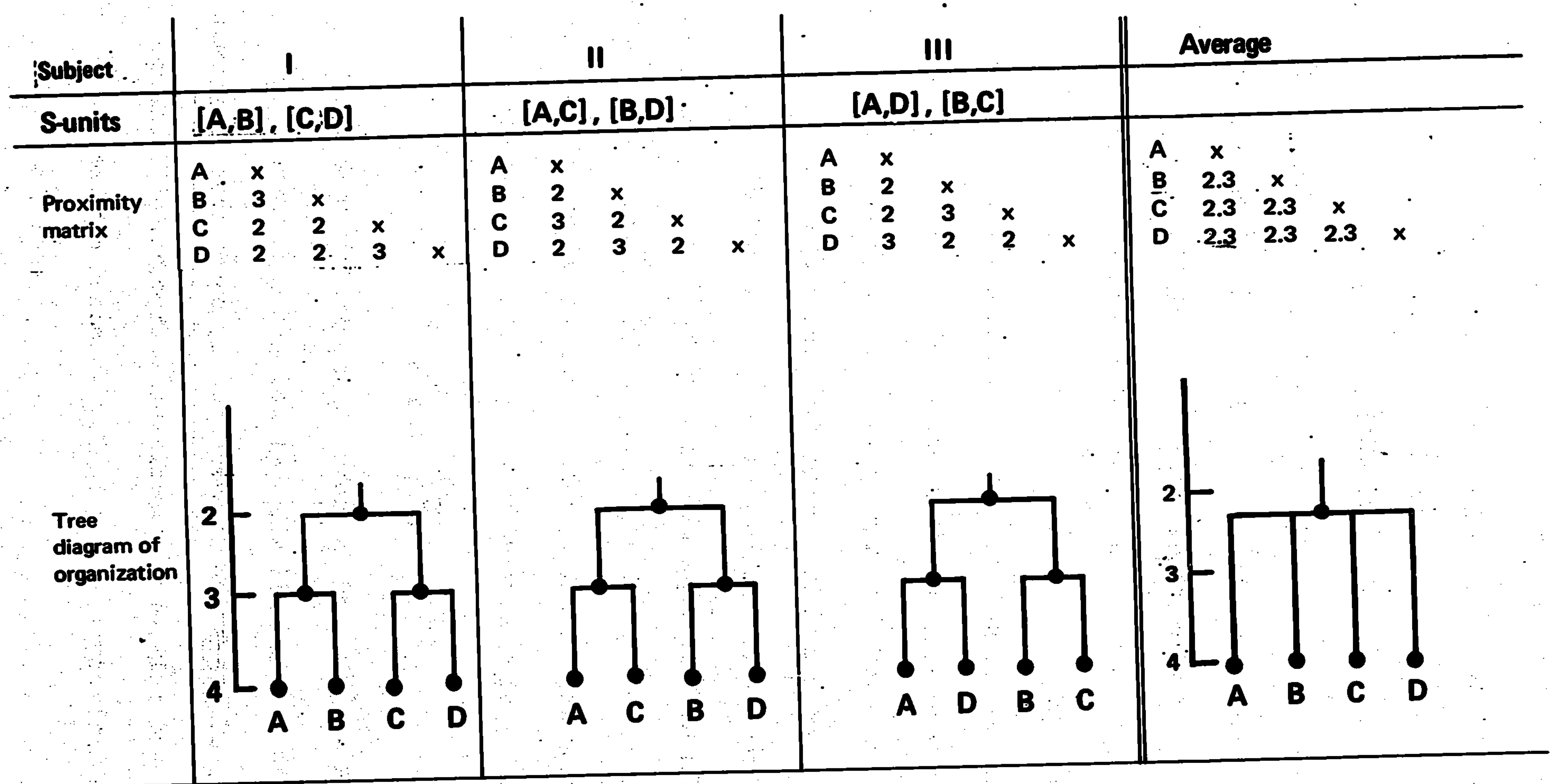


Fig. A1. Effects of averaging when subjects differ widely in organization.

differences between subjects which nonetheless are unpredictable, perhaps for lack of appropriate predictor variables, and (c) systematic differences which are predictable in practice. Only when differences in organization are mostly of type (c) is it possible a priori to sort subjects into homogeneous subsets and determine the organization for the "average individual" in each subset. When the varieties of grouping strategy employed by subjects are not predictable, it would be desirable to have some technique to determine empirically any systematic differences which do exist, and to determine the organization corresponding to each such empirical strategy.

