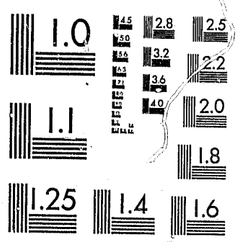
If you have issues viewing or accessing this file contact us at NCJRS.gov.

National Criminal Justice Reference Service

ncjrs

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D.C. 20531

08/04/82

MF-1

Final Report

MULTIVARIATE TAXONOMIC CLASSIFICATION FOR CRIMINAL JUSTICE RESEARCH

VOLUME 1

79121

OCTOBER, 1980

Prepared For:
National Institute of Justice
Washington, D.C.

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by
Public Domain
National Institute of Justice
to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

FINAL REPORT

MULTIVARIATE TAXONOMIC CLASSIFICATION FOR CRIMINAL JUSTICE RESEARCH

(Project No. 78-NI-AX-0065)

Volume 1: AN EVALUATIVE OVERVIEW OF CLASSIFICATION IN CRIMINAL JUSTICE

Principal Investigator: Tim Brennan Ph.D.*

Project Staff: Mary Downton
Randy Blair
Dick Williams Ph.D.

Behavioral Research Institute
Boulder, Colorado

October, 1980

Prepared for:

National Institute of Justice
Office of Research and Evaluation Methods
Department of Justice
Washington, D.C. 20531

*Now at Human Systems Institute, Inc., 1906 13th St., Suite 304, Boulder, Colorado, 80302.

NCJRS

JUN 2[illegible]1981

ACQUISITIONS

ACKNOWLEDGEMENTS

This project has benefitted from the input of many people. We thank particularly George Silberman of the National Institute of Justice (OREM) for his continued support, encouragement, and understanding of this project. Richard L. Linster (OREM), has also been extremely supportive of our work. We are indebted to Judge Horace B. Holmes and the staff members of the Boulder County Juvenile Probation department for allowing us access to selected data. Similarly, we acknowledge the support of Dr. Delbert S. Elliott in allowing us access to sub-samples of his larger national study of juvenile delinquency. In our analysis of prisoner classification, the staff at the Reception and Diagnostic Center at the Canon City Correctional facility was extremely helpful and kind. We thank particularly Lou Hesse, and Dr. Gideon Weiss, as well as Dr. Tom Crago. These persons were consistently supportive in our studies of prisoner classification. Dr. Herb Eber also provided invaluable help in regard to our analysis and understanding of the prisoner data.

Finally, I would like to acknowledge my appreciation to the project staff: Mary Downton, Dick Williams and Randy Blair. These people were a delight to work with and I am indebted to them and their intelligence, motivation and effort.

- T.B.

TABLE OF CONTENTS

Volume 1: AN EVALUATIVE OVERVIEW OF CLASSIFICATION IN CRIMINAL JUSTICE

	Page
Acknowledgements	
Chapter 1. Evaluative Criteria for Criminological Classification	1
Introduction	2
Structural Aspects of Classification Systems	4
The Attribute (Measurement) Space of Classifications	5
The Vertical Dimension of Classification	7
The Horizontal Dimension of Classification	8
Pragmatic Criteria for Classification Systems	15
Conclusion	26
Bibliography	27
Chapter 2. The Uses of Classification Within Criminology and Criminal Justice Systems	
Introduction	31
Uses of Classification in the Field of Research Criminology	32
Classification for Treatment	38
Classification in the Criminal Justice System	40
Conclusion	43
Bibliography	44
Chapter 3. The Failure of Classification in Criminology: Some Results of Inadequate Classification	50
Introduction	51
The Relation Between Inadequate Classification Research and the Failure of Classification in Practical Contexts	52
Ethical Issues in Applied Criminological Classification	61
Conclusion	69
Bibliography	70
Chapter 4. A Brief Examination of Decision Making and Classification in the Criminal Justice System	
Introduction	74
Conclusions	91
Bibliography	94

Chapter 5. Bibliographic Materials on the Use of Classification in Criminological Research and Its Use Within Criminal Justice Systems	96
Introduction	97
Bibliographic Materials	98
Volume 2: EVALUATING STATISTICAL CLASSIFICATION METHODS (MONTE CARLO STUDIES)	
Chapter 6. Concepts of Similarity as Used in Criminological Studies	112
Some General Problems Regarding Similarity as Used in Prior Criminological Research	113
Explanatory Theory as a Guide to the Specification of Similarity	122
Similarity in Criminological Studies Using Quantitative Classification Techniques	125
Conclusions	128
Bibliography	130
Chapter 7. Some Theoretical Observations on Distance Coefficients Versus Correlation Coefficients for Classification Studies	135
Introduction	136
Mahalanobis D^2: Orthogonality or Implicit Weighting of Attributes	137
Euclidean Distance vs. Correlation: Some Strong Opinions	141
Resolving the Paradox: A Brief Look at the Components of Similarity	142
Some Comments on Correlation	146
Distance Versus Correlation Coefficient: The 'Points-in-Space' Model Versus Profile Shapes	152
Other Ways of Measuring "Shape"	152
Bibliography	154
Chapter 8. Empirical Comparisons of Several Similarity Coefficients	156
Purpose	157
Method	157
Conclusions	169
Bibliography	178
Chapter 9. Evaluating Classifications: A Comparison of Five Coefficients for Partition Agreement	180
Method	185
Results	191
Conclusions	206
Bibliography	208

Chapter 10. Evaluation of Several Cluster Analysis Methods Using Artifical Data Sets 209

Introduction 210
Evaluation of Approaches 223
Results - Data Set D: Clusters of Different Shape and Separation 226
Data Set E: Differently Oriented Ellipses, With Minor Cluster Overlap 256
Data Set F: Overlapping Clusters of Different Compactness 291
Data Set U: A Uniform Distribution 331
Conclusions 353
Bibliography 367

Volume 3: APPLICATIONS OF NUMERICAL CLASSIFICATION METHODS IN CRIMINOLOGY AND GENERAL CONCLUSIONS

Chapter 11. Classification Analysis for the Epidemiologic Description of Crime Patterns 369

Introduction 370
Classification Approaches in the Epidemiology of Delinquency 377
A Behavioral Classification of Juvenile Violence 382
A Behavioral Classification of Juvenile Theft 400
A Behavioral Classification of Juvenile Drug Consumption Patterns 416
Discussion 426
Bibliography 430

Chapter 12. The I-Level Classification: An Evaluation by Taxometric Methods 432

Introduction 433
Methods 436
Data 439
Results 444
The Overlap Between Natural Clusters 458
Description of the Core Clusters 474
Cluster 1 474
Cluster 2 477
Cluster 3 478
Cluster 4 479
Cluster 5 480
Cluster 6 482
Conclusions 497
Bibliography 501

Chapter 13. A Classification of Probation Youth 503

Research Sequence 513
Results 515
Evaluation of ES Clustering Variables 517
Clustering Results for the HSPQ 523
Discussion 530
Bibliography 532

Chapter 14. Classification by Personality of a Prison Sample 533

Methods of Analysis and Data 537
Results 545
Prisoner Personality Types Based on the CAQ 569
Conclusions 595
Bibliography 598

Chapter 15. Making Predictions from Empirical Classifications Using Class Membership as a Dummy Variable 599

Method 601
Criteria 602
Predicting Violence 603
Validating the Predictions 609
Conclusions 611
Bibliography 614

Chapter 16. Concluding Remarks and Recommendations 615

The Diffusion and Utilization of Numerical Classification into the Criminal Justice System 616
Bibliography 625

CHAPTER 1

EVALUATIVE CRITERIA FOR CRIMINOLOGICAL CLASSIFICATION

INTRODUCTION

The goal of this introductory chapter is to lay out, in brief, a number of the major criteria which may be used in evaluating classification systems. Our examination of the practical uses of classification within criminology has led us to consider the inter-relationship between two basic sets of criteria i.e. pragmatic end-use criteria and the more statistical criteria which describe the structure of classification systems. We refer to these two sets as structural criteria and pragmatic criteria. It is important to realize that the structural criteria of classification systems have been discussed only minimally in the criminological (and sociological) literature, yet they have a profound influence upon the effectiveness and practical usefulness of classification systems. Attempts to improve the performance of applied classification will become more effective when guided by a better understanding of the relation between the structural aspects of classifications and the user functions to which classification is applied.

The evaluation of classification systems, particularly in an applied social psychological discipline such as criminology, is complex. The current published statements in criminology regarding the 'goodness' of classifications are somewhat vague, non-technical, and not particularly helpful (see Gibbons 1975, Solomon 1977, Kornfeld 1975, and others). The complexity of this evaluation

*A modified version of this chapter was presented at the International Differential Treatment Association Conference, May 1980, at Estes Park, Colorado, entitled Structural Criteria Required for Treatment Classifications by T. Brennan and M. Downton.

task stems from the fact that criminological classification systems have multiple purposes (particularly in applied settings), multiple structural criteria, and multiple evaluation criteria. In this chapter, therefore, we will review some of the main considerations which impinge upon the evaluation of taxonomies in criminology and the justice systems. The following two additional points should also be noted.

The Interdependence Between Evaluation and Purpose

Taxonomic systems, at base, represent a particular way of organizing complex multivariate information. This organization is usually done within the context of particular purposes. A critical first point is that without a statement of purpose, 'usefulness' cannot be judged. The value of the taxonomic system will always be related to some pre-existing purpose(s). Evaluative criteria for taxonomic work therefore must be chosen with particular purposes in mind.

The Interdependence Between Evaluative Criteria

A further complexity is that many of the evaluative criteria for taxonomic research are inter-dependent with one another. It may not be possible to jointly optimize certain of these criteria. For example, predictive accuracy may demand that descriptive accuracy be sacrificed. Similarly, reliability and precision are often inversely related, since precision is usually maximized at the lowest (sub-levels) of a hierarchical taxonomy while reliability is often optimized at the higher levels of a hierarchy.

We stress that the articulation of the inter-dependencies between different evaluative criteria and the various multivariate taxometric methods is a task

which is still at the earliest stages. Some tentative guidelines are evolving in regard to the optimization of particular classificatory goals (see Milligan 1980, Dubes and Jain 1979). These depend upon explicit specification of the purposes of the classification scheme as well as the nature of the available data, and a more complete understanding of the clustering methodologies available.

STRUCTURAL ASPECTS OF CLASSIFICATION SYSTEMS

In describing the structure of any taxonomic system, we will propose a limited number of fundamental criteria. Firstly, the classificatory domain (otherwise known as the measurement space, or attribute space) consists of the particular set of variables on which similarity between the objects is assessed. Secondly, the vertical dimension of classification may be understood as dealing with the level of inclusiveness and hierarchical organization of the classes. Highly inclusive (broad) categories may be useful for some purposes but not for others. Thirdly, the horizontal dimension of classification focusses upon the form and nature of the categorization that takes place at a particular level of inclusiveness. Thus, with the vertical dimension fixed, we may categorize the objects according to different grouping rules leading to different kinds of classes. Again, different purposes may be optimized by decisions made regarding horizontal level grouping. We will now examine these major principles in more detail.

THE ATTRIBUTE (MEASUREMENT) SPACE OF CLASSIFICATIONS

The nature of the attribute space (i.e., the set of classificatory variables) is critical in evaluating a taxonomic system. The selection of such classificatory variables will, in general, be primarily governed by the purpose of the classification and the theoretical orientation of the creator of the system. The following are some of the main criteria used in evaluating the attribute space.

Coverage: Comprehensive vs. Narrow. The set of classificatory variables may attempt to be descriptively comprehensive, covering a large number of salient features of the objects being classified. At the other extreme, only a few broad variables may be used. The attribute space of taxonomies which have a primary descriptive purpose may become extremely comprehensive. The limits of an attribute space must be set, usually by some theoretical argument. The main reviews of criminological taxonomies, e.g., Gibbons (1975), Hood and Sparks (1970), have complained that the sets of variables used in criminological taxonomies are not sufficiently comprehensive. This is an extremely complex issue (see Johnson 1968) which has not really been dealt with in the criminological literature.

Level of Measurement of Classificatory Variables. The attribute space can be evaluated according to the level of measurement of its variables. Opp (1973) criticized most prior criminological taxonomies on the basis of this issue. He claimed that the classificatory variables used in most earlier work were vague, unreliable and often non-operationally measured. Unreliability in the classificatory variables has devastating effects upon both the creation and use of taxonomies. Error and 'noise' are introduced, thus blurring both the class boundaries and central prototypes. The tasks of prediction, accurate diagnosis and assignment are undermined.

Theoretical Coherence of the Attribute Space. Any set of objects can be classified in an almost infinite number of ways since an almost infinite variety of possible classificatory variables can be specified (see Johnson 1968). Some theoretical gudiance is required in the selection (abstraction) of a set of classificatory variables (Toulmin, 1953). Every classificatory domain is somewhat artificial and partial in its coverage since it is abstracted from a total set of influences which impinge upon, or describe, the objects being classified. A useful way, therefore, of evaluating classification schemes is to examine the level of coherence or justification for the selection of classification variables. At one extreme, variables are chosen on hunch, intuition, or some other relatively implicit reason. At the other, they may be explicitly specified by implication from a clearly developed theory. The ad hoc, ill-defined atheoretical approach has a number of serious problems, including: high likelihood of spurious findings, inefficiency, a confusing proliferation of taxonomic systems, and endless disagreements and non-convergence of findings. Hood and Sparks (1970) indicate that in criminology the great majority of empirically constructed taxonomies have been based upon a relatively atheoretical specification of classificatory variables. Toulmin (1953) and Enc (1976) both eloquently argue for the mutual development of theories and classifications, each stressing the value of classificatory observations being influenced and controlled by reference to theoretical positions. The relatively poor state of theory development in criminology unfortunately provides inadequate guidance at the present time. We would argue that the atheoretical approach to the selection of classificatory variables should be avoided. The status and future development of empirical criminological classification would be enhanced by greater attention to the theoretical coherence of the classification

variables in relation to the purposes of classification.

THE VERTICAL DIMENSION OF CLASSIFICATION

Hierarchical classification systems organize classes at various levels of inclusiveness. Classes at the higher level will include those at the lower levels. The higher the inclusiveness of a class, the higher is its level of abstraction. Highly abstract systems generally include only a few broad general classes. They are parsimonious and simple. The lower levels of abstraction include narrower, more unique, complex, and particularized classes. Other synonyms for higher and lower levels of inclusion include the terms superordinate classes (higher) and subordinate classes (lower). Many specific issues may be grouped under the general discussion of the vertical structure of classifications. These issues include the following.

Contradictory Demands by Criminologists: Hood and Sparks (1970) illustrate this confusion in criminological classifications when they write ". . . the trouble is that there is no real way of knowing how rich in types, or how detailed, it should be." These writers, in discussing empirically derived etiological typologies vaguely suggest that such typologies should contain a "fairly large number" of types, while treatment typologies, are to be ". . . as rich in types as possible." Other criminologists, in contrast, desire that typologies should be parsimonious and clear.

Ambiguity vs. Accuracy: The superordinate class levels, while having the advantage of clarity, simplicity, and parsimony, inevitably suffer from greater ambiguity and within-class heterogeneity. This stems from the loss of too much information. The criminological researcher, guided by his particular purpose of classifying, must decide which of these directions should be optimized.

The 'Equivalence Principle' and the Dominance of Purpose. The equivalence principle, alluded to by Rosch (1978, pp. 28-29), asserts that with reference to the purpose of the classifier, the objects classified within a category are regarded as equivalent for that purpose. She writes:

> It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand.

Thus, the purpose of the classifier emerges as a critical determinant of the level of inclusiveness chosen. If further subdivision does not lead to improvement in the performance and usefulness of the classification vis-à-vis the user's purpose, then more particularized subordinate levels would be inefficient.

The 'Optimal Level' of Classification. Within the terms of a particular purpose a general guideline for choice of an optimal level is that level which maximizes the ratio of parsimony to accuracy. Ambiguity and error in relation to the user's purpose should be minimized while the classification, at the same time, attempts to be optimally economical. This issue is inevitably influenced by both statistical/analytical criteria (e.g., the 'stopping rules' in monothetic divisive classifications - see MacNaughton-Smith 1965, Sonquist and Morgan 1963), as well as social criteria such as 'fairness'. The latter criteria, for instance, would demand that ALL relevant criteria be included into the terms of the classification and thus would push the classification towards the particularizing, concrete, subordinate classes.

THE HORIZONTAL DIMENSION OF CLASSIFICATION

At a particular (vertical) level of inclusiveness, a classification can be evaluated according to a few major criteria: isolation criteria, compactness

criteria, goodness of fit to the raw data, and so on. These criteria are clearly interdependent with one another and must be jointly optimized for a useful methodology.

Isolation Criteria: Discontinuous or Over-Lapping Boundaries

There is much demand by criminological theorists for discontinuous, non-overlapping boundaries and isolated classes (Hood and Sparks 1970, Gibbons 1975, and others). Yet, boundary conditions are generally not clear-cut, and different approaches to the grouping and linking of cases have strong implications for the separation and kind of boundaries between classes. The data itself also determine whether the boundary conditions are fuzzy or clear-cut. Cognitive economy will generally influence the user of a classification to view the classes as well separated and clear. Our experience, with a variety of complex criminological data, is that clear-cut, well-separated boundaries between empirical classes are not common. Rosch (1978) indicates that the clarity of disjoint boundaries may be achieved through the use of formal, necessary and sufficient diagnostic criteria for class membership. The imposition of strict membership criteria upon relatively non-continuous phenomena is widely practiced in criminal justice classification.

Compactness Criteria

These criteria refer to the cohesion, or internal homogeneity of each class, in relation to its environment. Strong typological structure is usually associated with compact homogeneous classes. Homogeneity is a relative concept in that it varies continuously in relation to the immediate environment of the cluster. Classes are more heterogeneous at the upper levels of a hierarchical taxonomy, and more homogeneous at the lower levels.

There are many different ways of defining and measuring the homogeneity of a class (see Blashfield and Draguns 1976, Fleiss and Zubin 1969, and others). A serious problem is that there is currently no agreement on the best approach. Global homogeneity criteria for a complete classification partition include, for example, Wilk's Lambda Λ, and the various criteria stemming from multivariate statistics including: trace W, determinant W (when W is the within-class dispersion), Hotelling's criteria, and others (see for example Everitt 1974, McRae 1973, Friedman and Rubin 1967).

There are also many ways of examining the homogeneity of each class taken separately either globally across ALL classificatory variables; or across each specific classificatory variable. Tryon and Bailey (1970) suggest the homogeneity measure H_{ij}, which assesses the homogeneity of variable j in class i. This is defined by:

$$H_{ij} = 1 - \frac{s_{ij}^2}{s_j^2}$$

where n_i is the number of objects in a class i, and s_{ij}^2 is the variance of variable j in class i, and s_j^2 is the variance of j for the total sample. The overall homogeneity of a class across all variables may be assessed by:

$$H_i = \frac{\sum s_j^2 H_{ij}}{\sum s_j^2}$$

and the homogeneity of each particular variable j across all classes is assessed by:

$$H_j = \frac{\sum n_i H_{ij}}{H}$$

This latter coefficient is equivalent to the standard univariate correlation ratio based on class membership. These measures usually vary between 0 and 1, with high values associated with high heterogeneity.

For applied work in criminology, the form of class homogeneity is quite critical. The user of a taxonomic scheme must have a clear understanding of the kind of homogeneity that is offered by each class. The homogeneity requirement is related to practical issues such as predictive accuracy, ambiguity vs. precision, face validity, the problems of naming, language, and communication. Misunderstandings and misuse of classification schemes may result if the class homogeneities of a taxonomic scheme are either ignored or improperly assessed. Unfortunately, most criminological classification schemes do not have particularly clear discussions of the kinds of homogeneity implied by the classification, and many do not provide any operational approach to assessing homogeneity.

Internal Linkage Structure of Classes. Cases may be grouped together according to a large variety of linkage or membership criteria. Sneath and Sokal (1973), Anderberg (1973), Everitt (1974), Hartigan (1975) and others have described many of these approaches. Such approaches as single-link, comp'ete-link, centroid-sorting, average-linkage, and so on, have a profound impact upon both the shapes, internal homogeneities, and boundary conditions of the resulting classes. The articulation of the structure, form and statistical definition of a 'class' in relation to the conceptual meanings of the term 'class' found in the criminological literautre, has not been completed.

Similarity Assessment. All grouping proceeds according to mutual similarities. Objects which are sufficiently similar to each other are

classified together. Theoretically appropriate and valid similarity assessments are therefore the foundation for all classification. Numerical coefficients can be operationalized in a large variety of ways to measure different 'kinds' of similarity (see Green and Rao 1969, Morrison 1967, Sneath and Sokal 1973). ALL taxonomies must have some basis - either explicit or implicit - for the form of similarity that has governed their construction. The selection of a numerical coefficient to assess similarity may, in fact, have more influence on the structure of the resulting classification than the actual grouping procedure that is used. We deal with this issue in detail in our later chapters.

Reliability or Stability

The reliability of a classification can be conceptualized and assessed in a number of different ways. An important aspect of reliability is the degree to which a classification is reproducible across different classifiers and across different samples. Furthermore, the different specific structures of taxonomies may be more or less reliable. Different attribute-axes may vary in reliability. The central prototypes (centroids, homogeneity measures, etc.) of each class may show varying reliability, the boundaries between classes may vary in their stability. A further aspect of reliability is the degree of consensus between different classifiers in assigning a 'new' case into the existing classification scheme. Inter-observer agreement is absolutely critical in most applied settings since if there is high disagreement between observers/classifiers, the classifications will be unworkable and will have diminished value. There may be disagreements, for example, between clinicians regarding the use and frequency of diagnostic categories, different classification decisions may be reached regarding the

same person, etc. It can be seen that the issue of 'equity or fairness' is critically dependent upon reliability of classifications.

The reliability of a taxonomic system may be diminished by a number of different features: measurement of the classificatory variables may be inexact, certain relevant variables may be omitted, the diagnostic criteria used for class recognition may be ill-specified, etc. Most of the theoretically-generated psychological, psychiatric and sociological systems in criminology, to the extent that their theoretical terms were not tied to well defined operationally measureable diagnostic criteria, inevitably suffer from impaired reliability. Inter-classifier agreement requires that the classification be constructed by operationally objective methods. An enormous number of prior criminonological classifications did not comply with this criterion (see Ferdinand 1966).

A useful reliability index for inter-classifier agreement, particularly in those situations where one way of conducting the classification can be taken as 'correct', is Cohen's Kappa Coefficient (Cohen 1960). This provides a measure of 'agreement' between the classifications. It has the additional property of controlling for 'chance' levels of agreement, while a more complex version can account for different levels or seriousness of non-agreement. The basic form of Kappa is:

$$K = \frac{P_o - P_c}{1 - P_c}$$

when P_o = observed proportion of agreement, i.e., $P_o = \sum_i P_{ii}$, and is the sum of the proportions of cases appearing in the main diagonal of the square table with i rows and columns. P_c is the chance expected proportions of

agreement, $P_c = \sum_i P_{i+}P_{+i}$; and P_{i+} and P_{+i} are the marginal proportions in the ith row and column, respectively.

In examining the cross-classification of two partitions, weighted Kappa (K_w) takes account of the kinds of disagreements between the classifiers. Each ijth cell of the cross-classification is weighted (W_{ij}), where W varies between 0 and 1.0. This weight is used to indicate the seriousness of the disagreement. Weighted Kappa is:

$$K_w = \frac{P'_o - P'_c}{1 - P'_c}$$

when $P'_o = \sum_{ij} W_{ij}P_{ij}$ and $P'_c = \sum_{ij} W_{ij}P_{i+}P_{+j}$. For additional useful discussions, see Reynolds (1977), Everitt (1970), Fleiss, Cohen and Everitt (1969).

In our later chapters comparing the structures of classifications produced by clustering techniques we examine other partition-matching coefficients. These are described fully in later chapters.

Goodness-of-fit Criteria

At any particular hierarchical level classification structures can be assessed according to their degree of fit with the basic raw input data. Many approaches have been offered. The Cophenetic correlation (Sneath and Sokal 1973), the Kappa coefficient (Cohen 1960), Wilk's criterion, Rand's Coefficient (Rand 1971), Goodman and Kruskal's Lambda and others, have all been used to examine various aspects of the fit between input data and the classification structures reached by clustering methods. It may also be noted that some of these coefficients may be used to assess other criteria, e.g. agreement and reliability, and compactness of clusters, etc. There is a need to clarify the relative merits of these various approaches (see also Dubes and Jain 1976). In criminology much consideration should be given to the selection of appropriate coefficients to examine goodness of fit. Currently, this is a

neglected issue in this discipline.

Statistical Validity Criteria

Classification techniques will inevitably produce clusters even on random data. The dangers of artifactual results are, therefore, always a possibility. Many users of classification techniques will accept the results with only intuitive indicators of validity. In this sense face validity (if it supports prior conceptions) is the main aspect of acceptibility. Statistical tests have been developed (e.g. Engelman and Hartigan 1969, Hartigan 1975, Schultz and Hubert 1973, and others). However, the major difficulty is that they usually demand assumptions that are of dubious applicability for many real data situations (see Dubes and Jain 1979). The application of conventional standard significance tests is not a solution to the validity problem, since the null distribution assumptions required for clustering seem intractable (see Mountford 1970, Hartigan 1978). Classical distribution theory can be appropriately applied only to patterns that are classified a priori.

An enormous literature has developed around the various attempts to provide validity tests within the framework of clustering. Good reviews are available in Dubes and Jain (1979) and Blashfield, Aldenderfer and Morey (1978). Indices of cluster compactness, cluster isolation, goodness-of-fit coefficients and other approaches are brought to bear on the questions of the adequacy of a total partition as well as the adequacy of individual clusters. The issue of clustering tendency also enters into this growing literature.

PRAGMATIC CRITERIA FOR CLASSIFICATION SYSTEMS

In addition to the structural criteria mentioned above, classification systems may also be evaluated according to criteria stemming from the various

user purposes. Some of these purposes are extremely general (e.g., predictive validity) while others are more specifically tied to the requirements of criminology. The following criteria are considered:

(1) Predictive validity
(2) Descriptive validity
(3) Face (consensual) validity
(4) Theoretical validity and logical consistency
(5) Power and control: usefulness
(6) Precision
(7) Scope (or range)
(8) Generalizability
(9) Practicality and efficiency
(10) Acceptability to users

Predictive Validity

This is a critical evaluative criteria for all practical users of classification systems. Blashfield and Draguns (1976) indicate that "clinicians have often argued that prediction is the primary purpose of a classification." Predictive validity is at the heart of the choice of treatment for clients. Classifications should yield predictions that have clear relevance for the effectiveness of treatments for each client, e.g,., prison inmates probationers, etc. The predictive information that classification systems provide should enhance decision-making with reference to treatment intervention, release decisions and so forth. The work of Overall, et al. (1966), Paykel (1972) indicate the use of predictive validity within the context of the evaluation of differential treatment effectiveness.

Classification systems must have certain structural characteristics to be useful as predictive systems. First, the classes within the system must

be fairly well differentiated on the criterion variable to be predicted. Second, the classes within the system should have reasonably high within-type homogeneity on the predicted variable. Such homogeneity assessments can be examined, for instance, using Tryon and Bailey's H coefficient. Third, the input domain for attribute space of variables must be highly salient to the external criterion. Irrelevant variables, or unreliable variables, will detract from the predictive accuracy. Fourth, the lowest levels of the taxonomic hierarchy will, in general, maximize the predictive accuracy to the external variable. This allows larger numbers of interaction effects at the various concrete particular sub-types to come into play, allowing finer differentiation between types in regard to the criterion variable.

Descriptive Validity

Accurate description of a set of objects is one of the main general purposes of taxonomy construction. Classification for the purpose of description is usually aimed at gaining greater clarity of the particular set of phenomena. Numerical classification systems attempt to go beyond anecdotal, impressionistic descriptions, which are often neither clear, objective or reliable. Blashfield and Draguns (1976) stress the issues of the selection of both the set of classificatory variables and the sampling of objects. Both of these influence the descriptive validity of a classification system. Skinner (1980) argues that the descriptive validity of a classification is dependent upon both the convergent and discriminant properties of that classification system (see Campbell and Fiske, 1959). Convergent validity in this sense is the degree to which the objects are similarly classified across different attribute spaces, e.g., behavioral data, personality

characteristics, social history and so forth. Discriminant validity, on the other hand, examines the distinctiveness of the classes when alternative attribute spaces are used.

The most general criteria, however, for evaluating the descriptive validity of a classification is the degree to which the system is isomorphic with the empirical phenomena. It can be argued that the taxonomic system is never a perfect isomorph of an empirical situation. Some information is always deleted from a classification system. Descriptive validity, therefore, will always be somewhat incomplete. This may not cause difficulties where the information deleted is irrelevant to the purposes at hand. The level of descriptive accuracy that is acceptable is, therefore, again governed by the user's purpose.

The requirements of good descriptive accuracy include an appropriate and adequate sampling of both variables and subjects, as well as the selection of classificatory procedures which avoid the imposition of artifactual boundaries or misleading homogeneity structures. All numerical methods embody certain assumptions, and, particularly for the purposes of descriptive classification, such assumptions should be consistent with the nature of the data. A further critical consideration for the development of descriptively accurate classifications is the selection of the most appropriate level of inclusiveness. The issue of the correct number of classes becomes quite important if only one classification level is to be chosen from an overall hierarchical structure. Descriptive accuracy is usually enhanced by retaining the full taxonomic hierarchy rather than choosing one particular level. The limitation to one particular level incurs the danger of choosing the wrong level, i.e., too few, or too many classes (see Edelbrock 1979).

Face Validity

Face validity is sometimes referred to as consensual validity. This aspect of validity deals with the intuitive meaningfulness, reasonableness, and relevance of the classificatory constructs to users. Face validity contributes profoundly to the acceptance of the classification system by its users.

Theoretical Validity and Logical Consistency

Cromwell et al.(1975) have set forth a number of criteria that enhance the logical coherence of classification systems. These include (1) the designation of whether the system is dimensional, typological or a mixture of these; (2) whether the system is hierarchical and, if so, the clear specification of the nature of the vertical dimension; (3) whether the classes are mutually exclusive or overlapping; and (4) the inclusion of clear explicit and precise rules for assigning new cases into the system.

While all of these criteria obviously add to the clarity of the system, the theoretical and explanatory coherence of a classification can also be enhanced by other factors. First, the coherence of the attribute space can greatly clarify the meaning and focus of the system. Second, the elucidation of the form and structure of both similarity assessment and type-definition (agglomeration rules) will add to clarity and consistency. Third, attention should be paid to the logical consistency between the form of similarity, the type definitions, the purpose of classification and the nature of the data. (Enc (1976) probably gives the most enlightening treatment of the complex relations between empirical classification and theoretical development.)

Criminological empirical classifications have been consistently criticized for their lack of theoretical coherence (Opp 1973, Bottoms 1973, Gibbons 1975, and others). The major problems stemming from this lack of theoretical coherence is that the implications for intervention and control are fuzzy, weak, and unreliable (Meehan 1968). Strong coherent classificatory theory will usually imply clear interventions. Classifications tied to interventions, e.g., treatment classification, sentencing classification, etc., could be rendered far more effective if their theoretical coherence and validity could be improved.

Power and Control: Usefulness

Classification systems vary in regard to the amount of control they provide over the environment (Meehan 1968). Power or control is a function of the coherence of the explanation provided by a classification, the specificity of the intervention implied by this explanation, and the accuracy of the predictions which stem from the classification. Power is enhanced by precision of measurement, clarity of explanation, specificity of interventions, and validity of the predicted outcome. Strong explanations will lead to effective controlling interventions. The practical use of classification systems to enhance specific ends (e.g., rehabilitation), or to suppress others (e.g., recidivism) is dependent upon this concept of power/control. Classification systems in criminology have been evaluated by many reviewers as lacking in power. The control that they provide is, in fact, undermined by a variety of factors. These factors include incoherence of explanation, poor or weak precision of measurement, low predictive validity, inadequate coverage of criminological

populations, unreliability of the assignment/diagnostic process, and so on (see Gibbons 1975, Brennan and Huizinga 1976, Hood and Sparks 1970, and others).

Precision

Taxonomic systems vary in regard to the precision with which their class-concepts or categories are measured and related to real empirical events. Precision refers to the accuracy of measurement, the presence of explicit diagnostic criteria for each class, and the clarity of conceptualization of each type-concept. Precision is an important requirement of any applied classification system. Where operational diagnostic criteria are vaguely specified, and when reliability of measurement of classificatory variables is low or characterized by error, then the precision of the taxonomy will be low.

Precision also increases with greater information content. Thus, precision would be higher at the more particularized concrete subtype levels since these retain more information. The more generalized abstract classes at the higher levels of a hierarchy will usually be less precise, but more reliable, due to the smaller number of classes. In this manner precision and reliability are inversely related.

Precision is an absolutely critical aspect of applied social or psychological classification systems since these must deal accurately, efficiently, and unambiguously with persons, clients, etc. who must be identified into their correct categories.

Scope (or Range)

The scope of a classification system is the range of events to which it can be applied. The aim of many classification systems in criminology is to be

as wide in scope as possible. Classifications are regarded as better if they include a maximum number of offenders within their descriptive or explanatory range. Good systems should combine a broad range of relevance with high precision and power. However, widening the scope of classification systems usually means loosening the precision and coherence of the defining terms and thereby running the risk of ambiguity. For example, many of the earlier classificatory systems were loaded with psychiatric concepts (see Solomon 1977, Ferdinand 1966, for reviews) and were applied to a very broad range of criminal behavior. Yet, these systems generally offered only ambiguous class-terms, poor precision, and poor control. On the other hand, specific behavior or biographical classifications have relatively limited range and poor control but high precision. An obvious goal in criminological classification is to develop adequate theoretical/explanatory/predictive classifications which are fairly wide in scope but which do not sacrifice accuracy or precision.

Generalizability: The Universal - Local Dimension

Many classification systems are developed for specific populations. It is often important to examine the generalizability of classifications across different populations. Skinner (1980) proposes that the issue of generalizability is an important criteria for evaluating classifications systems. Earlier, McKinney (1966) introduced the problem of generalizability in terms of a dimension he referred to as universal vs. local. Most classification systems vary according to this continuum. In criminology, many current classification systems are limited and are tied to a specific locale. McKinney refers to such classification systems as "localized". The generalizability of a particular classification system is deeply influenced by the attribute criteria that are used. These may be idiosyncratic to a particular institution

or a particular cultural system. The attribute space for the I-level classification system stems from a relatively universal theory of ego development, and should, therefore, be conducive to a classification system which is universal in scope. It would be an advantage for criminologists to attempt to construct classification systems of the universal type, i.e., applicable "across the board". More explicit and theoretically generated attribute spaces would be a precondition for the development of such relatively universal classifications, as well as a more explicit approach to the sampling of cases. McKinney (1966) has indicated that the normative orientation of the social sciences (including criminology) is that research workers universalize classifications wherever possible. This demand, however, is usually offset by the desire of local social systems to optimize classification systems for their own specific purposes.

Practicality and Efficiency

A critical set of criteria for classifications used in practical institutional settings deals with convenience and practicality. Again, different structural criteria contribute to the overall practical feasibility of classifications. These practical criteria include:

- Costs of personnel training
- Time and costs involved for assessment of the person being classified. Some classifications may demand large amounts of complex data for the assignment process.
- Parsimony and face validity. Classification systems are efficient to the degree that they are easily understood, easily remembered and have high face validity.
- Misclassification rates. Classifications are unworkable in practice to the extent that either the basic reference classification is unreliable,

or there is unreliability in the assignment procedures for new cases. Unreliability in either of these two aspects will lead to higher levels of misclassification. Misclassification in turn leads to inequity, confusion, lowered confidence, cynicism, and disagreements between classification personnel.

- Coverage of Cases. Classifications are unworkable in situations where they deal only with a small segment of the full population being classified.

It should be clear from this brief discussion that the end-use consideration of practicality and efficiency is directly related to many of the methodological and structural criteria mentioned earlier in this chapter. A complete articulation of these relations and influences would be helpful in guiding the future development of criminal justice classifications. Similarly, the currently used classification systems can be evaluated in terms of these criteria.

Acceptability to Users .

There are many different aspects leading to acceptability. First, a classification system will tend to be accepted if the naming procedures that have been used are reasonable and accurately related to the phenomena being classified.

Second, if a classification system has good predictive accuracy leading to good treatment interventions, i.e., high clinical utility, then it will also tend to be accepted.

Third, a coherent theoretical structure will serve to increase acceptability.

Fourth, clear definitions, clear membership criteria, objective diagnostic decision rules, and an appropriate coverage of the objects being classified would all tend to increase the chances of any classification system gaining

wide acceptance.

In criminology, the different classification systems have gained various levels of acceptability. However, there is no classification system currently in use that has gained general and widespread acceptability among practitioners. The California I-level system, for example, had the advantages of an intuitively appealing theoretical structure and a simple and easily remembered set of names. It has some disadvantages, e.g., relative inefficiency, relative unreliability, marginal predictive validity, and a rather cumbersome approach to the assignment of new cases. Yet, for many years and among many practitioners it gained a high level of acceptance. Each classification system, therefore, could be evaluated in terms of the particular mix of features which lead into general acceptability.

One aspect of acceptability and face validity is the degree to which different observers agree upon the categorizations that have been made of the same set of objects. Katz (1965) studied the categorization of psychiatric patients from different vantage points, e.g., experienced clinicians, nurses, and family members. Consensual validity is related to the emergence of the same classification decisions from these different vantage points. The variability of different classifiers in regard to the same set of data can emerge from their different system of "relevances" and may lead to systematically different classification structures. When different classifiers agree on the same classification structures of the same raw data, then consensual validity can be claimed in terms of the level of agreement between such classifications (see Katz 1965). Coefficients such as the Kappa Coefficient, and Goodman and Kruskal's Lambda Coefficient, can be used to assess the overlap,

or consensual agreement, between the different classifiers.

CONCLUSION

This chapter has attempted to provide a general framework for the work of later chapters. There is a desperate need to upgrade the effectiveness of classification within criminology. It is our belief that this upgrading will be facilitated only to the extent that serious attention is paid to proper structural aspects of classification systems. This, in turn, demands a focus upon the methodology for creating and evaluating complex classification systems. The starting point for much of this work will be the end-use consideration, i.e., the practical purpose for which the classification has been created. Different uses, and purposes, will require quite different kinds of classifications, and this in turn will require different approaches to the creation and evaluation of these systems. Thus, we propose that a general goal for research in criminological classification is the elucidation of the relationships between (1) user-purposes; (2) structural criteria of classifications; (3) methods of creating different kinds of classifications; and (4) specific approaches to the evaluation of the performance of classifications. The work of the present chapter, and of this report generally, is aimed at these objectives.

BIBLIOGRAPHY

Anderberg, M.R.
1973 Cluster analysis for applications. Academic Press: New York.

Blashfield, R. K., and J.G. Draguns
1976 "Toward a taxonomy of psychopathology: The purpose of psychiatric classification." British Journal of Psychiatry, 129, 574-83.

Bottoms, A.E.
1973 "Methodological aspects of classification in criminology." In Collected studies in criminological research: Methodological aspects of classification in criminology. Council of Europe: Strasbourg, France.

Brennan, T., and D. Huizinga
1976 A review of typological research in delinquency. American Criminology Society Conference: Tucson, Arizona.

Campbell, D.T., and D.W. Fiske
1959 "Convergent and discriminant validation by the multitrait-multimethod matrix." Psychological Bulletin, 56, 81-105.

Cohen, J.
1960 "A coefficient of agreement for nominal scales." Educational and Psychological Measurement, 20, 37-46.

Cromwell, R.L., R.K. Blashfield and J.S. Strauss
1975 "Criteria for classification systems." In Issues in the classification of children. Jossey-Bass: San Francisco.

Dubes, R., and A.K. Jain
1976 "Clustering techniques: The user's dilemma." Pattern Recognition, 8, 247-260.

Dubes, R., and A.K. Jain
1979 "Validity studies in clustering methodologies." Pattern Recognition, 11, 235-254.

Edelbrock, C.
1979 "The accuracy of hierarchical clustering algorithms: The problem of classifying everybody." Multivariate Behavioral Research, 14, 367-384.

Enc, B.
1976 "Spiral dependence between theories and taxonomy." Inquiry, 19, 41-71.

Everitt, B.S.
1970 "A discussion of some of the problems of cluster analysis." Paper presented at the Conference of the Society for Multivariate Experimental Psychology, Free University of Brussels, Belgium.

Everitt, B.S.
1974 Cluster Analysis. Halsted Press: New York.

Ferdinand, T.N.
1966 Typologies of delinquency; a critical analysis. Random House: New York.

Fleiss, J.L., J. Cohen and B.S. Everitt
1969 "Large sample standard errors of Kappa and weighted Kappa." Psychological Bulletin, 72(5), 323-27.

Fleiss, J.L., and J. Zubin
1969 "On the methods and theory of clustering." Multivariate Behavioral Research, April, 235-250.

Friedman, H.P., and J. Rubin
1967 "On some invariant criteria for grouping data." American Statistical Association Journal, 62, 1158-78.

Gibbons, D.C.
1975 "Offender typologies - two decades later." British Journal of Criminology, 15(2), 140-56.

Green, P.E. and V.R. Rao
1969 "A note on proximity measures and cluster analysis." Journal of Marketing Research, 6, 359-64.

Hartigan, J.A.
1975 Clustering algorithms. John Wiley and Sons: New York.

Hood, R., and R. Sparks
1970 Key issues in criminology. McGraw-Hill Book Co.: New York.

Johnson, L.A.S.
1968 "Rainbow's end: The quest for an optimal taxonomy." Systematic Zoology, 19, 203-239 (1970).

Katz, M.D.
1965 "A phenomenological typology of schizophrenia." In Role and methodology of classification in psychiatry and psychopathology, pp. 300-322. Public Health Service Publication No. 1584: Washington, D.C.

Kornfeld, J.P., T.G. Eynon, W.A. Hetzner, M.E. Pfeiffer-Kornfeld, W. Bates, and T. Baume
1975 State-of-the-art of offender classification in the U.S.A. IIT Research Institute: Chicago.

MacNaughton-Smith, P.
1965 "Some statistical and other numerical techniques for classifying individuals." In Numerical taxonomy. W.H. Freeman & Co.: San Francisco.

McKinney, J.C.
1966 Constructive typology and social theory. Appleton-Century-Crofts: New York.

McRae, D.J.
1973 Clustering multivariate observations. Doctoral dissertation. University of North Carolina: Chapel Hill.

Meehan, E.J.
1968 Explanation in social science: A system paradigm. Dorsey Press: Homewood, Illinois.

Milligan, G.W., and P.D. Isaac
1980 A review of Monte Carlo tests of cluster analysis. Ohio State University: Columbus, Ohio.

Morrison, D.G.
1967 "Measurement problems in cluster analysis." Management Science, 13(12), 775-780.

Opp, K.D.
1973 "Problems of classification in criminology." In Collected Studies in Criminological Research: Methodological aspects of classification in criminology. Council of Europe: Strasbourg, France.

Overall, J.E., L.E. Hollister, M. Johnson, and V. Pennington
1966 "Nosology of depression and differential response to drugs." Journal of the American Medical Association, 195, 946-48.

Paykel, E.S.
1972 "Depressive typologies and response to amitriptyline." British Journal of Psychiatry, 120, 147-156.

Rand, W.M.
1971 "Objective criteria for the evaluation of clustering methods." Journal of the American Statistical Association, 66, 846-850.

Reynolds, H.T.
1977 The analysis of cross-classifications. The Free Press: New York.

Rosch, E.
1978 "Principles of categorization." In Cognition and categorization, pp. 28-49. Halsted Press: New York.

Skinner, H.A.
1980 Toward the integration of classification and theory and methods. Unpublished manuscript, University of Toronto.

Sneath, P.H.A., and R.R. Sokal
1973 Numerical taxonomy. W.H. Freeman and Company: San Francisco.

Solomon, H.M.
1977 Crime and delinquency: Typologies. University Press of America: Washington, D.C.

Sonquist, J.A., and J.N. Morgan
1963 "Problems in the analysis of survey data, and a proposal." In The detection of interaction effects (1964). University of Michigan: Ann Arbor, Michigan.

Toulmin, S.
1953 The philosophy of science. Harper & Row: New York.

Tryon, R.C., and D.E. Bailey
1970 Cluster analysis. McGraw-Hill: New York.

ADDITIONAL REFERENCES:

Blashfield, R.K., M.S. Aldenderfer, and L. Morey
1978 "Cluster analysis literature on validation." Paper presented at the Classification Society Meeting, May, 1978.

Mountford, M.D.
1970 "A test of the difference between clusters." In Statistical ecology, Vol. 3. Pennsylvania State University Press: University Park, Pennsylvania.

Schultz, J.R., and L.J. Hubert
1973 "Data-analysis and connectivity of random graphs." Journal of Mathematical Psychology, 10, 421-428.

CHAPTER 2

THE USES OF CLASSIFICATION WITHIN CRIMINOLOGY AND CRIMINAL JUSTICE SYSTEMS

INTRODUCTION

"The findings of classification specialists," writes Yepsen (1975), "are used as bases for every important phase of the offender's life while in the institution and after his release as well." Clinard and Quinney (1967), using a broader frame of reference, address the pervasiveness of classification in social research:

> The use of types and the ordering of the diversities of observed phenomena has been instrumental in the development of the social sciences. Types not only reduce phenomena to more systematic observation; they also assist in the formulation of hypotheses and serve as guides for research.

Not only is classification utilized at virtually every level of the criminal justice system, but a concern with classification informs a large portion of criminological research as well. There is simply no part of crime and delinquency research or processing which does not make some use of classification. Criminologists and those working in the criminal justice system have long recognized the need for valid and reliable classification systems to provide a basis for both "systematic observation" of criminal behavior and effective decision-making.

This necessity of creating classifications to adequately order the complex phenomena of criminal behavior has resulted, as Ferdinand (1966) notes, in a "torrent of typologies". These typologies reveal an attempt by criminologists to provide a clearer conceptual structure for a mass of quite confusing information. When confronted with the diverse phenomena of criminal behavior, classificatory systems are inevitably utilized most prominently in the

following three areas: a) the field of research criminology, where it is necessary for researchers to use some sort of classification for pursuit of various scientific ends; b) for offender treatment, which is granted a special emphasis because classification in this instance concerns both the criminological researcher and the criminal justice worker equally; and c) the criminal justice system itself, where decisions involving use of classification are made at nearly every stage of offender processing. These three categories are not meant to be absolute, but simply to provide a provisional organization of the ways in which classification is used. Each of these areas will now be examined in more detail.

USES OF CLASSIFICATION IN THE FIELD OF RESEARCH CRIMINOLOGY

Substantive Directions of Prior Taxonomic Research

The multiplicity of criminal typologies which have been generated may be provisionally broken down into six basic categories:

1) Sociological and social role systems, e.g. Schrag (1961), Sykes (1958), Garabedian (1964), Roebuck (1966), Kinch (1962), Lindesmith and Dunham (1941), Harary (1966), Dorn (1969), Marcus (1975), and Cohen (1971)

2) Psychological and psychiatric taxonomic systems, wherein emphasis is placed on the psychological or psychiatric characteristics of the deviant person. Examples may be found in Jenkins and Hewitt (1944), Redl (1956), Aichorn (1935), Reiss (1952), Argyle (1961) and Block and Flynn (1956), among others. Shafer (1969) and Ferdinand (1966) most likely provide the most comprehensive reviews of these psychological taxonomic systems.

3) Ego development and social interaction taxonomies. This approach provides a synthesis of psychological and sociological concerns. Specific focus is given to problems involving ego development and interpersonal interaction. The most visible of this sort of typology is the California I-level system, the explication of which has created an enormous amount of literature (e.g. Loevinger 1957; Warren 1966; Jesness 1974; and Molof and Jesness 1972). Critical reviews of the I-level system may be found in Gibbons (1970) and Beker and Hyman (1972). All of this literature exemplifies the ongoing controversy which exists concerning the validity and reliability of this taxonomic system.

4) Behavioral classifications. Surprisingly few of these exist in criminology (e.g. Kulik, Stein, and Sarbin 1968; Swanson and Mobley 1976; Hartjen and Gibbons 1969; and McCaghy, et al. 1976).

5) Physiological and constitutional systems. With the advent of criminology, a number of systems were created which classified criminals according to certain morphological, physiological and some mental characteristics. More recently, chromosonal structures have been offered as a basis for classification. Once again, there is controversy over the usefulness of these approaches.

6) Taxonomic systems for diverse particular crimes. Many extant classification schemes in the crime and delinquency literature concentrate on specific crimes. For example, the systems of Guttmacher (1960), Neustatter (1957), Megargee (1966), and Blackburn (1971), among others, deal exclusively with the problem of homicide. In the area of runaway youth, for example, there have been at least two dozen different typological systems offered to both explain and

describe this particular form of juvenile delinquency (see Brennan 1980). Theft, violence and drug abuse also have stimulated many different taxonomic systems (see Saunders, 1972).

It must again be stressed that the above represents only one particular form of structuring the substantive directions of criminological work utilizing classification.

Classification for the Epidemiological Mapping of Crime

Systematic mapping and classification of the incidence, prevalence, and distribution of criminal behavior provides an attempt to clarify the subject under consideration. This descriptive task comes under the heading of epidemiology, which, Gruenberg (1965) writes, "relates observed disorders to the environments in which people live." The epidemiology and enumeration of crime has had to contend with the complexity and multidimensionality of crime and delinquency. It is for this reason, among others, that epidemiology necessarily involves classification, or, as Gruenberg puts it, an epidemiologist needs "consistently applicable classifications of people...".

Good classification (or description of a phenomenon) is a critical part of an adequate epidemiology of the event to be counted, explained or controlled. Is the phenomenon increasing or decreasing, is its distribution even, random, discontinuous or structured in some other way? Reliable epidemiology further requires unbiased data and appropriate analytical methods. Glaser (1974) points out that clear classification of offenses is critical for facilitating the enumeration and observation of criminal cases. The study of the prevalence and spread of crime demands a classification framework of specific and well- differentiated units of enumeration. For many crimes, indicates Glaser, these classifications are either not available (i.e.

rape) or are currently in the process of being established (i.e. arson; see Inciardi 1970). This point underlines the importance of behavioral classification systems for particular crimes and additionally indicates some directions for further taxonomic research.

Many criminological investigators have recognized the importance of descriptive taxonomic research (Opp 1973, Kornfeld, et al. 1975), and most bemoan the fact that so few descriptive taxonomies have been forthcoming. They all point out that the descriptive taxonomic task in criminology has not been done sufficiently well. Wenk and Halatyn (1974) explicitly call for more 'behaviorally accurate' descriptions of crime, while Hood and Sparks (1970) indicate the failure of earlier attempts to construct valid descriptive taxonomies of criminal behavior. However most investigators continue to blur and undermine the epidemiological purpose by creating typologies which inextricably mingle descriptive, predictive, and theoretical elements; they do not see that one necessarily precedes the others.

Another major factor regarding the use of classification in epidemiology is the existence of complex interrelations between crimes. The domain of criminal behavior is complex in a multidimensional sense. Appropriate epidemiological mapping would by necessity have to address itself to this multidimensionality. The use of univariate percentages and bar-graphs oversimplifies the reality of the behavior being described. A multidimensional domain basically requires a multidimensional mapping if descriptions are to be realistic or accurate.

Classification for Explanatory Theoretical Research

Another dominant concern in criminology, a concern intimately related with the need for valid empirical descriptive taxonomies outlined above, is the use of taxonomic systems as a major strategy for the development of

scientific explanations (Glaser 1974). The need to utilize classification for explanatory theories of crime was explicity recognized by Sutherland in 1939:

> It is not likely that a general explanation of all crime will be sufficiently specific or precise to aid greatly in understanding or controlling crime. In order to make progress in the explanation of crime, it is desirable to break crime into more homogenous units.

Bottoms (1973) takes this point of view further, advocating the generation of theory "by working outwards from classifications of empirical data." The suggestion that empirical classification is an invaluable aid to the generation and testing of criminological theories is also made by Wenk and Halatyn (1974), Hood and Sparks (1970), Opp (1973), Ferdinand (1966), and Gibbons (1975), and others. Consideration of using classification for explanatory research has lead to much slightly inappropriate musing on the relative merits of 'theoretical' versus 'empirical' typologies (e.g. Ferdinand 1966). Much of this discussion stems from a pervasive misunderstanding of the interdependent relation between theoretical and empirical elements.

Explanatory classifications derive their practical significance from the fact that they are critically involved in the processes of control through treatment intervention and preventive intervention. Explanatory classification systems, if valid, can ideally be utilized for the development of intervention programs (see Cromwell, Blashfield, and Strauss 1975). Explanatory or etiological classifications not only describe criminological phenomena and generate predictions, but they try to provide an understanding of the processes involved and suggest ways in which the future course of events might (in principle) be controlled. Explanatory typologies and explanations implicitly suggest that by modifying the relationships and conditions which generate a prediction, the outcome itself may be changed. Strong explanatory systems should lead to strong control and appropriate

intervention. Conversely, weak and invalid explanatory systems may lead to weak or unreliable control and only marginally appropriate intervention.

At the present time, explanatory classifications and the predictive typologies they engender not only form a large body of work within criminology, but are used extensively in the criminal justice system, providing the basis for decisions ranging from parole and release risk to prison work assignments (Gottfredson and Wilkins 1978; Kornfeld, et al., 1975; Bottoms 1973). However, most explanatory typology research has been methodologically and conceptually weak (see Smiley 1977; Kornfeld, et al. 1975; Huizinga 1977; and Brennan and Huizinga 1976), a fact which calls into question decisions based on the classifications currently in use. In the above sense they provide only weak control and insufficient guidance for intervention or prevention programming.

Classification for Evaluation Research

There has been only minimal use of good classification methods in criminological evaluation research, yet their contributions could have a profound impact on the fruitfulness and procedures involved in this type of research. Examples of evaluation research incorporating this taxonomic perspective are the English - Borstal study (see Bottoms 1973), Sampson (1974), in his study of post-prison success, and the Behavioral Research Institute Study of Youth Diversion Processes (1978). Also relevant is Wood's (1969) general discussion of the homogeneity requirement for research into deviance and control. Recently, Glaser (1977) has argued that if correctional evaluation is to contribute to the development of practical knowledge, it should attempt to explain why particular types of programs should have any impact on particular types of clients. Hood and Sparks (1970) earlier point out the confusion and inconsistency in the results of this sort of research.

This underlines the need for the incorporation of taxonomic methods into evaluation designs. The heterogeneous mix encountered in particular treatment settings is entirely inappropriate for the development of unambiguous relevant treatment propositions. Any excess of diversity, especially when the unit of analysis is unclear, will lead to confusion and uncertainty regarding the propositions and results stemming from evaluation research. Scientific theory requires clear propositions in reference to well specified units of analysis.

CLASSIFICATION FOR TREATMENT

As mentioned above, treatment is accorded a special category in that it is equally a theoretical and practical institutional concern. It is also one of the categorizations which most intimately concerns the offender, since this sort of classification began as attention began to shift from focus on the offense to focus on the perpetrator of that offense (Yepsen 1975). The original assumption that all offenders are not alike, that they are heterogeneous in social, psychological and behavioral characteristics, lead to the further perception that offenders differ in their responses to the same treatment. It was increasingly regarded as a mistake to treat them all with a single global treatment approach. By the 1930's, classification plans began to be developed for prisoners which would provide for individualized treatment to meet their unique needs (Hippchen 1975). The totally individualized approach being clearly impractical, the search for more useful typologies has intensified in order to eliminate the problems which arise when like treatment is administered to offenders with different problems (See Bottoms 1973; Gibbons 1975: Kornfeld, et al. 1975). The problem, as defined by Quay (1975), is to find a way of "classifying heterogeneous populations of delinquents (or other criminals) into more treatment-relevant subgroups."

Classification for treatment implies the use of differential interventions for such purposes as rehabilitation, reform, and the minimization of recidivism. It implies the use of explanatory theory as a basis for making appropriate and meaningful interventions. There has consequently been an on-going search for adequate theoretical and explanatory classification systems. The published research offers many classification systems from the perspectives of psychiatric theory, sociological theory, psychological theory, interpersonal maturity theory, role theory, and so forth. There is, at this time, no generally accepted theoretical position regarding offender classification. Different causal theories have different treatment implications, and most of them perhaps have some partial validity for particular forms of criminality.

It is evident that classification systems resulting in valid treatment systems would have tremendous practical importance within the field of corrections (see Kornfeld, et al., 1975; Glaser, 1974; Warren, 1971, and others). Unfortunately the hoped for benefits from offender classification have consistently failed to be realized by criminal justice institutions. Weak and partially valid classification systems have seriously contributed to this failure. Orlando and Black (1975), for example, speak of the difficulty which juvenile courts have in attempting to modify effectively the behavior of children. This dilemma basically requires the creation of more valid treatment typologies. Kornfeld et al. (1975) point out that this task is made difficult by the "differential meaning" of treatment (rehabilitation, management, supervision, etc.). The problem of treatment classifications is also compounded by the controversy surrounding the relation of offender categories and differential treatment methods. As Hood and Sparks (1970) make clear, previous studies in this area have been inconclusive. This

inconclusiveness and related methodological confusions could be among the reasons why many prisoners have felt that classification makes little difference in treatment decisions and has indeed limited their possibilities for rehabilitation (Kornfeld, et al., 1975).

CLASSIFICATION IN THE CRIMINAL JUSTICE SYSTEM

Management Processes

Much classificatory activity in criminal justice institutions is aimed at the effective management of the institution. Most traditional classification schemes were not primarily based on the humanitarian goals of treatment, rehabilitation, and reform, but on custodial issues, escape risks, violence classifications, work assignments, segregation for preventing contamination of one type of offender by another, availability of facilities, and a variety of other practical considerations. Flynn (1975), in fact, suggests that former classification schemes have exhibited an over-concern with the issues of management and less concern with the welfare of the offender. Kornfeld, et al. (1975) also point out that classification in general has been used primarily for security and management, and only secondarily for treatment and rehabilitation.

Warren (1971) makes it abundantly clear that all management decisions "require an implicit or explicit classification system". However, as she also makes clear, an implicit system cannot be easily examined for reliability and validity. Thus, once again, an objective and operationally measurable typology is necessary for effective and humane institutional management. As Flynn (1975) has cogently written, "...classification, in the ideal sense, provides the administrator and the correctional planner with the best available decision-making tools."

Constraint, Escape Risks, and Violence

Many purely predictive classifications are utilized in criminal justice institutions to predict certain risks such as escape and violence, as well as indicating optimum levels for constraint on offenders. One of the more famous of these is Mountbatten's (1966) classification for Prison Security, which is critically discussed by Bottoms (1973). The reliability and validity of such predictive categorization systems is a critical issue. Bottoms (1973) suggests that there has been far too little knowledge of the reliability of such categorization systems. Predictive classifications may have absolutely no relationship to explanatory theory; their value is purely as a predictive device for management decision-making and its various purposes. Predictive accuracy, in this purpose, may be much more important than the theoretical meaning of the classification (see Simon 1972). Gottfredson (1975), in fact, considers prediction as the main component of justice decisions.

Parole and Release Decisions

Here again classification systems are concerned primarily with the prediction of recidivism. Such classifications may emphasize the minimization of the risk of recidivism rather than a concern with offender needs. Kornfeld, et al. (1975), express enthusiasm regarding the use of classification systems for the reduction of risk. They emphasize that this differs tremendously from the treatment or pathology model. They also suggest that taxometric methods and mathematical procedures should be heavily utilized to provide exact probability statements of risks of recidivism. Hood and Sparks (1970) reviewed classifications of offenders, emphasize offender needs in reference to reform and treatment, stating "we must now take this process of taxonomy a step further and consider some ways of classifying offenders in

ways relevant to the choice of treatment." Optimal treatment classifications, in any case, may potentially be more successful than purely predictive classifications regarding the reduction of recidivism. If an optimal treatment is provided for each type of offender there would theoretically be minimization of recidivism risk. However, we suggest that the different purposes of treatment intervention and the prediction of recidivism both require different kinds of classification systems if the two different purposes are to be optimized.

A further aspect of treatment classification concerns the degree to which classification procedures can be utilized to aid re-integration of the offender into society. Classification in this sense is concerned with the problem of minimizing the amount of control and supervision which is applied to different offenders, in order that no offender be held by constraints greater than his or her needs or risks demands. The probability of restoring the offender to society in a successful manner could be enhanced through the use of classification systems. An attempt can be made through the use of appropriate classifications to dramatically reduce the numbers of persons who are exiled from the community and to minimize the levels of this exile. Diversion, probation, and parole decisions represent attempts to minimize this exile. The use of classification here implies that differential levels of constraint and exile would go hand in hand with the search for new forms of treatment. The ultimate goal of this process would be the re-entry of the offender into the community, while simultaneously insuring high levels of public protection. Many criminological researchers see classification as an essential component in reaching these dual objectives.

Adjudication and Sentencing

Classification is pervasive throughout court proceedings. The courts make

innumerable classificatory differentiations, for example, the concept of classification by criminal responsibility. Glaser (1974) presents a thorough discussion of the issues surrounding classification via responsibility. Glaser states, "classification of criminals by their responsibility would appear to be a matter of deciding whether or not they are criminals." Various levels of responsibility may be designated. Glaser indicates that the major classification variables relevant to the concept of responsibility are age, insanity, and intelligence, and mental competence.

Gottfredson (1975) asserts that sentencing decisions are "influenced" by a number of conflicting goals. The justice system, he argues, is plagued by a "typical lack of attention to the classification and prediction problems inherent in the sentencing process." Hood and Sparks (1970) have also remarked that the effectiveness of differential sentencing could be greatly enhanced by utilizing the typological approach.

CONCLUSION

A concern for classification, as can be seen from this brief review, permeates most phases of both studying and dealing with criminal behavior. Despite this prevalent concern, however, few classification systems have been utilized extensively. This is due to a number of interelated factors, not the least of which is the sheer methodological inadequacy of most extant classifications.

BIBLIOGRAPHY

Aichorn, A.
1935 Wayward youth.

Argyle, M.
1961 "A new approach to the classification of delinquents with implications for treatment." California State Board of Corrections Monograph, 2, 15-26.

Beker, J., and D.S. Heyman
1972 "A critical appraisal of the California differential treatment typology of adolescent offenders." Criminology, May, 1-59.

Blackburn, R.
1971 "Personality types among abnormal homicides." British Journal of Criminology, 11(1), 14-31.

Block, H.A. and F.T. Flynn
1956 Delinquency. Random House: New York.

Bottoms, A.E.
1973 "Methodological aspects of classification in criminology." In: Collected studies in criminological research: Methodological aspects of classification in criminology, Vol. 10. Council of Europe: Strasbourg, France.

Brennan, T.
1980 "Structural criteria required for treatment classifications." Presented at the International Differential Treatment Association Conference, May, 1980, Estes Park, Colorado.

Brennan, T., and D. Huizinga
1976 "A review of typological research in delinquency." American Criminology Society Conference, Tucson, Arizona.

Clinard, M.B., and R. Quinney
1967 Criminal behavior systems, a typology. Holt, Rinehart, and Winston: New York.

Cohen, S.
1971 "Directions for research on adolescent group violence and vandalism." British Journal of Criminology, 11(4), 319-40.

Cromwell, R.L., R.K. Blashfield, and J.S. Strauss
1975 "Criteria for classification systems." In Issues in the classification of children. Jossey-Bass Publishers: San Francisco.

Dorn, D.S.
1969 "A partial test of the delinquency continuum typology." Social Forces, 47(3), 305-14.

Dunford, F. and B. Knowles
1978 "Diversion - a study of alternative processing practices." BRI Final Report Summary, NIMH (MH 26141), Feb., 1978.

Ferdinand, T.N.
1966 Typologies of delinquency - a critical analysis. Random House: New York.

Flynn, E.E.
1975 "Problems of reception and diagnostic centers." In Correctional classification and treatment: A reader. W.H. Anderson Co.: Cincinnati, Ohio.

Garabedian, P.G.
1964 "Social roles in a correctional community." Journal of Criminal Law, Criminology, and Police Science, 54, 338-47.

Gibbons, D.C.
1970 "Differential treatment of delinquents and interpersonal maturity levels theory: A critique." Social Services Review, March, 22-33.

Gibbons, D.C.
1975 "Offender typologies - two decades later." British Journal of Criminology, 15(2), 140-156.

Glaser, D.
1974 "The classification of offenses and offenders." In Handbook of Criminology, pp. 45-84. Rand McNally College Publishing Company; Chicago.

Glaser, D.
1977 "Concern with theory in correctional evaluation research." Crime and Delinquency, 23(2), 173-79.

Gottfredson, G.D.
1975 "Organizing crime: A classifactory scheme based on offense transitions." Journal of Criminal Justice, 3, 321-32.

Gottfredson, D.M., C.A. Cosgrove, L.T. Wilkins, J. Wallerstein, and C. Rauh
1978 Classification for parole decision policy. National Institute of Law Enforcement and Criminal Justice: Washington, D.C.

Gruenberg, E.M.
1965 "Epidemiology and medical care statistics." In The role and methodology of classification in psychiatry and psychopathology, pp. 76-98. U.S. Dept. of Health, Education and Welfare (PHS Publication #1584): Washington, D.C.

Guttmacher, M.S.
1960 The mind of the murderer. Farrar, Straus, and Giroux: New York.

Harary, F.
1966 "Merton revisited: A new classification for deviant behavior." Sociological Review, 31, 693-97.

Hartjen, C.A., and D.C. Gibbons
1969 "An empirical investigation of a criminal typology." Sociology and Social Research, 54, 56-62.

Hippchen, L.J.
1975 "Changing trends in correctional philosophy and practice." In Correctional classification and treatment: A reader, pp. 17-24. W.H. Anderson Co.: Cincinnati, Ohio.

Hood, R., and R. Sparks
1970 Key issues in criminology. McGraw-Hill Book Co.: New York.

Huizinga, D.
1977 Cluster analysis: Evaluation and dynamic typologies. Doctoral dissertation. University of Colorado: Boulder.

Inciardi, J.A.
1970 "The adult firesetter: A typology." Criminology, 8(2), 145-55.

Jenkins, N.L., and L. Hewitt
1944 "Types of personality structure encountered in child guidance clinics." American Journal of Orthopsychiatry, 14, 84-94.

Jesness, C.F.
1974 Sequential I-level classification manual. American Justice Institute: California Youth Authority.

Kinch, J.W.
1962 "Continuities in the study of delinquent types." Journal of Criminal Law, Criminology and Police Science 53, 323-328.

Kornfeld, J.P., T.G. Eynon, W.A. Hetzner, M.E. Pfeiffer-Kornfeld, W. Bates, and T. Baumer
1975 State-of-the-art of offender classification in the U.S.A. IIT Research Institute: Chicago.

Kulik, J.A. K.B. Stein, and T.R. Sarbin
1968 "Dimensions and patterns of adolescent anti-social behavior." Journal of Consulting and Clinical Psychology, 32, 375-82.

Lindesmith, A.R., and H.W. Dunham
1941 "Some principles of criminal typology." Social Forces, 19, 307-14.

Loevinger, J.
1957 "Objective tests as instruments of psychological theory." Psychological Reports, 3, 635-94.

Marcus, A.M.
1975 "The asocial personality - a classification." Canadian Journal of Criminology and Corrections, 17(1), 94-98.

McCaghy, C.M., P. Giordano, and T. Henson
1976 "Toward a typology of auto theft." Paper presented at The American Society of Criminology 28th Annual Meeting, Tucson, Arizona.

Megargee, E.I.
1966 "Undercontrolled and overcontrolled personality types in extreme antisocial aggression." Psychological Monographs, 80, no. 611.

Molof, M.J., and C.F. Jesness
1972 "Project sequil: The development of a sequential I-level classification system." Yearly report. American Justice Institute: California Youth Authority (Contract A 147-70).

Mountbatten, E.
1966 Report of the inquiry into prison escapes and security. Cmnd 3175 (H.M.S.O.).

Neustatter, W.L.
1957 The mind of the murderer. Philosophical Library: New York.

Opp, K.D.
1973 "Problems of classification in criminology." In Collected studies in criminological research: Methodological aspects of classification in criminology, Vol 10. Council of Europe: Strasbourg, France.

Orlando, F., and J. Black
1974 "Classification in juvenile court: The delinquent child and the child in need of supervision." Juvenile Justice, 25(1), 13-25.

Quay, H.C.
1975 "Classification in the treatment of delinquency and antisocial behavior." In Issues in the classification of children, Vol. 1, pp. 377-92. Jossey-Bass Publishers: San Francisco.

Redl, F.
1956 New perspectives for research on juvenile delinquency. Children's Bureau Publication No. 356.

Reiss, F.
1952 "Social correlates of psychological types of delinquency." American Journal of Sociology, 17, 710-718.

Roebuck, J.B.
1966 Criminal typology. Charles C. Thomas: Springfield, Illinois.

Sampson, A.
1974 "Post-prison success prediction - preliminary Florida study." Criminology, 12(2), 155-73.

Sanders, M. and J.H. Welton
1972 "Vandalism." Report prepared for Federal Railroad Administration. National Technical Information Service: Springfield, Virginia.

Schrag, C.C.
1961 "A preliminary criminal typology." Pacific Sociological Review, 4, 11-16.

Shafer, S.
1969 "The responsible criminal type." In Theories in criminology. Random House: New York.

Simon, F.H.
1972 "Statistical methods of making prediction instruments." Journal of Research in Crime and Delinquency, 9(1), 46-53.

Smiley, W.C.
1977 "Multivariate classification of male and female delinquent personality types." Doctoral dissertation. In Classification and delinquency: A review. W.C. Smiley (ed.).

Sutherland, E.
1939 Principles of criminology. J.B. Lippincott Co.: Philadelphia.

Swanson, R.M., and M.J. Mobley
1976 "Patterns of delinquent behavior: Dimensions of self-reported behaviors." Unpublished manuscript.

Sykes, G.M.
1958 The society of captives. Princeton University Press: Princeton, New Jersey.

Warren, M.Q.
1966 Classification of offenders as an aid to treatment. Prepared for the President's Commission on Law Enforcement and Administration of Justice, Task Force on Corrections: Washington, D.C.

Warren, M.Q.
1971 "Classification of offenders as an aid to efficient management and effective treatment." Journal of Criminal Law, Criminology, and Police Science, 62, 239-58.

Wenk, E.A., and T.V. Halatyn
1974 An analysis of classification factors for young adult offenders, Vol. 1. National Council on Crime and Delinquency: Davis, California.

Wood, A.
1969 "Ideal and empirical typologies for research in deviance and control." Sociology and Social Research, 53, 227-41.

Yepsen, L.
1975 "Classification: The basis for modern treatment of offenders." In Correctional classification and treatment. W.H. Anderson Company: Cincinatti, Ohio.

CHAPTER 3

THE FAILURE OF CLASSIFICATION IN CRIMINOLOGY:

SOME RESULTS OF INADEQUATE CLASSIFICATION

INTRODUCTION

Even a cursory review of the field of criminological classification will reveal that classification is often not utilized in a careful or systematic manner, either within research criminology or within the more practical contexts of criminal justice practice. Classification is, in many instances, either carried out in an extremely haphazard way or it is ignored entirely as an important component in decisions regarding the disposition and treatment of offenders. Yet, as Eynon (1975) has written,

> until we begin serious scientific research on classification and treatment, we will have to content ourselves with inflicting our ignorance upon the hapless offenders who have fallen into our clutches.

A similarly bleak picture regarding the current state of classification research and practice is also noted in a number of key reviews within the field (see, for example, Kornfeld, et al. 1975, and Gibbons 1975).

In a recent non-technical review of classification in criminology, Kornfeld et al. (1975) wrote that

> . . . the problem of developing a basic offender classification system, universally accepted and approved throughout the criminal justice system, remains to be completed.

The need for good classificatory research with clear, practical applications remains pressing. The promise of such work is still motivating many people to work with classificatory systems, but the factors which are undermining their efforts have not yet been examined and dealt with. We turn now to a brief examination of some of these factors.

THE RELATION BETWEEN INADEQUATE CLASSIFICATION RESEARCH AND THE FAILURE OF CLASSIFICATION IN PRACTICAL CONTEXTS

The failure of classification within the criminal justice system cannot be understood without examining the theoretical and research contexts in which this failure originates. Many ethical, political, and practical dilemmas will be resolved not through manifesting a more profound humanitarian and empathetic concern, but through the use of a more valid, thorough, and appropriate methodology. We believe that a concern with improved methodology may lead to more insight, fairness, and consequently to more humane treatment.

The relationship between "inadequate research" and "inadequate practice" can be seen as a series of levels in which failure at one level can be profoundly disruptive at the next level. The first "surface" level involves the everyday practice of classification within criminal justice institutions (jails, probation departments, courts, detention centers, diagnostic and assessment units, etc.). The next deeper level (level B) consists of the products of criminological research in classification. These products are basically various classification systems which generally include more or less explicit descriptions of each type contained within the system, and occasionally a set of rules for the identification and assignment of new cases into the system. The next level (level C) consists of the conceptual and methodological approaches to the creation and use of these systems. There are a number of issues present at each level which can interact with and affect the other levels. The provisional outline which follows is intended to provide an overview of the major issues which operate at each level. The interconnections between these levels will become progressively clearer later in the report.

Level A. Classification practice within criminal justice institutions (Practice).

1. Large amount of skepticism among criminal justice personnel regarding the validity of any current classification system (Eldefonso and Coffey, 1976).
2. Confusion among criminal justice personnel regarding the methods of classification and how they can be effectively implemented (see Gottfredson, 1975).
3. Many offender classification systems currently in use have severe operational shortcomings:
   a. Limited predictive validity. This problem means that the system in question will not be very useful for predictive decision making (see Bottoms, 1973).
   b. Lack of explicit allocation procedures. Many typologies do not include specific guidelines for the identification and assignment of new cases (see Beker and Heyman 1972 for criticism of the I-level in this regard).
   c. Partially due to the lack of explicit allocation procedures mentioned above, many cases are often mis-classified.
   d. Partially due to the unreliability and confusing instructions regarding the use of offender classification systems, many cases are often not classified at all.
4. Few offender typologies currently in use are connected in any clearly valid way to types of differential treatment. Thus they provide only a very minimal basis for intervention and control. (See Hood and Sparks 1970, Eynon, 1975).
5. Parole boards and other decision-making bodies within the justice system either generally ignore the need for effective classification systems or do not use currently operating classification systems in a rigorous manner.

6. Parole boards and other decision-making bodies continue to rely primarily on stereotyping and other forms of inexplicit and intuitive categorization when confronted with the exigencies of making a decision (Daudistel, Sanders, and Luckenbill 1979; Gottfredson 1975).

Level B: The products of classification research in criminology.

1. There has been an unsystematic proliferation of classification systems. These (often competing) systems are difficult to integrate with each other, often have different implications for decision-making, and there is usually no good reason for prefering one to another.
2. Most proposed classification systems have not been evaluated (see Hood and Sparks 1970, Gibbons 1970, Glaser 1974).
3. Many typological systems have no clear empirical and operational basis, and thus are difficult to effectively measure, evaluate or implement.
4. Many classification systems have never been replicated, which calls into question their reliability (see Beker and Heyman in relation to the I-level).
5. Most systems have no particularly obvious relationship to existing criminological theories, which limits their construct and theoretical validity (Gibbons 1975, Beker and Heyman 1972).
6. Most systems reveal high 'within-class' variability, and low 'between-class' differentiation. This implies that they have a very tenuous relations to actual structure of the data, and thus may not be particularly useful either for communicating about offenders or for storing and retrieving information.

7. There are almost no descriptive taxonomies in criminology. The descriptive taxonomic task has either been ignored or has not been well done. Across all the sciences it is almost axiomatic that careful and systematic descriptive research is a prerequisite to good explanatory and theoretical work (see Northrop 1947, Meehan 1968, Eysenck 1950, and others). Cattell's famous dictum that "nosology precedes etiology" asserts that etiological explanations are dependent upon the prior clarification of descriptive patterns. Most criminologists are now beginning to recognize that descriptive taxonomic research has been thoroughly neglected in criminology and must be pursued (see Wenk and Halatyn 1975, Hood and Sparks 1970, Sparks 1973, Gibbons 1975, etc.).
8. A failing related to the above has been the misguided attempt to find all-inclusive explanatory taxonomies. A number of criminologists have attempted to construct theoretical taxonomies which purport to explain all criminal behavior (see, for example, Ferdinand 1967). These theoretical systems suffer dramatically from the problem of oversimplification. They are simply inadequate to cope with the full complexity of criminal behavior.

Level C. Methodological and conceptual failures.

1. Conceptual confusion regarding structural aspects of classification. The concept of a 'class' can be operationally defined in a number of different ways. The most basic division may be that between monothetic and polythetic classes (see Sokal and Sneath 1963, Sneath and Sokal 1974, Bailey 1975). Additional variations in the creation of classes can stem from different linkage criteria embodied in various cluster analytic methods (e.g., single, average-link, complete link, etc.). Yet, although these considerations intimately affect the use of classification for a specific purpose, criminologists have, by and large, ignored them.

2. Failure to realize that different kinds of classification systems are appropriate for each distinct type of purpose. The nature of particular taxonomic systems can be varied to optimize a number of quite different purposes. For taxonomic research to be useful, it is necessary that the appropriate kinds of systems are generated for each particular purpose. There is thus an urgent need for criminological researchers to be cognizant of the different purposes of classification, and the manner in which different kinds of classification systems can be constructed to optimize a given purpose. Some of these different types of classification include:

   a. Predictive classifications: to optimize predictive accuracy in reference to a specified criterion (escape, recidivism, etc.).

   b. Descriptive classifications: for accurate description, enumeration, and epidemiology.

   c. Theoretical/explanatory classifications: to embody explanatory or theoretical processes within a taxonomic framework. Classification of any entity into this system would be equivalent to explaining its performance with relevance to some criterion variable.

   d. Data storage/retrieval classification systems: classification systems can operate for the orderly storage and retrieval of large masses of information (e.g., library classifications, the periodic table, etc.)

   e. Theory development classifications: these are used in exploratory research to restructure the data in ways that might reveal useful new hypotheses.

These general purposes reappear in a multiplicity of different practical situations within the criminal justice system. For example, predictive taxonomies can be used in regard to prediction of recidivism, escape risks, risks of violence, etc. Different purposes often imply very different kinds of classification structures, e.g., descriptive classifications would necessarily be isomorphic to the data whereas predictive, or data storage and retrieval systems do not necessarily have this property. Library classification schemes using alphabetical ordering procedures, for instance, have no isomorphism with the knowledge structures of the material classified.

3. Overabundance of non-empirical, and often non-operational, classification systems.

4. The attitude of some researchers that typology construction is an end in itself, with an accompanying disinterest in either follow-up evaluation or in non-research applications.

5. Oversimplification. This can happen in two basic ways. first, the classification may have too few categories. Second, it may be based on an insufficient number of descriptive variables. The problem of oversimplification leads directly to ethical criticisms of the type specified in level A, since one major argument is that many classification systems are overly reductionist and do not match the full cpmplexity of criminal behavior (see Kornfeld et al. 1975, Bottoms' 1973, Clausen 1965). Some criminologists are aware of this problem and there has been a call for more 'multi-dimensional' classification (Hood and Sparks 1970). These latter authors believe that a good typology should be "as rich in types as possible."

A different kind of problem emerging from the use of oversimplified classification systems is that of 'information overload.' An overly simple classification system will generally be unable to cope with the diversity and quantity of information which enters into real-life decision-making. They provide no effective way of synthesizing data in order to make effective decisions. A result of such overload is confusion and inordinate delay within the criminal justice system. A further result is that the system eventually becomes unacceptable to its users (i.e., it lacks 'face validity' and efficiency. The offender being classified and criminal justice personnel will both be skeptical about classifiation systems which do not match the complexity of the phenomena being dealt with.

6. Confusion, or carelessness, regarding adequate approaches to the evaluation of taxonomic systems. Occasionally a taxonomic system offered in the criminological literature, such as Blackburn's (1971) typology of homicides, is sensibly and rigorously evaluated. However, the performance of criminological researchers in regard to studying the reliability, validity, and general usefulness of their classification systems is dismal. A number of interrelated factors lie behind this generally inadequate approach to evaluation. The predominance of non-empirical and loosely defined theoretical typologies has been a disaster for systematic evaluation (see Ferdinand 1967, Solomon 1977). Most of these impressionistic systems are not open to empirical analytical evaluation, and have not been successfully operationalized. Another major problem is the conceptual confusion regarding the different purposes of taxonomic systems in relation to different modes of evaluation. Evaluation should be tied to a specific purpose. For example, predictive taxonomic systems

should be evaluated according to their predictive accuracy. Evaluative criteria therefore follow naturally from the purposes of the classifier. Unfortunately, a perusal of the criminological literature suggests that criminal taxonomists are either naive regarding the differential purposes of typologies or are quite unconcerned with them (again regarding taxonomy construction as somehow an end in itself). To the extent that purposes are unspecified, or are somehow unimportant, there will be no clear basis for evaluation.

7. Biased, partial, and erroneous taxonomy construction. This stems from many factors, including:

   a. Biased or partial sampling of cases. The creators of classification systems within the field of criminology have generally paid little attention to the representative nature of the samples on which they have based their taxonomic systems. An examination of the reviews by Ferdinand (1966), Rubenfeld (1967), and others, indicates an extremely unsystematic approach to sampling. A typical example of this problem can be found in Baer (1970), where a proposed typology of delinquents is constructed utilizing sixty male delinquents in one small area with no non-delinquent control groups.

   b. Incoherent or poor selection of variables. Use of irrelevant, non-operational or unreliable variables to construct a taxonomic system. Redundancy and overlap between variables can also create havoc in the construction of classification systems.

   c. The problem of "artificial boundaries." Each class is basically delineated by a central tendency and a boundary condition. In the absence of an adequate methodology, boundary conditions may be inappropriately drawn, thereby placing artificial or unrealistic

boundaries between classes. It is therefore possible for the proposed classification system to contain an artificial number of types, i.e., to impose a number of types which are not actually contained on or suggested by the data (see Hood and Sparks 1970). There may be either too many or too few categories in these typologies. In either case, there is a false conceptual imposition upon the phenomena being examined. Numerical taxonomy in particular has focused a very great amount of attention upon various methodological approaches to the process of delineation of boundaries between data modes. (see Everitt 1974, Huizinga and Brennan 1976).

8. The utilization of inappropriate taxonomic methodology for both the construction of classification systems and their evaluation, along with the mis-use and misunderstanding of currently available statistical methodology. A review of quantitative classification research in criminology indicates two serious problems at the level of methodological approaches to typology construction. Firstly, reliable quantitative methodology has not been widely available until recently. As late as the mid-1960's the great majority of research criminologists did not have access to quantitative taxometric methods. This largely accounts for the prevalence of non-empirical systems and over-simplified monothetic systems (created by the polar division of a few key variables).

With the advent of quantitative taxometrics and the increasing availability of sophisticated programs, criminologists have been increasingly utilizing more powerful and appropriate statistical methods. However, this increase of sophisticated methhodology has lead to the problem of their misunderstanding and mis-use. This unfortunately is to be expected since taxometric methods are extremely flexible, and involve

a large number of crucial decisions for their effective use, along with a large number of alternative ways to do similar tasks. Furthermore, the methods themselves remain by and large untested on criminological data. McKinney (1970) in particular is concerned about the issue of the general neglect by all the social sciences of typological methodology. He writes:

> . . . types and typologies are ubiquitous . . . everybody uses them, but almost no one pays any attention to the nature of their construction. Despite the omnipresence of typologizing in social enquiry it remains a relatively underdeveloped aspect of methodology generally.

This statement is particularly true of the field of criminology. Classification systems are employed pervasively by both criminological theorists and practitioners within ghe criminal justice system; yet, the appropriateness of their methodology remains generally suspect.

ETHICAL ISSUES IN APPLIED CRIMINOLOGICAL CLASSIFICATION

As Wilkins and Gottfredson point out, any decisions made which involve human beings inevitably contain a moral value content. Decisions which are made within the criminal justice system are thus inextricably entwined with what are essentially ethical problems, since these decisions may profoundly affect the offender's life. It is this human element which renders classification most difficult and which, it seems, is most often ignored. Poor classification practices inevitably lead to confusion, inefficiency, and injustice. The institution which perpetuates and imposes such poor classification decisions is, in a sense, behaving unethically. Ethical behavior would consistently attempt to minimize careless, over-simplified or confused classification decisions.