One technique which bridges the gap between the idiographic and nomothetic approaches is the individual differences model for multidimensional scaling developed by Tucker and Messick (1963). In this procedure the half-matrix of proximity values for each subject is strung out in a single column-vector of $N(N - 1)/2$ elements, for N items. The vectors for individuals are arranged side-by-side to form the group data matrix, G, a stimulus-pairs by subjects matrix. The matrix G is factored into principal components, treating subjects as variables. As a result of this factoring, G is approximated by the product of two matrices (Figure A2),

$$_{m}G_{S} = {}_{m}P_{r}Q'_{S} ,$$

where $m = N(N - 1)/2$ and the subscripts give the number of rows and columns of each matrix. If all subjects are organizing in the same way, then each pair of items should have roughly equal proximity values across all subjects, and only one "significant" component will emerge. The number, r, of substantial components actually obtained represents the number of ways in which subjects differ in the structure which they impose on the items in their recall, i.e., the number of distinct "organizational viewpoints."

The matrix Q gives the loading, or relative weight, of each subject on each of the r organizational viewpoint dimensions. Hence, each subject may be classified according to a (hopefully) small number of viewpoints. The matrix P contains the loadings of stimulus pairs on the dimensions of organization. Typically, the matrix P is rotated to a matrix P* for ease of interpretation. The rotation is usually performed according to some criterion, e.g., simple structure in the factor space of individuals, or by selecting some "idealized" individuals in this space. Each of the r columns of P* contains proximity estimates of the item pairs for a given empirical viewpoint, which can be arrayed in r separate N by N proximity matrices.

In the Tucker and Messick procedure, each of these viewpoint matrices is then analyzed by multidimensional scaling to yield a spatial representation for each viewpoint. With data from free recall protocols, however, the viewpoint proximities can be input to the cluster analysis procedure to determine the hierarchical structure for each dimension of organization. The analysis indicates that, of the S individual proximity matrices, only r of them represent different organizational schemes, and each subject's proximities can be given as a combination of these r viewpoints. If $r = 1$, i.e., only one viewpoint exists, then the first principal component of G will approximate the average proximities.

Because it is based on the linear, component analytic model, the Tucker-Messick procedure places strong metric restrictions on the data

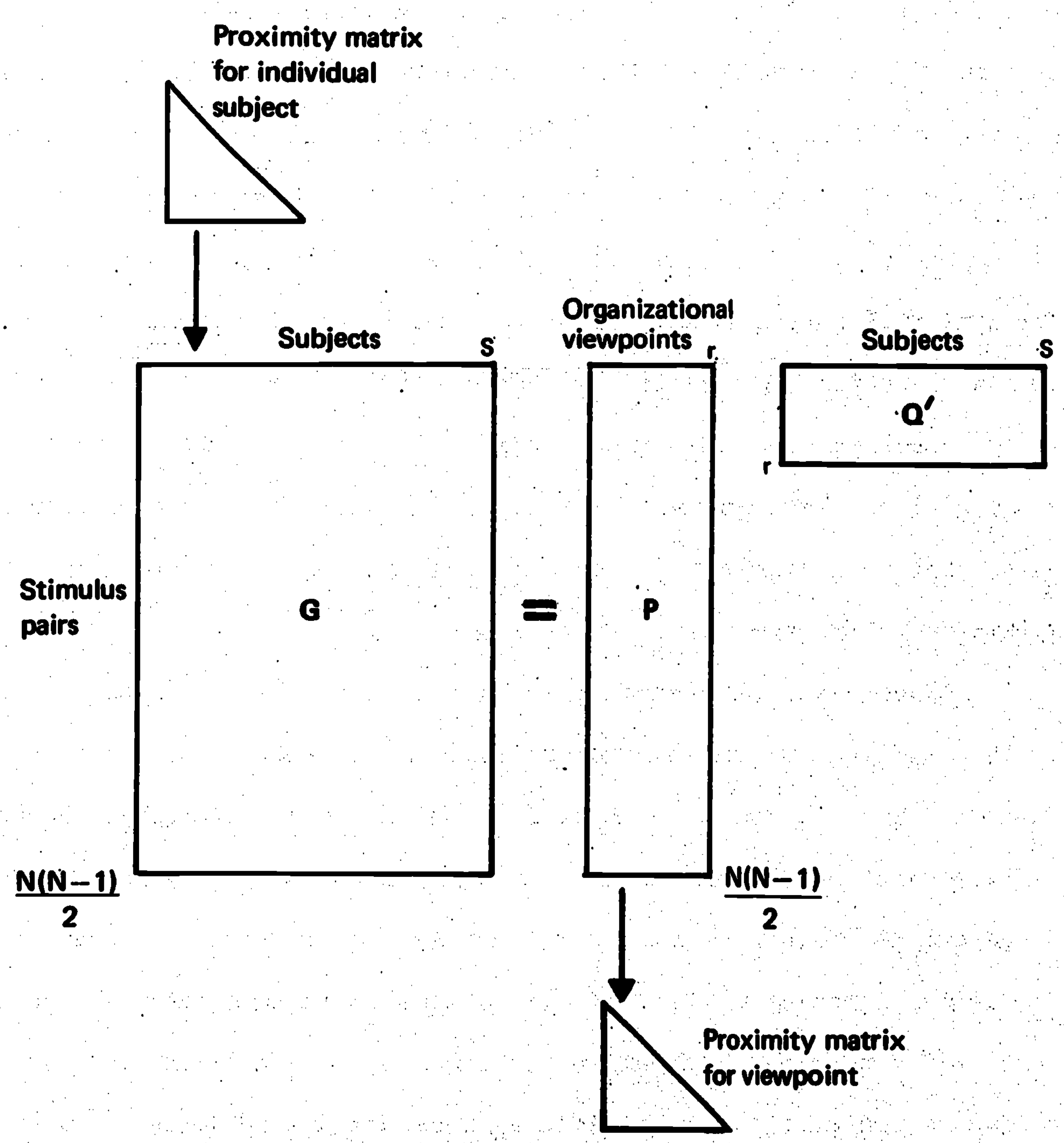


Fig. A2. Outline of Tucker and Messick's procedure for individual differences analysis.

which may not always be fulfilled in practice. In particular, the method requires that the input distance or proximity estimates be measured on a ratio scale. An alternative approach to individual differences, and one which is more directly related to hierarchical clustering is presented in Appendix B, where the general problem of comparing clustering solutions is considered.

A.3 Interword Response Times

The discussion in Chapter 2 was based on the idea that the organization displayed in free recall output could be indexed by interitem proximities. Methods of cluster analysis could then be used to locate groups of items which are highly proximal throughout recall. The cluster analysis gives an overall picture of the inferred organization of the list, but one in which the finer details can sometimes be discerned.

When recall is obtained in written form, the data for each subject consist of an ordered list of the items remembered on each trial. For lack of any additional information, the proximity of two items, both recalled on a given trial, was specified in terms of the number of items intervening between them. This is equivalent to assuming adjacent items to be equally spaced along some latent continuum of recall proximity. If, in the example given in A.1, items C and B formed one S-unit, while G and A formed another, the method would nevertheless assign the same proximity score to (B,G) as to (C,B) for that trial. This is all that can be done objectively, since written recall gives no information concerning the length of intervals between items.[1]

[1]It is for this reason that average proximity over some block of trials was suggested. If, in the example, (C,B) and (G,A) are really separate functional units in recall, then one would expect that C and B would usually