Complexity and oversimplification

One of the most important ethical considerations deals with the consequences of the omission of pertinent information or of attempting to condense all the significant aspects of a highly complex behavior (i.e., human behavior) into a manageable classification system (Clausen 1965). Human individuals exhibit an extraordinarily high level of complexity (or as Gottfredson and Wilkins, 1978, put it, human intelligence is a "very high variety generator") and certain approaches to classification can (and do) over-simplify the behavioral, social, and psychological phenomena surrounding, criminal justice. Any classification system will result in certain omissions and distortion, and the moral implications of this distortion must be explored (Wenk and Halatyn 1974). All relevant information must be considered; the omission of any relevant criteria when dealing with human beings may render a classification system unethical. Poor classification systems may severely over-reduce the complexity of human behavior by compressing multidimensional information into essentially unidimensional categories (Kornfeld, et al. 1975).

Since it is evident that no single explanation or classification can account for ALL offenses and behaviors, some way must be found to adequately cope with the heterogeneity of deviant behavior. An initial impractical response to this problem tends to be a call for individualized programs for offenders (Yepsen 1975); however, this solution is not practically nor theoretically feasible, and usually leads to increased confusion. No fully individualized policy could hope to deal with the 'infinite variety' of offender behaviors. Some condensation and categorization is always necessary.

These and similar problems find their expression in what Ashby (1962) calls "the law of requisite variety", which simply states that any attempt to describe, control, or explain a phenomenon must match the phenomena in terms of complexity and variety. As mentioned above, one way to meet this requirement would be to create a unique category for each individual: an undertaking which is at once impossible and undesirable. Another way of conforming to Ashby's law would be to decrease the complexity of the system while increasing its ability to handle information. This is essentially what classification schemes attempt to do, i.e. to match the variety of behavior with a system that has 'considerable information handling capacities' and a high 'variety of response' (Gottfredson and Wilkins 1978). Classifications may contain an element of injustice, since they will necessarily truncate and distort the holistic nature of human behavior. However, it is only when the ability of the classification system to handle variety falls significantly below the ability of the offender to generate it, that decisions made using that system are open to the accusation of unethical practice. This criterion could be used in evaluating many of the extant classification systems which permeate criminal justice systems.

Stigma and Labeling Stemming from Poor Classification

Another ethical issue relevant to criminal classification is the problem of labeling and the stigma which often accompanies it. Any sort of label inevitably affects the way in which an individual is perceived and understood by others (Orlando and Black 1975). Labeling is in a sense related to the problem of oversimplification outlined above in that a label may obscure much more than it communicates: a categorization unavoidably involves a loss of information. Many prisoners feel that a classificatory label entails an extreme dehumanization (Kornfeld, et al. 1975). This viewpoint

reveals a fear that the complexity of a whole being is being ignored and that the prisoner is being reduced to a label. This feeling further exacerbates the suspicion that classification is being used predominantly for institutional needs and not the prisoner's (Ibid.). Clausen (1975) provides this necessary caution in reference to the power of a (misapplied) label:

> A diagnostic label too often serves as an excuse for not looking at the person and his situation, but this is the fault of the labeler, not the label.

It is not the classification per se, therefore, which can be so destructive, but poor classification and confused utilization of it.

The label 'delinquent' can have profound and upsetting consequences, though as Quay (1975) points out, it is a legal label and does not necessarily imply psychological, sociological, or behavior homogeneity. Orlando and Black (1975) state that alternative naming devices such as "child in need of supervision", "unruly children", etc., are being used to minimize the stigma attached to the label "delinquent". All of these, they stress, are nonscientific legislative definitions. In the Gault decision of 1967, which required more procedural safeguards for children involved with the law, it was pointed out that identification as a 'delinquent' produces "only slightly less stigma than the term 'criminal'". In the 1946 decision of Jones vs. Commonwealth, the Virginia Supreme Court had this to say:

> ...the stigma of conviction will reflect upon him for life. It hurts his self-respect. It may, at some inopportune, unfortunate moment, rear its ugly head to destroy his opportunity for advancement, and blast his ambition to build up a character and reputation entitling him to esteem and respect of his fellow man. (Ibid.)

With even more melodramatic flair, Justice Musmanno (In re Holmes, 1954) wrote:

> The grim truth is that a juvenile court record is a lengthening chain that its riveted possessor will drag after him through childhood, youthhood, adulthood, and middle age. (Ibid.)

As a result of the above considerations, the concept of 'spoiled public identity' gained widespread attention. Some theorists have suggested that the official response to delinquency may push a youth further into criminal behavior (Ibid.). For example, Gold and Williams (1969) suggest that apprehension (and thus the label 'delinquent') itself leads to further delinquency. The label is seen as a self-fulfilling prophecy.

The stigma associated with what are basically legalistic or social categories is a pervasive theoretical, practical and political dilemma. This is nicely illustrated by categorization of offender 'dangerousness'. As Kornfeld, et al. (1975) point out, danger is a social perception. What constitutes danger and risk is a social decision and thus is fraught with political overtones (Bottoms 1973). Nonetheless, no matter how relative the label utilized, the stigma tends to become generalized across a number of social contexts. Good classification practice and systems can avoid the problems of stigma and premature or innappropriate labeling by being sufficiently accurate, sufficiently complex, and by careful 'naming' of the resultant classes. The 'naming' of a classification system, and of its component sub-classes is a critically important component of the construction and practical use of classification. Procedures for 'naming' classes will vary according to the purpose of the system. Different purposes will command more or less public concern and public scrutiny, e.g., predictive systems for release and parole decisions will be more critically scrutinized than descriptive systems used for enumeration and epidemiology.

The President's Commission on Law Enforcement of 1967 suggested that

written guides and standards be implemented in the course of in-service training for criminal justice workers, in order to prevent abusive labeling by the intake staff (Orlando and Black 1975). Similar guidelines for construction and implementation of classification systems might be helpful, and would hopefully begin to combat the current loss of information, dehumanization, and stigma inherent in many extant classificatory systems.

Offender Rights and Needs in Relation to Classification

Kornfeld, et al. (1975) point out, in reference to the use of classification, that the offender may have a legal right to refuse information to classifiers; this is essentially questioning the extent to which an attempt to classify an offender involves an untoward and illegal invasion of his or her privacy. How much information is an offender required to give? Similarly, there is a certain question as to whether the offender has the right to refuse treatment; treatment attempts could be construed as unlawful intervention into the offender's life. This raises the further question of whether or not a classificatory label can be considered a legal status, and whether boards which make decisions utilizing classification schemas are subject to judicial restraints.

We should note here that under certain conditions offenders react positively to classification. Kornfeld, et al. (1975) found that offenders were positive about classification if they felt that it could 1) validly determine appropriate treatment, 2) accurately predict recidivism, 3) provide more understanding of the offender and if 4) it was used for constructive purposes. It is thus interesting to reflect that a positive appraisal on the part of the person being classified seems to depend on essentially the same critical factors which are important to the user, i.e. issues of validity and reliability in the classification system per se, and whether it is used

intelligently for constructive (presumably rehabilitative) purposes. Offenders also accepted classification more positively if he/she was given some explanation of the results.

Related to consideration of the offender's legal rights is the necessarily problematic issue which arises in reference to the offender's needs, particularly since the needs of the offender are often at variance with the interests and needs of the community (Orlando and Black 1975). It is evident that much classification is being performed with insufficient attention to offender need. One possible way of assuring this increased relevance would be to further involve offenders in the classification process itself. Kornfeld, et al. (1975) stress the need for offender participation by stating that offender deletion of erroneous information could lead to a more initially accurate classification. They further suggest that perhaps no meaningful classification can take place "without the informed cooperation of offenders". This participation may also help to alleviate a problem which occurs in the use of many current classification systems: the tendency to be overconcerned with the requirements of institutional management at the expense of concern with "offender re-integration" (Flynn 1975).

There is much confusion about factors which have to be taken into account if classification systems are to adequately deal with offender needs. Bottoms (1973), for example, without clear specification stresses the need for a typology to take into account the offender's "situational environment". On the other hand, Kornfeld, et al. (1975) suggest that it is actually unrealistic to consider offender need as a criteria for classification, and offer 'reduction of risk' as a more adequate criteria since it can be handled and evaluated mathematically. This latter position seems to reflect some confusion about purposes. It blurs the distinction between classification for treatment and classification for prediction.

Nonetheless, it is imperative that offenders be treated fairly; and this requires, among other things, a more consistent and valid approach to classification (Flynn 1975). This revolves around the problem of equity; that is, whether or not "similar offenders in similar circumstances, are given similar sentences" (Gottfredson and Wilkins 1978). In this context, unfairness implies dissimilar treatment for similar offenders in similar frameworks, which further implies the use of haphazard or widely unreliable classification systems. Another issue which relates to the concept of equity requires that the offender not receive more help than is required and not be kept in a more "secure condition" than is required by potential risk (Flynn 1975). Thus, the equity issue converges here with the issues of efficiency and diversion.

Accountability for Poor Classification Decisions

Kornfeld, et al. (1975) formulate the problem of accountability for poor classification bluntly: if a person placed on probation commits a murder, who or what is to be held responsible? The question of accountability arises whenever the effects of a dubiously rendered decision are felt. These poor classification decisions may result from a faulty system of classification or careless use of a valid system. Currently there is a certain trend towards making those responsible for a bad decision legally liable. This, as Kornfeld, et al. (1975) point out, could very well lead to a rise in incarcerations since decision makers will be reluctant to place themselves on the line by making release decisions. More accurate classificatory systems would help to determine risk of recidivism, violence, and escape, among other things.

It is also likely that increased accountability could lead to an attempt at clarifying the criteria used for making decisions, and to make explicit

those tacit categorizations which are often used in decision making. Gottfredson and Wilkins (1978) suggest that parole boards adjust their focus in reference to the tenor of the attacks directed towards them. This implies that increasing accountability could very well implement an attempt to create more adequate and reliable guidelines for classification and decision making. Gottfredson and Wilkins (1978) also suggest that using models to test decision making methods may decrease the quantity of bad decisions: that is, situations usually assumed to involve moral value choices may be simulated (possibly by computer).

CONCLUSION

The above outline is meant to be provisional and does not suggest a dogmatic view of the various relationships examined. The interrelations between different issues at different levels are complex. For instance, the absence of a theoretical foundation for taxonomy construction in criminology has been partially responsible for the rampant proliferation of overlapping partial classification systesm. This in turn has deluged the criminal justice worker with too many competing systems. This has lead to confusion and finally rejection on the part of many criminal justice personnel. Needless to say, the absence of predictive validity, theoretical coherence, and good guidelines for the operationa use of typological systems have also contributed to the negative or antipathetic attitudes of many criminal justice workers. Thus it should be evident that practical difficulties at one level are intimately related to research and methodological difficulties at another level. Our purpose in the present Chapter is simply to present an initial delineation of these different levels and to briefly examine some of the connections between them.

BIBLIOGRAPHY

Ashby, W.R.
1956 "Self-regulation and requisite variety." In Systems thinking, pp. 105-24 Penguin Books

Baer, D.J.
1970 "Taxonomic classification of male delinquents from autobiographical data and subsequent recidivism." Journal of Psychology, 76(1), 27-31.

Bailey, K.
1975 "Cluster analysis." In Sociological methodology. Jossey-Bass: San Francisco.

Beker, J., and D.S. Heyman
1972 "A critical appraisal of the California differential treatment typology of adolescent offenders." Criminology, May, 1-59.

Blackburn, R.
1971 "Personality types among abnormal homicides." British Journal of Criminology, 11(1), 14-31.

Bottoms, A.E.
1973 "Methodological aspects of classification in criminology." In Collected studies in criminological research: Methodological aspects of classification in criminology, Vol. 10. Council of Europe: Strasbourg, France.

Clausen, J.A.
1965 "Some implications of classification in sociology for problems of classification in psychiatry and psychotherapy." In Classification in psychiatry and psychopathology, pp. 168-72. U.S. Dept. of Health, Education, and Welfare (PHS-NIMH): Chevy Chase, Maryland.

Daudistel, H.C., W.B. Sanders, and D.F. Luckenbill.
1979 Criminal justice situations and decisions. Holt, Rinehart and Winston: New York.

Eldefonso, E., and A.R. Coffey
1976 Process and impact of the juvenile justice system. Glencoe Press: Beverly Hills, California.

Everitt, B.
1974 Cluster analysis. Halsted Press: New York.

Eynon, T.G.
1975 "New roles of research in classification and treatment." In Correctional classification and treatment: A reader, pp. 70-74. W.H. Anderson Co.: Cincinnati, Ohio.

Ferdinand, T.N.
1966 Typologies of delinquency - a critical analysis. Random House: New York.

Flynn, E.E.
1975 "Problems of reception and diagnostic centers." In Correctional classification and treatment: A reader, pp. 28-34. W.H. Anderson Co.: Cincinnati, Ohio.

Gibbons, D.C.
1970 "Differential treatment of delinquents and interpersonal maturity levels theory: A critique." Social Services Review, March, 22-33.

Gibbons, D.C.
1975 "Offender typologies - two decades later." British Journal of Criminology, 15(2), 140-56.

Glaser, D.
1974 "The classification of offenses and offenders." In Handbook of criminology, pp. 45-84. Rand McNally College Publishing Company: Chicago.

Gottfredson, G.D.
1975 "Organizing crime: A classifactory scheme based on offense transitions." Journal of Criminal Justice, 3, 321-32.

Gottfredson, D.M., C.A. Cosgrove, L.T. Wilkins, J. Wallerstein, and C. Rauh
1978 Classification for parole decision policy. National Institute of Law Enforcement and Criminal Justice: Washington, D.C.

Hood, R., and Sparks, R.
1970 Key issues in criminology. McGraw-Hill Book Co.: New York.

Huizinga, D., and T. Brennan
1976 "Some issues in the methodology of type construction." Paper presented at the American Society of Criminology 28th Annual Meeting, Tucson, Arizona.

Kornfeld, J.P., T.G. Eynon, W.A. Hetzner, M.E. Pfeiffer-Kornfeld, W. Bates, and T. Baumer
1975 State-of-the-art of offender classification in the U.S.A. IIT Research Institute: Chicago.

McKinney, J.C.
1966 Constructive typology and social theory. Appleton-Century-Crofts: New York.

Meehan, E.J.
1968 Explanation in social science, a system paradigm. The Dorsey Press: Homeward, Illinois.

Northrup, F.S.C.
1947 The logic of the sciences and the humanities. MacMillan: New York.

Orlando, F., and J. Black
1974 "Classification in juvenile court: The delinquent child and the child in need of supervision." Juvenile Justice, 25(1), 13-25.

Quay, H.C.
1975 "Classification in the treatment of delinquency and antisocial behavior." In Issues in the classification of children, Vol. 1, pp. 377-392. Jossey-Bass Publishers: San Francisco.

Rubenfeld, S.
1967 Typological approaches and delinquency control - a status report. Center for Studies of Crime and Delinquency, National Institute of Mental Health. HEW: Washington, D.C.

Sokal, N.R., and P.H.A. Sneath
1963 Principles of numerical taxonomy. Freeman: London.

Solomon, H.M.
1977 Crime and delinquency: Typologies. University Press of America: Washington, D.C.

Sparks, D.N.
1973 "Euclidean cluster analysis." Applied Statistics, 22(1), 126-30.

Wenk, E.A., and T.V. Halatyn
1974 An analysis of classification factors for young adult offenders, Vol. 1. Research Center, National Council on Crime and Delinquency: Davis, California.

Yepsen, L.
1975 "Classification: The basis for modern treatment of offenders," In Correctional classification and treatment. W.H. Anderson Co.: Cincinatti, Ohio.

CHAPTER 4

A BRIEF EXAMINATION OF DECISION MAKING AND CLASSIFICATION IN THE CRIMINAL JUSTICE SYSTEM

INTRODUCTION

Kornfeld, et. al. (1975), in conducting a limited institutional survey, discovered that virtually none of the inmates questioned were aware that they had been subjected to a process of classification. This is all the more surprising in light of the pervasive use of classification at each stage of the criminal justice system. Most writers agree that use of clasificatory processes is unavoidable in justice decision making (see Gottfredson and Wilkins 1978). Most decisions concerning an offender require that the case first be interpreted "as an instance of some larger categories" (Daudistel, Sanders, and Luckenbill 1979). Often these categorization decisions are highly intuitive (see Hood and Spenks, Kornfeld 1975). However, in this form they represent manifestations of "implicit policy" and merely serve to underscore the ubiquitous influence of classification in the justice process (Bottoms 1973).

The National Advisory Commission of Criminal Justice Standards and Goals suggest that "classification is conceptualized as a system or process by which a correctional agency, unit, or component determines differential care in handling of offenders (Megargee and Bohn 1979)." Classification, then, is utilized to differentiate between types of offenders in order to provide equitable and appropriate treatment; a faulty or unreliable classification system can therefore produce disastrous effects whose ripples are felt to the outermost reaches of criminal justice. A poor classification decision, whether it results from use of an inadequate and misunderstood method or from sheer carelessness, implies a series of moral, theoretical, and pragmatic

issues each of which must be dealt with by criminal justice workers if the system is to be made to operate both humanely and effectively. It seems prudent to re-examine the use of classification in decision making within criminal justice institutions.

In light of the above considerations, the purpose of this paper is twofold: first, to identify the decision making points in the criminal justice system where classification usually occurs, and second, to identify the basic problems related to classification in current justice decision making practice. It is hoped that this analysis will help to clarify the importance of classification in regard to fair and consistent decision making. "Until we begin serious scientific research on classification and treatment," writes Eynon (1975), "we will have to content ourselves with inflicting our ignorance upon the hapless offenders who have fallen into our clutches."

Decision Making Junctures

It is often not at all clear, even to those who are intimately involved, how certain decisions are made in the criminal justice system. However, even though some of these mechanisms by which decisions are made remain largely hidden, it is possible to identify the points within the system where explicit decision making is prominent and to examine ways in which classification is involved at those points. As mentioned earlier, decision making inevitably entails the interpretation of a case as an instance of some larger more general categories. Daudistol, Sanders and Luckenbill (1979) have argued that because penal code categories are "inherently abstract," criminal justice workers develop "typifications" to represent categories for the processing of cases; these "typifications" include many properties learned through "organizational experience," which involves knowledge of the offender's

psychological and sociological background, etc. (ibid.) "Typifications" such as these may be extensively utilized whenever a decision needs to be made, yet at the same time they may remain only vague formulations. The lack of explicit classification systems serves to underscore the essentially arbitrary nature of many decisions and the need for explicating the classificatory schemas which are embedded within this process.

The following section attempts to identify the major decision making junctures in the criminal justice system and to at least partially delineate the way in which classification is used in each of them. A brief glance at Figure 1 should make it clear that while the adult and juvenile justice systems exhibit a basically similar structure, they also diverge at several important states. This is partially because the juvenile justice system conceives of itself as being essentially different from the adult system. The explicit goal is to help the youthful offender and to divert him or her from the system (though the reality is often otherwise). In fact, until the Gault decision of 1967, juveniles essentially had no rights of due process, since the court conceived of itself in a parental role. It follows that decision making in the juvenile system might assume a more informal manner, which indeed it does. Even more is left up to the personal discretion of the judge and the juvenile probation officer, and there is much more personal interaction with the offender. However, use of classification (particularly intuitive and subjective typifications) seems to be pervasive, and, given the legal and moral complexity of juvenile delinquent behavior, the need for valid classification may be more pressing.

Factors Common To Both Adult and Juvenile Systems

1. Contact With Police.

a. Decision: What action, if any, is necessary (arrest, warning, etc.).

b. Classification is usually used here on an intuitive level, utilizing a number of relevant variables (age, sex, seriousness, etc.) to "typify" a case in order to make an appropriate decision. The major difference between treatment of adults and juveniles at this stage revolves around the fact that the police are constrained to make a choice which least restricts the freedom of the minor (Eldefonso and Coffey 1976). After arrest of a juvenile, the police can handle a case within the department, release the offender to parents or guardians without referral, or refer the case to a non-authoritative treatment agency.

Adult System

1. Charging / First Appearance.

a. Decision: Those suitable for charging must be screened, and then a charge must be chosen. At first appearance, an offender may be sentenced (in the event of a guilty plea) or released with or without bond, in or out of custody.

b. Classification is necessary to categorize alleged offenders into separate classes which would suggest whether specific charges should be brought up and how the offenders should be processed (remand to jail, release on personal bond or payment of bail, etc.).

2. Preliminary Hearing.

a. Decision: Concerning the probable cause to hold a defendant for trial.

b. Classification at this stage is usually designed to reflect degrees of seriousness of the crime.

3. Plea Negotiation / Arraignment.

a. Decision: What plea to enter.

b. Classification is used if there is negotiation between the defense attorney and the prosecutor, e.g. they "typify" a case: "To 'typify' a concrete case is to define it as an instance of a 'commonsensical' class of

DECISION-MAKING POINTS IN THE CRIMINAL JUSTICE SYSTEM

WHERE CLASSIFICATION IS IMPORTANT

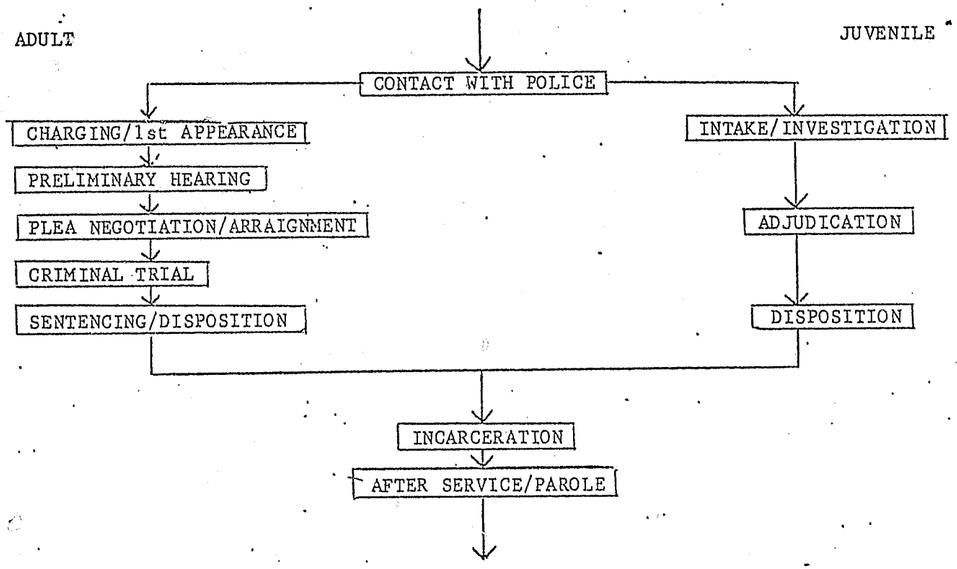


Figure 1.

events for which there are routine dispositions deemed by courthouse members as appropriate and just (Daudistel, Sanders, and Luckenbill 1979)."

4. Criminal Trial (in the event of not guilty plea).

a. Decision made with regard to guilt.

5. Sentencing / Disposition.

a. Decision needs to be made regarding disposition (institution, probation, etc.). This is essentially a placement decision. (Gottfredson, 1975, stresses that sentencing decisions are "guided unsystematically": "It would be difficult to find other decision problems affecting critically the liberty and future lives of large numbers of people in which decisions are made with so little knowledge of their results.")

b. Classification is necessary to provide valid guidelines for a correspondence between <u>type of offender</u> and <u>type of sentence / placement</u> (this essentially needs to be a diagnostic classification); however, as suggested above, this use of classification is also usually intuitive.

<u>Juvenile System</u>

1. Intake / Investigation.

a. Decision: Does the court have jurisdiction over the act? Should there be court action or referral to another agency? Is detention or release in order?

b. Classification is used extensively at this stage. The probation officer makes a 'diagnosis' of the juvenile which later informs the decision of the court. The juvenile must be classified as a delinquent, status offender, or neglected - abused child (Eldefonso and Coffey 1976). Responsibility classifications and seriousness classifications, and rehabilitation classifications are often dominant in this process. This is an important stage, since "the decisions made at the intake level can

characterize the juvenile in the eyes of other community agencies . . . (ibid.)." Phelps (1976) points out that 50% of the cases do not move beyond the intake level.

2. Adjudication.

a. Decision: Whether or not the juvenile will be found legally delinquent.

b. Classification is used here in a diagnostic way to determine how the case should be handled; particularly since the court emphasizes therapeutic concerns (Katkin, Hyman, and Kramer 1976). Classification here is still predominantly intuitive, though Eldefonso and Coffey (1976) have pointed out that strong attempts have been made to establish "standardized practices" of categorization.

3. Disposition.

a. Decision: needs to be made regarding the best course of action (probation, institutional commitment, or complete dismissal).

b. Classification is again imperative; many authors stress the need for classes which are empirically related to specific dispositions.

Both Systems

1. Incarceration.

a. Decision: What sort of institution would be appropriate for a specific inmate? Which inmates within a specific institution are to be used for which tasks? What would an appropriate treatment be? (Would parole be appropriate?)

b. Classification (both diagnostic and predictive) is obviously needed to provide an appropriate placement for each type of offender. Different types of offenders also must often be separated within the institution. Treatment and rehabilitation related classification systems are required at this point.

Eldefonso and Coffey (1976), among others, warn that here there is often a danger of "gearing classifications to institutional requirements."

2. After Services / Parole.

a. Decision: Who to release and under what circumstances.

b. Classification system helpful at this point to provide prediction of recidivism. Predictive classification systems are emphasized along with rehabilitation systems.

It is necessary to stress that decision making points may vary somewhat in their order (charging, for example, may follow the first appearance). Procedure in the juvenile court is even more subject to adjustments and revisions. The judicial processes, for examples, are 'informal' to the point where treatment is often prescribed without the benefit of a full hearing (this is known as "informal adjustment"). This is in harmony with the stated goal of the juvenile court which is emphatically rehabilitation. Or, as Katkin, Hyman, and Kramer (1976) put it, the juvenile court is partially a court of law and partially a social service agency. Thus, rehabilitation needs and potentiality may dominate classification decisions.

Basic Problems In Current Decision Making Practice

A. Substantive Problems

1. Inconsistency of decisions made. The lack of explicit and widely implemented classification systems has lead to a bewildering multiplicity of standards regarding decisions to be made. In the realm of juvenile justice, for example, "...any ten juvenile probation officers are likely to describe any ten cases in nearly a hundred different ways. (Eldefonso and Coffey 1976)." Furthermore, each state, local jurisdiction, probation officer, etc., has a "different standard for determining classification matters (ibid)."

The same situation obtains among the practitioners of adult criminal justice. Decisions tend to be made intuitively and often depend upon varying and perhaps unexamined ideas about classification. In making decisions regarding disposition, for instance, choosing among the alternatives "is a matter largely within the discretion of judges (Katkin, Hyman, and Kramer 1976)." Sentencing decisions, writes Gottfredson (1975), are usually "guided unsystematically" by a number of conflicting goals. This is partially a result, he further states, of the "typical lack of attention to the classification and prediction problems inherent in the sentencing process."

Most classifications used by criminal justice workers are, as mentioned above, intuitively constructed. Some are even implicit in the decision to the point that those using them are unaware of having used them. When decisions are made by the police, for instance, an implicit classification schema is almost always at work, wherein individuals are categorized according to informal cues (race, dress, etc.), seriousness of crime, context of arrest, etc. (Daudistel, Sanders, and Luckenbill 1979). Even in a particular justice system where decisions concerning disposition or treatment are based upon the use of explicit classification, other decisions (involving charges to be made, the preliminary hearing, or the sentencing itself) may remain on an intuitive level. An effective and fair system would require that the presuppositions operating behind all decisions be made explicit and consistent through the use of valid classificatory tools. This may be the only way to effectively organize and systematically accumulate the mass of data produced in each specific ase and to choose and finally evaluate amont many alternatives present in each decision.

There have been many attempts, of course, to create and use more explicitly defined classification systems within the justice system. The I-level theory, for example, has been used extensively throughout the juvenile system in California to classify offenders for treatment needs, possibility of probation, etc. A slightly more unusual way in which classification can be used is reported by Gottfredson (1975). The prosecuting attorney's office in Washington D.D., in order to deal with the sheer volume of incoming work, employs a classification system "based upon an extensive collection of objective data obtained from each case" to provide a "daily ranking of cases." All in all, however, intuitive and unsystematic use of classificatory ideas far outweighs the presence of well-defined and consciously utilized classification systems. Inconsistent decision making, then, is due not only to the lack of good classification systems, but also to the varying ideas about classification held by personnel in the criminal justice system. Inconsistency is an inevitable consequence when inexplicit categories and non-objective methods are used in decision making.

2. Case Overload. The sheer number of cases which must be handled in most parts of the country creates two crucial problems. First, the massive amount of information generated by each offender often renders decision making extremely difficult. Unless this data can be effectively ordered and conceptualized, offenders may be dealt with in an arbitrary manner. In the face of serious information overload, decisions tend to be made unsystematically, with the decision-maker utilizing only bits and pieces of the relevant data. The judge presiding over a case is naturally confronted with a great mass of

data; and while the courts keep a record of decisions made, they do not keep a record of their eventual outcomes. Thus, there is little real data about the relevance of case information to decisions for disposition (Gottfredson 1975). This serves to underscore the need for more systematic and valid classification systems, both in order to organize the data and to emphasize the most relevant information in reference to the most appropriate treatment.

The second dilemma related to case overload concerns overcrowding. If institutions are overcrowded or there is a severe lack of funds, existing facilities may not be adequate for ensuring the appropriate treatment of each type of class of offender. This point is relevant to the problem of practical constraints on good decision making, which will be examined later.

3. Skepticism. The lack of effective and valid classification schemas, as Eldefonso and Coffey (1976) point out, engenders a certain cynicism among criminal justice workers regarding the number of poorly conceived approaches to classification with which they are expected to work. As a consequence, the systems which do exist tend to be used haphazardly, if at all. For the most part, criminal justice personnel distrust classification systems (and are highly suspicious of the social sciences in general). This is not surprising, given the fact that most available classificatory schemas suffer from major deficiencies and are often difficult to implement. Social scientists and criminologists are not always aware of the needs which exist in real-life situations.

B. Methodological Problems In Available Classification Systems. Many of the classification systems now in use are plagued with methodological inadequacies which seriously call into question their validity and reliability in actual operation. A valid classification system would help to provide more reliable

guildelines and help counteract the effects of prejudice and other extraneous influences. Without this, the tenor of decisions made will fluctuate widely, giving them a decidedly arbitrary cast. "The problem of equity," states Gottfredson (1975), "implies a classification problem." To be fair and equitable, decision making must at least be consistent. One way to help ensure consistent decisions regarding offenders is to base them on a reliable and valid offender classification; a classification preferably based on observable behavior, not offender feelings or perceptions (they are, after all, "not in trouble for what they perceive but for what they do. Eldefonso and Coffey (1976). This statement emphasizes the importance of behaviorally based taxonomic systems.

The following list provides a partial delineation of the shortcomings contained in almost every criminal classification now in use. These deficiencies in taxonomies not only lead to confused and inappropriate decisions, but encourage criminal justice workers to forego the use of any explicitly constructed classification system and to rely on ad hoc intuitively derived categories. (A more detailed consideration of these problems may be found in our earlier project report, i.e. Brennan, et al. 1979)

1. Limited Predictive Validity. There is a great need in the criminal justice system for valid predictive classification schema, since many decisions rest on the ability to accurately predict recidivism, etc. However, most offender classification systems have not been designed or evaluated with regard to predictive validity, thus limiting their usefulness.

2. Lack of Explicit Guidelines for the Identification and Assignment of New Cases. Criminal justice workers find it difficult to use classification systems that provide no clear method with which offenders may be assigned to the categories they include. This confusion concerning assignation renders even the most elegant taxonomy unusable.

3 Lack of Reliability. Most offender classification systems have never been replicated, which leads to serious doubts about their reliability. Furthermore, where replication studies have been attempted, many proposed taxonomies are found to be unreliable.

4. Cases are Often Mis-Classified. Lack of clarity concerning the assignment of cases to an existing taxonomy, along with inadequate training and lack of interest on the part of criminal justice personnel, can lead to offenders being dealt with according to their placement in an inappropriate category.

5. Cases Are Often Not Classified At All. If classes in the taxonomy are ill-defined, if they do not correspond adequately to the actual domain of criminal behavior, or if it is unclear how cases are to be assigned to the categories, criminal justice workers often forego the explicit classification entirely and return to a reliance on their more or less intuitive ideas about types of offenders.

6. Current Systems Are Not Related to Actual Criminal Behavior. There is an astonishing lack of offender typologies based on a systematic description of criminal behavior. Many existing systems exhibit a confused mixture of social, psychological, and behavioral approaches.

7. Misunderstanding of the Conceptual Properties of Classification Systems. Taxonomies may be structured in many different ways (polythetic vs. monothetic, etc.), and may be constructed for many different purposes

(description, prediction, etc.). Each structure and purpose leads to a specific kind of classification system, and failure to be clear about these properties (or mixing them unsystematically) may render the resultant taxonomy inappropriate for the task at hand.

8. Oversimplification. Current offender typologies often severely reduce the complexity of criminal behavior to the extent that their relation to the phenomena being classified is slight. Usually this oversimplification stems from the use of too few variables in the construction of the system or too few categories in the system itself.

9. Biased Sampling Used in Construction. If the sample utilized in creating the classification system is somehow too strictly limited by the investigator's bias or by other unexamined contingencies, the resultant taxonomy will not contain the full range of criminal behavior.

10. Deficient Evaluation. Most of the current classifications are deficient in terms of being evaluated. Many of them have, in fact, not been evaluated at all. This provides another good reason why criminal justice personnel are hesitant to use them.

11. Misunderstanding and Inappropriate Use of New Classification Methodology. Many problems in current taxonomies may stem from a deficient understanding of statistical and taxometric methods for creating taxonomies. The new quantitative taxometric techniques are complex, many have severe limitations, or are appropriate only to specific tasks. If this is not recognized, the taxonomies created using these methods will be sub-optimal or erroneous.

C. Contextual Problems: The Justice System. Improving the methodology used in constructing classification schemas, and ensuring their relevance to the purposes for which they are designed, is only part of the necessary reform for

CONTINUED

1 OF 9

the use of classification in justice decision making. It would be useless to have valid and reliable taxonomies if they could not be effectively implemented, or if there were no way to control their use. We will briefly outline some of the problems in the criminal justice system as a whole which might impair or undermine the use of a good classification schema.

1. Homeostasis. This term refers to the tendency of any complex system to move toward stability. Any variable in a stable operating system which is changed may cause other variables to adjust and restore a state of equilibrium. Thus, in effect, nothing changes. The criminal justice system, regardless of where it is prodded, may always adjust back to its current set of working procedures.

   As a concrete example of this tendency toward operational equilibrium, let us briefly address the phenomena of procedural delay. Most observers agree that delays in the criminal justice system are deplorable and should be eliminated. Delay is seen, mistakenly, as a abberation of the system rather than an intrinsic and required part. Many solutions are proposed to handle this problem: rigid time limits for different aspects of the trial, elimination of certain laws which tend to keep the courts full (those that deal with drunkenness, pornography, etc.), more consistent refusal of bail, etc. (Wright and Fox, 1978).

   This type of solution reveals a fundamental minunderstanding of the dilemma. First, these are ad hoc measures which would be applied to one area of the system and cause more problems in another, thereby allowing delay to again proliferate. Second, delay is not necessarily an annoying condition to certain members of the justice system. Delay is seen as a fundamental, structural component of the

system and is used as such. Defense attorneys, for instance, use time delays to keep their clients on the street or to try to force a more lenient sentence by delaying the trial until interest in it is low (Rosett and Cresey, 1976). What does this imply? For one thing, it reveals that most suggestions for improving the criminal justice system (such as Nagel, 1975) may be doomed to failure because they are being posited at the wrong level. Solutions proposed at the operational level of the system might simply be absorbed into it and undermined in the process of maintaining the basic system equilibrium. Any viable improvement must be posed at a level above and outside the system in order to effectively implement true structural changes. An actor in a play cannot propose improvements in the production while performing on stage, since they would simply be seen as part of the play. He must step outside of the context in order to change it. This means that a classification system, regardless of its validity and viability, may have little effect if simply plugged into the existing system.

2. Political and Economic Constraints. The contours of this category should be obvious, but they must be mentioned. While Levin (1977) maintains that judges, for example, are not strongly affected by local politics, their decisions may certainly be much affected by particular political biases. In parts of the country where elected judges have party affiliations, the affect of political viewpoint is even more constraining and complex. However, political influence in the justice system extend far beyond personal bias. As Wright and Fox (1978) point out, criminal justice in complex societies rests upon a political base. This means, for one thing, that those with

political connections (which usually means those that are socially and economically advantaged) wield significant power in the courts and elsewhere. It also means that policy in the courts and in the prisons is dependent on the current socio-political context, e.g., classifications for release in relation to violence and recidivism.

Economic constraints are intimately related to political considerations. A conservative political climate, for example, may lead to lack of funds for liberally oriented treatment programs. Scarcity of money also leads to a severe competition within the justice system for resources: each group (police, courts, correctional institutions) will consider its own needs of the highest priority. The ways in which economic constraints can undermine the effective functioning of a good classification system are obvious. If an offender is assigned to a class which suggests that a specific disposition would be most appropriate, and the facility or program is either overcrowded or non-existent, then the point of valid and reliable classification is lost.

3. Bias. There is more to bias than a political content. Whatever its sources, it can have a very concrete effect on classificatory decision making. Levin (1977), for example, found that in Pittsburgh and Minneapolis whites consistently receive a greater percentage of probation than blacks (this in spite of the fact that more blacks are arrested). In this sense, a good classification system might help over-ride bias, since it would presumably not discriminate on the basis of color. However, if its use were not monitored, it could be misapplied or used in such a way as to conform to its user's prejudices. This would be particularly the case where the criminal justice worker's bias is against explicit classification in general.

4. Conflicting Goals. Each state and section of the criminal justice system has a different purpose. Prosecuting and defense attorneys, probation officers, judges, etc., all have different goals regarding the fate of a particular offender. This interplay of conflicting immediate purposes is set in a context of continuing debate over the general goals of the justice system (punishment, rehabilitation, isolation from society?). Each group has a certain tendency to want to influence prevailing classifications, stereotypes and decisions in accordance with their own purposes and biases. The police, for example, though they are not directly involved, can influence the process of charging merely by the way in which they present information and by what classificatory implications they choose to suggest. It must also be remembered that the police informally classify (screen) those suitable for charging anyway, so there is ample opportunity for their influence to be included in the eventual decision (Daudistel, Sanders, and Luckenbill 1979). These sorts of considerations obviously affect the tenor of decisions made, most often leading to inconsistency and confusion.

CONCLUSIONS

These general comments are not intended to be an exhaustive critique of the criminal justice system, nor are they a prelude to proposing specific reform measures; clearly that does not fall within the scope of this paper. The intention of this brief overview is to make explicit some of the less obvious constraints on the effective use of classification in its operational context. If some of the constraints which inhibit valid and equitable classification are removed, then the problem of larger system constraints on

the use of classification must be addressed. Neither of these changes will be effective without the other; decision making is dependent on good classification, but good classification may be less effective without an improved context for implementation.

Decision making in the criminal justice system requires both that the law be followed and that appropriately humane action be taken in the case of each offender. "Therefore," wrote Melanchthon in his Ethics, "an interpretation should be applied to every law that would bend it to more humane and lenient decisions." This is in agreement with Aristotle's definition of equity (epiekeia) as the correction of law, or its adjustment to fit a very transient human reality. As H. G. Gadamer (1975) states, "the law is always imperfect in itself, but, in comparison with the ordered world of law, human reality is necessarily imperfect and hence does not allow of any simple application of the former." This again implies that real cases will almost never fit the generalized legalistic classes which form the essence of the written legal code, and that there will always have to be fine adjustments between such real unique cases and each legal sub-classification. "Consequently," write Daudistel, Sanders, and Luckenbill (1979), "when members (of the justice system) are confronted with concrete cases, they approach them from the perspective of their typifications, not from the literal meanings of the statutes."

How do the mechanics of classificatory decision making deal with the complexity of human behavior in reference to the written law? As was stressed previously, at each point where a decision is to be made within the criminal justice system, classification is necessary to order the diverse kinds of information generated by an offender, and to suggest appropriate treatment, disposition, etc. Decisions made including subjective choices are never

arbitrary, but depend on the use of some sort of more or less obvious classification schema. These schemas, however, usually depend on intuitive ideas about types of people and types of crime. This subjectivity may cause decisions to vary widely across the system of criminal justice. Where explicit classifications have been utilized, they have, on the other hand, tended to be woefully inadequate, both methodologically and in terms of strategies of use. Development of valid and reliable classification systems, along with a program for their effective implementation in the larger context, seems imperative if decisions are to be of maximal value to the offender and provide maximal efficiency for the operation of criminal justice.

BIBLIOGRAPHY

Casper, J. D.
1972 American criminal justice: the defendant's perspective. Prentice-Hall, Inc.: New Jersey.

Daudistel, H. C., W. B. Sanders, and D. F. Luckenbill
1979 Criminal justice: situations and decisions. Holt, Rinehart and Winston: New York.

Eldefonso, E., and A. R. Coffey
1976 Process and impact of the juvenile justice system. Glencoe Press: Beverly Hills.

Gadamer, H. G.
1975 Truth and method. Seabury Press: New York.

Gottfredson, D. M.
1975 "Prosecution and sentencing decisions." In Decision-Making in the Criminal Justice System (Gottfredson, ed.), NIMH Center for Studies of Crime and Delinquency: Rockville, Maryland.

Gottfredson, D. M., et al.
1978 Classification for parole decision policy. National Institute of Law Enforcement and Criminal Justice: LEAA.

Katkin, D., D. Hyman, and J. Kramer
1976 Juvenile delinquency and the juvenile justice system. Duxbury Press: Massachusetts.

Kornfeld, J. P., et al.
1975 State-of-the-art of offender classification in the U.S.A. ITT Research Institute: Chicago, Ill.