The situation is different, however, when subjects are required to produce oral recall. In this case, it is often noticed that subjects typically recall items in bursts, i.e., groups of words whose interword response times (IRTs) are substantially smaller than the IRTs of immediately preceding and following words. In a study of the associative structure of items which are recalled in bursts, Pollio, Kasschau, and DeNise (1968) concluded that "when Ss are asked to recall highly structured word sets, the temporal characteristics of individual recalls are markedly irregular, with fast recall sequences containing highly similar and associatively related words and with slower recall sequences containing less similar and more weakly connected words" (p. 196). Similar conclusions may be derived from the studies of Bousfield and Sedgewick (1944) on continued associations to category labels, e.g., names of animals. For data averaged over a group of subjects, the cumulative number of responses to a category name as a function of time describes a smoothly increasing (negative exponential) curve. The corresponding curves of individuals, however, reveal that individual subjects typically respond in bursts, composed of items from some subclass of the category, e.g., wild animals, household pets, etc.

The IRT between two recalled words, then, can be taken as a measure inversely related to the probability that the words belong to a functional unit or chunk. The shorter the time interval between production of the items, the more likely it is that the items have been chunked. Therefore,

---

be quite proximal in a series of trials, as would G and A, whereas, e.g., B and G might be close together in recall on some trials and distant on other trials. When averages are taken over some block of trials (Eq. 2.4), the pairs (C,B) and (G,A) would have high proximity, whereas the proximity of (B,G) would be lower.

the organizational proximity between items assessed in terms of the response-time scale should provide a more adequate representation of the way in which the items have been organized by the subject.

The use of IRTs corresponds to a transformation from the scale of output position to one of response time, τ. On the former, intervals between items are presumed, but are unlikely to be, equal; on the time scale, intervals on the scale itself are (at least physically) equal, whereas intervals between items are assumed to reflect their proximity in the sense used here. By obtaining oral recall and using IRTs in the analysis of organization, it may be possible to gain information that is ignored when recall is written.[2]

This transformation of the scale is carried over to the proximity measure. For use with IRTs, the proximity between any pair of items, (i,j), may be expressed (cf. Eq. 2.4) as

$$P_{ij}(\tau) = \frac{\sum_{k=1}^{k} \phi_{ijk} \cdot \{T - |\tau_{ik} - \tau_{jk}|\}}{n_{ij}}$$

where T is the total time allowed for recall (constant from trial to trial), τ_{ik} is the latency of recall of item i on trial k, defined just in case $\phi_{ik} = 1$. The origin of the time scale may be taken arbitrarily at the start of the recall period.

---

[2]Studies by Craik (1969, 1970) and Murray (1965) have shown a small but consistent superiority of written over spoken response in FR. This difference according to response mode appears to be independent of mode of presentation (Craik, 1970). However, if recall scores are broken down into PM and SM components (Waugh & Norman, 1965), the output modality effect appears only in the primary memory component. Secondary memory, presumed to be the locus of subjective organization (Tulving, 1968), is independent of output modality.

An analysis of the organization of items over a large number of trials is unlikely to be very different whether we measure proximity along a scale of ordinal position or response time. It is predicted, however, that the latter scale will give a clearer picture of the change or development of organization over smaller blocks of trials.

APPENDIX B

COMPARISON OF CLUSTERING SOLUTIONS

The fact that clustering procedures of the type discussed here are discrete, nonprobabilistic methods requiring relatively weak conditions on the data has in part led to their appeal to investigators in diverse fields. For substantive and theoretical reasons, hierarchical representations find application in a variety of research efforts concerned with verbal behavior (Martin, 1970; Miller, 1969). One may therefore predict an increasing use of hierarchical clustering methods in psychology. While such techniques have great usefulness as exploratory, hypothesis-generating methods, the same properties responsible for their appeal apparently reduce their utility as confirmatory, hypothesis-testing methods. That is, because clustering techniques are discrete, giving rise to a nonprobabilistic structural description of data which is treated as error-free, they are well suited to exploratory work. Given these properties, however, it is difficult to see how the problem of "significance" of results may even be discussed, no less solved.