Levin, M. A.
1977 Urban politics and the criminal courts. University of Chicago Press: Chicago.

Megargee, E. I., and M. J. Bohn, Jr.
1979 Classifying criminal offenders. Safe Publications: Beverly Hills.

Nagel, S. S.
1975 Improving the legal process. Lexington Books: Massachusetts.

Orlando, F. A., and J. P. Black
1975 "The juvenile court." In Issues in the classification of children (Hobbs, ed.), Jossey-Bass Publishers: San Francisco.

Phelps, T. R.
1976 Juvenile delinquency: a contemporary view. Goodyear Publishing Company, Inc.: Pacific Palisades, CA.

Pospisil, L.
1974 Anthropology of law: a comparative theory. HRAF Press: New Haven.

Rosett, A., and D. R. Cressey
1976 Justice by consent: plea bargains in the American courthouse. J. P. Lippincott Company: Philadelphia.

Viano, E.
1975 Criminal justice research. Lexington Books: Massachusetts.

Wilkins, L. T.
1975 "Perspectives on court decision-making." In Decision-Making in the Criminal Justice System (Gottfredson, ed.), NIMH Center for Studies of Crime and Delinquency: Rockville, Maryland.

Wright, B., and V. Fox
1978 Criminal justice and the social sciences. W. B. Saunders Co.: Philadelphia.

CHAPTER 5

BIBLIOGRAPHIC MATERIALS ON THE USE OF CLASSIFICATION IN CRIMINOLOGICAL RESEARCH AND ITS USE WITHIN CRIMINAL JUSTICE SYSTEMS

INTRODUCTION

As part of our research program on the development and evaluation of taxonometric methods for use within the field of criminology we have compiled an extended bibliography on classification within criminology. This bibliography will be up-dated during the course of our research program. It should be noted that this bibliography covers general crime and delinquency studies. It does not focus at all upon the larger general field of taxonomic methods. Those studies which are marked with an asterisk have included some application of taxometric methods.

We would be grateful for any response regarding classification papers in criminology which might have been overlooked for this bibliography. We are particularly interested in locating additional applied criminological studies which utilize multivariate taxometric techniques.

BIBLIOGRAPHIC MATERIALS

Adams, R. 1969a "Research note: delinquency typologies and correctional treatment." Criminologica, 7(2), 51-58.

1969b Delinquency typologies and correctional treatment; the Stonewall Jackson Project, North Carolina Board of Juvenile Correction. American Society of Criminology.

Allan, Harry, et al. 1971 "Hostile and simple sociopaths: an empirical typology." Criminology, 9(1), 27-47.

Andrew, J. M. 1974 "Delinquency, the Wechsler P W sign, and the I-Level system." Journal of Criminal Psychology 30(3), 331-335.

Argon, W. F. 1943 "A proposed functional classification of criminal behavior." Journal of Criminal Psychopathology, 687-701, April.

Argyle, M. 1961 "A new approach to the classification of delinquents, with implications for treatment." California State Board of Corrections Monograph, 2, 15-26.

Asbury, G. O. 1971 Preinstitutional criminal offender types and parole performance: a study of a typology. (University Microfilms).

Aultman, M. G. 1976 "A typological comparison of theoretical explanations of delinquency." Paper presented at 1976 meeting of the American Society of Criminology, Tucson, Arizona.

Babst, D. V., and Koval, M. 1972 "Relationship of time served to parole outcome for different classifications of burglars based on males paroled in fifty jurisdictions in 1968 and 1969." Journal of Research in Crime and Delinquency, v. 9(2), 99-116.

*Baer, D. J. 1970 "Taxonomic classification of male delinquents from autobiographical data and subsequent recidivism." Journal of Psychology, 76(1), 27-31.

Ballard, K. B., and Gottfredson, D. M. 1963 "Predictive attribute analysis and prediction of parole performance." Social Agency Effectiveness Study, California Medical Facility, Vacaville, California.

Bannan, P. E., et al. Caseload classification and supervision approaches. Pennsylvania Juvenile Court Judges Commission, Pennsylvania Department of Justice.

Beker, J., and Heyman, D. S. 1972 "A critical appraisal of the California differential treatment typology of adolescent offenders." Criminology, May.

*Asterisk denotes the use of multivariate methods

Bell, R. R. 1971 "Delinquent subcultures (female delinquents)." In Social Deviance, Dorsey Press.

Bennett, L. A. 1972 "Classification requirements in the process of social change." Paper presented at National Institute of Law Enforcement and Criminal Justice seminar.

Bersani, C. A. 1970 Crime and Delinquency (chapter on classification). New York: Macmillan.

Blackburn, R. 1971 "Personality types among abnormal homicides." British Journal of Criminology, 11(1), 14-31.

Blakenship, R. L., and Martin, G. H. 1972 "Some factors related to differential labeling of juvenile offenders." Unpublished manuscript.

Blakenship, R. L. and Singh, B. K. 1976 "Differential labeling of juveniles: 2 multivariate analyses." Criminology, 13(4), 471-490.

Blanc, M., and Susini, J. 1968 "Typology of offenders and typology of treatments." Collected Studies in Criminological Research, Vol. III, 70-128.

Block, H. A., and Flynn, F. T. 1956 Delinquency. New York: Random House.

Blos, P. 1969 "Three typical constellations in female delinquency." In Family Dynamics and Femal Sexual Delinquency, Palo Alto: Science and Behavior Books, Inc., 99-137.

Borjeson, B. 1968 "Type of treatment in relation to type of offender." Collected Studies in Criminological Research, Vol. III, Council of Europe, 173-236.

Bottoms, A. E. 1972 "Methodological aspects of classification in criminology." Paper presented to the Tenth Conference of Director of Criminological Research Institutes.

Breed, A. F. 1967 Significance of classification procedures to the field of correction. U. S. President's Commission on Law Enforcement and Administration of Justice.

Brennan, T., and Huizinga, D. 1976 "A review of typological research in delinquency." American Criminology Conference, Tucson, Arizona.

1977 Typological studies in delinquency. NIMH final report.

*Brown, L. D. 1978 "Development of parole classification system using discriminant analysis." Journal of Research in Crime and Delinquency, v. 15(1).

Brungardt, T. M. 1973 Self-reported delinquent behavior: an analysis of selected distribution and causal variables. Dissertation, University of Michigan.

Buikhuisen, W., and Jongman, R. W. 1970 "A legalistic classification of juvenile delinquents." British Journal of Criminology, 10, 109-123.

Bureau of Prisons. 1970 Differential treatment: a way to begin. Washington, D. C.

Butler, E. W., and Adams, S. W. 1966 "Typologies of delinquent girls: some alternative approaches." Social Forces, 44, 401-407.

Carlson, L. A. 1968 "The relationship of delinquent types of an industrial school setting to personality - motivation profiles." Dissertation Abstracts International, 29(5-A), 1446.

*Carlson, K. A. 1971 "A multivariate classification of reformatory inmates." Unpublished doctoral dissertation, University of Western Ontario.

1972 "Classes of adult offenders." Journal of Abnormal Psychology, 79, 84-93.

Carr-Hill, R. A. 1973 "Victims of our typologies: on the violent offender, reality or illusion?" Journal of Research in Crime and Delinquency, 75(4).

Cavan, R. S., and Ferdinand, T. N. 1975 Juvenile Delinquency (chapter on typologies of delinquency). New York: J. B. Lippincott Co.

Clinard, M. B., and Quinney, R. 1967 Criminal behavior systems: a typology. New York: Holt, Rinehart, and Winston.

Cloward, R. A., and Ohlin, L. E. 1969 "Types of delinquent subcultures." In Juvenile Delinquency, Cavan (ed.), Philadelphia: J. B. Lippincott Co., 62-83.

*Collins, H. A., Burger, G. K., and Taylor, G. A. 1976 "An empirical typology of heroin abusers." Veterans Administration Hospital, Psychology Services, St. Louis, Mo.

Cormier, B. M., Kennedy, M., Sangowicz, T., and Trottier, M. 1959 "Presentation of a basic classification for criminological work and research in criminology." Canadian Journal of Corrections, 1(4), 21-34.

Cowden, J. E., Pacht, A. R., and Bodemer, O. A. 1970 "The 16PF vs. the MCl in a group testing program of reformatory inmates." Journal of Clinical Psychology, 26(4), 510-513.

Cowden, J. E., with Pacht, A. R., and Peterson, W. 1969 "The MCl vs. the Jesness Inventory as a screening and classification instrument at a juvenile correctional institution." Journal of Clinical Psychology, 25(1), 57-60.

Cowden, J. E., with Schroeder, C. R., and Peterson, W. 1971 "The CPl vs. the 16PF at a reception center for delinquent boys." Journal of Clinical Psychology, 27(1), 109-111.

Davies, M. 1969 "Offense behavior and the classification of offenders." British Journal of Criminology, 9(1), 39-50.

Devies, R. K. 1975 "Use of the differential behavioral classification system of the juvenile offender to distinguish probation successes from probation failures." (University microfilms).

Donovan, J. 1976 "A typological study of self-reported deviance in a national sample of adolescents." Dissertation Prospectus, April 28.

1977 "A typological study of self-reported deviance in a national sample of adolescents." Ph.D. thesis, University of Colorado.

D'orban, P. T. 1976 "Child-stealing: a typology of female offenders." British Journal of Criminology, v. 16(3), 275-281.

Dorn, D. S. 1969 "A partial test of the delinquency continuum typology: contracultures and subcultures." Social Forces, 47, March.

Downe, W. N. 1968 "Delinquent boys: a treatment of classification and approach." Canadian Journal of Corrections, 10(2), 346-351.

Esman, A. H. 1955 "Diagnostic categories of delinquency." NPPA Journal, October, 113-117.

Eysenck and Eysenck. 1970 "Crime and personality: an empirical study of the three-factor theory." British Journal of Criminology, 10, 225-239.

Faust, F. L. 1973 "Delinquency labeling: its consequences and implications." Crime and Delinquency, 41-48.

Felice, M., and Offord, D. R. 1972 "Three developmental pathways to delinquency in girls." British Journal of Criminology, 12(4), 375-389.

Felthous, A. R. 1977 Approaching a comparative typology of assaultive female offenders. William Alanson White Psychiatric Foundation.

Ferdinand, T. N. 1966 Typologies of delinquency: a critical analysis. New York: Random House.

Field, E. 1967 Types of delinquency and home background; a validation study of Hewitt and Jenkins' hypothesis. Home Office Research Unit Report No. 10, London, H.M.S.O.

*Fildes, R., and Gottfredson, R. M. 1973 "Cluster analysis in a parole sample." Journal of Research in Crime and Delinquency, 2-11.

Gibbons, D. C. "Prospects and problems of delinquency typology." Sociological
1962 Inquiry, 32, 235-244.

1975 "Offender typologies: two decades later." British Journal of Criminology, 15(2), 140-156.

Gibbons, D. C., and Garrity, D. L. "Definition and analysis of certain
1962 criminal types." Journal of Criminal Law, Criminology and Police Science, March, 27-35.

Gibbons, D. C., and Garrity, D. L. "A preliminary typology of juvenile
1958 delinquents." Unpublished manuscript.

Gibbs, J. P. "Needed: analytic typologies in criminology." Southwestern
1960 Social Science Quarterly, 40, March.

Gibson, H. B. "The factorial structure of juvenile delinquency: a study of
1971 self-reported acts." British Journal of Social and Clinical Psychology, 10(1), 1-9.

Gilbert, J. "Delinquent (approved school) and non-delinquent
1972 (secondary-modern school) girls." British Journal of Criminology, 12, 325-356.

Glaser, D. "The classification of offenses and offenders." In The Handbook
1974 of Criminology, 45-83. Rand McNally College Pub. Co.: N.Y., N.Y.

Glueck and Glueck. "Varieties of delinquent types." British Journal of
1965 Criminology, 5, 236-248, 388-405.

1972 Toward a typology of juvenile offenders: implications for therapy and prevention. New York: Grune and Stratton.

Gottfredson, D. M. "Diagnosis, classification, and prediction."
1975 Decision-Making in the Criminal Justice System: Reviews and Essays, NIMH, 1-9.

Gottfredson, D. M., et al. "Association analysis in a prison sample and
1963 prediction of parole performance." Social Agency Effectiveness Study, California Medical Facility, Vacaville, California.

Gottfredson, D. M. with Ballard. "The validity of two parole prediction
1965 scales." California Medical Facility, Vacaville, California.

Gottfredson, D. M., et al. Classification for parole decision policy.
1978 National Institute of Law Enforcement and Criminal Justice, July.

Gottfredson, G. D. "Organizing crime: a classificatory scheme based on
1975 offense transitions." Journal of Criminal Justice, Vol. 3, 331-332.

Grant, M. Q. "Interaction between kinds of treatment and kinds of
1961 delinquents." Board of Corrections Monograph, no. 2, Sacramento: State of California Printing Division, Documents Section, July.

Gregory, R. J. "The actuarial description of delinquency: seven MMPI code
1972 types." Unpublished doctoral dissertation, University of Minnesota.

1974 "Replicated actuarial coordinates for three MMPI types in juvenile delinquency." Journal of Clinical Psychology, 30(3), 390-393.

Harary, F. "Merton revisited: a new classification for deviant behavior."
1966 American Sociological Review, 693-697.

Hartjen, C. A., and Gibbons, D. C. "An empirical investigation of a criminal
1969 typology." Sociology and Social Research, 54, October, 56-62.

Hathaway, S. B., and Monachesi, E. D. Analyzing and predicting juvenile
1953 delinquency with the MMPI. Minneapolis: University of Minnesota Press.

Hayner, N. S. "Characteristics of five offender types." A.S.R., 26, 96-102.
1961

Hedlund, C. S. "A validational study of a program utilizing a differential
1970 treatment of juvenile delinquency." Unpublished master's thesis, West Virginia University.

Hellervik, L. W., et al. Model assessment and classification system for men
1974 and women in correctional institutions. Personnel Decisions, Inc.

Her Majesty's Prison, Wormwood. Borstal typology study. London, England.
1971

Hess, H. "Repressive crime and criminal typologies: some neglected
1977 types." Contemporary Crises, 1(1), 91-108.

Hezel, J. D. "Some personality correlates of dimensions of delinquency."
1969 Dissertation Abstracts, 29(413), 3087.

Hippchen, L. J. (ed.) Correctional Classification and Treatment.
1975 Cincinnati, Ohio: The W. H. Anderson Company

*Hindelang, M. J. and Weis, J. "Personality and self-reported delinquency:
1972 an application of cluster analysis." Criminology, November.

1972 "The BCTry cluster and factor analysis system: personality and self-reported delinquency." Criminology, 10, 286-294.

Hood and Sparks "The classification of crimes and criminals." In Key
1970 Issues in Criminology. New York: McGraw-Hill Book Co.

Hooke, J. F. 1970 "Rating delinquent behavior." Psychological Reports, 27(1) 155-158.

Hunt, D. E., and Hardt, R. H. 1965 "Developmental stage delinquency and differential treatment." Journal of Research in Crime and Delinquency, 20-31.

Illinois State Training School for Boys. 1953 "Report on diagnostic categories." Sheridan: Illinois State Training School for Boys Treatment Committee.

Inciardi, J. A. 1970 "The adult firesetter: a typology." Criminology. August, 145-155.

IIT Research Institute. 1975 "Development of a model offender classification system." Executive summary, prepared for Department of Justice, LEAA.

Jackson Memorial Hospital. 1972 Behavior and attitude modification in a jail setting. Metropolitan Dade County

James, W. G. 1977 Diagnosis, classification, and treatment of the Santa Clara County Jail System: Summary of a study and design for change. Santa Clara County Office of the Sheriff.

Jenkins, R. L., and Boyer, A. 1968 "Types of delinquent behavior and background factors." International Journal of Social Psychology, 14, 65-78.

Jesness, C. F. 1971 "The Preston typology study." Youth Authority Quarterly, 26-38.

1974a "Sequential I-level classification manual." American Institute: California Youth Authority.

1974b Classifying juvenile offenders; the sequential I-level classification manual. Palo Alto, California: Consulting Psycholgists Press.

Jesness, D. F. 1962 "The Jesness Inventory: development and evaluation." California Youth Authority Research Report, 29.

Kinch, J. W. 1962a "Continuities in the study of delinquent types." Journal of Criminal Law, Criminology and Police Science, 53, 323-328.

1962b "Self-conceptions of types of delinquents." Sociological Inquiry, V. 32, 228-234.

Kissling, E. J. 1969 "A comparison of delinquent typologies and their relationship to age and race." Dissertation Abstracts International, 30(413), 38-70.

Kornfeld, J. P., et al. 1975a State-of-the-art of offender classification in the U. S. A. ITT Research Institute, Chicago, Ill.

1975b Development of a model offender classification system. Chicago: Illinois Institute of Technology.

Kratcoski, P. C., and Kratcoski, J. E. 1975 "Changing patterns in the delinquent activities of boys and girls: a self-reported delinquency analysis." Adolescence, 5(37), 83-91

Kraus, T. 1973 "Judicial labels as a typology of offences committed by male juveniles." British Journal of Criminology, 13(3), 269-274.

1977 "Classification of juvenile offenders and ecological factors." Australian Journal of Social Issues, 12(3), 224-231.

*Kulik, Stein, and Sarbin. 1968 "Dimensions and patterns of adolescent anti-social behavior." Journal of Consulting and Clinical Psychology, 32, 375-82.

Lambert, L. R., and Birkenmayer, A. C. 1972 Assessment of the classification system for placement of wards in training schools: factors related to classification and community adjustment. Ontario Department of Correctional Services.

Landau, S. F. 1974 "Type of homicide and pathologies among homicide offenders: some cultural profiles." In Crime and Delinquency, New York: Praeger Press.

*LeBlanc, M., and Caplan, A. 1976 "Measuring delinquency: the homogeneous-heterogeneous issue." Groupe de Recherche sur L'inadaptation Juvenile, University of Montreal.

LeBlanc, M., and Biron, L. 1978 "Psychological and sociological correlates of a behavioral typology of self-reported delinquency." Groupe de Recherche sur L'inadaptation Juvenile, University of Montreal.

Levinson, R. B. 1975 "Differential treatment and adult typology." American Correctional Association.

Lindesmith, A. R., and Dunham, H. W. 1941 "Some principles of criminal typology." Social Forces, 19.

Lopreato, J., and Alston, L. 1970 "Ideal types and idealization." American Sociology Review, 12, 68-75.

Lovegrove, S. A. 1938 "The significance of three scales identifying a delinquent orientation among youth Australian males." Aust. and New Zealand Journal of Criminology, 29, 343-392.

Mack, J. L. 1969 "Behavior ratings of recidivist and non-recidivist delinquent males." Psychological Reports, 25, 260.

Maltz, M. D. 1976 "On defining 'organized crime'-the development of a definition and a typology." Crime and Delinquency, 22(3), 338-346.

Marshall, T. F. "An investigation of the delinquency self-concept theory of
1973 Reckless and Dinitz." British Journal of Criminology, 13(3), 227-236.

McClintock, F. H., and Avison, N. H. "Ways of classifying offenses for
1964 criminological research." Paper presented at the First National Conference on Research and Teaching in Criminology.

McCaghey, et al. "Towards a typology of auto theft." Paper presented at
1976 American Criminology Society, Tucson, Arizona.

McKenna, J. J., Jr. Empirical testing of a typology of adult criminal
1972 behavior. (University Microfilms)

Meade, A. C. "The labeling approach to delinquency: state of the theory
1974 as a function of method." Social Forces, 53(1), 83-91.

Megargee, E. I. "Undercontrolled and overcontrolled personality types in
* 1966 extreme antisocial aggression." Psychology Monograph, 611(80).

* 1976 "Revision and refinement of an MMPI-based typology of youthful offenders." FCI Technical and Treatment Notes, 6(1).

* 1977a "Revision and refinement of the classificatory rules." Criminal Justice and Behavior, 4(2), 125-148.

* 1977b "New classification system for criminal offenders." Criminal Justice and Behavior, 4(2).

* 1977c "Need for a new classification system." Criminal Justice and Behavior, 4(2).

*Meyer, J., and Megargee, E. I. "Development of an MMPI-based typology of
1972 youthful offenders." Florida Correctional Institute, Res. Rep. 4(2).

Miller, M. J., McKelthan, E., and Stein, J. M. "Allegheny County jail-
? diagnostic and classification project - evaluation report." Pennsylvania Governor's Justice Commission, Harrisburg, PA.

Miller, R. C. Personality patterns among delinquent behavioral types.
1958 Washington, D. C.: Catholic University American Press.

Molof, M. J., and Jesness, C. F. Project sequil - the development of a
1972 sequential I-level classification system: yearly report, February 1971 - January 1972. American Justice Institute Sacramento, California.

Morrison, W. A. "Criminal homicide and the death penalty in Canada - time
1973 for re-assessment and new directions - toward a typology of homicide." Canadian Journal of Criminology and Corrections, 15(4), 367-396.

Mukherjee, S. K. A typological study of school status and delinquency. Ann
1971 Arbor: University of Michigan Microfilms.

National Institute of Law Enforcement and Criminal Justice. Classification
1973 of criminal behavior. National Technical Information Service.

Nettler, G. Explaining Crime. New York: McGraw-Hill Book Co.
1974

Nye, F. D., and Short, J. F. "Scaling delinquent behavior" American Sociolo-
1957 gical Review, 22, 326-31.

Oberlander, L. "Quantitative tools for criminal justice planning." Abstract
1975 of LEAA published document, crime and delinquency literature, 8(2).

Oliver, R. M. Maine state prison: model classification and tracking system.
1977 Maine State Prison.

Opp, K. D. "Problems of classification in criminology." Collected Studies
1973 in Criminological Research, Vol. X, Council of Europe.

Orlando, F., and Black, J. "Classification in juvenile court: the delinquent
1974 child and the child in need of supervision." Juvenile Justice, 25(1), 13-25.

Palmer, T., and Werner, E. Review of I-level reliability and accuracy in the
1972 California Community Treatment Project. California Department of the Youth Authority.

Panton, J. H. "MMPI profile configurations among crime classification groups."
1958 Journal of Clinical Psychology, 14, 305-308.

PMS Limited Puerto Rico - Evaluation of the classification diagnostic and
1977 treatment center: final report. Puerto Rico Crime Commission.

Podboy, J. W. Diagnosis of specific learning disabilities among a juvenile
1977 delinquent population. Sonoma County Probation Department.

*Quay, H. C. "Dimensions of personality in delinquent boys as inferred from
1964a the factor analysis of case history data." Child Development, 35, 379-484.

* 1964b "Personality dimensions in delinquent males as inferred from factor analysis of behavior ratings." Journal of Research in Crime and Delinquency, 1, 33-37.

1966 "Personality dimensions in pre-adolescent delinquent boys." Educational and Psychological Measurement, 26, 99-110.

Quay, H. C. and Parsons, L. B. The differential behavioral classifications
1971 of the juvenile offender. Federal Bureau of Prisons, Washington, D. C.

Quay, H. C. and Parsons, L. B. "The differential behavioral classifications
1971 of the juvenile offenders." U. S. Department of Justice.

1975 "Classification in the treatment of delinquency and antisocial behavior." From Issues in the classification of children (N. Hobbs, ed.), Washington: Jossey-Bass Publishers.Rada, R. T.

Rada, R. T. 1978 Classification of the rapist. New York: Grune and Stratton, Inc.

Reiss, F. 1952 "Social correlates of psychological types of delinquency." American Journal of Sociology, 17.

Robin, G. 1963 "Gang member typology: its extent, sequence and typology." Journal of Criminal Law, Criminology, and Political Science, 55, 59-69.

Robinson, C. D. 1971 "Personality variables related to types of offense committed." Australian and New Zealand Journal of Criminology, 4(3), 181-186.

Roebuck, J. 1963a "Negro drug addict as an offender type." Journal of Criminal Law, Criminology, and Police Science, 53(1), 35-43.

1963b "Negro numbers man as a criminal type: the construction and application of a typology." Journal of Criminal Law, Criminology and Police Science, 54(1), 48-60.

1966 Criminal typology. Springfield, Ill.: Charles C. Thomas.

Sarbin, T. R. 1969 "The myth of the criminal type". Monday Evening Papers #18, Center for Advanced Studies, Wesleyan University.

Schrag, C. A. 1944 "Social types in a prison community." Unpublished master's thesis, University of Washington.

1961 "A preliminary criminal typology." Pacific Sociological Review, 4, 11-16.

Schuessler, K., and Slatin, G. 1964 "The MMPI characteristics of incarcerated females." Crime and Delinquency, 10, 119-126.

Search Group Inc. 1972 "Classification requirements in the process of social change." Paper presented at National Institute of Law Enforcement and Criminal Justice Seminar.

Sellin, T., and Wolfgang, M. E. 1964 The measurement of delinquency. New York: Wiley.

Severy, L. J., Bebeau, E. C., and Elliott, D. S. ? "Dimensions of delinquency proneness, empirically derived groups, and their relation to actual delinquent and dropout behavior." Unpublished manuscript funded by University of Colorado's Graduate School Computing Center and NIMH grants.

Shafer, S. 1969 "The responsible criminal type." In Theories in Criminology, New York: Random House.

Shoham, S. and Rahar, G. 1973 "Classification of prisoners in Israel" ?

Shinohara, M., and Jenkins, R. L. 1967 "MMPI study of three types of delinquents." Journal of Clinical Psychology, 23-156-63.

Simon, F. H. ? "Statistical methods in making prediction instruments." Journal of Research in Crime and Delinquency.

Sirico, L. J. 1972 "Prisoner classification and administration decision making." Texas Law Review, 50(6), 1229-2154.

Smiley, W. C. ? "Classification and delinquency: a review."

1977 "Multivariate classification of male and female delinquent personality types." Unpublished doctoral dissertation, University of Western Ontario, London, Canada.

Smith, David E. 1974 "Relationships between the Eysench and Jesness personality inventories." British Journal of Criminology, 14(4), 376-384.

Smith, P. M., and Austin, H. R. 1974 "Socialization as related to delinquency classification." Psychological Reports, 34, 677-678.

Smyth, R. A., and Ingram, G. 1970 "Relationship between type of offender and reasons for seeking medical care in a correctional setting." Nursing Research, 9, 456-458.

Solomon, H. M. 1977 Crime and delinquency: typologies. Washington: University Press of America.

Sparks, R. F. 1968 "Types of treatment for types of offenders." Collected Studies in Criminological Research, Vol. III, Council of Europe.

*Stein, K. B., Gough, H. G., and Sarbin, T. R. 1966 "The dimensionality of the CPI socialization scale and an empirically derived typology among delinquent and non-delinquent boys." Multivariate Behavior Research, 197-208.

Stein, K. B., Sarbin, T. R., and Kulik, J. P. 1971 "Further validation of antisocial personality types." Journal of Consulting and Clinical Psychology, 36(2), 177-182.

Stein, K. B., and Rozynko, V. 1974 "Psychological and social variables personality patterns of drug abusers." International Journal of the Addictions, 9(3), 431-446.

Sullivan, C., Grant, J. D., and Grant, M. Q. 1956 "The development of interpersonal maturity: application to delinquency." Psychiatry, 20, 373-385.

*Swanson, R. M., and Mobley, M. J. 1976 "Patterns of delinquent behavior: dimensions of self-reported behaviors." Unpublished manuscript.

Taylor, R. W. 1974 Narcotic traffickers: A typology. National Institute of Law Enforcement and Criminal Justice, Washington, D. C.

Thomas, C. W., and Williams, J. S. 1974 Deterrence of crime - a reformulation of Chambliss' typology of deterrence. College of William and Mary, Williamsburg, VA.

Tiffany, T. L., Peterson, D. R., and Quay, H. C. 1961 "Types and traits in the study of juvenile delinquency." Journal of Clinical Psychology, 17, 19-24.

Tsubouchi, K., and Jenkins, R. L. 1969 "Three types of delinquents: their performance on MMPI and PCR." Journal of Clinical Psychology, 25(4), 353-358.

Wardrop, K. R. 1967 "Delinquent teenage types." British Journal of Criminology, 7(4), 371-380.

Warren, M. A., and Palmer, T. B. 1965 "Community treatment project: an evaluation of community treatment for delinquents." CTP Research Report, 6, October.

Warren, M. Q. 1966 Interpersonal maturity level classification: juvenile diagnosis and treatment of low, middle, and high maturity delinquents. Community Treatment Project: California Youth Authority.

1970 "Classification of offenders as an aid to efficient management and effective treatment." Journal of Criminal Law, Criminology, and Police Science, 62, 239-258.

1972 Classification for treatment. Washington, D. C.: LEAA.

Wenk, E. A., and Halatyn, T. V. 1974 An analysis of classification factors for young adult offenders. National Council on Crime and Delinquency: Davis, California.

Werner, E. 1975 "Psychological and ethnic correlates of interpersonal maturity among delinquents." British Journal of Criminology, Delinquency, and Deviant Social Behavior, 15(1), 51-68.

Whittier, J. L. 1966 "The empirical classification of delinquent boys." Unpublished doctoral dissertation, University of Oregon.

Wilgosh, L., and Paitaich, D. 1974 "Juvenile offenders, grouped according to type of delinquent behavior and their parents: intelligence, achievement, and family interaction." Canadian Journal of Criminology and Corrections, 22, 68-76.

Wilkins, L., and MacNaughton-Smith, P. 1964 "New prediction and classification methods in criminology." Journal of Research in Crime and Delinquency, 1, 19.

Wilkinson, Spotts, and Doyle. 1972 The predelinquent intercept project. The greater Kansas City Mental Health Foundation.

Williams, K. 1975 Criminal incidents and defendants involved in them: an empirical analysis based on a new crime classification system. Washington, D. C.: Institute for Law and Social Research.

Winch, R. F. 1947 "Heuristic and empirical typologies: a job for factor analysis." American Sociological Review 12, 68-75.

Yablonsky, L. 1961 "The classification of gangs." From The Violent Gang, New York: the Macmillan Co.

Final Report

MULTIVARIATE TAXONOMIC CLASSIFICATION FOR CRIMINAL JUSTICE RESEARCH

VOLUME 2

OCTOBER, 1980

79121
NCJRS
JUN 26 1981
ACQUISITIONS

Prepared For:
National Institute of Justice
Washington, D.C.

FINAL REPORT

MULTIVARIATE TAXONOMIC CLASSIFICATION FOR CRIMINAL JUSTICE RESEARCH

(Project No. 78-NI-AX-0065)

Volume 2: EVALUATING STATISTICAL CLASSIFICATION METHODS (MONTE CARLO STUDIES)

Principal Investigator: Tim Brennan Ph.D.*

Project Staff: Mary Downton
Randy Blair
Dick Williams Ph.D.

Behavioral Research Institute
Boulder, Colorado

October, 1980

Prepared for:

National Institute of Justice
Office of Research and Evaluation Methods
Department of Justice
Washington, D.C. 20531

*Now at Human Systems Institute, Inc., 1906 13th St., Suite 304, Boulder, Colorado, 80302.

TABLE OF CONTENTS

Volume 1: AN EVALUATIVE OVERVIEW OF CLASSIFICATION IN CRIMINAL JUSTICE

	Page
Acknowledgements	
Chapter 1. Evaluative Criteria for Criminological Classification	1
Introduction	2
Structural Aspects of Classification Systems	4
The Attribute (Measurement) Space of Classifications	5
The Vertical Dimension of Classification	7
The Horizontal Dimension of Classification	8
Pragmatic Criteria for Classification Systems	15
Conclusion	26
Bibliography	27
Chapter 2. The Uses of Classification Within Criminology and Criminal Justice Systems	
Introduction	31
Uses of Classification in the Field of Research Criminology	32
Classification for Treatment	38
Classification in the Criminal Justice System	40
Conclusion	43
Bibliography	44
Chapter 3. The Failure of Classification in Criminology: Some Results of Inadequate Classification	50
Introduction	51
The Relation Between Inadequate Classification Research and the Failure of Classification in Practical Contexts	52
Ethical Issues in Applied Criminological Classification	61
Conclusion	69
Bibliography	70
Chapter 4. A Brief Examination of Decision Making and Classification in the Criminal Justice System	
Introduction	74
Conclusions	91
Bibliography	94

Chapter 5. Bibliographic Materials on the Use of Classification in Criminological Research and Its Use Within Criminal Justice Systems	96
Introduction	97
Bibliographic Materials	98

Volume 2: EVALUATING STATISTICAL CLASSIFICATION METHODS

(MONTE CARLO STUDIES)

Chapter 6. Concepts of Similarity as Used in Criminological Studies	112
Some General Problems Regarding Similarity as Used in Prior Criminological Research	113
Explanatory Theory as a Guide to the Specification of Similarity	122
Similarity in Criminological Studies Using Quantitative Classification Techniques	125
Conclusions	128
Bibliography	130
Chapter 7. Some Theoretical Observations on Distance Coefficients Versus Correlation Coefficients for Classification Studies	135
Introduction	136
Mahalanobis D^2: Orthogonality or Implicit Weighting of Attributes	137
Euclidean Distance vs. Correlation: Some Strong Opinions	141
Resolving the Paradox: A Brief Look at the Components of Similarity	142
Some Comments on Correlation	146
Distance Versus Correlation Coefficient: The 'Points-in-Space' Model Versus Profile Shapes	152
Other Ways of Measuring "Shape"	152
Bibliography	154
Chapter 8. Empirical Comparisons of Several Similarity Coefficients	156
Purpose	157
Method	157
Conclusions	169
Bibliography	178
Chapter 9. Evaluating Classifications: A Comparison of Five Coefficients for Partition Agreement	180
Method	185
Results	191
Conclusions	206
Bibliography	208

Chapter 10. Evaluation of Several Cluster Analysis Methods Using Artifical Data Sets 209

Introduction 210
Evaluation of Approaches 223
Results - Data Set D: Clusters of Different Shape and Separation 226
Data Set E: Differently Oriented Ellipses, With Minor Cluster Overlap 256
Data Set F: Overlapping Clusters of Different Compactness 291
Data Set U: A Uniform Distribution 331
Conclusions 353
Bibliography 367

Volume 3: APPLICATIONS OF NUMERICAL CLASSIFICATION METHODS IN CRIMINOLOGY AND GENERAL CONCLUSIONS

Chapter 11. Classification Analysis for the Epidemiologic Description of Crime Patterns 369

Introduction 370
Classification Approaches in the Epidemiology of Delinquency 377
A Behavioral Classification of Juvenile Violence 382
A Behavioral Classification of Juvenile Theft 400
A Behavioral Classification of Juvenile Drug Consumption Patterns 416
Discussion 426
Bibliography 430

Chapter 12. The I-Level Classification: An Evaluation by Taxometric Methods 432

Introduction 433
Methods 438
Data 439
Results 444
The Overlap Between Natural Clusters 458
Description of the Core Clusters 474
Cluster 1 474
Cluster 2 477
Cluster 3 478
Cluster 4 479
Cluster 5 480
Cluster 6 482
Conclusions 497
Bibliography 501

Chapter 13. A Classification of Probation Youth 503

Research Sequence 513
Results 515
Evaluation of ES Clustering Variables 517
Clustering Results for the HSPQ 523
Discussion 530
Bibliography 532

Chapter 14. Classification by Personality of a Prison Sample 533

Methods of Analysis and Data 537
Results 545
Prisoner Personality Types Based on the CAQ 569
Conclusions 595
Bibliography 598

Chapter 15. Making Predictions from Empirical Classifications Using Class Membership as a Dummy Variable 599

Method 601
Criteria 602
Predicting Violence 603
Validating the Predictions 609
Conclusions 611
Bibliography 614

Chapter 16. Concluding Remarks and Recommendations 615

The Diffusion and Utilization of Numerical Classification into the Criminal Justice System 616
Bibliography 625

CHAPTER 6

CONCEPTS OF SIMILARITY AS USED IN CRIMINOLOGICAL STUDIES

"All the real knowledge which we possess depends on methods by which we distinguish the similar from the dissimilar."
Linnaeus, 1737

Similarity judgments are critical to the conceptualization and construction of all typological and classification systems across ALL disciplines. Entities grouped together into a class are regarded as being mutually 'similar' to each other in some sense, and different from other classes in the same system. The structure of any typological system is profoundly dependent upon the basis by which 'similarity' and 'dissimilarity' are assessed.

Within the field of criminology, however, the methods by which items are designated as similar rarely, if ever, come into a realm of conscious explication. The basis of similarity in most criminal typologies usually remains on an entirely intuitive level, and tends to be influenced by certain unstated a priori assumptions which emerge from the investigator's professional bias (behavioral, sociological, psychodynamic, etc.)

I. SOME GENERAL PROBLEMS REGARDING SIMILARITY AS USED IN PRIOR CRIMINOLOGICAL RESEARCH

Some criminological theorists give the impression of asking for an explicit definition of ideas about similarity as a necessary prerequisite to any good offender typology. Kinch (1962), for example, stresses that a criminological theory must specify the factors which define separate types in a classification scheme. Gibbons (1975) similarly states that in a good offender typology the characteristics used for classifying types must be made clear. Frankenstein (1970) speaks of the need for comparing "behavioral units as clearly separated from each other as possible." Rubenfeld (1967) calls for

the "classification of subjects into conceptually similar subtypes," while Hood and Sparks (1970) ask that the criteria for similarity be made clear to facilitate "easy assignment" of subjects to specified categories. None of these vague statements, however, recognize similarity as a separate and definable problem; usually it is inextricably confused with concepts of 'type'. Conceptual ideas of similarity remain on the level which Sparks (1968) thought was relevant only to 'empirical' typologies; that is, most investigators proceed

> ...simply by grouping together individuals according to their most obvious apparently relevant features, so that each group contains members which are as similar as possible to each other and as different as possible from all other groups.*

The general utilization of 'direct similarity judgments' which "appear to be based more on intuition than empirical results" (Summers and MacKay 1976) leads to what Bannister (1968) has called the "relative failure of the diagnostic category clearly to imply other decisions." Failure to explicate the manner in which similarity is assessed invariably seems to lead to a number of logical confusions in the creation of criminal typologies. One consequence of intuitive similarity judgments is that the criteria by which such judgments are made may change over time with a single subject. Thus we find an investigator like Kinch (1962) oscillating between behavioral ('offense patterns'), psychological ('self-concept'), and sociological

---

*We acknowledge here that the structural definition of the concept of a 'type' or 'class' can also profoundly affect the structure of any empirical classification system. These are usually defined by internal linkage criteria and/or boundary conditions - see, for example, Wishart (1969), Cattell and Coulter (1966) and others.

('reference groups') bases for the judgment of similarity; or Frankenstein (1970), who begins by speaking about 'behavioral units' but ends up positing similarity in terms of psychological variables.

A problem therefore in the intuitive use of similarity is that occasionally more than one concept of similarity is operative in a single typology. Cavan and Ferdinand (1975), for instance, create a typology using social contexts as a basis for similarity, but then advocate using psychological factors in conjunction with the initial typology. It also appears that when investigators criticize each other's typologies, it is often due to a misunderstanding concerning the focus of similarity judgments. When Gibbons (1975) refers to multi-factor analysis as 'explanatory porridge' and 'causal nihilism', the disagreement seems to reside in differing ideas concerning the basis of similarity.

There are criminological studies extant which have used multivariate cluster analysis to generate typologies, but few of these, surprisingly enough, state the substantive focus of similarity assessments, or what similarity measure has been chosen, and why (see section V).

Thus, the major problems concerning the concept of similarity in criminological studies seem to be:

1. Intuitive use of the concept of similarity;
2. Merging the concept of similarity with the concept of a type --- i.e., failure to recognize similarity as a distinct problem;
3. Shifting the basis of similarity from one focus to another (behavioral to psychological, etc.) without making this shift explicit (this may lead to an unsystematic use of more than one concept of similarity in a single typology);

4. Failure to realize when two opposing typologies are operating on different assumptions of similarity;
5. When multivariate cluster analytic methods are used, failure to state what similarity measure is being used, the substantive focus of similarity, and why they were chosen.

II. SUBSTANTIVE BASES OF SIMILARITY IN CRIMINOLOGICAL STUDIES

1. Behavioral Similarity

Similarity conceived in terms of actual criminal behavior could provide a strict basis for criminal behavior taxonomies. However, a problem exists in the fact that variables from other domains (psychological, social, physiological, etc.) inevitably seem to creep into the measurement system. It is almost impossible to find any good behavioral taxonomies within the criminological literature. Very few writers have focused upon behavioral similarity as a basis for taxonomy studies. Consequently (and incredibly!) the entire field of criminology is almost devoid of good criminal behavior taxonomies.

The reasons for this omission are not hard to find. Criminological taxonomists seem to be fixated on the explanatory task and uninterested in the epidemiological/descriptive task. Opp (1973) for example dismisses behavioral classifications on the grounds that such systems provide no information about why the offender committed the offense in the first instance. We see here the misguided demand that behavioral classification systems should be 'explanatory'. Opp seems to ignore the fact that classification research can have a number of different purposes, explanation being only one such purpose. Smiley (1977) more recently makes the same mistake, suggesting that descriptive behavioral taxonomies be dismissed because they are not useful for either differential diagnoses or treatment.

Aside from misguided appreciation of the use of behavioral taxonomies of crime, a number of other methodological problems have undermined studies of behavioral similarity. Some researchers have emphasized formal statistics using institutional data such as arrest records, sentencing information, and other legalistic data. A number of well known examples of this approach to behavioral similarity are available - see Buikhuisen and Jongman (1970), Roebuck (1963), Gibbons (1968), Hartjen and Gibbons (1969), and others. Each of these published studies differs in the detail of their conceptualization and assessment of similarity. However, they each emphasize behavioral similarity - primarily assessed utilizing offical data records. Issues of the reliability, validity, cross-sectional sampling of behavior, and so forth, of such official data, become extremely important in assessing the worth of these efforts. We applaud the effort of such researchers in moving towards well-defined taxonomies of criminal behavior; however, we would agree with Hood and Sparks (1970) that official statistics probably provide only an extremely thin and unreliable slice of the overall deviant behavior of any person. Gibbons himself broadens the basis of the similarity assessment by offering the concept of the criminal role career. We suggest that the assessment of self-reported behavioral similarity , across a multiplicity of possible crimes, using fairly long time intervals, would have more defensible levels of reliability and content validity than most of the approaches mentioned above. Currently, there do not seem to be any good studies of criminal behavior taxonomy making use of extended self-reported behavioral checklists,

good sampling, and defensible taxometric methodology. A large number of studies have approached the 'criminal-type' problem through factor analysis of behavioral ratings (Quay 1966, Gold 1970, and others). We adopt the position of Cattell (1965), Lorr, Jenkins and Medland (1955) and others, that factor analytic methodology is 'fundamentally inappropriate' for the discovery of taxonomic systems. If these above researchers were attempting to discover 'types of criminals' with their factor analytically orientated studies they were utilizing innappropriate methodology and were probably unaware of the earlier issues raised by Lorr, Cattell, and others, regarding the usefulness of factor analysis for the discovery of 'types'.

2. Sociological and Social Structural Bases of Similarity

The investigators in this category see similarity in terms of social contexts and social processes. Kinch (1962), for example, uses the delinquent's orientation to society and to a 'reference group'. Lindesmith and Dunham (1941) also see similarity in terms of the criminal's relation to society, as does Harary (1966) in his reformulation of Merton's typology of social deviance. Dorn (1969) takes a similar tack, concentrating on the criminal's attitude toward society, and using the Yinger-Cavan concept of sub vs. contracultural delinquency. Clark (1964), Schrag (1961), Sykes (1958), Garabedian (1964), and Cohen (1971) use various 'social role' models of similarity, while Smith and Austin explore the concept of socialization. Cavan and Ferdinand (1975), along with Yablonsky (1961), also see similarity in terms of relation to society and to a delinquent peer group.

The utilization of sociological and social structural variables as the basis for similarity assessment in criminological and deviance studies is primarily justified by the argument that the definition of deviance involves

social consensus and that deviance is inherently a social phenomenon. Social class, social roles, socialization processes, and so forth, are viewed as critical factors within any valid explanatory classification system for social deviance. A majority of sociologically orientated typological systems have been theoretical in nature, and therefore have assessed similarity only on a very few broad social structural variables (see Cloward and Ohlin 1969; Merton 1938; Harary 1966; Cohen and Short 1958; Thrasher 1963; and others). The theoretical nature of these systems has meant that most of the major issues around 'similarity' were never really addressed. The concept could essentially remain implicit and theoretical. The statistical and conceptual issues which might have come to light in its operationalization could be overlooked.

The main criticism of basing similarity only on social phenomena is the failure to provide an adequate treatment of individual differences, and an inadequate treatment of learning processes. There is the danger that certain critical differentiations will remain hidden within the typology. The resulting typology will be completely unable to make certain critical differentiations. The omission of such differentiations will result in serious errors of over-generalization, and the emerging types will logically contain an undue amount of within-type heterogeneity. This will, in general, seriously damage the predictive accuracy of any typological system. In commenting on sociologically-based classification systems of offenders Bottoms (1973), for instance, comments that "...within this sociological framework, however, I believe that psychological contributions are ignored at peril."

3. Psychological Similarity

Sullivan, Grant and Grant (1957) may be the most representative of this group, positing similarity in terms of personality 'maturity levels'

(I-level). Butler and Adams (1966) operate on the same assumptions in subjecting I-level data to Q-factor analysis (although no relation between Q-types and the I-level types were found). Frankenstein (1970) uses psychological etiology as a basis for similarity. Meyer and Megaree (1972) develop a typology based on the MMPI, while Smith (1974) compares the Eysenck and Jesness personality inventories.

Psychological orientations to the assessment of similarity have been criticized for being overly narrow in focus (primarily for the omission of social influences). Smiley (1977) argues that typological systems based on multi-trait personality inventories such as the 16PF, CPI, and others, have '... only limited value in discriminating between different levels and type of delinquent behavior.' Erickson and Roberts (1966), Smith and Austin (1974) and Hood and Sparks (1970) have also made this criticism within the criminological literature.

The primary assumption of the use of a psychological focus for criminal typological studies is that criminals are emotionally or psychologically disturbed in certain ways. Motivational patterns, intellectual variables, personality traits, beliefs, psycho-analytically based variables, and so forth, enter into this perspective. This orientation has had the strongest popular appeal, and there are innumerable theoretical and empirical taxonomies which base their similarity assessments on psychological characteristics of the criminal population. The proliferation of innumerable (mostly overlapping, non-operational, partial and untested) psychological systems has lead to confusion within this field of study. An overabundance of partial typological systems based on some unsystematic psychological hypothesis is possible. Shafer (1969), pp. 163-164, for example, writes,

> "...to catalogue all psychological typologies would be close to impossible. This is especially the case since the Freudians got into the act."

4. Social Psychological Bases of Similarity

Within this perspective, a mixture of social and psychological factors are taken as the basis for similarity. Glueck and Glueck (1965) see similarity in terms of personal etiology and social context. Kinch (1962) uses a scale of self-assigned personality attributes as well as attitudes towards society as a whole. Gibbons (1975) uses the concept of a 'role-career' which is partially socially defined and personally generated. Felice and Offord (1972) assess similarity in terms of family and larger social contexts, while Marshall (1973) uses Reckless' containment theory (i.e., scale of internal and external 'containment').

The basic problem is that all of the earlier classificatory type-criteria are only partially descriptive of the persons being classified. It has been suggested that for a much more stable, reliable, and powerful explanatory typology there should be a more comprehensive representation of the domain of etiological forces. Warren hints at this position in her 1966 paper:

> "Clearly the last word on typologies has not yet been written. Sociologists continue to accuse psychological typologies of taking insufficient cognizance of environmental factors; psychologists continue to accuse sociological typologists of having insufficient regard for intra-psychic factors. Nevertheless it is now possible to find investigators who are attempting to link theoretically the sociological, psychological and situational variables which are all relevant to a completely satisfactory taxonomy."

Ferdinand (1966) similarly endorses an approach which brings together sociological and psychological variables in the construction of a general explanatory typology of delinquents. Focusing largely on the explanation of juvenile delinquency, Ferdinand provides reviews of dozens of different typological schemes. These are classified into two large classes termed

"social typologies" and "psychological typologies". Ferdinand finally proposes that there must be a convergence of these two general classes of type-criteria if definitive explanatory typologies are to be constructed.

A number of more recent criminological studies have utilized this broadly based social psychological perspective in conjunction with taxometric approaches (see, for example, Brennan and Huizinga 1975, Donovan 1977, Brennan et al. 1978, and others).

III. EXPLANATORY THEORY AS A GUIDE TO THE SPECIFICATION OF SIMILARITY

For certain purposes (e.g. particularly prediction, model testing, and the creation of explanatory-theoretical typologies) it is clear that prior theory will provide guidelines for the specification of attributes which will enter into similarity assessment for taxonomy construction. The assessment of similarity therefore will often be directly guided by the prior theory. The theory will provide the rationale and justification for the relevant attributes of the offenders on which they are assessed as similar or dissimilar.

Theoretical justification, however, although preferable is only possible when good theories are available. If there are a large number of poorly developed, ill-tested, and partially valid theories, the usefulness of the theoretical development within a discipline will be dramatically reduced. The opposite extreme from carefully justified theoretical relevance for the attribute space for similarity assessment is often the ad hoc 'shot-gun' approach in which intuition, guesswork or the simple availability of data, govern the selection. Almost random mixtures of demographic, social, psychological and behavioral data are lumped together in the assessment of similarity (see for example the typologies of Baer 1970, Glueck and Glueck 1965, Roebuck 1967, and others.) The more theoretically orientated researchers are

horrified by this practice and a lively debate on this issue has continued for over a decade within the criminological literature. The ad hoc atheoretical approach has been dubbed as the 'multi-factor approach' (see Hood and Sparks 1970, Reuterman 1973, and others). The question immediately arises 'similarity of what?' There is no clear answer to this question when one considers the often chaotic collection of variables which some researchers wish to utilize in criminological taxonomic studies.

The main lines of the debate regarding theoretical vs. multi-factor approaches to the specification of variables for similarity assessment have been outlined by a number of researchers and we will not review this material further (see Bottoms 1973, Cohen 1957, Gibbons 1975, Reuterman 1967, and others). However, from the perspective of similarity assessment and taxonomic analysis, it seems that a number of critical issues should be mentioned.

1) The problem of irrelevant variables: The assessment of similarity and the consequent taxonomic classification may be seriously undermined by the presence of irrelevant variables. Ball (1965, 1970), Wishart (1969), and others, have demonstrated the manner in which the presence of irrelevant variables may distort the similarity assessments and blur the boundaries between classes. We suggest that the multi-factor approach is much more likely to lead to serious distortions of the taxonomic process as a result of irrelevant information than a more theoretically orientated approach - since in the former there is usually no clear knowledge of the relevancy of many of the included variables.

2) The problem of non-cumulative proliferation of research: As mentioned elsewhere, there is a serious problem in taxonomic research of integrating the findings from different studies, so that there may be some coherency in the development of the discipline. Research which is characterized by an ad hoc

mixing of concepts, variables, assumptions (which are often contradictory) and languages from multiple theoretical approaches, is often impossible to integrate. Popper's (1961) view of scientific progress as the development, clarification, and validation of a line of argument becomes almost impossible to the extent that the 'line of argument' becomes utterly blurred within the multi-factor approach. Coherency and the cumulative improvement and validation of studies would appear to require a stronger theoretical focusing of the substantive bases (the attribute space) for similarity and dissimilarity assessments in taxonomic studies of crime and delinquency.

3) The problem of infinite proliferation of taxonomies: The scope of the taxonomic task in criminology is already large and does not really require endless expansion. There is an infinity of possible combinations of sub-sets of the different domains of social, psychological, psycho-dynamic, behavioral, and physiological variables which could potentially enter into taxonomic research in this area. Careful theoretical work would be required to prevent an endless multiplication of different, partially overlapping, sets of classificatory variables. The ad hoc 'shotgun' approach to the specification of variables for similarity assessment could clearly lead to such an undisciplined expansion in the number of studies. The situation is already serious, as Ferdinand (1966) implies by referring to '...a veritable torrent of typologies...'.

In conclusion, although we would generally agree with Bottoms (1973), Hood and Sparks (1970), and Ryan (1970) regarding the sorry state of theoretical development in criminology and the social sciences in general, it seems imperative to optimally utilize available theoretical development in the further development of taxonomies for crime and delinquency.

IV. SIMILARITY IN CRIMINOLOGICAL STUDIES USING QUANTITATIVE CLASSIFICATION TECHNIQUES

The utilization of any sort of multivariate clustering method would seem to automatically imply an explicit search for the most appropriate similarity assessment. It would naturally be expected that studies involving cluster analysis would make explicit mention of the similarity coefficient chosen and offer justification for the substantive focus of this assessment. However, in most published studies the basis for an assumed similarity is not mentioned. This again suggests that the importance of similarity judgments is not recognized even where the objective classification method requires an explicit consideration of it. Occasionally, a statistical method rigidly imposes the use of a specific similarity measure (as is the case of Q-factor analysis, in which the correlation coefficient is inherent). Nonetheless, a thorough understanding of the appropriateness of various similarity measures; and their effects on the data when used with different methods, seems imperative if quantitative classification analysis is to be used with maximum effectiveness.

The following diagram specifies the method, similarity measure utilized, and substantive focus of eleven studies which use quantitative classification techniques. Also provided are any statements made by the authors which attempt to justify either the similarity assessment made or the substantive focus of the variables. As can be seen, virtually none of the studies contain reasons for the use of a particular similarity coefficient, though a few provide justification for the use of a method which invariably uses a specific measure. Of these studies, only two (Fildes and Gottfredson 1973; and Hindelang and Weis 1972) made explicit mention of the similarity coefficient utilized; in the other cases, the measure of similarity was deduced from the method used.

STUDY	METHOD	SIMILARITY	SUBSTANTIVE FOCUS	JUSTIFICATION STATEMENT
Gilbert, J. 1972 "Delinquent (approved school) and non-delinquent (secondary-modern school) girls"	Predictive attribute analysis	monothetic similarity	socio-psychological	"...study of the institution of the family should be at the core of any sociological study of deviancy." No methodological justification.
Collins, Burger and Taylor 1976 "An empirical typology of heroin abusers,"	Q-factor analysis	correlation coefficient	psychological (use of MMPI)	None
LeBlanc and Biron 1978 "Psychological and sociological correlates of a behavioral typology of self-reported delinquency."	correlation of prevalence diversity scores for one scale; discriminant function analysis performed on other scale		behavioral (one scale) and socio-psychological (other scale)	cites lack of self-report behavioral typologies; does not state why they should be desirable. No methodological justification.
Kulik, Stein, and Sarbin 1968 "Dimensions and patterns of adolescent antisocial behavior"	BCTry	Euclidean distance	behavioral	states that Q-typing in BCTry system "provides an objective means of forming a typology"
Hindelang and Weis 1972 "Personality and self-reported delinquency,"	BCTry	Euclidean distance	two scales: one behavioral and one psychological (attempt to correlate them) (test of Eysenck personality scale)	method used produces clusters which are defined by variables that are "highly interrelated, have the greatest possible generality, and are as independent of each other as possible." BCTry allows the user to "evaluate the clustering process from step to step."

STUDY	METHOD	SIMILARITY	SUBSTANTIVE FOCUS	JUSTIFICATION STATEMENT
Blackburn 1971 "Personality types among abnormal homicides."	Lorr and McNair method	Q-correlation	psychological (MMPI)	use of psychological variables necessary to understand criminal behavior
Baer 1970 "Taxonomic classification of male delinquents from autobiographical data and subsequent recidivism."	Hyvarinen (progressive deviation from a mean)		Mixture (unclear)	None
Fildes and Gottfredson 1973 "Cluster analysis in a parolee sample."	Association analysis (and Gower's method)	comparison of Cramer's statistic and the multiple correlation coefficient	Institutional (parole) record	different clustering methods produce different structures; thus methods must be tested as to their validity and effectiveness relative to purpose
Jenkins and Boyer 1968 "Types of delinquent behavior and background factors."	an agglomerative technique		behavioral and sociological	None
Wilkinson, Spotts, and Doyle 1972 "The predelinquent intercept project."	BCTry	Euclidean distance	Mixture	None
Swanson and Mobley 1976	BCTry	Euclidean distance	Behavioral	

V. CONCLUSIONS

This review of published criminological studies which utilize quantitative classificatory techniques indicates a remarkable dearth in the treatment of similarity. This is remarkable for two fundamental reasons. Firstly, the structure of any taxonomic system or classification is profoundly dependent upon the mode of assessment of similarity. Q-correlation, Euclidean distance, cosine, the dot product, and so on, have been shown to measure quite different structural aspects of similarity and to give different taxonomic structures when applied to the same data sets (Downton and Brennan 1979, Wishart 1970). This finding holds true for both artificial data sets and real criminological data. It seems, therefore, that the lack of attention to the treatment of similarity (both statistically and substantively) is a very serious oversight on the part of criminological researchers. Secondly, the trend towards the use of quantitative statistical methods in taxonomy construction is associated with the general move away from subjective, intuitive, and non-systematic methods. Methodological explicitness and objectivity are the hallmarks of this newer quantitative approach. We are clearly faced with the paradoxical realization that the apparently objective quantitative approaches to taxonomy construction actually contain a number of relatively subjective decision points. It is important that criminological research users of these methods are fully aware of, and informed about, the implications of these decisions. Needless controversy, inconsistent findings, and non-replication will continue to plague the field of criminological classification to the extent that these critical decisions continue to be fogged over and treated implicitly. At the very least such decisions (e.g. the focus of similarity, the choice of a statistical similarity coefficient or the particular 'stopping rule' used to determine the number of types in a taxonomic system) should be stated

explicitly. This would move the state of the art of criminological classification further towards the goal of objectivity and greatly enhance the possibility of good evaluation and replication in classificatory research.

BIBLIOGRAPHY

Baer, D. J.
1970 "Taxonomic classification of male delinquents from autobiographical data and subsequent recidivism." Journal of Psychiatry, 76(1), 27-31.

Ball, G. H.
1965 "Data analysis in the social sciences: what about the details?" Proceedings of the Fall Joint Computer Conference, Stanford Research Institute: Menlo Park, CA.

1970 Classification analysis: technical report. Stanford Research Institute: Menlo Park, CA.

Blackburn, R.
1971 "Personality types among abnormal homicides." British Journal of Criminology, 11(1), 14-31.

Bottoms, A. E.
1973 "Methodological aspects of classification in criminology." Collected Studies in Criminological Research, Vol. X, Council of Europe: Strasburg.

Brennan, T., and Huizinga, D.
1975 "Multivariate testing of social psychological models." Society for Multivariate Experimental Psychology: Chicago.

Buikhuisen, W., and Jongman, R. W.
1970 "A legalistic classification of juvenile delinquents." British Journal of Criminology, 10(2).

Butler, E. W., and Adams, S. W.
1966 "Typologies of delinquent girls: some alternative approaches." Social Forces, 44, 401-407.

Cattel, R. R.
1965 "Factor analysis: an introduction to essentials. II: the role of factor analysis in research." Biometrics, 2, 405-435.

Cavan, R. S., and Ferdinand, T. N.
1975 Juvenile Delinquency. J. P. Lippincott: New York.

Clark, S. M.
1964 "Similarities in components of female and male delinquency: implications for sex-role theory." In Interdisciplinary Problems in Criminology (Reckless and Newman, eds.), American Society of Criminology: Columbus, Ohio, 217-227.

Cloward, R. A., and Ohlin, L. E.
1969 "Types of delinquent subcultures." Chap. 7 in Juvenile Delinquency (Cavan, ed.), J. P. Lippincott Co.: Philadelphia.

Cohen, A. K., and Short, J. F.
1958 "Research on delinquent subcultures." Journal of Social Issues 3, 20-37.

Cohen, S.
1971 "Directions for research on adolescent group violence and vandalism." British Journal of Criminology, 11(4), 319-340.

Collins, H. A., Burger, G. K., and Taylor, G. A.
1976 "An empirical typology of heroin abusers." British Journal of Criminology, 12, 325-356.

Donovan, J. E.
1977 "A typological study of self-reported deviance in a national sample of adolescents." Ph.D. Thesis, University of Colorado: Boulder.

Dorn, D. S.
1969 "A partial test of the delinquency continuum typology." Social Forces, 47(3), 305-314.

Downton M., and Brennan, T.
1979 "An empirical comparison of nine similarity coefficients." Paper to be presented at American Classification Society Conference, April.

Erikson, R. V., and Roberts, A. H.
1966 "An MMPI comparison of two groups of institutionalized delinquents." Journal of Protective Techniques and Personality Assessment, 30(3), 163-166.

Felice, M., and Offord, D. R.
1972 "Three developmental pathways to delinquency in girls." British Journal of Criminology, 12(4), 375-389.

Ferdinand, T. N.
1966 Typologies of delinquency; a critical analysis. Random House: New York.

Fildes, R., and Gottfredson, R. M.
1973 "Cluster analysis in a parolee sample." Journal of Research on Crime and Delinquency, 11(2).

Frankenstein, C.
1970 Varieties of juvenile delinquency. Gordon and Breach: New York.

Garabedian, P. G.
1964 "Social roles in a correctional community." Journal of Criminal Law, Criminology, and Police Science, 54, 338-347.

Gibbons, D. C.
1968 Society, crime, and criminal careers. Prentice-Hall: Englewood Cliffs, N. J.

1975 "Offender typologies - two decades later." British Journal of Criminology, 15(2), 140-156.

Gilbert, J.
1972 "Delinquent (approved school) and non-delinquent (secondary-modern school) girls." British Journal of Criminology, 5, 236-248.

Glueck and Glueck
1965 "Varieties of delinquent types." British Journal of Criminology, 5, 236-248.

Gold, M.
1970 Delinquent behavior in an American city. Brooks/Cole Publishing Company: Belmont, CA.

Harary, F.
1966 "Merton revisited: a new classification for deviant behavior." Sociological Review, 31, 693-697.

Hartjen, C. A., and Gibbons, D. C.
1969 "An empirical investigation of a criminal typology." Sociology and Social Research, 54, October, 56-62.

Hindelang, M. J., and Weis, J.
1972 "Personality and self-reported delinquency: an application of cluster analysis." Criminology, November.

Hood and Sparks
1970 "The classification of crimes and criminals." In Key Issues In Criminology, McGraw-Hill: New York.

Jenkins, R. L., and Boyer, A.
1968 "Types of delinquent behavior and background factors." International Journal of Social Psychiatry, 14, 65-76.

Kinch, J. W.
1962 "Continuities in the study of delinquent types." Journal of Criminal Law, Criminology, and Police Science, 53, 323-328.

Kulik, J. A., Stein, and Sarbin.
1968 "Dimensions and patterns of adolescent antisocial behavior." Journal of Cons. and Clinical Psychology, 32, 375-382.

LeBlanc, M., and Biron, L.
1978 "Psychological and sociological correlates of a behavioral typology of self-reported delinquency." Groupe de Recherche sur L'inadaptation Juvenile: University of Montreal.

Lindesmith, A. R., and Dunham, H. W.
1941 "Some principles of criminal typology." Social Forces, 19.

Lorr, M., Jenkins, R. L., and Medland, F. F.
1955 "Direct vs. adverse factor analysis: a comparison of results." Educational and Psychological Measurements, 15, 441-449.

Marshall, T. F.
1973 "An investigation of the delinquency self-concept theory of Reckless and Dinitz." British Journal of Criminology, 13(3), 227-236.

Merton, R. K.
1938 "Social structure and anomie." American Sociology Review, 3, 672-682.

Meyer, J., and Megargee, E. I.
1972 "Development of an MMPI-based typology of youthful offenders." FCI Research Reports, 4(2), 1-24.

Opp, K. D.
1973 "Problems of classification in criminology." Collected Studies in Criminological Research, Council of Europe: Strasbourg.

Popper, K.
1961 The logic of scientific discovery. Science Editions: New York.

Quay, H. C.
1966 "Personality dimensions in pre-adolescent delinquent boys." Educational and Psychological Measurement, 26, 99-110.

Reuterman
1967 A new multiple-factor approach to delinquency and its application to types of juvenile offenders. Ph. D. Thesis, University of Colorado.

Reuterman, N. A., Love, M. J., and Fiedler, F.
1973 "A partial evaluation of Bloch and Niederhoffer's theory of gang delinquency." Criminology, 10(4), 415-425.

Roebuck, J.
1963 "The negro numbers man as a criminal type: the construction of a typology." Journal of Criminal Law, Criminology and Police Science, 54, 48-60.

1967 Criminal typology. Charles C. Thomas: Springfield, Ill.

Rubenfeld, S.
1967 Typological approaches and delinquency control. U. S. Dept. of Health, Education, and Welfare: Washington, D. C.

Ryan, A.
1970 The philosophy of the social sciences. The Macmillan Press LTD: London.

Schrag, C. C.
1958 "A preliminary criminal typology." P. Sociological Review, 4, 11-16.

Shafer, S.
1969 "The responsible criminal type." Theories in Criminology, Random House: New York.

Smith, D. E.
1974 "Relationships between the Eysenck and Jesness personality inventories." British Journal of Criminology, 14(4), 376-384.

Smith, P. M., and Austin, H. R.
1974 "Socialization as related to delinquency classification." Psychological Reports, 34, 677-678.

Sparks, R. F.
1968 "Types of treatment for types of offenders." Collected Studies in Criminological Research, Council of Europe: Strasburg.

Sullivan, C. E., Grant, M. Q., and Grant, J. D.
1957 "The development of interpersonal maturity: application to delinquency." Psychiatry, 20, 373-385.

Summers, J. O., and MacKay, D. B.
1976 "On the validity and reliability of direct similarity judgments." Journal of Market C., 13(3), 289-295.

Swanson, R. M., and Mobley, M. J.
1976 "Patterns of delinquent behavior: dimensions of self-reported behaviors." LEAA report.

Sykes, G. M.
1958 The society of captives. Princeton University Press.

Thrasher, F. M
1963 The gang: a study of 1,313 gangs in Chicago (1927). University of Chicago Press: Chicago, Ill.

Warren, M. Q.
1966 "Classification of offenders as an aid to efficient management and effective treatment." Journal of Criminal Law, Criminology, and Police Science, 62, 239-258.

Wilkinson, Spotts, and Doyle.
1972 "The predelinquent intercept project." The Greater Kansas City Mental Health Foundation.

Wishart, D.
1969 "Numerical classification method for deriving natural classes." Nature, 221, 97-98.

1970 Some problems in the theory and application of the methods of numerical taxonomy. Ph. D. thesis, St. Andrews University Dept. of Computational Sciences.

Yablonsky, L.
1961 "The classification of gangs." From The Violent Gang, The Macmillan Co.: New York.

CHAPTER 7

SOME THEORETICAL OBSERVATIONS ON DISTANCE COEFFICIENTS VERSUS CORRELATION COEFFICIENTS FOR CLASSIFICATION STUDIES

INTRODUCTION

An overwhelming variety of similarity and dissimilarity measures has been proposed, but many are minor variations of the two standard popular ones: Euclidean distance d and the correlation coefficient r_{ij}. (r_{ij} refers to the Pearson product-moment correlation coefficient used to measure the similarity between objects i and j, as opposed to its original purpose, as r_{xy}, to measure the similarity between variables x and y. This measure of correlation between objects is often called Q-correlation.)

Some similarity measures attempt to overcome specific weaknesses of d and r_{ij}. For example, r_{ij} is defective in that it is sensitive to changes in the direction of measurement of the variables. Cohen (1969), Howard and Diesenhaus (1967) and others, point out that on rating scales like extraversion-introversion or liberalism-conservatism, or on a semantic differential scale like good-bad, the direction of scoring may be totally arbitrary - either side of the scale could be scored "high" and the other "low". Yet reversing the direction of one item can alter the value of r_{ij} so much that the computed similarity is changed from a high positive value to a negative value. Thus arbitrary decisions in an experimental design may drastically affect the computed similarities and the classification systems which are based on those similarities. To overcome this defect, Cohen devised a modification which he calls r_c which has the same descriptive properties as r_{ij} but is invariant under reflection of the variables (reversal of the scales).

Cattell (1949) developed a similarity measure r_p which is a function of d^2, has the descriptive properties of d, but is expressed on a scale of -1 to +1 like a correlation coefficient. It is particularly valuable for certain purposes because significance levels for r_p have been computed (Horn, 1961). Thus it is possible to determine whether the r_p similarity between two individuals is statistically significant or not. However, there are problems in using these tables of significance when measuring the similarity of two groups (Cattell, 1969; Cattell, Coulter and Tsujioka, 1966). As a result, for clustering purposes it would have little or no advantage over Euclidean distance. In fact, Rohlf and Sokal (1965) point out that within the range of d observed in numerical taxonomic studies there is practically a linear relationship between d and r_p, therefore classifications based on these two coefficients would be nearly identical.

MAHALANOBIS D^2 : ORTHOGONALITY OR IMPLICIT WEIGHTING OF ATTRIBUTES

A difference of opinion on whether it is essential for variables to be orthogonal leads to more options in the choice of a similarity measure. Cattell, Coulter, and Tsujioka (1966) emphasize the need for variables to be orthogonal before clustering is attempted, for if they are not then some factors receive an implicit but unknown weighting. For this reason, factor scores rather than test scores are recommended in computing Cattell's r_p. Other authors (Morrison 1967; Overall 1964) caution against the accidental overweighting of some factors when intercorrelated variables are used to compute Euclidean distance, recommending Mahalanobis' generalized distance in preference.

The difference between Euclidean and Mahalanobis' distances can best be seen in matrix notation. Let the individual i be represented in m-dimensional

space by the vector $X_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{im} \end{bmatrix}$. The vector of differences between X_i and X_j on each variable may be written $X_{ij} = \begin{bmatrix} x_{i1} - x_{j1} \\ \vdots \\ x_{im} - x_{jm} \end{bmatrix}$

Then $d_{ij}^2 = \sum_{k=1}^{m} (x_{ik} - x_{jk})^2 = X'_{ij} X_{ij}$.

Mahalanobis' distance is $D_{ij}^2 = X_{ij}\, C^{-1} X_{ij}$ where C^{-1} is the inverse of the covariance matrix for the m variables. If the original raw variables are orthogonal and have been standardized then C^{-1} is a scalar multiple of the identity matrix, in which case D^2 and d^2 will be the same except for a scaling factor.

In effect, Mahalanobis D^2 transforms the individual's scores into weighted versions of the orthogonal components which would be produced by a principal components analysis, then computes d^2 on these components. This may be seen as follows. Let A be the m x m matrix containing <u>all</u> of the principal component vectors for the set of m variables, and let Λ be the diagonal matrix containing the principal component variances λ_i :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & \cdots & & \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix}$$

The covariance matrix C can be expressed as $C = A \Lambda A'$, and $AA' = I$, thus the λ_i are the eigenvalues of C. (Background for this may be found in Overall and Klett, <u>Applied Multivariate Analysis</u>, 1972.)

point out, the last components may be almost entirely due to error of measurement, so equal weighting implies that unreliable components may in effect be overweighted. This is generally not a problem if the purpose is to measure differences between relatively large groups, because such groups can be expected to show negligible differences on factors which consist largely of error. However, differences between individuals or small groups could be affected greatly by a factor that represents mainly measurement error.

For the purpose of detecting clusters within classification analysis, equal weighting of the components implies that clusters in component space will be stretched in the direction of the minor components and shrunk in the direction of the major ones. Although the number of clusters would not actually be changed, their shape and separation could change a great deal thereby altering the effectiveness of the various clustering methods.

In short, the numerical taxonomist is faced with two alternatives, neither of which is entirely satisfactory. Using Mahalanobis D^2 is likely to result in overweighting of unreliable components. But using d on the original variables leads to an implicit weighting of factors associated with those variables. Cronbach and Gleser, in the article cited above, argue for the use of d rather than Mahalanobis D^2, maintaining that d can be interpreted in a meaningful way even when it is applied to correlated variables. When the variables have been standardized each is given equal weight in the computation of d; hence a factor represented in several correlated variables will receive extra weight. In effect, d weights factors according to their representation in the original attribute space. Cronbach and Gleser assert that often this weighting occurs because the factor is considered especially important to the problem under investigation. This approach, however, requires that the original set of variables be chosen with great care!

EUCLIDEAN DISTANCE VS. CORRELATION: SOME STRONG OPINIONS

While Cattell's r_p and Mahalanobis' D^2 may offer improvements on d for some purposes, both measure the same basic aspects of resemblance between objects that d does. Similarly, although Cohen's r_c eliminates one defect of r_{ij}, it was devised to measure the same aspects of resemblance that r_{ij} does. A more critical contrast between measures of resemblance is that between d and r_{ij} themselves - the two most popular and frequently used measures.

Many researchers have compared the results of the two, using both natural and artificially generated data, and come up with strong and, as might be expected, opposing preferences as a result. Findings on artificial data sets are interesting because the "correct" clustering outcomes are known. Edelbrock (1978) compared the clustering results obtained from four different hierarchical methods, using d and r_{ij} on 10 of Blashfield's artificial data sets. He found that r_{ij} produced more accurate clusters than d with all four of the clustering methods. Wishart (1972) similarly examined both d and r_{ij} on an artificial data set using a k-means algorithm and found that while d performed well, the results obtained using r_{ij} were very unsatisfactory, leading Wishart to conclude that its "future use is not recommended". It will be seen below that r_{ij} MUST fail on the particular cluster configuration in Wishart's data set.

Studies comparing d and r_{ij} on real data sets have demonstrated that the relationship between similarity matrices produced by the two measures is definitely non-linear and that clustering results are quite different depending on which similarity measure is chosen (Rohlf and Sokal, 1965; Moore and Russell, 1967; Green and Rao 1969). Since the "correct" clustering solution in using real data is usually unknown, the results of such comparisons are often

confusing and require an understanding of both the data and the similarity measures. Sokal and Sneath (1973) cite the comparative studies on similarity coefficients by Boyce (1969) on hominoid data and Sokal and Michener (1967) on bee data, both of which conclude with a preference for r_{ij} and the group-average clustering method. These preferences are strongly based on these authors' past experience with their data groups, not on conformity with externally defined criteria.

The defects of r_{ij} as a similarity measure have been pointed out persuasively by a number of authors, notably Cronbach and Gleser (1953), Cattell (1949), and Cattell, Coulter and Tsujioka (1966). It is not difficult to construct artificial data sets in which clusters cannot be differentiated and absurd results are produced by r_{ij} (see Wishart 1970, Eades 1968). Yet paradoxically, many researchers continue to find that it produces the kind of classifications they are looking for. Sokal and Sneath (1973), referring to applications in biology, recommend r_{ij} as the most useful similarity coefficient, suggesting that it provides "the purest measure of shape of the commonly used resemblance measures". Similar views have been expressed by some researchers in psychology; e.g. Cohen (1969) in proposing his revised version.

RESOLVING THE PARADOX: A BRIEF LOOK AT THE COMPONENTS OF SIMILARITY

The disagreement described above is not as paradoxical as it may seem. After all, different research purposes require focusing on different aspects of similarity; different features of the objects under study may be important for different purposes. The features are reflected both in the choice of variables and in the choice of a similarity measure.

The measured similarity between two objects can be separated into 3 distinct parts, which Cronbach and Gleser refer to as shape, elevation, and scatter.*

Elevation (frequently called "level") of a profile is the within-profile mean; that is, the mean of all the attribute values in that profile. When comparing two profiles, the effect of a difference in elevation can be eliminated by subtracting the respective elevations from all of the attribute values in the two profiles.

Scatter (sometimes referred to as "accentuation" or "variance") of a profile is the within-profile standard deviation; i.e., the standard deviation of the attribute values in that profile. When comparing two profiles, the effect of a difference in scatter can be eliminated by dividing the attribute values in each profile by a scalar multiple of their respective within-profile standard deviations.

Shape of a profile is the configuration of attribute values that is left after elevation and scatter have been removed. (Often usage of the term is less specific, sometimes meaning that only elevation has been removed, as in the "Shape component of distance" described by Sokal and Sneath, 1973, and ascribed to Penrose, 1954.)

Sokal and Sneath use the term "size" loosely in reference to elevation, scatter, or vector length, not differentiating between them. Elevation and scatter taken together represent fairly well the intuitive idea of the physical size of an object. A possible analogy in studies of social deviance or psychopathology might be the "seriousness" of the disorder.

---

*Other ways of dividing similarity into components are quite possible; however, this particular approach is the most useful for describing the effect of the correlation coefficient.

Returning to the issue of comparing d and r_{ij}, their differences can be described in the above terms. Euclidean distance lumps shape, elevation, and scatter together in a single measure of overall dissimilarity between two objects, or profiles. If the relative "size" difference between two objects is great, then similarity in shape may be virtually unrecognizable. If the purpose of the research is something akin to grouping the mouse and the peccary together in a cluster called "rodents" then there is good reason to attempt to reduce the impact of size on the comparison. Otherwise the mouse might seem more closely related to the frog! The impact of size can be reduced by choosing variables in the form of ratios, however, to eliminate as much correlation between variables as possible (Rohlf and Sokal, 1965). For example, leg length could be expressed as a proportion of total body length.

The correlation coefficient totally eliminates elevation and scatter from the comparison of objects. A look at the equation for computing r_{ij} will demonstrate this. When m=the number of variables, the correlation between points i and j is

$$r_{ij} = \frac{\sum_{k=1}^{m} (x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{m} (x_{ik}-\bar{x}_i)^2 \sum_{k=1}^{m} (x_{jk}-\bar{x}_j)^2}}$$

subtracts elevation

divides by scatter

An interesting difference of opinion has arisen in psychology over whether size effects should be eliminated. Cohen (1969) argues that differences in

elevation should not contribute to a measure of profile similarity, saying "...unless correlations between elements are positive and non-trivial, no meaning can be attached to a difference in level. Thus, a level difference would have meaning for intelligence subtests, but not for the subscales of a comprehensive personality test (such as the 16PF)." But Cattell (1969), who in fact developed the 16PF test, attacks the correlation coefficient specifically because it compares only the shape of two profiles and ignores elevation and scatter. He deliberately devised a similarity measure based on d^2 which combines shape, elevation, and scatter, and has used it extensively to measure the similarity between two individuals on the 16PF personality factors.

A novel comparison between numerical measures of similarity and the more subjective assessments of similarity by human experts was reported by Muldoon and Ray in 1958. It may help to explain the popularity of r_{ij} among psychologists. Similarity rankings of 20 profiles based on temperament scales were obtained using 6 different numeric similarity measures and the opinions of 11 clinical psychologists. The rankings obtained from these 17 sources were correlated by the rank-order technique and factor analyzed. Eight of the 11 psychologists agreed closely with that similarity coefficient which measured only shape, eliminating elevation and scatter. Although the agreement was not perfect, shape of the profile was the main determining factor in the psychologists' assessments of similarity. It is interesting to note that the other 3 psychologists produced rankings quite different from the 8 who emphasized shape. The clinician who took the longest time to rank the profiles had a factor structure practically identical to Euclidean distance (represented in this experiment as Osgood's D).

SOME COMMENTS ON CORRELATION

If the researcher determines that shape really is the key feature of similarity for his or her experimental purpose, then it is still appropriate to be aware of the errors that may be introduced, and the clusters that may be overlooked, by using the correlation coefficient. Many of the defects of r_{ij} have been discussed in detail in the classic article by Cronbach and Gleser (1953) and will not be reiterated here. Rather, we propose to re-examine some of the mathematical properties of r_{ij}, and their application to specific examples.

It helps to ask first which objects will be called "identical" by a similarity measure. Under Euclidean distance, two profiles are identical (that is, $d = 0$) only if their measurement on every variable is the same, i.e. $x_{ik} = x_{jk}$ for $k = 1, \ldots m$. Under correlation, the value $r_{ij} = 1$ indicates "perfect similarity", i.e. that the profiles i and j are identical for the purposes of the research. We will look at the conditions under which $r_{ij} = 1$.

Let $X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im} \end{bmatrix}$ represent the vector of scores in profile i.

A <u>linear transformation</u> of a vector X is written $Y = aX + bI$ where a and b are scalar constants and I is the unit vector containing all 1's.

Suppose that the vector X_j is a linear transformation of the vector X_i. Then $X_j = aX_i + bI$ for some constants a, b, and the vector X_j

looks like this

$$X_j = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jm} \end{bmatrix} = \begin{bmatrix} a x_{i1} + b \\ a x_{i2} + b \\ \vdots \\ a x_{im} + b \end{bmatrix}.$$

Furthermore, the elevation of the vector X_j may be written

$$\bar{x}_j = \frac{1}{m}\sum_{k=1}^{m} x_{jk} = \frac{1}{m}\sum_{k=1}^{m}(a x_i + b) = \frac{a}{m}\sum_{k=1}^{m} x_{ik} + b = a\bar{x}_i + b.$$

The correlation between X_i and X_j will be

$$r_{ij} = \frac{\Sigma (x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)}{\sqrt{\Sigma (x_{ik}-\bar{x}_i)^2 \, \Sigma (x_{jk}-\bar{x}_j)^2}}$$

$$= \frac{\Sigma (x_{ik}-\bar{x}_i)(a x_{ik}+b-a\bar{x}_i-b)}{\sqrt{\Sigma (x_{ik}-\bar{x}_i)^2 \, \Sigma (a x_{ik}+b-a\bar{x}_i-b)^2}}$$

$$= \frac{a \, \Sigma (x_{ik}-\bar{x}_i)^2}{\sqrt{a^2 \left(\Sigma (x_{ik}-\bar{x}_i)^2\right)^2}}$$

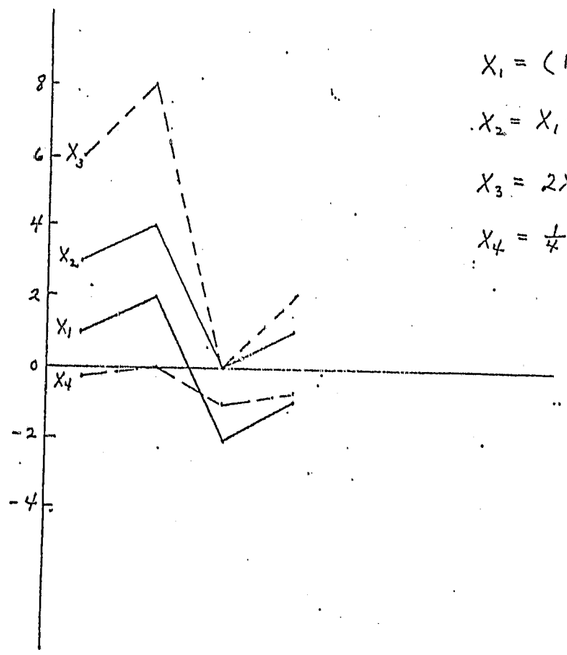
$$= \frac{a}{\sqrt{a^2}} = \begin{cases} +1 & \text{if } a > 0 \\ -1 & \text{if } a < 0 \end{cases}$$

Thus any linear transformation of the vector X such that $a > 0$ will correlate perfectly with X. Furthermore, any linear transformation of X such that $a < 0$ will have perfect negative correlation with X and hence is considered by r_{ij} to be the diametrical opposite of X.

In the equation $X_j = aX_i + bI$ $(a > 0)$, the scalar b represents the difference in elevation between profiles i and j, and the scalar a represents the relative scatter of profiles i and j.

The four profiles shown in Figure I all correlate perfectly: $r_{ij} = 1$ for each pair. They have identical "shape" as judged by the correlation coefficient. (Incidentally, X_1 and X_2 differ only in elevation; X_2 and X_3 differ only in scatter.)

FIGURE I



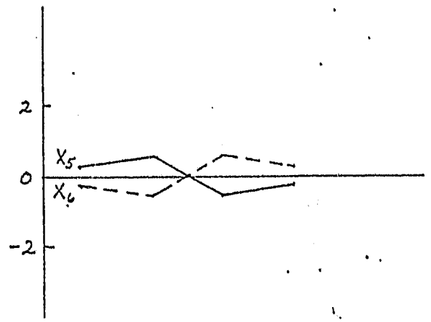
$X_1 = (1, 2, -2, -1)$

$X_2 = X_1 + 2I = (3, 4, 0, 1)$

$X_3 = 2X_1 + 4I = (6, 8, 0, 2)$

$X_4 = \frac{1}{4}X_1 - \frac{1}{2}I = (-\frac{1}{4}, 0, -1, -\frac{3}{4})$

FIGURE II



$$X_5 = \tfrac{1}{4} X_1 = (\tfrac{1}{4}, \tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{4})$$

$$X_6 = -\tfrac{1}{4} X_1 = (-\tfrac{1}{4}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{4})$$

Profile X_5 in Figure II correlates perfectly with all of the profiles in Figure I, but is diametrically opposed to profile X_6: $r_{56} = -1$. Thus although X_5 and X_6 are both small in "size" they have opposite shapes.