The explicit concern of this report is with the application of cluster analysis to the study of free recall, and not with such problems specific to cluster analysis per se. Nonetheless, it is inevitable that such questions be raised, if only to determine, by internal criteria, the success of this application.

Two basic problems are of interest. The first concerns the comparison of clustering solutions for different subjects or groups, and the second concerns the goodness-of-fit of a given set of data to a hierarchy.

B.1 Comparison of Clusterings

Given two hierarchical clusterings, H_1 and H_2, we desire to express their congruence or similarity. One approach to this problem may be indicated as follows. From Johnson's analysis of hierarchical clustering schemes we know that there is an exact correspondence between a hierarchical tree structure on a set $\underset{\sim}{S}$ of N objects and a distance matrix on $\underset{\sim}{S} \times \underset{\sim}{S}$ satisfying the ultrametric inequality. Hence the comparison of two tree structures may be reduced to that of expressing the similarity between two distance matrices which satisfy the UMI.

A hierarchical clustering, $H = [\{C\},\delta]$, consists of a sequence of partitions, or clusterings, $\{C_\ell\} = C_0, C_1, \ldots, C_N$ such that each cluster in $C_{\ell+1}$ is obtained by merging the clusters of C_ℓ, together with a set of numerical levels, δ_ℓ, $\ell = 0,1,\ldots,N$ whose values represent the diameter of the clusterings C_ℓ. The ultrametric corresponding to H may be defined by a matrix $D = D(i,j)$ given by

$$D(i,j) = \sup_\ell(\delta_\ell) \quad , \quad i,j \in C_\ell \quad .$$

In words, the ultrametric distance between objects i and j is the diameter of the clustering C_ℓ in which they are first joined in the same cluster.

Let d_1, d_2 be real valued symmetric matrices of dissimilarities on $\underset{\sim}{S} \times \underset{\sim}{S}$. In the present application, d_1 and d_2 could be derived via Eq. (2.5) from the proximity matrices for two subjects in free recall. If H_1 and H_2 are hierarchical clusterings derived from d_1 and d_2, respectively, the congruence of H_1 and H_2 may be assessed in terms of some measure of the similarity of D_1 to D_2. Let $\underset{\sim}{M}_D$ be the set of all dissimilarity metrics d on the set of objects, $\underset{\sim}{S} \times \underset{\sim}{S}$; similarly define $\underset{\sim}{M}_D$ as the set of all ultrametrics D induced on $\underset{\sim}{S} \times \underset{\sim}{S}$ by the clustering of d . Then $\underset{\sim}{M}_D$ may be regarded as a subset of $(N/2)(N - 1)$ - dimensional space, in which D_1 and D_2 are represented by points. A measure of the distance between these points, and hence of the discrepancy between H_1 and H_2 is provided by

$$\rho(D_1,D_2) = \left\{\frac{1}{N(N - 1)} \sum_{i \neq j} \sum [D_1(i,j) - D_2(i,j)]^2 \right\}^{\frac{1}{2}} ,$$

that is, the root mean square discrepancy between corresponding elements. Hence ρ is a metric on $\underset{\sim}{M}_D$.

Given a group of s subjects, and clusterings $H_1,H_2,\ldots,H_s$, we may compute $\rho_{gh} = \rho(D_g,D_h)$ for all pairs of subjects and array these values in a matrix $R = \{\rho_{gh}\}$. Then R may be analyzed by suitable techniques, e.g., multidimensional scaling, or a "second-order" cluster analysis, to determine individual differences in the structure of the hierarchies.

An alternative approach has been suggested by Gruvaeus and Wainer (personal communication)[1] who have proposed as a measure of similarity between two hierarchies the element-by-element correlation of D_1 and D_2 , and have written a computer program to perform these calculations. Our preference for the distance measure given above stems from an uneasiness regarding the ability of a correlation coefficient to discriminate adequately among degrees of similarity between hierarchies in the range typically of interest. In the present application, subjects will tend to have some

[1]The present development owes much to the work of Gruvaeus and Wainer.

degree of overlap in the way in which they organize a set of verbal items, which tendency increases over repeated free recall trials (Tulving, 1962a) and as a function of the a priori relatedness of the stimuli. In such situations it is not clear (a) that the elements of D_1 and D_2 will be linearly or even monotonically related and (b) that the correlation between elements of D_1 and D_2 will reflect only the degree to which the corresponding hierarchies are similar, and not other, irrelevant aspects of their distributions. The vagueness of the preceding comments should be taken to imply that both suggested measures be considered tentative until their behavior in situations of interest is better understood. The nature of the problem suggests the need for Monte Carlo work.

B.2 Goodness of Fit

In section 2.5 we raised the question of the degree to which a given empirical distance matrix conforms to the ultrametric inequality, i.e., to what extent the matrix has exact tree structure representation. Johnson's minimum and maximum methods will give identical results whenever the UMI is satisfied, and interpretation of the results would involve no choice between the two. With real data, some deviations from the UMI are to be expected, if only due to random error, and in some situations, a tree structure representation may be grossly inappropriate. Since Johnson's clustering procedure will always find a hierarchical solution for any set of data, it is obviously desirable to be able to express a degree of confidence in the adequacy with which a given set of data is so represented.

A variety of ad hoc solutions may be proposed to deal with goodness of fit. Some possible approaches are: (a) The UMI states a relation to be satisfied by the distances between all sets of three objects. A rough indicant of the degree to an empirical dissimilarity matrix satisfies this relation would be given by the proportion of triples in which the UMI holds. (b) When a data matrix has exact tree structure, all nodes in the hierarchy will be identical under the diameter and connectedness method solutions. Miller (1969) has proposed counting the number of nodes in common between two solutions as an index of fit to a hierarchy. In assessing the value obtained for a given set of data, he compared the result for real data with that found for "statistical subjects."

The approach of the previous section can, however, be expanded to deal with the problem of goodness of fit. Again, let d be a matrix of dissimilarities and D the ultrametric imposed on d by fitting a hierarchical clustering scheme. A measure of the distortion introduced in representing d by the tree structure may be given by

$$\rho(d,D) = \frac{\|d - D\|}{\|d\|} = \left\{ \frac{\sum\sum_{i \neq j} (d_{ij} - D_{ij})^2}{\sum\sum_{i \neq j} d_{ij}^2} \right\}^{\frac{1}{2}} .$$

While no distribution theory has been worked out to allow precise tests of fit, experience with this statistic suggests that values under 0.1, or 10%, may be regarded as adequate. An unnormalized form of ρ has been proposed by Hartigan (1967), who considered the further problem of finding trees to minimize the mean square discrepancy over M_D .

For some purposes it may be desired to determine whether the maximum method or the minimum method represents a better fit to a given set of data. In the maximum method, each merged cluster becomes more distant from some other clusters and nearer to none. Under the minimum method, the reverse obtains. Therefore,

$$D_{min}(i,j) \leq d(i,j) \leq D_{max}(i,j) \quad , \quad \text{for all} \quad i,j\varepsilon S \ ,$$

where D_{min} and D_{max} are the ultrametrics induced by the two methods. Thus the maximum method effects an expansion of the system of objects, while the minimum method causes the system to contract. It is conjectured that D_{min} is the largest ultrametric such that $D \leq d$; and D_{max} the smallest ultrametric for which $d \leq D$ holds, size measured in the sense of $\|D\|$.

Thus, a measure of the goodness of fit of the maximum method relative to that of the minimum method is provided by the <u>expansion ratio</u>,

$$E = \frac{\rho^2(d,D_{max})}{\rho^2(d,D_{min})} = \frac{\|d - D_{max}\|^2}{\|d - D_{min}\|^2}$$

$$= \frac{\Sigma\Sigma[d(i,j) - D_{max}(i,j)]^2}{\Sigma\Sigma[d(i,j) - D_{min}(i,j)]^2} \ .$$

Values of $E > 1$ indicate a smaller mean square error in fitting the dissimilarities by the maximum method than by the minimum method, and vice versa for $E < 1$.