This would continue to hold, no matter how small their size became. That is, if $X_7 = \frac{1}{10^{10}} X$ and $X_8 = \frac{-1}{10^{10}} X$, then $r_{78} = -1$, so these two profiles are as different as they can possibly be from the perspective of r_{ij}. When the scatter of a profile becomes close to zero, it is quite possible for the error of measurement to blot out whatever slight amount of shape is present, producing a "shape" based on error alone. Assignment to clusters would clearly be meaningless for such cases, and yet they will not be detected unless there is some special test or unless the scatter is precisely zero (in which case r_{ij} cannot be computed). If such a situation is likely to arise in the data, then a provision should be included to remove profiles which are nearly flat (perhaps assigning them to a cluster of their own, or declaring them "outliers").

A practical (somewhat facetious) example will indicate that r_{ij} may be appropriate if one is primarily interested in grouping individuals by kind and not by degree. Consider the following (abbreviated) questionnaire:

How many times in the last year have you done the following?	Responses			
	Person A	Person B	Person C	Person D
1. Snuck into a movie theatre	0	0	1	3
2. Stolen a car	1	18	0	2
3. Robbed a little old lady	0	0	0	2

Conclusions based on correlations:

(1) Persons A and B correlate perfectly: $r_{AB} = 1$. We can cluster them into a group called "car thieves".

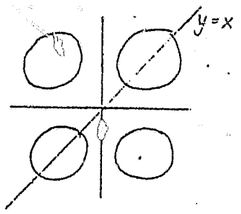
(2) $r_{AC} = -1/2$ and $r_{BC} = -1/2$. So both person A and person B must be quite different from person C, and to the identical extent.

(3) $r_{CD} = 1$, so person C is exactly like person D! (We admit that in practice, with a large number of variables, it would be an extremely unusual coincidence to get a match like this, however.)

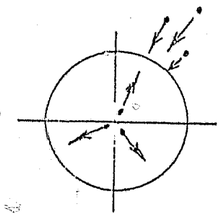
It is apparent that r_{ij} will separate the "specialists" like person B from the "generalists" like person D. It will also separate each of the "specialists" into an appropriate category, perhaps creating groups of specialties that tend to go together. Such groupings could be useful in classifying crimes, but leave much to be desired where classification of individuals is concerned. (It should be noted that standardizing the variables in the above example would considerably alter the shapes of the profiles and allow r_{ij} to recognize some difference between persons C and D. It is generally agreed that variables should always be standardized before using r_{ij}.)

There are other pitfalls that the user of r_{ij} should be on guard against. Wishart's artificial data set, referred to above, is an extreme example. His data set, depicted in Figure III, involved only two variables. It can easily be shown that if m = 2 then r_{ij} has only two possible values: $r_{ij} = \pm 1$. Therefore only two distinct clusters can be found. Furthermore, they will always partition the space along the line y=x (barring minor deviations caused by the clustering method).*

The effect is less blatantly obvious when m > 2, but it still exists. Cronbach and Glaser argue that the removal of elevation and scatter effectively reduces the dimensionality of the space by 2. The effect on points near the origin when scatter is removed is particularly important. All points are projected onto an (m-1)-dimensional hypersphere of radius 1, as depicted in Figure IV. Therefore, a cluster centered at the origin could not be detected using r_{ij}. If there is a cluster at the origin representing a random sample from some population then there will be small random differences between the points, which could take on all possible profile shapes. The points will be separated by r_{ij} according to these infinitesimal random shapes.



m=2
Figure III



m > 2
Figure IV

*The partition of a 2-dimensional space into two clusters reflects the fact that if m=2 the profile diagram can have only two possible "shapes" (not counting the completely flat profile):

1. slanting up
2. slanting down

This can be taken one step further, for when elevation is removed clusters having a mean of the form $\bar{X} = bI = (b, b, \cdots, b)$ will all be projected together into a single cluster with center at the origin. These clusters also will be broken up according to random profile shapes.

Any cluster of nearly flat profiles, then, will be split up by r_{ij} on the basis of random deviations and measurement errors. If such clusters are likely to occur in the data, we suggest that they should be removed by preliminary processing.

DISTANCE VERSUS CORRELATION COEFFICIENT: THE 'POINTS-IN-SPACE' MODEL VERSUS PROFILE SHAPES

Clusters which appear very compact and well-separated, both to the eye and to the Euclidean distance measure, may not appear so at all when compared by the correlation coefficient. Likewise, r_{ij} may clearly find two clusters on the basis of profile shape which would look like one cluster if the data could be viewed as points in space.

These two similarity coefficients offer two distinct models. Euclidean distance d represents a "position in space" model; the correlation coefficient r_{ij} represents a "shape of profile" model. The model to be used must be based on the purpose of the research.

OTHER WAYS OF MEASURING "SHAPE"

Although the correlation coefficient is the most commonly used measure of space, others have been used with some success. Two, in particular, may be useful when a "shape of profile" model of similarity is needed.

1. Shape component of distance (Penrose, 1954)

$$C_\rho^2 = \frac{1}{m}\sum_{k=1}^{m}(x_{ik}-x_{jk})^2 - \frac{1}{m^2}\left[\sum_{k=1}^{m}(x_{ik}-x_{jk})\right]^2$$

$$= \frac{1}{m}\sum_{k=1}^{m}\left[(x_{ik}-\bar{x}_i)-(x_{jk}-\bar{x}_j)\right]^2$$

(subtracts elevation from each point)

This is equivalent to 1/m times the Euclidean distance between points i and j after elevation has been subtracted from each of the points.

2. Cosine, or normalized vector product (Sokal and Sneath, 1973; Boyce, 1965, 1969)

$$\cos\theta_{ij} = \frac{\sum_{k=1}^{m} x_{ik}x_{jk}}{\sqrt{\left(\sum_{k=1}^{m} x_{ik}^2\right)\left(\sum_{k=1}^{m} x_{jk}^2\right)}}$$

(divides by vector length)

This is similar to the correlation coefficient r_{ij} except that elevation is not subtracted from the points. θ_{ij} is the angle between the vectors from the origin to points i and j. Profiles i and j have "similar shape", as measured by the cosine, if their values on all of the variables are in nearly the same proportion. "Size" of a profile is here represented by the vector length, and the "shape" is what is left after vector length has been removed.

BIBLIOGRAPHY

Boyce, A.J. 1965 — The methods of quantitative taxonomy with special reference to functional analysis. Doctoral dissertation. Oxford University.

Boyce, A.J. 1969 — "Mapping diversity: A comparative study of some numerical methods." In Numerical taxonomy. W.H. Freeman and Co.: San Francisco.

Cattell, R.B. 1949 — "R_P and other coefficients of pattern similarity." Psychometrika, 14, 279-98.

Cattell, R.B. 1969 — "The profile similarity coefficient in vocational guidance and diagnostic classification." British Journal of Educational Psychology, 39, 131-42.

Cattell, R.B., M.A. Coulter, and B. Tsujioka 1966 — "The taxonometric recognition of types and functional emergents." In Handbook of multivariate experimental psychology. Rand McNally: Chicago.

Cohen, J. 1969 — "R_c: A profile similarity coefficient invariant over variable reflection." Psychological Bulletin, 71, 281-84.

Cronbach, L.J., and G.C. Gleser 1953 — "Assessing similarity between profiles." Psychological Bulletin, 50(6), 456-573.

Eades, D.C. 1965 — "The inappropriateness of the correlation coefficient as a measure of taxonomic resemblance." Systematic Zoology, 14, 98-100.

Edelbrock, C. 1978 — "Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody." National Institute of Mental Health, American Classification Society Conference.

Green, P.E., and V.R. Rao 1969 — "A note on proximity measures and cluster analysis." Journal of Marketing Research, 6, 359-64.

Horn, J.L. 1961 — "Significance tests for use with r_p and related profile statistics." Educational and Psychological Measurement, 21, 363-70.

Howard, K.I., and H.I. Diesenhaus
1967 "Direction of measurement and profile similarity." Multivariate Behavioral Research, 2, 225-37.

Moore, A.W., and J.S. Russell
1967 "Comparison of coefficients and grouping procedures in numerical analysis of soil trace element data." Geoderma, 1, 139-58.

Morrison, D.G.
1967 "Measurement problems in cluster analysis." Management Science, 13(B), 755-80.

Muldoon, J.F., and O.S. Ray
1958 "A comparison of patterns similarity as measured by six statistical techniques and eleven clinicians." Educational and Psychological Measurement, 18, 775-781.

Overall, J.E.
1964 "Note on multivariate methods for profile analysis." Psychological Bulletin, 61, 195-98.

Overall, J.E., and C.J. Klett
1972 Applied multivariate analysis. McGraw-Hill Book Co.: New York.

Penrose, L.S.
1954 "Distance, size, and shape." Annual Eugenics, 18, 337-43.

Rohlf, F.J., and R.R. Sokal
1965 "Coefficients of correlation and distance in numerical taxonomy." University of Kansas Science Bulletin, 45, 3-27.

Sokal, R.R., and C.D. Michener
1967 "The effects of different numerical techniques on the phenetic classification of bees of the complex." (Megachilidae) Proceedings, Linn. Soc. Lond., 178, 59-74.

Sokal, R.R., and P.H.A. Sneath
1963 Principles of numerical taxonomy. Freeman: London.

Wishart, D.
1970 Some problems in the theory and application of the methods of numerical taxonomy. Doctoral dissertation. St. Andrews University: St. Andrews, Scotland.

Wishart, D.
1972 "A general tripartite clustering method and similarity generating function." Statistics 1 Division Report No. R-31. Civil Service Department: Whitehall, London.

CHAPTER 8

EMPIRICAL COMPARISONS OF SEVERAL SIMILARITY COEFFICIENTS

PURPOSE

Some measure of resemblance is essential in every classification method. This fact can be overlooked if a resemblance measure is built into the clustering algorithm so that no choice is offered to the user. For example, some iterative (k-means type) techniques treat the objects to be clustered as points in m-dimensional Euclidean space and attempt to minimize the Euclidean distance d from the points of a cluster to its centroid--d is implicitly used as the similarity coefficient.

Many coefficients besides Euclidean distance have been tried, however, and the choice of similarity measure can profoundly affect the clusters which result. In fact, some comparisons have indicated that the choice of similarity coefficient makes more difference than the choice of clustering method in the clusters found (Edelbrock 1978; Moore and Russell 1967). It could be argued, as Sokal and Sneath (1973) point out, that

> "Different coefficients should lead to different taxonomic structures, for it can be shown that different coefficients estimate different aspects of the taxonomic relationship."

Therefore, before comparing clustering algorithms, we propose to look at the resemblance measures upon which they are based.

METHOD

Using three small data sets with quite different characteristics, we have compared the similarity matrices produced by ten similarity coefficients. The Pearson product-moment correlation between each pair of similarity matrices was computed in a search for similarity measures which are linearly related, or nearly so, and can therefore be expected to produce clusterings which would

be very nearly the same. Additionally, since some cluster algorithms (e.g. single and complete link) deal only with the rank-order and not the magnitude of similarities, the Spearman rank-order correlation between pairs of similarity matrices was computed. If two similarity coefficients have a rank-order correlation of ± 1.0 then, for example, a complete-link clustering based on them would be identical. One might expect, although it remains to be tested, that a high rank-order correlation (above .95, say) between similarity matrices would lead to similar clusters using other hierarchical clustering techniques as well.

All ten of the similarity coefficients were used on standardized variables. In addition, to see the effect of standardization upon similarity, we have included in the comparison similarity matrices for six of the coefficients computed on unstandardized data.

Similarity Coefficients

Criminological data is primarily ordinal or interval-level; therefore, similarity measures intended exclusively for binary or qualitative data have not been considered in this study. Most of the coefficients selected for comparison are intended for quantitative variables only. However, for cases in which a few qualitative variables must be combined with the quantitative ones, a general coefficient which provides for both types may be useful. Therefore we have included Gower's general similarity coefficient in our comparison (although only the part of it which deals with quantitative data could be compared with our other coefficients).

The coefficients indicated below were used. In the equations, m=number of variables or characteristics, x_{ik} = value of object i on variable k. All sums are from k = 1 to m. References to Sokal and Sneath are from their 1973 text.

Standardized data

S1. Euclidean distance squared: $d_{ij}^2 = \sum (x_{ik} - x_{jk})^2$

(Sokal and Sneath, p. 124)

S2. Q-correlation: $r_{ij} = \dfrac{\sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum (x_{ik} - \bar{x}_i)^2 \sum (x_{jk} - \bar{x}_j)^2}}$

(Sokal and Sneath, pp. 137-140)

S3. Cattell's r_p: $r_{p(ij)} = \dfrac{2\chi_m^2 - d_{ij}^2}{2\chi_m^2 + d_{ij}^2}$

Where d_{ij}^2 is the squared Euclidean distance between points i and j, and χ_m^2 is the median Chi-square value for m degrees of freedom.

(Cattell, 1949 and 1969; Sokal and Sneath, p. 125)

S5. Dot Product: $S_{D(ij)} = \frac{1}{m} \sum x_{ik} x_{jk}$

(Orloci, 1966)

S6. Cosine, or normalized vector product: $\cos\theta_{ij} = \dfrac{\sum x_{ik} x_{jk}}{\sqrt{(\sum x_{ik}^2)(\sum x_{jk}^2)}}$

(Sokal and Sneath, p. 172)

S8. Shape component of distance: $C_{p(ij)}^2 = \frac{1}{m} \sum (x_{ik} - \bar{x}_i)^2 - \left[\frac{1}{m} \sum (x_{ik} - x_{jk})\right]^2$

(Sokal and Sneath, p. 170) $= \frac{1}{m} \sum \left[(x_{ik} - \bar{x}_i) - (x_{jk} - \bar{x}_j)\right]^2$

S9. Dispersion, or Q-covariance: $S_{cov(ij)} = \frac{1}{m} \sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$

(Orloci, 1966)

S12. Gower's general coefficient (computed for quantitative data only in this experiment): $S_{G(ij)} = \frac{1}{m} \sum \left(1 - \dfrac{|x_{ik} - x_{jk}|}{\text{Range}(k)}\right)$

where range (k) = range of variable k

(Gower, 1971; Sokal and Sneath, pp. 135-136)

S13. City block metric: $d_{CB(ij)} = \sum |x_{ik} - x_{jk}|$

(Sokal and Sneath, p. 125)

S15. Euclidean distance: $d_{ij} = \sqrt{\sum (x_{ik} - x_{jk})^2}$

(Sokal and Sneath, p. 124)

Unstandardized data

U1. Euclidean distance squared

U2. Q-correlation

U8. Shape component of distance

U12. Gower's general coefficient

U13. City block metric

U15. Euclidean distance

Coefficients #2, 5, 6, and 9 are "angular" measures based on the scalar product (or dot product) of two vectors:

$$X_i \cdot X_j = |X_i||X_j| \cos\theta$$

where θ is the angle between the two vectors at the origin. The above equation expressed in terms of the components of the vectors is

$$\sum_{k=1}^{m} x_{ik} x_{jk} = \sqrt{(\sum x_{ik}^2)(\sum x_{jk}^2)} \cdot \cos\theta$$

These coefficients can be strongly affected by a change in the origin. They are based on the cosine of the angle between two vectors from the origin, and of course the location of the origin determines that angle. Generally, it

has been recommended that data be centered at the origin before using such coefficients, either by standardizing the variables (Rohlf and Sokal; Sokal and Sneath) or by subtracting the mean for each variable (Orloci, 1966).

Several pairs of coefficients differ only in whether elevation (within-profile mean) is subtracted from the points. The shape component of distance is the same as squared Euclidean distance computed after the elevation has been removed from each point (S8 and S1; U8 and U1). Similarly, Q-covariance is the same as dot product with elevation subtracted (S9 and S5) and Q-correlation is the same as cosine with elevation subtracted (S2 and S6). Hence we can expect that these pairs of similarity matrices would be highly correlated whenever the points of a data set have elevations near zero. In practice this may happen rather frequently when the number of variables is large and data has been standardized since high values in one variable may be balanced by low values in others. Rohlf and Sokal (1965) checked this in their data on species of bees and mosquitoes and found that the observed within-profile means, after variables had been standardized, were very close to zero.

Data Sets

Three data sets were used, two of them artifically generated, the third randomly selected from a nationwide study of self-reported delinquent behavior among adolescents. The data sets were kept to a size of 60 points or less to conserve computer time and space.

Data set M consists of 55 points measured on 4 uncorrelated variables. These were randomly generated in the form of two multivariate normal clusters, one more compact than the other, plus a third group of uniformly distributed points to simulate "noise" in the data. The two clusters were centered at (0, 0, 0, 0) and (-3, 3, -3, 3), and the uniformly distributed noise was centered

at (-2, 2, -2, 2). Hence the average elevation for points was relatively low both before and after standardization. The ranges of the variables did not differ much, the smallest being 9.02, the largest 11.46. Standardization therefore would not be expected to have much effect on the configuration of points.

Data set N consists of 60 points measured on 4 uncorrelated variables. These also were randomly generated in the form of two multivariate normal clusters. In this set, however, the clusters were centered at (2, 20, 20, 50) and (5, 40, 20, 200) so that many points would have large elevations both before and after standardization. Furthermore, the ranges of the variables differed greatly, from 6.5 for variable 1 to a range of 496.7 for variable 4. Hence standardization could be expected to change the measurement scale drastically and make a great deal of difference in computed distances between points.

Data set R, the only "real" data in this experiment, contains information on 60 teenagers drawn at random from a nationwide probability sample of American families which was collected in 1977.* The variables are self-reported levels of participation in 38 delinquent behaviors measured on a scale of 1 to 9, where 1 = never and 9 = more than once a day. Like much data on deviance, the distribution of the variables is highly skewed. Ranges of the variables differ from 1 to 8. Additionally, a number of completely flat profiles (all responses = 1) occur in the sample.

---

*In research conducted by D. S. Elliott, Behavioral Research Institute, Boulder, CO, under grant #5 R 01 MH27552-1 from the National Institute of Mental Health.

RESULTS

Correlations between the similarity matrices for each of the data sets are shown in Tables 1-6. Pearson correlations $|r_{xy}| \geq .95$ have been underlined, for at that level $r_{xy}^2 > .90$, implying that over 90% of the variance in one similarity coefficient corresponds to variance in the other. With such correspondence, any clustering algorithm might be expected to give very similar results for either similarity coefficient and hence the choice of clustering method becomes much more important than the choice of coefficient. In a later development of this work we plan to examine whether this assumption holds true, by subjecting selected matrices to cluster analysis

Perfectly correlated similarity coefficients.

The few perfect correlations (± 1.0) that were found can be shown to be true for all data sets, not just the ones tested here. A linear correlation of $r_{xy} = 1.0$ was found between Gower's coefficient for standardized and unstandardized data. It can easily be shown that Gower's coefficient is algebraically identical for standardized and unstandardized data; in a sense, division by the range takes the place of standardization. Thus standardizing the variables has no effect on Gower's coefficient and hereafter we will not specify whether data has been standardized in reference to this coefficient.

Perfect rank-order correlation ($r_s = \pm 1.0$) was found between Cattell's r_p and Euclidean distance (both d and d^2). This means these coefficients are jointly monotonic and would produce identical cluster results when used with a clustering algorithm that considers only the order of similarities, such as single or complete link. Of course, a clustering algorithm which uses the similarities in arithmetic computations may not give the same results unless there is a strong linear relationship as well. Looking at the

Pearson (linear) correlation between r_p and d we find it to be greater than .95 for all three data sets. That this relationship will hold for all data sets can be verified by plotting r_p as a function of d. Within the expected range of standardized data, r_p is nearly a linear function of d. So we can expect clustering results to be quite similar for these two similarity coefficients for any clustering algorithm.

Other highly correlated coefficients.

Other relationships are more dependent on the data set being used. Data set M produced many highly correlated similarity matrices. In particular the close relationships between Q-correlation and cosine (S2 and S6), dot product and Q-covariance (S5 and S9), and d^2 and shape component of distance (S1 and S8; U1 and U8) can be attributed to the relatively low elevation of points in this data set. These relationships disappear in data set N where the points have high elevations, except for the surprisingly high correlation between U1 and U8. We are unable to explain the latter relationship theoretically and suspect it reflects a quirk in this particular data set since a somewhat lower correlation between U1 and U8 was found in data set R where the elevation of the unstandardized data is more moderate.

Several other pairs of similarity matrices are highly correlated for data set M but not for data set N: S1 and U1, S8 and U8, and S13 and U13. These all demonstrate the effect that standardization of variables can have on the computed similarities, since each pair involves a single distance coefficient (Euclidean, Shape Component, and City Block, respectively) measured on standardized and unstandardized versions of the data set. In set M where variances were similar to begin with, standardization made little difference.

But in set N where variances differed by a factor of 80, standardization made a great difference in the computed distances between points, thus some clustering algorithms could be expected to produce profoundly different results depending upon whether the data is standardized or not.

Some relationships are strongly evident in all three data sets, even though the correlations are not perfect. The city block metric correlates highly ($r_{xy} > .96$) with Euclidean distance for all three data sets, when standardization or non-standardization is held constant. Furthermore, Gower's coefficient correlates very highly ($r_{xy} > .98$) with the city block metric used on standardized data for all of the data sets. Indeed it now appears that for standardized data, 3 of our similarity coefficients are so closely related to unsquared Euclidean distance as to be (nearly) interchangeable: Cattell's r_p, Gower's coefficient, and the city block metric.

Coefficients which are distinctly different.

The only coefficient based on distance which sometimes appears to measure something genuinely different from Euclidean is the shape component (coefficient #8). The only occasion to prefer it over Euclidean distance would be when the elevation of some points is large and there are theoretical reasons for removing elevation from the study (for example, a desire to remove "overall size" effects from the variables).

The angular coefficients show much less consistent results than do those based on distance. It is clear that they are not linearly or monotonically related to the distance measures, or even to each other except in certain special cases. However, Q-correlation and cosine (S2 and S6) will be very much alike when the elevation of the points is low. This can be seen in both data sets M and R.

The dot product (S5) and Q-covariance (S9) have seldom been used in applied cluster analytic studies and are difficult to interpret geometrically. They were proposed by Orlóci (1966) as an improvement on the correlation coefficient when using Q-type principal components analysis as an ordination technique. They were included in the present study to see whether they would show any relationship to Q-correlation and cosine, but since they did not we will not pursue them further.

Standardization of variables in relation to similarity assessment.

The primary effects of standardization are to:

a) move the origin to the centroid of the data set, and

b) change the scales of measurement to give all variables the same variance.

Influences of transformation of the origin. This will make no difference in the various coefficients which are based on distances -- all are invariant under linear transformation. The angular measures, on the other hand, can be greatly affected. In particular, the Q-correlation and cosine coefficients cannot detect a cluster which is centered at the origin because, in effect, they project all points onto a unit hypersphere centered at the origin, losing all information about proximity to the origin. So if the raw data contains a cluster surrounding the initial origin, then standardization may improve the situation by moving the origin away from the cluster. This is the case in our data set M where the difference between standardized and unstandardized versions of Q-correlation (S2 and U2) is obvious - their linear correlation is only $r_{xy} = .447$, their rank-order correlation only .471. Conversely, if the raw data contains a cluster surrounding the grand centroid, certainly a possible event in social science data, then standardization would make it impossible to detect that cluster in one piece.

A particular difficulty in using Q-correlation shows up in our real data set (R). This coefficient is undefined for completely flat profiles -- cases which have the same score on every variable. While unstandardized measurements are often based on an absolute scale, standardized scores are measured relative to other members of the data set and are much less likely to produce flat profiles. Of the 60 teenagers in set R, 9 answered "never" to all of the delinquent behaviors, thus had profiles containing all 1's. For unstandardized variables, then, it was impossible to compute similarities for these 9 cases. But once the variables were standardized the profiles were no longer flat, hence similarities could be computed.

Influence of changes in scale of measurement. This affects both the shape and separation of clusters. We have seen in data set N that the change in scale produced by standardization can strongly affect the similarity coefficients based on distance. Similar effects can be seen, to a lesser extent, in our real data set (R). There all variables have a potential score from 1 to 9 but some in actuality only vary from 1 to 2. Thus, after standardization, an extreme value of 8 or 9 in one variable will receive the same weight in the distance measures as a value of 2 in another variable. Change of scale will alter the angular similarity measures as well. It will be worthwhile to consider how these changes may affect the clustering process.

After standardization, clusters viewed as masses of points in space will be stretched in some dimensions and compressed in other dimensions. Thus a spherical cluster may be stretched into a cigar-shape or compressed into a pancake. Likewise, a cigar-shaped cluster might be squashed into something more closely resembling a sphere. Such changes in the shapes of the clusters may profoundly affect the performance of some clustering methods. For

example, certain clustering methods are strongly biased toward spherical clusters, e.g. the iterative methods which strive to minimize trace W, or Ward's minimum variance method. Wishart (1969) criticizes the minimum variance technique for this reason, pointing out that the introduction of an irrelevant variable will stretch spheres into ellipsoids when data is standardized (see Figure I); yet it may not be possible to determine in advance which variables are relevant. An example of the difficulty one clustering method has in correctly identifying parallel ellipses is provided by Everitt



Before standardization After standardization

FIGURE I

(1974). He found that a k-means algorithm minimizing trace W partitioned two parallel elliptical clusters "crosswise" into two clusters that were more nearly circular - each containing about half of each of the original clusters. In the comparison of clustering methods in later chapters, we will begin to assess the ability of various clustering methods to properly detect ellipsoidal clusters.

The "stretching" effect may not be extreme enough in the real data set R to have a significant impact on clustering, however, for our experiment shows linear correlations above .9 between standardized and unstandardized versions for both d^2 and city block. This can be contrasted with the much more extreme differences in range in data set N, where correlation between standardized and unstandardized versions of d^2 was only .54.

Certainly resemblance measures and the resulting clusters should be independent of the units of measurement of the variables. Most researchers recommend standardizing as a means to achieve this (e.g. Cattell, 1949; Rohlf and Sokal, 1965). Sokal and Sneath (1973) point out that no profile element should be allowed to assume excessive weight thereby dominating the resemblance measure. They suggest various scaling methods to bring the variables into a reasonable balance and recommend standardizing quantitative variables so that the variance in each is given equal weight in the resemblance measurement. But they point out some potential problems, notably that a variable with extremely small variance may be reflecting mostly measurement error which would then be given equal weight with the genuine differences expressed by other variables.

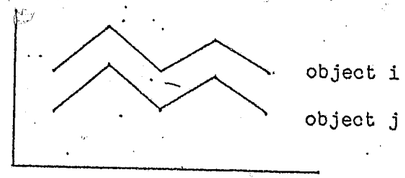
CONCUSIONS

Throughout numerical taxonomy, Euclidean distance d and Q-correlation r_{ij} are by far the most commonly used resemblance measures for continuous quantitative data. Thus any comparison of clustering methods should include those two coefficients. Our experiment suggests that there would be no advantage in using Cattell's r_p or the city block metric because they measure virtually the same features of resemblance as d. Gower's coefficient does not differ greatly from d either. It suffers as well from the fact that division by the range takes the place of standardization and the range is much more subject to measurement error, being based on only two points, than is the standard deviation. Yet Gower's coefficient offers a means for combining qualitative and quantitative variables in one analysis. For this reason we retain it for further study.

The two remaining coefficients, cosine and shape component of distance, do appear to measure somewhat different aspects of resemblance than r_{ij} and d.

Q-correlation subtracts the elevation from the points and cosine does not; the shape component subtracts elevation from the points while Euclidean distance does not. Thus in both cases differences will only show up in data sets where elevation of points is considerably different from zero, as in data set N. This becomes less likely when there is a large number of variables and data is standardized.

Both the cosine and the shape component have been proposed as measures of similarity between "shapes of profiles". The shape component takes on a value of zero, indicating identical "shape" only if the difference between two objects is constant for all of the variables, as in Figure II.



Object i and j have the same "shape component" but different elevations.

FIGURE II

However, researchers interested in comparing shapes are more likely to want to measure similarity in proportions, rather than a constant difference. Rohlf and Sokal (1965) and Boyce (1964, 1965) both found that Q-correlation was the better measure of their concept of similarity in shape.

The cosine, on the other hand, takes a value of 1 indicating perfect similarity when two objects have exactly the same proportions. According to Sokal and Sneath, Boyce (1965, 1969) found that the cosine performs well as a measure of similarity in proportion.

The purpose of the research should determine which similarity measure is chosen. If all aspects of similarity are to be included in the comparison then Euclidean distance is a logical choice because it is easy to understand and interpret geometrically. Overemphasis on size can be reduced by appropriate scaling of the variables or by expressing some variables in the forms of ratios. If the purpose of the research is to compare proportions, however, then Q-correlation or cosine may be most appropriate.

TABLE 1

PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET M

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U12	U13
S1	1.000	-.713	-.913	-.765	-.729	.977	-.747	-.956	.959	.973	.970	-.351	.948	-.956	.948
S2	-.713	1.000	.792	.856	.971	-.744	.874	.775	-.771	-.771	-.722	.447	-.749	.775	-.779
S3	-.913	.792	1.000	.709	.813	-.885	.691	.960	-.957	-.982	-.909	.312	-.879	.960	-.958
S5	-.765	.656	.709	1.000	.881	-.775	.979	.750	-.748	-.749	-.763	.546	-.770	.750	-.750
S6	-.729	.971	.813	.881	1.000	-.730	.860	.791	-.785	-.791	-.742	.438	-.739	.791	-.795
S8	.977	-.744	-.885	-.775	-.730	1.000	-.790	-.938	.940	.946	.956	-.397	.978	-.938	.932
S9	-.747	.874	.691	.979	.860	-.790	1.000	.736	-.735	-.731	-.745	.561	-.785	.736	-.737
S12	-.956	.775	.960	.750	.791	-.938	.736	1.000	-.999	-.980	-.949	.332	-.928	1.000	-.997
S13	.959	-.771	-.957	-.748	-.785	.940	-.735	-.999	1.000	.980	.942	-.330	.922	-.999	.994
S15	.973	-.771	-.982	-.749	-.791	.946	-.731	-.980	.980	1.000	.957	-.339	.930	-.980	.975
U1	.970	-.722	-.909	-.763	-.742	.956	-.745	-.949	.942	.957	1.000	-.358	.977	-.949	.957
U2	-.351	.447	.312	.546	.438	-.397	.561	.332	-.330	-.339	-.358	1.000	-.402	.332	-.334
U8	.948	-.749	-.879	-.770	-.739	.978	-.785	-.928	.922	.930	.977	-.402	1.000	-.928	.936
U12	-.956	.775	.960	.750	.791	-.938	.736	1.000	-.999	-.980	-.949	.332	-.928	1.000	-.997
U13	.948	-.779	-.958	-.750	-.795	.932	-.737	-.997	.994	.975	.957	-.334	.936	-.997	1.000

TABLE 2

SPEARMAN RANK-ORDER CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET M

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U12	U13
S1	1.000	-.777	-1.000	-.773	-.799	.969	-.762	-.980	.980	1.000	.984	-.335	.957	-.980	.977
S2	-.777	1.000	.777	.887	.960	-.821	.905	.788	-.786	-.777	-.774	.471	-.818	.788	-.788
S3	-1.000	.777	1.000	.773	.799	-.969	.762	.980	-.980	-1.000	-.984	.335	-.957	.980	-.977
S5	-.773	.887	.773	1.000	.919	-.780	.978	.776	-.771	-.773	-.778	.547	-.784	.776	-.779
S6	-.799	.960	.799	.919	1.000	-.797	.894	.806	-.803	-.799	-.801	.454	-.799	.806	-.807
S8	.969	-.821	-.969	-.780	-.797	1.000	-.800	-.956	.955	.969	.956	-.384	.988	-.956	.954
S9	-.762	.905	.762	.978	.894	-.800	1.000	.768	-.764	-.762	-.766	.561	-.803	.768	-.771
S12	-.980	.788	.980	.776	.806	-.956	.768	1.000	-.999	-.980	-.971	.326	-.950	1.000	-.998
S13	.980	-.786	-.980	-.771	-.803	.955	-.764	-.999	1.000	.980	.965	-.323	.944	-.999	.995
S15	1.000	-.777	-1.000	-.773	-.799	.969	-.762	-.980	.980	1.000	.984	-.335	.957	-.980	.977
U1	.984	-.774	-.984	-.778	-.801	.956	-.766	-.971	.965	.984	1.000	-.337	.971	-.971	.979
U2	-.335	.471	.335	.547	.454	-.384	.561	.326	-.323	-.335	-.337	1.000	-.385	.326	-.328
U8	.957	-.818	-.957	-.784	-.799	.988	-.803	-.950	.944	.957	.971	-.385	1.000	-.950	.957
U12	-.980	.788	.980	.776	.806	-.956	.768	1.000	-.999	-.980	-.971	.326	-.950	1.000	-.998
U13	.977	-.788	-.977	-.779	-.807	.954	-.771	-.998	.995	.977	.979	-.328	.957	-.998	1.000

TABLE 3

PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET N

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U15*	U13
S1	1.000	-.412	-.921	-.788	-.702	.697	-.519	-.924	.946	.976	.543	-.262	.525	.564	.649
S2	-.412	1.000	.462	.531	.575	-.507	.777	.394	-.391	-.448	-.237	.207	-.255	-.249	-.272
S3	-.921	.462	1.000	.773	.789	-.630	.489	.924	-.941	-.983	-.517	.219	-.499	-.575	-.655
S5	-.788	.531	.773	1.000	.904	-.430	.649	.803	-.789	-.795	-.490	.188	-.462	-.528	-.584
S6	-.702	.575	.789	.904	1.000	-.356	.540	.774	-.755	-.763	-.463	.156	-.435	-.526	-.575
S8	.697	-.507	-.630	-.430	-.356	1.000	-.668	-.557	.606	.674	.321	-.100	.364	.315	.376
S9	-.519	.777	.489	.649	.540	-.668	1.000	.457	-.469	-.512	-.270	.207	-.295	-.270	-.307
S12	-.924	.394	.924	.803	.774	-.557	.457	1.000	-.988	-.943	-.600	.223	-.576	-.662	-.741
S13	.946	-.391	-.941	-.789	-.755	.606	-.469	-.988	1.000	.963	.552	-.233	.528	.614	.706
S15	.976	-.448	-.983	-.795	-.763	.674	-.512	-.943	.963	1.000	.540	-.244	.521	.582	.665
U1	.543	-.237	-.517	-.490	-.463	.321	-.270	-.600	.552	.540	1.000	-.258	.996	.956	.933
U2	-.262	.207	.219	.188	.156	-.300	.207	.223	-.233	-.244	-.258	1.000	-.267	-.269	-.279
U8	.525	-.255	-.499	-.462	-.435	.364	-.295	-.576	.528	.521	.996	-.267	1.000	.947	.921
U15*	.564	-.249	-.575	-.528	-.526	.315	-.270	-.662	.614	.582	.956	-.269	.947	1.000	.989
U13	.649	-.272	-.655	-.584	-.575	.376	-.307	-.741	.706	.665	.933	-.279	.921	.989	1.000

* Gower's coefficient for unstandardized data (U12) is not shown in this table because it is identical to S12. Instead, Euclidean distance for unstandardized data (U15) is shown.

TABLE 4

SPEARMAN RANK-ORDER CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET N

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U15*	U13
S1	1.000	-.450	-1.000	-.813	-.774	.674	-.495	-.945	.967	1.000	.579	-.332	.549	.579	.661
S2	-.450	1.000	.450	.533	.578	-.589	.920	.413	-.410	-.450	-.237	.331	-.255	-.237	-.268
S3	-1.000	.450	1.000	.813	.774	-.674	.495	.945	-.967	-1.000	-.579	.332	-.549	-.579	-.661
S5	-.813	.533	.813	1.000	.944	-.400	.575	.833	-.817	-.813	-.544	.269	-.504	-.544	-.599
S6	-.774	.578	.774	.944	1.000	-.391	.571	.799	-.781	-.774	-.540	.247	-.503	-.540	-.592
S8	.674	-.589	-.674	-.400	-.391	1.000	-.640	-.557	.591	.674	.306	-.340	.341	.306	.362
S9	-.495	.920	.495	.575	.571	-.640	1.000	.460	-.461	-.495	-.252	.327	-.274	-.252	-.290
S12	-.945	.413	.945	.833	.799	-.557	.460	1.000	-.988	-.945	-.674	.283	-.640	-.674	-.747
S13	.967	-.410	-.967	-.817	-.781	.591	-.461	-.988	1.000	.967	.632	-.308	.598	.632	.716
S15	1.000	-.450	-1.000	-.813	-.774	.674	-.495	-.945	.967	1.000	.579	-.332	.549	.579	.661
U1	.579	-.237	-.579	-.544	-.540	.306	-.252	-.674	.632	.579	1.000	-.229	.991	1.000	.990
U2	-.332	.331	.332	.269	.247	-.340	.327	.283	-.308	-.332	-.229	1.000	-.250	-.229	-.250
U8	.549	-.255	-.549	-.504	-.503	.341	-.274	-.640	.598	.549	.991	-.250	1.000	.991	.976
U15*	.579	-.237	-.579	-.544	-.540	.306	-.252	-.674	.632	.579	1.000	-.229	.991	1.000	.990
U13	.661	-.268	-.661	-.599	-.592	.362	-.290	-.747	.716	.661	.990	-.250	.976	.990	1.000

*Gower's coefficient for unstandardized data (U12) is not shown in this table because it is identical to S12. Instead, Euclidean distance for unstandardized data (U15) is shown.

TABLE 4

SPEARMAN RANK-ORDER CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET N

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U15*	U13
S1	1.000	-.450	-1.000	-.813	-.774	.674	-.495	-.945	.967	1.000	.579	-.332	.549	.579	.661
S2	-.450	1.000	.450	.533	.578	-.589	.920	.413	-.410	-.450	-.237	.331	-.255	-.237	-.268
S3	-1.000	.450	1.000	.813	.774	-.674	.495	.945	-.967	-1.000	-.579	.332	-.549	-.579	-.661
S5	-.813	.533	.813	1.000	.944	-.400	.575	.833	-.817	-.813	-.544	.269	-.504	-.544	-.599
S6	-.774	.578	.774	.944	1.000	-.391	.571	.799	-.781	-.774	-.540	.247	-.503	-.540	-.592
S8	.674	-.589	-.674	-.400	-.391	1.000	-.640	-.557	.591	.674	.306	-.340	.341	.306	.362
S9	-.495	.920	.495	.575	.571	-.640	1.000	.460	-.461	-.495	-.252	.327	-.274	-.252	-.290
S12	-.945	.413	.945	.833	.799	-.557	.460	1.000	-.988	-.945	-.674	.283	-.640	-.674	-.747
S13	.967	-.410	-.967	-.817	-.781	.591	-.461	-.988	1.000	.967	.632	-.308	.598	.632	.716
S15	1.000	-.450	-1.000	-.813	-.774	.674	-.495	-.945	.967	1.000	.579	-.332	.549	.579	.661
U1	.579	-.237	-.579	-.544	-.540	.306	-.252	-.674	.632	.579	1.000	-.229	.991	1.000	.990
U2	-.332	.331	.332	.269	.247	-.340	.327	.283	-.308	-.332	-.229	1.000	-.250	-.229	-.250
U8	.549	-.255	-.549	-.504	-.503	.341	-.274	-.640	.598	.549	.991	-.250	1.000	.991	.976
U15*	.579	-.237	-.579	-.544	-.540	.306	-.252	-.674	.632	.579	1.000	-.229	.991	1.000	.990
U13	.661	-.268	-.661	-.599	-.592	.362	-.290	-.747	.716	.661	.990	-.250	.976	.990	1.000

*Gower's coefficient for unstandardized data (U12) is not shown in this table because it is identical to S12. Instead, Euclidean distance for unstandardized data (U15) is shown.

TABLE 5

PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET R

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U12	U13
S1	1.000	-.325	-.795	-.376	-.407	.913	-.293	-.864	.938	.930	.907	**	.745	-.864	.924
S2	-.325	1.000	.674	.617	.962	-.485	.780	.669	-.565	-.553	-.337	**	-.429	.669	-.538
S3	-.795	.674	1.000	.457	.750	-.920	.468	.905	-.902	-.959	-.820	**	-.838	.905	-.914
S5	-.376	.617	.457	1.000	.720	-.385	.818	.600	-.521	-.435	-.373	**	-.364	.600	-.490
S6	-.407	.962	.750	.720	1.000	-.547	.778	.736	-.642	-.637	-.431	**	-.504	.736	-.624
S8	.913	-.485	-.920	-.385	-.547	1.000	-.426	-.898	.934	.962	.899	**	.874	-.898	.939
S9	-.293	.780	.468	.818	.778	-.426	1.000	.573	-.479	-.403	-.290	**	-.363	.573	-.442
S12	-.864	.669	.905	.600	.736	-.898	.573	1.000	-.982	-.931	-.823	**	-.779	1.000	-.961
S13	.938	-.565	-.902	-.521	-.642	.934	-.479	-.982	1.000	.965	.876	**	.789	-.982	.982
S15	.930	-.553	-.959	-.435	-.637	.962	-.403	-.931	.965	1.000	.902	**	.836	-.931	.966
U1	.907	-.337	-.820	-.373	-.431	.899	-.290	-.823	.876	.902	1.000	**	.935	-.823	.935
U2	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**
U8	.745	-.429	-.838	-.364	-.504	.874	-.363	-.779	.789	.836	.935	**	1.000	-.779	.874
U12	-.864	.669	.905	.600	.736	-.898	.573	1.000	-.982	-.931	-.823	**	-.779	1.000	-.961
U13	.924	-.538	-.914	-.490	-.624	.939	-.442	-.961	.982	.966	.935	**	.874	-.961	1.000

** The correlation coefficient r_{ij} could not be computed for some of the profiles in data set R because r_{ij} is undefined for completely flat profiles.