DISTRIBUTION LIST

NAVY

4 Director, Personnel and Training
Research Programs
Office of Naval Research
Arlington, VA 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Director
ONR Branch Office
536 South Clark Street
Chicago, IL 60605

1 Office of Naval Research
Area Office
207 West 24th Street
New York, NY 10011

6 Director
Naval Research Laboratory
Washington, D.C. 20390
ATTN: Library, Code 2029 (ONRL)

6 Director
Naval Research Laboratory
Washington, D.C. 20390
ATTN: Technical Information
Division

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, VA 22314

1 Behavioral Sciences Department
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

1 Chief
Bureau of Medicine and Surgery
Code 513
Washington, D.C. 20390

1 Chief
Bureau of Medicine and Surgery
Research Division (Code 713)
Department of the Navy
Washington, D.C. 20390

1 Commanding Officer
Naval Medical Neuropsychiatric
Research Unit
San Diego, CA 92152

1 Director
Education and Training Sciences
Department
Naval Medical Research Institute
National Naval Medical Center
Building 142
Bethesda, MD 20014

1 Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

1 Chief of Naval Training
Naval Air Station
Pensacola, FL 32508
ATTN: Capt. Allen E. McMichael

1 Mr. S. Friedman
Special Assistant for Research
& Studies
OASN (M&RA)
The Pentagon, Room 4E794
Washington, D.C. 20350

1 Chief, Naval Air Reserve Training
Naval Air Station
Box 1
Glenview, IL 60026

1 Chief
Naval Air Technical Training
Naval Air Station
Memphis, TN 38115

1 Chief of Naval Air Training
Code 017
Naval Air Station
Pensacola, FL 32508

1 Chief of Naval Operations (Op-98)
Department of the Navy
Washington, D.C. 20350
ATTN: Dr. J. J. Collins

2 Technical Director
Personnel Research Division
Bureau of Naval Personnel
Washington, D.C. 20370

2 Technical Library (Pers-11B)
Bureau of Naval Personnel
Department of the Navy
Washington, D.C. 20360

1 Technical Director
Naval Personnel Research and
Development Laboratory
Washington Navy Yard, Building 200
Washington, D.C. 20390

3 Commanding Officer
Naval Personnel and Training
Research Laboratory
San Diego, CA 92152

1 Chairman
Behavioral Science Department
Naval Command and Management Division
U.S. Naval Academy
Luce Hall
Annapolis, MD 21402

1 Superintendent
Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

1 Commanding Officer
Service School Command
U.S. Naval Training Center
San Diego, CA 92133

1 Research Director, Code 06
Research and Evaluation Department
U.S. Naval Examining Center
Building 2711 - Green Bay Area
Great Lakes, IL 60088
ATTN: C.S. Winiewicz

1 Technical Library
Naval Ordnance Station
Indian Head, MD 20640

1 Mr. George N. Graine
Naval Ship Systems Command (SHIP 03H)
Department of the Navy
Washington, D.C. 20360

1 Technical Library
Naval Ship Systems Command
National Center, Building 3 Room 3
S-08
Washington, D.C. 20360

1 Col. George Caridakis
Director, Office of Manpower Utilization
Headquarters, Marine Corps (A01H)
MCB
Quantico, VA 22134

1 Col. James Marsh, USMC
Headquarters, Marine Corps (A01M)
Washington, DC 20380

1 Dr. James J. Regan, Code 55
Naval Training Device Center
Orlando, FL 32813

ARMY

1 Behavioral Sciences Division
Office of Chief of Research and
Development
Department of the Army
Washington, D.C. 20310

1 U.S. Army Behavior and Systems
Research Laboratory
Commonwealth Building, Room 239
1320 Wilson Boulevard
Arlington, VA 22209

1 Director of Research
US Army Armor Human Research
Unit
ATTN: Library
Bldg 2422 Morande Street
Fort Knox, KY 40121

1 Commanding Officer
ATTN: LTC Cosgrove
USA CDC PASA
Ft. Benjamin Harrison, IN 46249

1 Director
Behavioral Sciences Laboratory
U.S. Army Research Institute of
Environmental Medicine
Natick, MA 01760

1 Division of Neuropsychiatry
Walter Reed Army Institute of
Research
Walter Reed Army Medical Center
Washington, D.C. 20012

AIRFORCE

1 AFHRL (TR/Dr. G. A. Eckstrand)
Wright-Patterson Air Force Base
Ohio 45433

1 AFHRL (TRT/Dr. Ross L. Morgan)
Wright-Patterson Air Force Base
Ohio 45433

1 AFSOR (NL)
1400 Wilson Boulevard
Arlington, VA 22209

1 Lt. Col. Robert R. Gerry, USAF
Chief, Instructional Technology
Programs
Resources & Technology Division
(DPTBD DCS/P)
The Pentagon (Room 4C244)
Washington, D.C. 20330

1 Personnel Research Division
(AFHRL)
Lackland Air Force Base
San Antonio, TX 78236

1 Director
Air University Library (AUL-8110)
Maxwell Air Force Base,
Alabama, 36112

1 Commandant
U.S. Air Force School of
Aerospace Medicine
ATTN: Aeromedical Library
Brooks AFB, TX 78235

1 Director of Manpower Research
OASD (M&RA) (M&RU)
Room 3D960
The Pentagon
Washington, D.C.

OTHER GOVERNMENT

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh Street, S.W.
Washington, D.C. 20591

1 Dr. Alvin E. Goins, Chief
Personality and Cognition Research
Section
Behavioral Sciences Research Branch
National Institute of Mental Health
5454 Wisconsin Ave., Room 10A01
Bethesda, MD 20014

MISCELLANEOUS

1 Dr. John Annett
Department of Psychology
Hull University
Hull
Yorkshire, ENGLAND

1 Dr. Richard C. Atkinson
Department of Psychology
Stanford University
Stanford, CA 94305

1 Dr. Bernard M. Bass
University of Rochester
Management Research Center
Rochester, NY 14627

1 Dr. Lee R. Beach
Department of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Mats Bjorkman
University of Umea
Department of Psychology
Umea 6, SWEDEN

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts & Sciences
River Campus Station
Rochester, NY 14627

1 Dr. Marvin D. Dunnette
University of Minnesota
Department of Psychology
Elliot Hall
Minneapolis, MN 55455

1 Dr. David Weiss
University of Minnesota
Department of Psychology
Elliot Hall
Minneapolis, MN 55455

1 Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
2001 "S" St. N.W.
Washington, DC 20037

1 Dr. Albert S. Glickman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. Bert Green
Department of Psychology
Johns Hopkins University
Baltimore, MD 21218

1 Dr. Duncan N. Hansen
Center for Computer Assisted Instruction
Florida State University
Tallahassee, FL 32306

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
11428 Rockville Pike
Rockville, MD 20852

1 Dr. M.D. Havron
Human Sciences Research, Inc.
Westgate Industrial Park
7710 Old Springhouse Road
McLean, VA 22101

1 Human Resources Research Organization
Library
300 North Washington Street
Alexandria, VA 22314

1 Human Resources Research Organization
Division #3
Post Office Box 5787
Presidio of Monterey, CA 93940

1 Human Resources Research Organization
Division #4, Infantry
Post Office Box 2086
Fort Benning, GA 31905

1 Human Resources Research Organization
Division #5, Air Defense
Post Office Box 6021
Fort Bliss, TX 77916

1 Human Resources Research Organization
Division #6, Aviation (Library)
Post Office Box 428
Fort Rucker, AL 36360

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
Santa Barbara Research Park
6780 Cortona Drive
Goleta, CA 93017

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Psychological Abstracts
American Psychological Association
1200 Seventeenth Street, N.W.
Washington, D.C. 20036

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
University Park
Los Angeles, CA 90007

1 Dr. George E. Rowland
Rowland and Company, Inc.
Post Office Box 61
Haddonfield, NJ 08033

1 Dr. Robert J. Seidel
Human Resources Research Organization
300 N. Washington Street
Alexandria, VA 22314

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, PA 19087