TABLE 6

SPEARMAN RANK-ORDER CORRELATIONS BETWEEN 15 SIMILARITY MATRICES

DATA SET R

	S1	S2	S3	S5	S6	S8	S9	S12	S13	S15	U1	U2	U8	U12	U13
S1	1.000	-.644	-1.000	-.647	-.676	.992	-.590	-.914	.953	1.000	.923	**	.914	-.914	.957
S2	-.644	1.000	.644	.800	.959	-.649	.893	.775	-.731	-.644	-.577	**	-.585	.775	-.684
S3	-1.000	.644	1.000	.647	.676	-.992	.590	.914	-.953	-1.000	-.923	**	-.914	.914	-.957
S5	-.647	.800	.647	1.000	.887	-.626	.895	.789	-.749	-.647	-.579	**	-.571	.789	-.695
S6	-.676	.959	.676	.887	1.000	-.669	.894	.811	-.769	-.676	-.617	**	-.615	.811	-.724
S8	.992	-.649	-.992	-.626	-.669	1.000	-.596	-.906	.943	.992	.924	**	.922	-.906	.946
S9	-.590	.893	.590	.895	.894	-.596	1.000	.735	-.693	-.590	-.521	**	-.530	.735	-.635
S12	-.914	.775	.914	.789	.811	-.906	.735	1.000	-.988	-.914	-.850	**	-.841	1.000	-.951
S13	.953	-.731	-.953	-.749	-.769	.943	-.693	-.988	1.000	.953	.881	**	.872	-.988	.976
S15	1.000	-.644	-1.000	-.647	-.676	.992	-.590	-.914	.953	1.000	.923	**	.914	-.914	.957
U1	.923	-.577	-.923	-.579	-.617	.924	-.521	-.850	.881	.923	1.000	**	.996	-.850	.943
U2	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**
U8	.914	-.585	-.914	-.571	-.615	.922	-.530	-.841	.872	.914	.996	**	1.000	-.841	.935
U12	-.914	.775	.914	.789	.811	-.906	.735	1.000	-.988	-.914	-.850	**	-.841	1.000	-.951
U13	.957	-.684	-.957	-.695	-.724	.946	-.635	-.951	.976	.957	.943	**	.935	-.951	1.000

** The correlation coefficient r_{ij} could not be computed for some of the profiles in data set R because r_{ij} is undefined for completely flat profiles.

BIBLIOGRAPHY

Boyce, A. J.
1965 The Methods of Quantitative Taxonomy with Special Reference to Functional Analysis, Ph.D. Thesis, Oxford University.

1969 "Mapping Diversity: A Comparative Study of Some Numerical Methods". In Numerical Taxonomy (A. J. Cole, ed.), Academic Press, London.

Cattell, R. B
1949 "rp and other coefficients of similarity." Psychometrika, 14 279-298.

1969 "The profile similarity coefficient in vocational guidance and diagnostic classification." British Journal of Educational Psychology, 39.

Cattell, R. B., Coulter, M. A., and Tsujioka, B.
1966 "The taxonomic recognition of types and functional emergents." In Handbook of Multivariate Experimental Psychology (Cattell, ed.), Rand McNally: Chicago.

Cohen, J.
1969 "R_c: a profile similarity coefficient invariant over variable reflection." Psychological Bulletin, 71, 281-284.

Cronbach, L. J., and Gleser, G. C.
1953 "Assessing similarity between profiles." The Psychological Bulletin, 50(b).

Edelbrock, C.
1978 "Mixture model tests of hierarchical clustering algorithms: the problem of classifying everybody." American Classification Society Conference.

Everitt, Brian
1974 Cluster Analysis, John Wiley and Sons, Inc.: New York.

Gower, J. C.
1971 "A general coefficient of similarity and some of its properties." Biometrics, 27.

Green, P. E., and Rao, V. R.
1969 "A note on proximity measures and cluster analysis." Journal of Marketing Research, VI, August.

Howard, K. J., and Dissenhaus, H. I.
1967 "Direction of measurement and profile similarity." Multivariate Behavioral Research, 2, 225-237.

Moore, A. W., and Russell, J. S.
1967 "Comparison of coefficients and grouping procedures in numerical analysis of soil trace element data." Geoderma, 1, 139-158.

Morrison, D. G.
1967 "Measurement problems in cluster analysis." Management Science, 13(B), 755-780.

Muldoon, J. F., and Ray, O. S.
1958 "A comparison of pattern similarity as measured by six statistical techniques and eleven clinicians." Educational and Psychological Measurement, 18, 775-778.

Overall, J. E.
1964 "Note on multivariate methods for profile analysis." Psychological bulletin, 61, 195-198.

Overall, J. E. and Klett, C. J.
1972 Applied Multivariate Analysis. McGraw-Hill: New York.

Orloci, L.
1966 "Geometric models in ecology." Journal of Ecology, 53, 193-215.

Penrose
1954 "Distance, size, and shape." Annual Eugenics, 18, 337-343.

Rohlf, F. J., and Sokal, R. R.
1965 "Coefficients of correlation and distance in numerical taxonomy." University of Kansas Science Bulletin, 45, 3-27.

Sokal, R. R., and Sneath, P. H. A.
1973 Principles of numerical taxonomy. W. H. Freeman and Co.: San Francisco.

Wishart, D.
1969 "Mode Analysis: A Generalization of Nearest Neighbor which Reduces Chaining Effects." In Numerical Taxonomy (Cole, Ed.), Academic Press: New York, 282-311.

1972 "A general tripartite clustering method and similarity generating function." Statistics 1 Division Report No. 4-31, Civil Service Dept., Whitehall: London.

CHAPTER 9

EVALUATING CLASSIFICATIONS: A COMPARISON OF FIVE COEFFICIENTS FOR PARTITION AGREEMENT

In evaluating classifications, particularly classifications which have been generated by numerical procedures, it is often the case that we wish to examine the association or overlap between two classifications of the same objects. Often, where a set of data has been repeatedly analyzed by different cluster analytic techniques, the analyst may wish to examine the similarities and differences between the various competing classifications as a way of understanding the characteristics of each method (Blashfield, 1976; Edelbrock, 1978, Milligan, 1978, and others). Coefficients such as the Rand Coefficient (Rand, 1971) have frequently been used in comparative studies of clustering methods as a measure of "cluster recovery" and for assessing in a more general sense the adequacy of the solutions offered by the various clustering methods.

It has become increasingly clear that there are various ways of assessing the relationships between different classifications. To the extent that these different coefficients assess different components of the similarity between classifications, the reliance upon a single criteria to assess the performance of different methods of classification may be folly. Mezzich (1978), in fact, criticizes a number of the prior comparative studies of classification techniques for their reliance upon a single evaluative criteria.

A general situation in the evaluation of various classifications occurs when there is a known "correct" or "valid" classification. In this type of comparison the evaluative coefficients will generally be based on some function of the proportion of points misclassified. The method with the smallest proportion of misclassified points will be evaluated as "best." This situation logically requires that each class in the experimental classification be "matched" with a class in the known correct classification.

A second general situation occurs in which the investigator has no "known correct" classification. This instance occurs, for example, when the results

of different clustering methods are being compared to each other. Comparisons between any two partitions may be used to determine the general levels of agreement and specific kinds of agreement. The most specific approach providing a maximum of information is to form a simple contingency table, showing the cross-classification of data points in the two classifications. This approach was suggested by Borko et al. (1968) and described by Anderberg (1973).

In the present chapter we will evalute several of the most common approaches to examining the overlap between classifications and will look at their differences, strengths, and weaknesses.

Partition-matching Coefficients to be Evaluated

The following partition-matching coefficients were selected for inclusion in this study. A more complete discussion of these is given earlier in our evaluation review chapter.

Rand's Coefficient and Jaccard's Coefficient. In cross-classifying two partitions against each other we consider that for any pair of subjects there are four possibilities:

Partition 1

Partition 2	Placed in same cluster	Placed in different clusters
Same Cluster	a	b
Different Cluster	c	d

This matrix indicates the four possibilities for joint and different classification of each pair of subjects in the two partitions. Various combinations of a, b, c, and d can be formed into a large variety of different coefficients, as follows:

$$\text{Rand Coefficient} = \frac{a + d}{a + b + c + d}$$

$$\text{Jaccard's Coefficient} = \frac{a}{a + b + c}$$

It can be seen that Rand's Coefficient represents the proportion of pairs of data points that are treated alike in the two partitions. The effect of negative matches (i.e., "d") is often seen as problematic, and it is often excluded. This is done in the Jaccard Coefficient.

Cramer's V. This is based on the chi-square statistic and is defined as:

$$V = \sqrt{\frac{X^2}{N \cdot \min (r-1)\ (c-1)}}$$

where N = number of cases, and r and c represent the numbers of rows and columns respectively in the contingency table. The coefficient ranges from 0 to 1, with a high score indicating close agreement.

Goodman and Kruskal's Lambda. The asymmetric and symmetric versions of this coefficient are based on the concept of proportional reduction in error. The asymmetric versions of Lambda measure the effectiveness of one partition as the predictor for the other (see Reynolds, 1977, p. 48-52). It is based on the assumption that the best strategy for prediction is to select the category with the most cases, since this will minimize the number of wrong guesses.

Suppose the objective is to guess class membership in partition 2, based upon knowledge of its class membership in partition 1. If its membership in partition 1 were not known, the best one could do would be to choose the partition 2 class with the largest marginal total, max $f_{\cdot k}$. But if the subject is known to belong to class j in partition 1, then only row j of the contingency table is of interest and the best guess is the class in partition 2 corresponding to the largest entry in row j, having frequency max f_{jk}.

Then the proportional reduction in error achieved by using partition 1 classes to predict membership in partition 2 classes is measured by asymmetric Lambda:

$$L_2 = \frac{\sum_j \max_k f_{jk} - \max_k f_{\cdot k}}{N - \max_k f_{\cdot k}}$$

where N is the number of subjects in the data set.

Analogously, using partition 2 classes to predict membership in partition 1, the proportional reduction in error would be

$$L_1 = \frac{\sum_k \max_j f_{jk} - \max_j f_{j\cdot}}{N - \max_j f_{j\cdot}}$$

Symmetric Lambda is a kind of average of the two asymmetric values L_1 and L_2. It is computed by

$$L = \frac{\sum_j \max_k f_{jk} + \sum_k \max_j f_{jk} - \max_k f_{\cdot k} - \max_j f_{j\cdot}}{2N - \max_k f_{\cdot k} - \max_j f_{j\cdot}}$$

According to Anderberg, L is "the relative decrease in error probability due to the use of predictor classes when the directions of prediction are equally important."

Values of L_1, L_2, and L lie between 0 and 1. A value of 1 indicates perfect predictive ability, while a value of 0 indicates no improvement in predictive ability through the use of predictor classes. The value of L always lies between L_1 and L_2 inclusive, and is indeterminant only if the entire data set lies in a single cell.

Symmetric Lambda was judged to be more appropriate than its assymetric version as a general measure of partition agreement because it gives equal weight to prediction in each direction. Furthermore, it will only attain the

maximum score of 1 when there is an exact match between the two partitions. (Asymmetric Lambda can take on a value of 1 when the predictor classification contains more categories than the criterion classification, as long as each predictor class is a subset of one of the criterion classes.)

METHOD

Several data sets were structured for the examination of the performance of different classification methods. Each of these methods may be regarded as a "classifier". The different classifiers; i.e., cluster analytic methods, have been applied to these four data sets. Four of the data sets are artificial and the fifth data set is the real data set. The above four agreement coefficients were used to compare the different classifier partitions with each other for each of the data sets and also used to compare the classifier solutions with the "correct" classes.

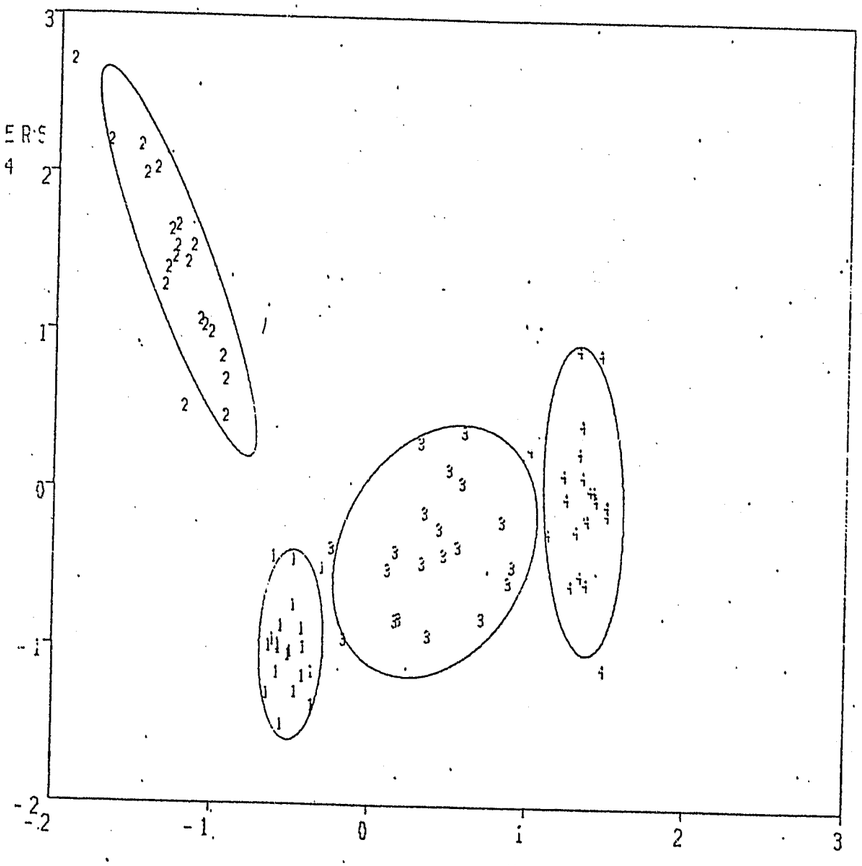
Monte Carlo Data Sets

These official data sets were deliberately constructed to include a variety of classification problems that are encountered in real behavioral science data. A full description of these data sets is given in the previous chapter dealing with the comparison of cluster analytic techniques. Briefly then, the following data sets are as follows:

Data Set D. This included an N of 80 points assessed on eight variables. This is the easiest data set to cluster. The clusters vary in terms of compactness and separation. All of the classifiers (cluster analytic methods) gave fairly good results on this data. The data set is illustrated in the plot below for data set D.

CLUSTER GRAPH

DATA SET 0
GENERATED-CLUSTERS
X=VAR 1 Y=VAR 4



Data Set E. This contains 180 points assessed on three variables. A substantial amount of noise was added to blur the boundaries between the classes. The classes themselves contain a certain amount of overlap, thereby making it more difficult to correctly classify cases at the boundaries. There was a slight overlap between clusters 1 and 3. The data set can be seen in the plot below for data set E.

Data Set F. This also consisted of 180 points assessed on four variables with a substantial amount of added noise. This data set was very difficult for the classifiers because of the closeness between the different classes, the overlap between boundaries, and the added noise. The classes are very close together and have fairly different densities and shapes. The data set can be seen below as the cluster graph for data set F.

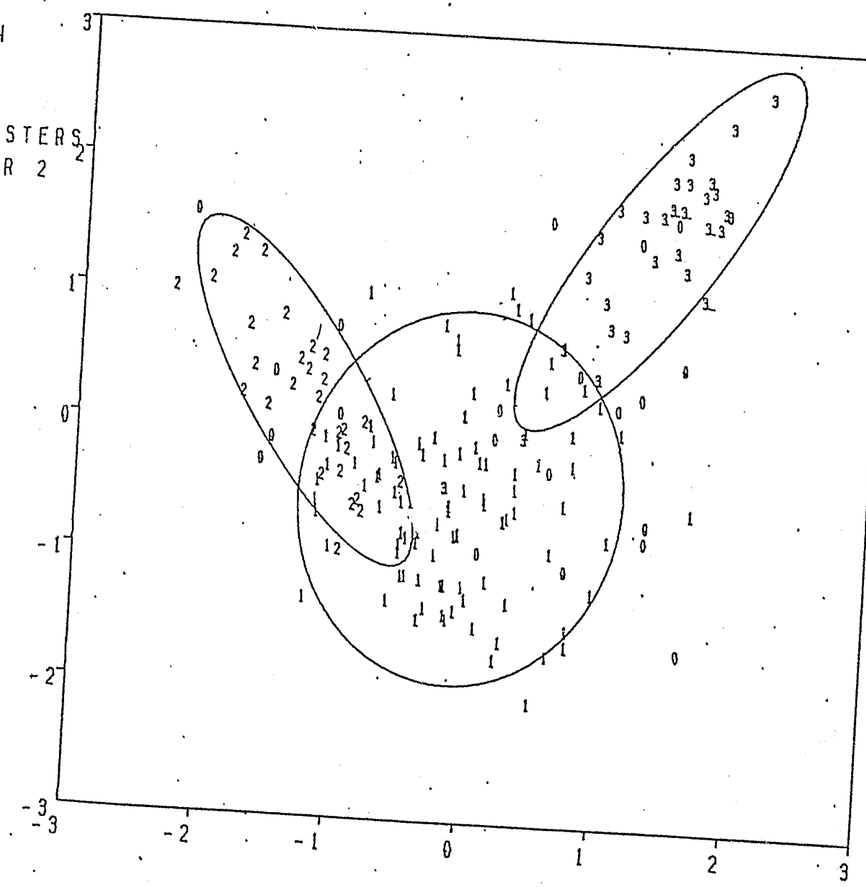
Data Set U. This was a single uniform distribution and eight variables with 150 points. There were no classes within this data set and thus there are no boundaries. The value of this particular data set is that it may give an indication of baseline levels for the agreement coefficients.

Real Data

This is a set of data in which 199 youth have been assessed on the Jesness Psychological Inventory. This Inventory includes 11 scales measuring different personality characteristics. The Jesness Inventory is primarily used to separate levels of interpersonal maturity which are theoretically believed to be related to different levels of delinquent and antisocial behavior. The data set had been extensively clustered by different analytical methods to find personality types that had some intrinsic structure within the data set. The cluster graph below for Jesness Subset 1 indicates the stable classes that were found when this data set was analyzed by a variety of classification

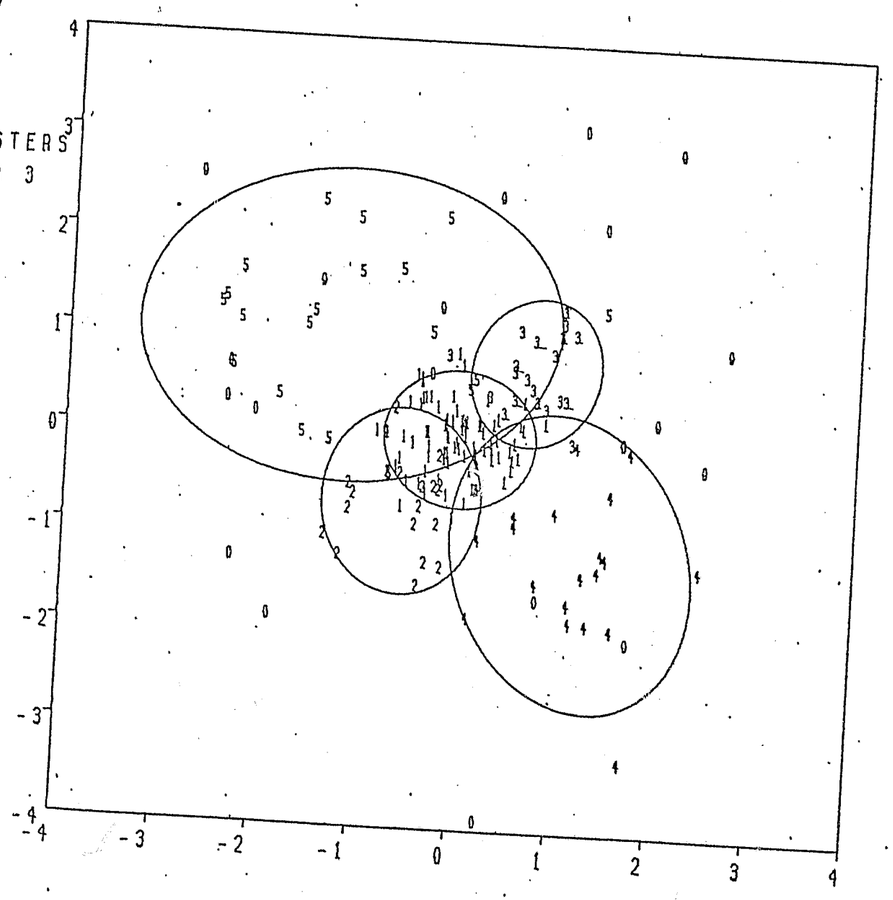
CLUSTER GRAPH

DATA SET E
GENERATED CLUSTERS
X=VAR 1 Y=VAR 2



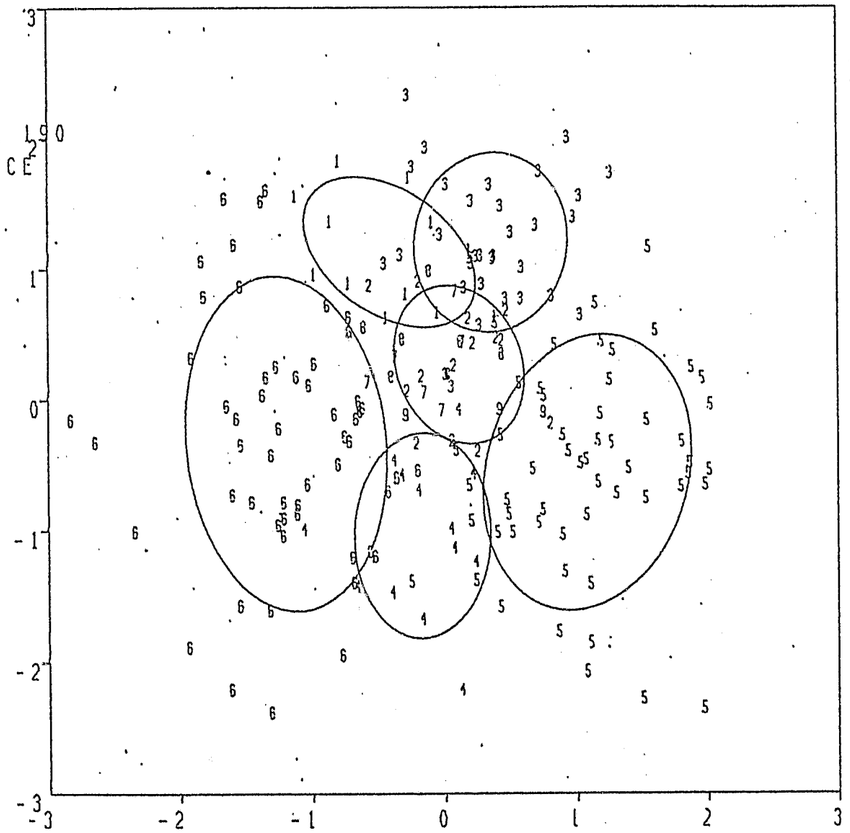
CLUSTER GRAPH

DATA SET F
GENERATED CLUSTERS
X=VAR 1 Y=VAR 3



CLUSTER GRAPH

JESNESS SUBSET 1
UPGMA-COS. LEVEL L90
DISCRIMINANT SPACE



techniques. This particular solution was obtained from the UPGMA cluster method with cosine as a similarity coefficient. It can be seen that there are six clusters in this data. The full psychological description and derivation of these clusters is given in the chapter devoted to the analysis of the Jesness I-level data. A critical difference between the present real data exercise and the above Monte Carlo data exercises is that in the present situation there is "no correct" partition. In the present situation, therefore, we are comparing classifier partitions with each other.

Classification Methods

The different classifications of the above data sets were obtained by the use of five different cluster analysis methods and various combinations with five similarity coefficients. We did not systematically use every cluster analytic method with every similarity coefficient because of the fact that certain methods; e.g., the K-means method, are restricted to a single one or two similarity coefficients. This produced 11 to 13 different "classifiers" for each of the data sets. The methods of clustering are described elsewhere. However, they included the following methods:

K-means (the MIKCA method), Ward's hierarchical method, the group average (UPGMA) method, Wishart's hierarchical mode search method, and in certain circumstances, the N-mode method for discovering natural clusters.

RESULTS

Comparison of Classifier Solutions with "Correct" Classifications

This particular type of comparison, i.e., with correct classes, is restricted to data sets D, E, and F of the above five sets of data.

Data Set D. The results of the comparisons of the different ways of classifying data set D are shown in Table 1. Eleven different classifications of the data are shown in the first column. These are ranked into seven

Measures of Partition Agreement on Data Set D
with Success Ratings for the Four-Group Partitions

4-group partitions, in order of success in duplicating the clusters as generated	# of points misclassified (minimum)	Points left as outliers	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
1. UPGMA, cosine and MIKCA, det W, Mahal. D	0	0	1.000	1.000	1.000	1.000
2. MIKCA, det W, Euclidean d	1	0	.988	.950	.984	.983
3. UPGMA, r	6	2	.927	.736	.912	.870
4. HMODE, cosine	9	0	.912	.703	.891	.838
5. HMODE, r	10	6	.884	.620	.856	.782
6. UPGMA, both d and Gower; HMODE, both d and Gower; MIKCA, trace W, d	25	0	.850	.591	.816	.750
7. WARD, d^2	31	1	.731	.410	.720	.581
NMODE, d; no 4-group partition						

Other partitions, ordered by # of groups; within that, ranked according to Lambda	# of Groups	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
HMODE, r	2	.620	.388	1.000	.500
MIKCA, det W and trace W; UPGMA with d, Gower, and cos; HMODE with d, Gower, and cos	3	.873	.655	1.000	.800
NMODE, d	3	.850	.593	.801	.750
UPGMA, r	3	.765	.464	.717	.642
WARD, d^2	3	.597	.349	.707	.438
UPGMA, cosine	5	.975	.896	1.000	.958
HMODE, d	5	.964	.854	.984	.941
UPGMA, d	5	.963	.846	.989	.933
MIKCA, det W, both D and d	5	.961	.840	.985	.917
HMODE, cosine	5	.960	.837	.984	.916
MIKCA, trace W, d	5	.821	.512	.816	.680
WARD, d^2	5	.713	.371	.720	.505
HMODE, Gower	6	.899	.637	.920	.816

general levels according to the number of points which were misclassified. It can be seen that two of the "classifiers", i.e., the UPGMA with cosine and the K-means method using Mahalanobis distance similarity, and minimizing the determinant W, achieved perfect matches with the correct classification. The top half of table 1 is quite informative in that all of the partition matching coefficients indicate a perfect score for the two classifiers which showed perfect matching with the correct classification. Secondly, there is complete agreement in terms of rank ordering the various classifiers according to their general level of performance. Although the actual magnitudes of the values of each coefficient may vary, they rank order the classifiers in the same way.

The bottom half of Table 1 examines classifications which did not reach the correct four group level (recall that in data set D there were four input classes). Two general points may be noted in this section of Table 1. First, Cramer's V coefficient assigns a score of 1 in a number of cases to classifications which were clearly erroneous, particularly the poor two group partitions. Cramer's V in this instance shows dramatic disagreements with the other coefficients, particularly the Jaccard and Symmetric Lambda coefficients. A second general point is that some of the five group partitions were actually ranked higher than some of the four group partitions. Paradoxically, we would agree with these indications from the evaluation coefficients, particularly since some of the four group classifications had merged incorrect clusters and bore no relationship to the correct four classes of the input data.

Data Set E. In ranking the different classifications and classifiers of this data set, we took two factors into account. First, the actual number of points misclassified--as seen in column 2 of Table 2. However, this, in some cases, was modified to take account of the number of outliers in the

Table 2

Measures of Partition Agreement on Data Set E
with Success Ratings for the Three-Group Partitions

3-group partitions, in order of success in duplicating the clusters as generated	# of points misclassified (minimum)	cluster pts. left as outliers	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
1. MIKDA, Euclidean d (both det W and Trace W)	2	0	.897	.778	.904	.752
2. HMODE, d	4	10	.904	.778	.787	.729
3. HMODE, Gower's	6	9	.891	.754	.777	.712
4. UPGMA, d - (4th grp. (considered outliers)	9	1	.843	.690	.722	.660
5. UPGMA, Gower's	10	0	.835	.680	.825	.631
6. MIKCA, Mahalanobis D^2	17	0	.803	.587	.800	.584
7. HMODE, cosine	15	7	.795	.554	.679	.578
8. UPGMA, cosine	19	0	.790	.562	.798	.576
9. NMODE, d	19	5	.789	.608	.702	.536
10. WARD, d^2	38	0	.650	.503	.635	.416
11. UPGMA, r	52	0	.636	.326	.537	.301
12. HMODE, r	52	1	.637	.321	.445	.297

Other partitions, ordered by # of groups; within that, ranked according to Lambda	# of groups	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
MIKCA, Euclidean d, det W	2	.630	.495	.855	.436
MIKCA, Mahalanobis D^2, det W	2	.619	.446	.735	.352
MIKCA, Mahalanobis D^2, det W	4	.766	.465	.725	.557
MIKCA, Euclidean d (both det & trace)	4	.766	.462	.725	.554
UPGMA, Gower's	5	.829	.604	.743	.591
UPGMA, cosine	5	.747	.420	.695	.567
MIKCA, d, trace W	5	.793	.511	.723	.533
HMODE, cosine	6	.802	.537	.731	.612
WARD, d^2	6	.840	.677	.743	.592
UPGMA, d	6	.800	.510	.756	.565

classification. We made the judgment that it was worse to misclassify a point than to leave it as an outlier. Thus, in certain cases there will be reversals away from the strict number of misclassified points. These occur when points which were basically outliers were actually erroneously classified into the system. A general point to be noted is that most of the coefficient values are lower than in the above data set D because of the slightly greater difficulty that the classifiers had in organizing this particular data set and in reaching the "correct" classification. This was primarily a result of the added amount of noise points and the closer or overlapping boundaries between the classes. A second general point to be noticed is that Cramer's V again prefers the two group partition, ranking them with the best of the correct three group partitions.

The column of values for the Rand coefficient indicates that this particular coefficient generally tends to increase with an increasing number of classes. This is so despite the fact that this particular data set correctly contained only three clusters. Thus, going strictly by the numerical values over the Rand coefficient, the cluster analysis user would tend to select the five or six level classification of this data set. This is less noticeably true of the other coefficients although it seems that both the Jaccard and Symmetric Lambda coefficients show a less noticeable tendency to increase in value with increasing number of classes.

Data Set F. Table 3 indicates the values of the various coefficients of partition matching for the classifications of data set F. It should be recalled that this was a most difficult set of data for the classifiers to deal with and there will be a larger number of misclassifications and erroneous mergings and splitting of classes. None of the classifiers could particularly

Table 5

Measures of Partition Agreement on Data Set F
with Success Ratings for the Five-Group Partitions

5-group partitions, in order of success in duplicating the clusters as generated	# of points misclassified (minimum)	cluster pts. left as outliers	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
1. MIKCA, D(centroids given)	15	0	.887	.642	.825	.695
2. MIKCA, d(centroids given)	17	0	.877	.616	.811	.675
3. UPGMA, d(branch lines individually selected)	21	5	.864	.611	.772	.681
4. UPGMA, Gower	20	23	.841	.559	.700	.585
5. UPGMA, d(all branches cut at same level)	34	16	.811	.536	.667	.562
6. MIKCA, d(random start)	53	0	.777	.419	.699	.500
7. MIKCA, D(random start	40	0	.844	.549	.647	.470
8. HMODE, d	37	23	.748	.443	.563	.434
9. HMODE, Gower	38	21	.757	.469	.579	.444
10. UPGMA, cosine	75	0	.702	.228	.557	.304
HMODE, cosine	*	34	.685	.212	.479	.252
UPGMA, r	*	0	.649	.167	.362	.125
HMODE, r	*	25	.639	.169	.347	.140
WARD, d^2: No reasonable 5-group partition						
NMODE, d: No 5-group partition						

Other partitions, ordered by # of groups; within that, ranked according to Lambda	# of groups	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
UPGMA, d	3	.724	.467	.821	.577
UPGMA, Gower	3	.718	.462	.825	.574
MIKCA, Euclidean d, det W	3	.744	.466	.851	.514
MIKCA, Mahalanobis D, det W	3	.642	.402	.823	.425
WARD, d^2	3	.318	.221	.246	.017
MIKCA, Mahalanobis D, det W	4	.814	.550	.844	.624
MIKCA, Euclidean d, det W	4	.810	.547	.828	.608
HMODE, d	4	.755	.473	.629	.447

* In these partitions, no meaningful match with the generated clusters can be defined.

separate clusters 1 and 2 (without external intervention by the analyst). Thus, the first three partitions all involved the special knowledge of the data analyst in making certain analytical decisions; i.e., the input of the correct K-level for the MIKCA analyses. Our attempt to provide a "best ranking" procedure was much less clearcut in this instance than for the above data sets D and E. We again used the primary measure of the number of points which had been misclassified and the number of outliers which were erroneously classified within the system. However, on certain decisions, we attempted to introduce subjective ideas that different kinds of misclassifications may be more or less serious. This was based on our knowledge of the correct structuring of these artificial data sets. For instance, with the classifiers that are ranked 6 and 7; i.e., the two MIKCA analyses using random starts, we rank the Euclidean distance MIKCA as somewhat superior to the Mahalanobis MIKCA in spite of the fact that it produced 53 misclassifications whereas the Mahalanobis analysis produced 40 misclassifications. However, the errors of the 40 were particularly serious (see Table 4). It is interesting to note that this subjective judgment receives the support of both the Lambda and the Cramer coefficients. The actual data for this "subjective" judgment can be seen in Table 4, sections (a) and (b). The MIKCA analysis with Mahalanobis distance is indicated in Table 4(a) and the analysis with Euclidean distance in indicated in 4(b). Although the solution with Mahalanobis distance contains fewer misclassified points, it is less successful in locating the various clusters than is the solution with Euclidean distance. The latter, however, does in fact contain more points misclassified.

Table 3 indicates an extremely disturbing aspect of the use of the various partition matching coefficients. This is that they disagree quite markedly

Table 4

Cluster Overlap Matrices for Data Set F
MIKCA with Mahalanobis D: 3, 4, and 5 Group Partitions

(a) det W, D, 5 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	(73)	4	1	3	1	0	82
2	4	0	8	5	3	3	23
3	3	16	0	10	0	4	33
4	0		11	2	1	7	21
5	0	0	0	0	(15)	6	21
Total	80	20	20	20	20	20	

(b) det W, d, 5 groups

1	0	0	0	0	(15)	5	20
2	19	0	5	2	4	2	32
3	(61)	20	0	1	0	2	84
4	0	0	0	(16)	0	2	18
5	0	0	(15)	1	1	9	26
Total	80	20	20	20	20	20	

(c) det W, D, 5 groups (started with correct centroids)

1	(72)	1	1	2	0	1	77
2	8	(19)	0	0	0	3	30
3	0	0	(19)	0	3	8	30
4	0	0	0	(18)	0	3	21
5	0	0	0	0	(17)	5	22
	80	20	20	20	20	20	

(d) det W, d, 3 groups

1	(76)	20	4	2	0	5	107
2	4	0	16	0	20	12	52
3	0	0	0	(18)	0	3	21
	80	20	20	20	20	20	

regarding the relative rankings of the different classifiers. Some of these are fairly large differences, indicating that the different coefficients are eliciting very different meanings of "partition agreement." It can be seen that the Mahalanobis solution is judged better by the Rand coefficient and the Jaccard coefficient while the Euclidean K-means solution is judged better by the Cramer coefficient and the Lambda coefficient.

Table 3 also examines the various classifications which were found at classes other than the five correct class solutions. A first point to note is that the four class partitions were rated higher than their five class counterparts by all of the coefficients, with one exception. This, in fact, matches our own assessment since classes 1 and 2 were not separated very well. The collapsing of these classes, thus giving a four group partition, was generally a far more successful analysis than the analysis which correctly identified the five class solution but at the cost of much larger numbers of misclassifications and generally erroneous boundaries between the five classes. It can again be noted that Cramer's V rates the three class partition the highest, again showing a certain bias toward smaller numbers of classes in the partition. This again deviates from the correct level.

Discussion of the Above Results

These results suggest that Cramer's V tends to disagree with the other partition matching coefficients and with our own conception of good partition matching in that it tends to give a higher score and a higher ranking to partitions with smaller numbers of clusters. A second criticism of this coefficient is that it may rank classifications at the 1.0 level even when the match is not perfect.

The three other coefficients generally produced rather similar rankings even when the number of classes in the partition differed slightly. Examining specifically the Rand coefficient, it can be noted that it always has misleadingly high values. These values range between 0.6 and 1.0. Its effective range becomes very narrow for larger class partitions. It also tends to increase in value with larger numbers of classes in a partition. With this coefficient it seems difficult to find guidelines for "good agreement" and it seems as though such values would definitely vary with partition size in terms of number of classes.

Finally, it seems that the Goodman and Kruskal Lambda coefficient consistently agreed with our own rankings of goodness of classification more often than did the Rand and Jaccard coefficients.

Comparison of Two Classifications without a "Known" Correct Classification

Four of the data sets deal with this situation. Data set U basically consists of completely nonclassifiable data. The real data set, consisting of the Jesness material, is such that any real or "correct" solution is unknown and must be discovered. Data sets E and F are "difficult" data sets and the amount of overlap and noise that has been added renders a known correct lution somewhat problematic.

Data Set U, a Uniform Distribution. In this situation classification methods as "classifiers" must find clusters where, in fact, there are none. Table 5 indicates the results of comparing four of the different classification approaches. It may be noted that, once again, the Rand coefficient values are extremely high, particularly for the second of these classification comparisons. Similarly, Cramer's V and Lambda reach fairly high or moderate levels of agreement. These results raise the question of

UNIFORMLY DISTRIBUTED DATA

Table 5

Cluster Overlap Matrices for Data Set U
Matches between HMODE and UPGMA Partitions

(a) UPGMA vs. HMODE, both with Euclidean d

HMODE-d 7 groups	UPGMA-d, 7 groups 1	2	3	4	5	6	7	Outliers	Total
1	4	0	0	0	0	9	0	3	16
2	0	5	0	0	0	0	0		5
3	2	(35)	5	11	2	9	9	2	75
4	0	0	0	0	(12)	0	0		12
5	0	1	1	0	0	10	0		12
6	0	5	0	0	0	0	0	1	6
7	0	7	3	0	0	0	0		10
Outliers		3			4		7		14
Total	6	56	9	11	18	28	16	6	

R = .681 J = .213 V = .492 L = .284

(c) MIKCA-D vs. UPGMA-d, 6-group partitions

UPGMA-d 6 groups	MIKCA-D, 6 groups 1	2	3	4	5	6	Total
1	5	0	2	0	1	1	9
2	5	1	8	(30)	1	11	56
3	2	0	3	4	0	0	9
4	1	1	(26)	0	1	0	29
5	12	(14)	0	1	0	1	28
6	0	1	0	1	(16)	1	19
Total	25	17	39	36	19	14	

R = .791 J = .330 V = .603 L = .532

how much agreement can be taken as implying that there are, in fact, clusters present. Does moderate partition agreement imply that there are clusters? The results of this analysis suggest strongly that such moderate agreement cannot be taken as sufficient evidence. On the other hand, it seems clear that low agreement may be a stronger indication that clustering structure does "not exist" in the data. Table 6 indicates the full set of comparisons when the various classifiers were applied to this uniform data set. This table summarizes the results for the uniform data set (DAT-U) and for the structured artificial data sets (DAT-E and DAT-F). This table is instructive in giving an indication of the ranges that the various partition matching coefficients take when various levels of structure are present in the data. The two K-means partitions with Euclidean distance and Mahalanobis distance seem clearly to have converged upon highly similar partitioning of the uniform data set. The Symmetric Lambda of .86 and the Rand coefficient of .93 are both very high, indicating a high degree of similarity in the partitionings. All of the other results have a much lower level of overlap between the various classifiers. The results for the structured data sets E and F are substantially higher, indicating that the classifiers in these cases have reached partitions which show much higher general levels of class overlap.

Table 7 indicates some tentative guidelines that emerge from this research regarding the sizes and numerical values for partition agreement coefficients based upon the artificial data sets. Again, since certain of the coefficients show a tendency to larger values based purely upon the number of classes present, we reiterate that the present artificial data sets contained between one and five classes. We have divided the ranges into excellent, moderate, questionable, and poor levels of agreement. The relative sizes of the Rand

Table 6

Measures of Partition Agreement for Pairs of Partitions Produced by Cluster Analysis on the Same Data

Partition 1	Partition 2	# of groups p. 1	# of groups p. 2	# of outliers p. 1	# of outliers p. 2	Rand's coef.	Jacc. coef.	Cramer's coef.	Symmetric Lambda
Data Set U									
MIKCA-D	vs MIKCA-d	6	6	0	0	.933	.688	.880	.867
MIKCA-D	vs MIKCA-D	6	4	0	0	.748	.297	.598	.430
MIKCA-D	vs UPGMA-d	6	6	0	0	.791	.330	.603	.532
HMODE-d	vs UPGMA-d	7	7	14	6	.681	.213	.492	.284
UPGMA-cos	vs UPGMA-d	7	7	13	6	.782	.238	.613	.451
HMODE-d	vs HMODE-cos	6	7	19	27	.687	.150	.524	.259
Data Set E									
MIKCA-D	vs UPGMA-d	3	4	0	0	.911	.826	.893	.800
HMODE-d	vs UPGMA-d	3	4	24	0	.885	.770	.806	.765
UPGMA-cos	vs UPGMA-d	3	4	0	0	.761	.566	.774	.576
HMODE-d	vs HMODE-cos	3	3	24	11	.783	.545	.680	.530
Data Set F									
MIKCA-D	vs UPGMA-d	3	3	0	20	.884	.808	.867	.637
MIKCA-D	vs UPGMA-d	4	5	0	33	.875	.717	.796	.564
HMODE-d	vs UPGMA-d	5	5	40	33	.886	.743	.741	.732

(cosine and d partitions not comparable because of cluster centered at origin)

Table 7

Tentative Guidelines for Judging Partition Agreement
Based on Artificial Data Sets Containing 1-5 Clusters

Agreement Level	Rand's Coefficient	Jaccard's Coefficient	Symmetric Lambda
Excellent	Above .9	Above .75	Above .75
Moderate	.8 - .9	.6 - .75	.6 - .75
Questionable	.7 - .8	.5 - .6	.5 - .6
Poor	Below .7	Below .5	Below .5

Two factors should be taken into account in using these guidelines. First, Rand's coefficient tends to increase as the number of groups in the partitions increases. Therefore, the Rand guidelines should be raised for partitions containing more than 5 groups. Second, the existence of a large number of outliers may affect coefficient values, raising them if the two partitions have similar outlier groups, lowering them if one partition has a large outlier category while the other assigns most of the outliers to clusters.

coefficient, the Jaccard coefficient, and Symmetric Lambda are provided in Table 7. These should give some general indications of tentative guidelines regarding the interpretation of these coefficients in comparative studies of cluster analytic methods and in validity studies against correct prior classification systems.

Comparing Classifications for Cluster Analytic Studies of Real Data: An Unknown "Correct" Classification

Our final data set illustrates the multiple analysis of a real data set by many classification approaches and the comparison of the empirically produced classifications. In this case, there is no "known correct" classification solution. The analyst is confronted with a situation in which he must deduce from the various cluster analytic results the best or most likely "correct" classification. Table 8 shows the results of a number of comparisons of classification results obtained from this real data. Virtually all of these involve the use of the UPGMA and the HMODE cluster analysis methods with either distance coefficients or cosine similarity coefficients. Additionally,

TABLE 8

Measures of Partition Agreement for JES1
Comparing UPGMA and HMODE Partitions

Partitions (same sim. coef.)	# Clusters*	# Outliers	Rand	Jaccard	Lambda
UPGMA-d (190) vs. HMODE-d (6)	7x6	4, 46	.787	.457	.463
UPGMA-d (190) vs. HMODE-d (6)	4x5	18, 51	.786	.468	.482
UPGMA-cos (190) vs. HMODE-cos (6)	8x8	3, 40	.882	.527	.664
UPGMA-cos (193) vs. HMODE-cos (6)	6x8	0, 40	.844	.476	.616
UPGMA-r (191) vs. HMODE-r (6)	8x6	0, 42	.742	.258	.385
Partitions (different sim. coef.)					
UPGMA-d (190) vs. UPGMA-cos (193)	4x5	18, 6	.731	.321	.466
UPGMA-cos (190) vs. UPGMA-r (192)	9x7	0, 0	.813	.321	.517

*The number of clusters at a particular tree-level depends in part on the minimum number of points required to make a cluster. Smaller groups are thrown into the "outlier" category. In this table, where the number of groups for a given partition changes it is because the minimum cluster size was changed. (Treatment of such small numbers of points has little impact on the agreement coefficients.)

the correlation coefficient in its Q analysis sense was used as a similarity coefficient. A first point to note about this analysis is the very high levels assumed by the Rand coefficient, particularly when eight clusters were present in both classifications. This high value would be especially encouraging to an analyst, although it is clear from our prior results that the Rand coefficient tends to increase with an increasing number of classes. The classification solutions based on distance coefficients are all slightly under a .5 cutoff on both Jaccard and Lambda coefficients. The coefficients based on correlation with the same cluster analysis methods are substantially lower. This would lead the analyst to judge that these positions are particularly unstable. Two of the Lambda coefficients on this real data set are between .6 and .75 and would thus indicate "moderate" overlap in the partitions produced. It might be noted here that the substantive interpretations of the clusters produced by these two analyses were intuitively highly similar in the sense that they gave largely the same kind of psychological profile for each class produced. These were the classifications that were produced by both cluster analytic methods when they used the cosine similarity coefficient.

CONCLUSIONS

This research suggests that the Jaccard and Symmetric Lambda coefficients appear to be potentially useful. We suggest that these coefficients would be worth comparing in a much larger and more systematic study of classification overlap. The Lambda coefficient may be marginally preferable as a result of the clearer expositions of its meaning (see Reynolds, 1977). Lambda has further advantages in that it seems less affected by variations in number of classes in a partition. It is also clearly associated with a contingency table

and the joint use of this coefficient with the full contingency table is very valuable for the direct perception of the types of overlap and disagreements between the classes of each partition. This contingency table, in fact, was found to be extremely useful in the discovery of the particular "styles" of each of the classification approaches that have been used. The Rand coefficient seems to be less desirable, partially because of its generally very high and potentially misleading levels, its greater susceptibility to being enlarged as a result of an increased number of classes in a partition, and its very narrow range.

BIBLIOGRAPHY

Anderberg, M.R.
1973 Cluster analysis for applications. Academic Press: New York.

Blashfield, R.K.
1976 "Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods." Psychological Bulletin, 83, 377-88.

Borko, H., D.A. Blankenship, and R.C. Burket
1968 On-line information retrieval using associative indexing. RADC-TR-68-100, AD 670195. Systems Development Corporation: Santa Monica, California.

Edelbrock, C.
1978 "Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody." National Institute of Mental Health, American Classification Society Conference.

Mezzich, J.E.
1978 "Evaluating clustering methods for psychiatric diagnosis." Biological Psychiatry, 13, 265-81.

Milligan, G.W.
1978 "An examination of the effect of six types of error perturbation on fifteen clustering algorithms." Working Paper Series. College of Administrative Science, Ohio State University: Columbus, Ohio.

Rand, W.M.
1971 "Objective criteria for the evaluation of clustering methods." Journal of the American Statistical Association, 66(336), 846-850.

Reynolds, H.T.
1977 The analysis of cross-classifications. The Free Press: New York.

CHAPTER 10

EVALUATION OF SEVERAL CLUSTER ANALYSIS METHODS

USING ARTIFICIAL DATA SETS

Introduction.

Criminological data generally involves measurements of social, behavioral and psychological variables which may be highly skewed, intercorrelated, or simply irrelevant to the desired classification. In this chapter, we will look at the characteristics of criminological data that are likely to make the search for clusters difficult. We will then attempt to duplicate these characteristics in artificial data sets, to create data of known structure which poses the same challenges. Clustering techniques which perform well on such data can then be selected for use in clustering "real" criminological data.

Monte Carlo studies have frequently been used by numerical taxonomists to test and compare clustering algorithms. A set of data is artificially created by drawing random probability samples from several multivariate normal populations. Thus the data contains a mixture of points from several populations, and a clustering method can be judged by its success in finding clusters which match the original populations. Knowing the "correct" cluster assignment for each point, it is possible to rate the clustering methods on the basis of a criterion such as "number of points misclassified".

Such artificial data sets often have been created to test a particular clustering method developed by the author (e.g. McRae's MIKCA, Ball's ISODATA, Huizinga's mode search). Everitt (1974) used artificial data to compare a variety of hierarchical and k-means techniques, and David Wishart (1972) compared a number of similarity measures using one clustering algorithm. Both Everitt and Wishart confined their artificial data to two dimensions, however, which is convenient for visual display of results but hardly approximates the actual data found in research situations.

Blashfield (1976), on the other hand, developed 50 artificial data sets with a wide variety of characteristics, differing in number of variables, size of data set, sizes and shapes of clusters, separation of clusters, and amount of correlation between variables. He used them to compare four hierarchical clustering methods and has made them available for use by other researchers. The features of his data sets were selected randomly within specified ranges, however, so the problems posed by each data set were not directly related to problems in any particular type of real data.

In using data sets drawn from multi-normal distributions it should be kept in mind that they are based on a 'points in space' model of similarity. They would generally not be appropriate for comparing clustering methods when the research purpose is to compare shapes of profiles.

Inventing artificial data to represent a 'shape of profiles' model is more difficult because the concept of 'shape' varies according to the topic of research. One such data set was created by Bartko, Strauss and Carpenter (1971) to compare clustering techniques which might be used to classify psychiatric patients. They argue: "Since there is no statistical test of significance of emerging clusters, the choice among various techniques and emerging clusters has tended to depend on the meaningfulness and usefulness of the actual clusters produced." They fabricated psychiatric symptoms for 100 archetypal patients, representing five diagnostic categories with 20 patients per category. Thus the clusters were designed to duplicate descriptions used in clinical practice; clustering techniques were evaluated based upon how well they would duplicate clinical diagnoses. (It is interesting to note that the 'best' clustering method of those tried was complete link with Q-correlation

as the resemblance measure. Indeed, it was the only technique which reproduced the original input groups. The same clustering method used with Euclidean distance produced many misclassifications.)

A number of sets of real (as opposed to factitious) data, which have known 'right answers' based on theory, have been used to compare clustering methods. A simple but popular one is the Fisher-Kendall Iris data, containing four measurements made on 150 specimens from three species of Iris (Ball, 1967; Bezdek, 1973, McRae, 1973). Of course, finding the right answer is partly dependent on using a measure of resemblance that is appropriate for both the data and the theory.

Ideally, it would be good to create a great many artificial data sets, as Blashfield has done, to compare the ability of the clustering methods to handle a variety of easy and difficult situations. Variations in shape, size and separation of clusters will affect different algorithms in different ways, for example. The amount of 'noise' in the data and the amount of correlation in the variables are also likely to have a significant effect.

For the purposes of this project, four artificial data sets were specifically designed to simulate the data structures which are most likely to occur in social, behavioral, and psychological data.

Challenges In Criminological Data

Measurements of human behavior often do not fall into neatly normal distributions. Distributions of deviant behavior, for example, are often extremely skewed. The clusters within may be very peaked or rather flat. Thus artificially generated clusters, which are drawn from multi-normal distributions, do not necessarily duplicate the real-life situation. The relationship between clusters also concerns us.

For detecting an underlying structure in a set of data classification methods must be able to detect two basic kinds of clusters:

(a) Compact, separated clusters whose boundaries are clearly definable.

(b) Modal clusters, which may be close enough together that their boundaries overlap, yet which can be distinguished by the existence of more than one highly dense area (mode). Boundaries between modal clusters are fuzzy, not clearly defined.

The danger is that a clustering method may instead identify artificial clusters based on its own method of cluster-building, imposing artificial boundaries that are not inherently contained in the data.

We will focus on the following problems and data structures:

1. Determining The Correct Number Of Clusters, If Any. It is possible that a data set may contain no clusters, having only minor variations in frequency and, perhaps, a single large mode. If the purpose is to uncover a natural structure, a clustering method should not artificially partition a continuous distribution into several arbitrary blocks. It may be equally inappropriate to break up some of the natural clusters into smaller segments or tack two adjacent modes together. Which error is most acceptable depends on the purpose of the classification.

2. Unclassifiable Individuals. Even when there are a variety of clusters, a few individuals are likely to fit in none of the categories, appearing instead as peripheral points or points in the region between clusters. A clustering method should recognize the unclassifiable "noise" points in some way, rather than forcing them into one of the clusters when the fit is poor. Furthermore, it should not be misled by a few points between clusters into merging distinct classes.

3. Differing Size And Compactness Of Clusters. Data in criminal behavior frequently contains a very large compact cluster of "average" individuals, with much smaller and sparser categories of "deviant" individuals. The number of points and the cluster diameters interrelate to produce clusters that may range from very compact (many points within a small territory) to very sparse (a few points spread over a rather large territory but still identifiable as a cluster because of a background that is even more sparse).

A clustering method should be able to detect clusters of widely different size and compactness within one data set. Tree diagrams from the hierarchical methods should contain signals which help in identifying compact and sparse clusters in the same data, especially when two compact clusters merge before a sparse cluster has completely formed.

4. Small Separation Or No Separation Between Clusters. Even compact, separated clusters may be difficult to distinguish if the separation between them is small relative to their diameters, particularly if a few noise points happen to fall in the gap between.

Much more difficult is the task of distinguishing significant modes when there is no empty gap between, only a lessening of density. Yet this will occur when two populations occupy territories that overlap slightly. We would like to be able to recognize distinctive-but-overlapping populations, while ignoring the minor variations in density which are inevitable in a random sample.

5. Different Cluster Shapes And Orientations In Space. This study will focus on clusters whose territories can be described as hyper-spheres and hyperellipsoids, ignoring the more esoteric possibilities such as hyper-horseshoes, doughnuts, and teacups.

The ellipticity of a cluster is determined by the variances and covariances of the variables on points in that cluster. It is reasonable to expect that covariances may be different for distinct subpopulations within one data set. Thus it is important that a clustering method be able to detect ellipsoidal clusters at a variety of orientations in space.

Everitt (1974) argues that certain clustering techniques (minimum variance, for example) will perform optimally with spherical clusters and will have difficulty recognizing ellipsoidal clusters.

6. Irrelevant Variables. The categories within a classification system are defined entirely on the basis of the variables which were used to develop that system. There is a great temptation, when searching for clusters to throw in every variable on which measurements are available in hopes of finding a classification system that will distinguish categories on as many dimensions as possible. (Numerous examples of such folly can be found in the criminological literature: Baer, 1970; Jenkins and Boyer, 1967).

The inclusion of extra variables which are irrelevant to the "natural" clusters that presumably lie hidden in the data will usually make those clusters more difficult to find. Suppose, for example, that 5 variables are being used when clear-cut clusters would occur on 3 of the variables, while the other 2 variables are uniformly distributed across the entire data set. Two points which are very close together on the 3 clusterable variables may, simply by chance, be far apart on the 2 extra variables and thus appear fairly distant in the 5-dimensional space. The inclusion of extraneous variables distorts boundaries between clusters, making them less distinct. Of course, there may be no way of telling in advance which variables will be "relevant" to those hidden clusters. But caution should be exercised against including too many or too diverse a collection of variables.

7. Redundant Variables. If two variables are highly correlated over the whole data set (say, r> .7) they are measuring essentially the same characteristic. If both are used in the clustering process that characteristic will, in effect, be given double weight. It is quite possible that one of the variables should be judged redundant and not be used in clustering. Preliminary principal components analysis is sometimes advisable to determine how much redundancy exists. Of course, two partially redundant variables could be deliberately retained if the user judges that the characteristic they measure in common is so important that it deserves extra weight (see, particularly, Morrison 1967).

(This problem has not been addressed in our study of artificial data, but will be considered in the applications to real data.)

Challenges Built Into The Artificial Data Sets

In the following descriptions of our artificial data sets we will indicate the particular data structures that are embedded in each data sets. Four artificial data sets are constructed, DAT-D, DAT-E, DAT-F and DAT-U. The starred challenge in each case is the major structure of each data set.

Numerical information on the exact statistical structure of each of the data sets is included later in the results section. The cluster graphs shown later in this chapter indicate the structure of the three sets D, E, and F.

Starred challenges (*) indicate the major focus of each data set.

1. Set D: Four separated clusters, one of which is spherical in shape, the others ellipsoidal with different orientations in space. No overlap between clusters on at least one variable and no added noise.

*Challenges #3, differing compactness; and #5, different cluster shapes: One cluster is much less compact than the others and is elliptical in shape with a great deal of within-cluster covariance. It is quite well separated from the other clusters.

Challenge #4: small separation between clusters. Two of the clusters are quite close together, but they do not overlap in the sense that all points can be correctly allocated by discriminant analysis.

2. Set E: Three ellipsoidal clusters at different orientations in space, with additional "noise" points.

Challenge #2: Unclassifiable noise points.

Challenge #3: Differing # of points in clusters creating somewhat different degrees of compactness.

*Challenge #4: No separation, a slight overlap between two of the clusters at the end of their ellipsoidal territories.

*Challenge #5: Differently shaped ellipsoids, at different orientations in space.

3. Set F: Five spherical clusters which differ greatly in compactness, with additional "noise" points. Three of the clusters overlap greatly to simulate modal clusters with fuzzy boundaries.

Challenge #2: Unclassifiable noise points.

*Challenge #3: Large difference in size and compactness of clusters. One cluster contains many more points than the others within a small territory, while two of the other clusters cover very large territories.

*Challenge #4: No separation, considerable overlap of two cluster-pairs. This can be expected to interact with challenge #3, because the two better separated but less compact clusters will not be fully formed in a tree diagram until after the three compact and overlapping clusters have merged into one.

4. Set U: One cluster of points drawn from a uniformly distributed population.

*Challenge #1: Determining the number of clusters in a data set. Which clustering methods will tell the user that this data set should not be partitioned into subsets? Results can be used as a basis for comparing with clustering results on other data sets. Determining the number of clusters will be a recurring challenge in all data sets.

Limitations On The Artificial Data

Size Of Sample: Sample size is limited by the memory capacity of the computer to be used and by the cost of computer time. The hierarchical clustering programs used in this study can handle a maximum of 200 points. Programs for the other clustering techniques can process data sets of 500 points and most, with minor modifications, could handle up to 1,000 points. Because of the size limitation on hierarchical techniques, all of the artificial data sets in this study contain 200 points or less.

Number Of Variables: Clustering of real criminological data will frequently involve 15 to 20 variables, or more. However, it is difficult to duplicate that situation in artificial data and maintain an accurate perception of the characteristics of the data set. Therefore, artificial data sets in this study involve from 3 to 8 variables.

Interval-Level Variables: Most of the clustering methods are designed to be used on continuous variables measured on an interval-level scale, so our artificial data is all of that type. It may be possible to use these methods on ordinal data when distances between categories can be assumed to be roughly equivalent. Nominal (categorical) variables have not been included in this project, although the use of Gower's coefficient as a similarity measure would make it possible to do so.

Standardization Of Variables: This study does not address the question of whether to standardize. All variables were standardized before clustering.

"Points In Space" Model Of Similarity: By drawing artificial data from multinormal populations in m-dimensional space we automatically define the clusters within a "points in space" model (see the earlier discussion of similarity). Hence this study will not test the ability of the methods to cluster objects according to similarity in their profile shapes. As a result, the potential usefulness of the correlation coefficient, which most closely measures similarity in profile shape, is likely to be under-rated.

Clustering Methods

Clustering methods to be compared were selected from those most widely used in numerical taxonomy, plus two mode-seeking methods that are less well-known but appeared to be appropriate for detecting modal clusters.

K-means methods: The MIKCA computer program (McRae, 1973) was selected with two criteria to be optimized. Minimizing trace W, where W is the within-groups covariance matrix, is the most common optimization criterion, but Everitt (1974) however argues that it tends to break up ellipsoidal clusters into smaller, more-or-less spherical parts. According to Everitt, minimizing determinant W is more able to detect an ellipsoid because the covariance within clusters is taken into account by the clustering criterion. Indeed, McRae found that MIKCA performed better when minimizing det W than when minimizing trace W. Thus we will use det W as the clustering criterion, in addition to selected tests of trace W.

One drawback of K-means methods is that they must be told in advance how many clusters to search for. With artificial data one has the rare avantage of knowing how many clusters there should be. Since that is usually not the

case with real data, when the artificial data contains k clusters we will request k-1, k, and k+1 clusters to see how MIKCA allocates points when it is given the wrong number of clusters.

Hierarchical Methods: Of the many methods that have been devised for agglomerating points into clusters in a hierarchical fashion, two stand out as the most successful in the comparisons that have been done: the unweighted group average method (UPGMA) and Ward's minimum variance technique. (See, for example, Edelbrock, 1978; Blashfield, 1976; Delabre, 1973; Cunningham and Ogilvie, 1972). Hence those two have been chosen for use in this project. Both are options within the HCLUST computer program, which contains hierarchical clustering subroutines offered by Anderberg (1973). The result of these methods is a dendrogram (tree diagram) which must be interpreted by the user, who selects one or more levels of the tree as the basis of a classification system. The number of clusters, k, need not be known in advance; in fact, it is to be hoped that some clue to the number of clusters can be gleaned from the tree diagram.

Mode-Seeking Methods: The high incidence of noise in criminological data, and the possibility of overlapping modal clusters, creates a need for clustering methods which will recognize separate modes (regions of high density) within an area of moderate point-density. Wishart (1973) developed such a method, involving a hierarchical linking of modes, which we have implemented in a program called HMODE. This produces a dendrogram showing linkages between modes, or dense-points, rather than between all points. The number of dense-points provides a maximum for the number of clusters, k, which the user may then revise downward if it is determined that some dense-points are so close together they probably represent only minor variations in density.

The second natural mode-seeking algorithm defines modes on the basis of density within a fixed radius R, then assigns points to modes using a hill-descending approach. This method, developed by Huizinga (1978) in program NMODE, produces an unequivocal statement of the number of modes, k, in the data--a potentially very useful feature, if it proves to be reliable. It also recognizes the possibility of "fuzzy" boundaries between clusters by printing a measure of the likelihood of each point belonging in each cluster.

Similarity Measures

The need to match the measure of resemblance to the purpose of research was discussed in Chapters 7 and 8. Because our artificial data is based on the concept of points in space, it would seem most appropriate to use the distance-based coefficients of similarity (or dissimilarity), because they utilize all of the information in the data. Our earlier comparisons of similarity indicate that there is very little difference between the various coefficients which are based on distances between points in space. Hence in the comparison of clustering methods that follows, Euclidean distance will be most heavily used. Two other distance-based coefficients will be used for limited purposes: Gower's general coefficient because of its potential value in handling a combination of qualitative and quantitative variables, and Mahalanobis' distance in MIKCA when min (det W) is the clustering criterion because it is more appropriate mathematically than Euclidean distance (see Everitt, 1974).

Other researchers have reported success in using the Pearson correlation coefficient on artificial data created according to a "points in space" model (e.g. Edelbrock, 1978). We doubt the general applicability of Q-correlation because of its sensitivity to the position of the origin. Particularly in

sociological and behavioral data there is likely to be a large cluster of "average" individuals centered fairly close to the population mean. Such a cluster could not be detected in one piece by any of the angular similarity coefficients. However, because of the popularity of the correlation-coefficient it will be included in our tests. Its potential usefulness would probably be better demonstrated on data in which clusters were to be defined on the basis of similarity of profile shape.

The cosine, or normalized vector product, measures similarity in angular position of points relative to the origin. Its meaning can be easily visualized in a "points in space" model. It also can serve as a measure of similarity in proportion. Although it has not been extensively compared with other similarity coefficients, it is the only angular coefficient that is easily interpretable as a measure of similarity between points in space (see earlier discussion of similarity).

The table below shows the combinations of clustering method and similarity measure which are used in this study. Some clustering algorithms require the use of a particular distance measure; for example, d^2 is an integral part of the minimum variance technique. In other cases, the possible combinations of clustering method and similarity coefficient were limited by the computer programs available. Formulas used to compute the selected similarity measures can be found in our earlier chapters discussing similarity.

CLUSTERING METHODS AND SIMILARITY MEASURES

USED ON ARTIFICIAL DATA SETS

	Correlation	Cosine	Gowers	Euclidean	Mahal.
	r			d	D^2
K-means methods (MIKCA)					
min det W				X	X
min trace W				X	
Hierarchical methods					
Group average (UPGMA)	X	X	X	X	
Minimum variance (WARD)				X	
Mode-seeking methods					
Wishart's hierarchical (HMODE)	X	X	X	X	
Huizinga's (NMODE)				X	

EVALUATION OF APPROACHES

Determining The Number Of Clusters

Each clustering method requires the user to make a decision which determines the number of clusters and may greatly affect the clusters that result. MIKCA's k-means algorithm requires that the user set the number of clusters directly, and it does not change that number. It does occasionally produce a one- or two-point cluster, which may indicate that k has been set too large or that the singleton is an extreme outlier. Hierarchical methods produce a tree diagram; the user selects the level of the tree to be used, thereby determining the number of clusters to be considered. (This applies to HMODE as well as WARD and UPGMA.) NMODE requires the user to set the dense-point radius R. Although the user does not have any direct control over the number of clusters, the choice of R will affect that number, frequently determining whether any clusters are detected at all.

Prior comparative studies using artificial data have frequently adopted a standard rule which determines cluster selection. For example, for a data set containing g clusters, McRae (1973) automatically sets k=g when testing MIKCA, and Blashfield (1976) looks only at the last g branch lines of the trees he is comparing. This is certainly not the procedure that would be followed by a researcher with real data, where g is generally unknown.

Although our evaluation, too, must focus primarily on the "correct" number of clusters, some attention will be given to the types of errors that will occur when a different number of clusters is selected. MIKCA will be repeated for several values of k, NMODE for several values of R. Tree branches will be "cut" keeping in mind the judgment of the interpreter of real data.

Interpretation Of Tree Diagrams

Usually, interpretation of tree diagrams must be done with little prior knowledge about the correct number of clusters. It is based, instead, on such clues as relative lengths of branch lines and numbers of points in the developing clusters. In many past studies the classification 'level' may be chosen by a 'straight cut' across the tree. This produces a series of partitions at various levels of abstraction; i.e., K, K-1, K-7, ...K-N, etc.

While the "straight cut" technique may appear to be more systematic and objective, it ignores a valuable resource: the judgment of the user. Thus when the best cutting point in a tree is not obvious we will try several possible cuts rather than applying a standard rule. The objective will be to get a good match with the correct classification by cutting at points which appear naturally in the tree.

In addition, very small clusters (e.g. containing 1 to 4 points in a data set of 160 points) will not be considered as clusters: their points will be

classified as outliers. (Edelbrock, 1979, points out the unreasonableness of calling a singleton a cluster in comparative studies. He demonstrates that classification may be more accurate when some points are left as outliers than when every point is required to be classified.) This is an example where 'user judgement' may modify the strict application of the 'straight cut' technique.

Comparing The Selected Partitions: Criteria For Evaluation

In a number of studies clustering artificial data, comparison of results has been reduced to a rating of each partition based on a single coefficient of agreement such as Rand's coefficient (Rand, 1971) or kappa (Blashfield, 1976) or simply the number of points misclassified (McRae, 1973). Each of these represents a slightly different definition of "agreement". Furthermore, both counting the number of points misclassified and computing kappa requires prior matching between the clusters in the correct partition and clusters in the partition to be tested. This match is not unique, nor is a "best" match always intuitively obvious.

Recognizing that some clustering mistakes may be considered to be more serious than others, and wishing to discover which clustering methods have trouble with which "challenges", we shall not reduce our comparison to a single coefficient of agreement. Rather, we will analyze the cluster overlap patterns in some detail and relate the errors of classification to the characteristics of the data set involved.

Values of several coefficients of association (described in Chapter 10) have been computed: Rand's, Jaccard's, Cramer's V, and Goodman-Kruskal's Lambda. The number of points misclassified has also been noted, when a reasonable determination was possible. An assessment of the usefulness of

these summary statistics will be delayed until after our judgments on the partitions have been pronounced.

Comparison With Discriminant Analysis

Discriminant analysis on the "correct" clusters will provide a measure of the difficulty of the clustering task in each data set. Where populations overlap, some misclassification of points is to be expected. Therefore the number of points misclassified according to discriminant analysis will provide an "expected" number of misclassifications. Points with less than a .9 probability of being correctly classified will be considered to fall in a fuzzy boundary region between clusters where some misclassification by the clustering methods is likely and should be tolerated.

(It should be cautioned that discriminant analysis is based on measurement of distances between points. Its use is advisable only when a points-in-space model of similarity is desired. It could be quite misleading where classification is to be based upon similarity in profile shape.)

Evaluation Of Individual Clusters

Analysis of within-cluster structure and homogeneity is not generally done on artificial data where structure is known. However, it may provide a useful comparison with real data clusters in later chapters. Therefore, individual clusters from a few selected partitions will be examined.

RESULTS

DATA SET D: CLUSTERS OF DIFFERENT SHAPE AND SEPARATION

Data set D contains 80 points measured on 8 variables. Twenty points were randomly drawn from each of four multivariate normal distributions. Thus the data set was generated in the form of 4 clusters, all containing the same

number of points. Two of the distributions, represented by clusters #1 and 3, had no correlation between the variables. The other two, clusters #2 and 4, both had some correlation between variables.

Specific parameters for the 4 sub-populations are:

Cluster 1

Centroid: (0, 0, 0, 0, 0, 0, 0, 0)

Standard deviation = 1 on each variable

Cluster 2

Centroid: (-8, 8, 8, 8, 8, -8, -8, -8)

Standard deviation = 2 on each variable

Correlation matrix:

1.0	.8	.8	-.8	.8	-.8	.8	.8
.8	1.0	.8	-.8	.8	-.8	.8	.8
.8	.8	1.0	-.8	.8	-.8	.8	.8
-.8	-.8	-.8	1.0	-.8	.9	-.8	-.8
.8	.8	.8	-.8	1.0	-.8	.8	.8
-.8	-.8	-.8	.9	-.8	1.0	-.8	-.8
.8	.8	.8	-.8	.8	-.8	1.0	.8
.8	.8	.8	-.8	.8	-.8	.8	1.0

Cluster 3

Centroid: (9, -2, -2, 2, 2, 0, 0, 0)

Standard deviations: $\sigma_1 = 3; \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 1; \sigma_6 = \sigma_7 = \sigma_8 = 2$

Cluster 4

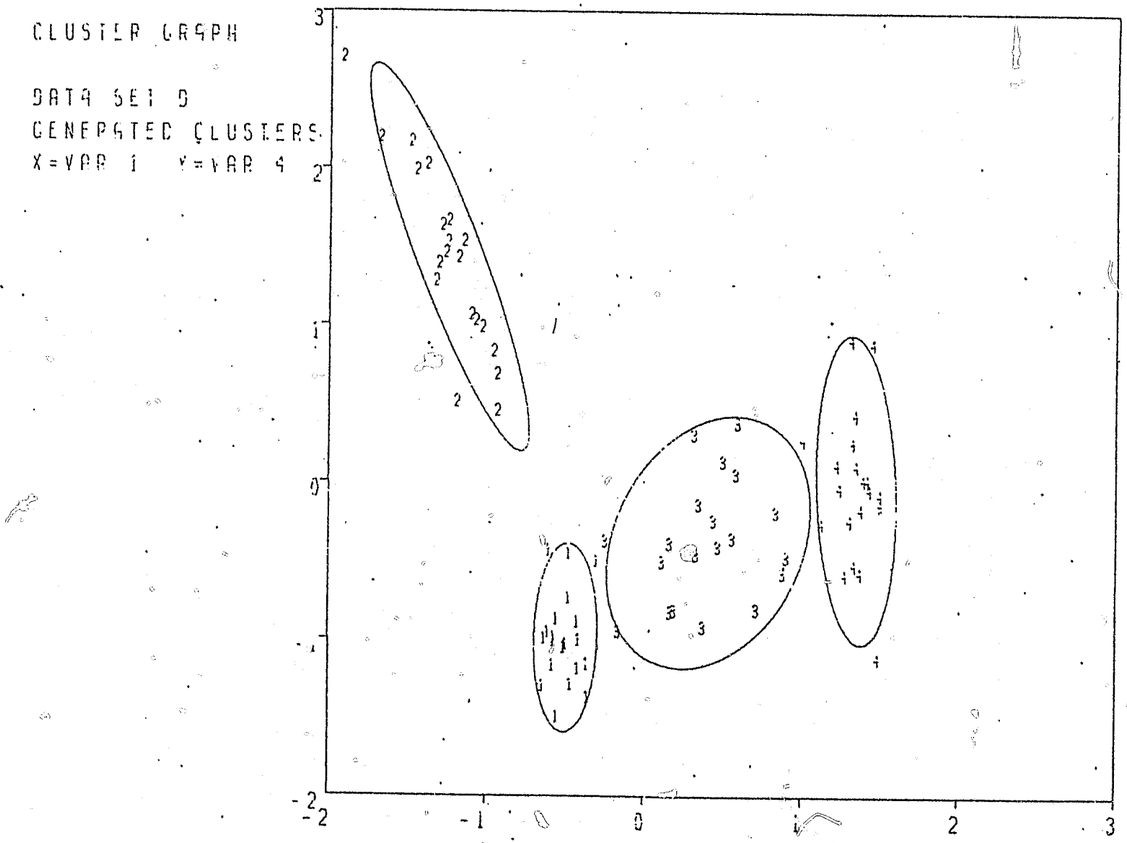
Centroid: (18, 4, 4, 4, 4, 4, 4, 4)

Standard deviations: $\sigma_1 = 1; \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 2; \sigma_6 = \sigma_7 = \sigma_8 = 1$

Correlation matrix:

1.0	0	0	0	0	0	0	0
0	1.0	0	0	0	0	0	0
0	0	1.0	.9	.9	.3	.6	.9
0	0	.9	1.0	.9	.3	.6	.9
0	0	.9	.9	1.0	.3	.6	.9
0	0	.3	.3	.3	1.0	.5	.3
0	0	.6	.6	.6	.5	1.0	.6
0	0	.9	.9	.9	.3	.6	1.0

Figure 1D


CLUSTER GRAPH
DATA SET D
GENERATED CLUSTERS
X=VAR 1 Y=VAR 4
3
2
1
0
-1
-2
-2
-1
0
1
2
3

To completely picture this data set based on all pairs of variables would require 28 separate graphs. Only a limited impression can be obtained from one; however, two dimensions showing typical separation between clusters are plotted earlier in Figure 1D. In the graph, the ellipse drawn for each cluster is based on a multiple of the mean Mahalanobis D^2 from the points to the cluster centroid; therefore the ellipse encloses most, but not all, of the points in the cluster.

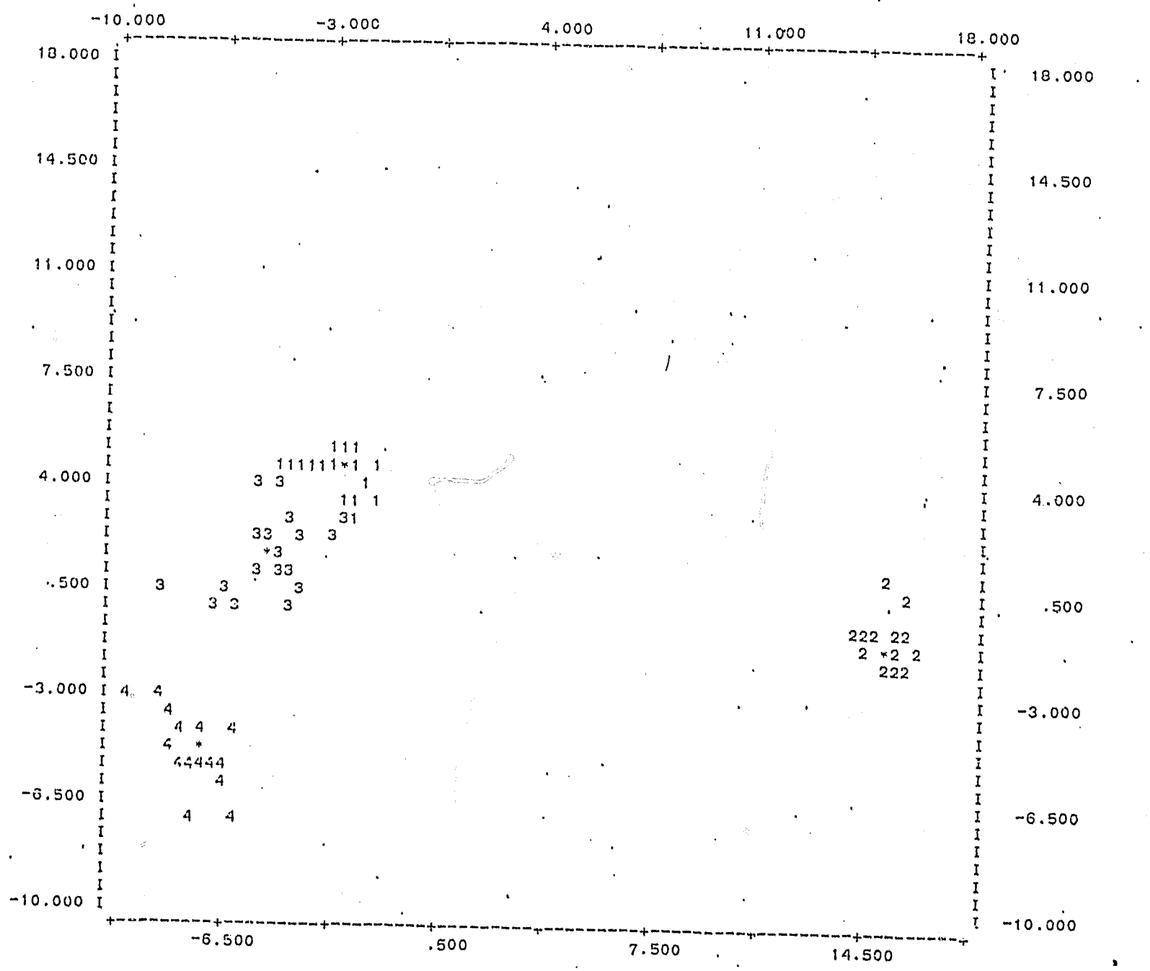
Shapes: In unstandardized space, cluster 1 is drawn from a population which is shaped like a hypersphere, whereas the other populations are shaped like hyperellipsoids. Because the variables are uncorrelated for cluster 3, we may think of it as a hyperellipsoid whose axes are parallel to the coordinate axes. Clusters 2 and 4, though, are oriented at odd angles in space.

Densities: The clusters vary in compactness, with cluster 1 being the most compact and cluster 2 the least compact. All contain the same number of points.

Separation: The four clusters can be clearly distinguished on variable 1 (see graph) with clusters 1 and 3 being fairly close together on that variable. In all of the other variables, some clusters overlap. The greatest amount of overlap is between clusters 1 and 3. Cluster 2 is the most clearly separated in all 8 variables.

Discriminant analysis based on the four correct clusters was able to clasify 100% of the cases correctly. Although clusters 1 and 3 are close together only one point of cluster 3 showed less than a 90% chance of belonging there. Figure 2D is a plot of the data in discriminant space, using the first two discriminant functions.

PLOT OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CENTROID.



Clustering Results (Data set D)

In this 80-point data set we require that a group in a partition must contain at least 3 points. All other points will be designated "outliers" in that partition.

MIKCA, minimizing determinant W: Partitions containing 3, 4, and 5 groups were requested. Mahalanobis' D and Euclidean d produced identical results in the 3- and 5-group cases, and differed in the placement of only one point in the 4-group case. Euclidean d misclassified one point, while Mahalanobis' D produced a perfect 4-group solution!

Table 1D shows cross-tabulations of the MIKCA partitions with the "correct" partition, i.e., the clusters as generated. It can be seen from the table that the 3-group partition merged clusters 1 and 3 into a single group, while the 5-group det W partition split cluster 3 between two groups (along with one point from cluster 4). Similar cluster overlap tables will be used to compare cluster analysis partitions with the "correct" one throughout this chapter.

MIKCA, minimizing trace W: Partitions obtained using trace W are also shown in Table 1D. At the 3-group level, partitions obtained from trace W and det W are identical. Indeed, it will be seen below that most of the clustering alogorithms produced the same 3-group partition. The 4- and 5-group partitions from trace W are clearly inferior to those from det W, however. The trace W criterion was unable to separate clusters 1 and 3 in any of the partitions.

WARD: Figure 3D shows the tree diagram from Ward's minimum variance method. Many short branch lines occur until the 5-group level. At that time branch lines become longer, so a user would be most likely to select that

Table 1D

Cluster Overlap Matrices for Data set D
MIKCA: 3, 4, and 5 Group Partitions

(a) All MIKCA, 3-group partitions (det W and trace W)

MIKCA Partition	Generated Partition 1	2	3	4	Total
1	20	0	20	0	40
2	0	20	0	0	20
3	0	0	0	20	20
Total	20	20	20	20	

(b) det W, Mahalanobis D, 4 groups

MIKCA Partition	Generated Partition 1	2	3	4	Total
1	20	0	0	0	20
2	0	20	0	0	20
3	0	0	20	0	20
4	0	0	0	20	20
Total	20	20	20	20	

(c) trace W, Euclidean d, 4 groups

MIKCA Partition	Generated Partition 1	2	3	4	Total
1	20	0	20	0	40
2	0	5	0	0	5
3	0	15	0	0	15
4	0	0	0	20	20
Total	20	20	20	20	

(d) det W, Mahal. D and Euclid. d, 5 groups

MIKCA Partition	Generated Partition 1	2	3	4	Total
1	20	0	0	0	20
2	0	20	0	0	20
3	0	0	8	1	9
4	0	0	12	0	12
5	0	0	0	19	19
Total	20	20	20	20	

(e) trace W, Euclidean d, 5 groups

MIKCA Partition	Generated Partition 1	2	3	4	Total
1	20	0	20	0	40
2	0	5	0	0	5
3	0	15	0	0	15
4	0	0	0	13	13
5	0	0	0	7	7
Total	20	20	20	20	

Table 2D

Cluster Overlap Matrices for Data Set D
Hierarchical Methods with Distance-Based Similarity Coefficients

(a) WARD-d^2, 5 groups

WARD Partition	Generated Partition 1	2	3	4	Total
1	20	0	20	7	47
2	0	7	0	0	7
3	0	4	0	0	4
4	0	8	0	0	8
5	0	0	0	13	13
Outliers		1			1
Total	20	20	20	20	

(b) UPGMA-d, 5 groups*

UPGMA Partition	Generated Partition 1	2	3	4	Total
1	20	0	0	0	20
2	0	15	0	0	15
3	0	4	0	0	4
4	0	0	20	0	20
5	0	0	0	18	18
Outliers		1		2	3
Total	20	20	20	20	

(c) WARD-d^2, 4 groups

WARD Partition	Generated Partition 1	2	3	4	Total
1	20	0	20	7	47
2	0	15	0	0	15
3	0	4	0	0	4
4	0	0	0	13	13
Outliers		1			1
Total	20	20	20	20	

(d) UPGMA-d and Gower, and HMODE-d and Gower, 4 groups

UPGMA/HMODE Partition	Generated Partition 1	2	3	4	Total
1	20	0	20	0	40
2	0	5	0	0	5
3	0	15	0	0	15
4	0	0	0	20	20
Total	20	20	20	20	

(e) WARD-d^2, 3 groups

WARD artition	Generated Partition 1	2	3	4	Total
1	20	0	20	20	60
2	0	5	0	0	5
3	0	15	0	0	15
Total	20	20	20	20	

(f) UPGMA and HMODE, using d and Gower coefficients: 3 group partitions identical to those for MIKCA in Table 1D (a).

* Other UPGMA & HMODE 5-group results are very similar. Differences are:
UPGMA-Gower had two more outliers.
HMODE-Gower misclassified 6 points of cluster 3 with group 1.
HMODE-d misclassified 1 point of cluster 3 with group 1.

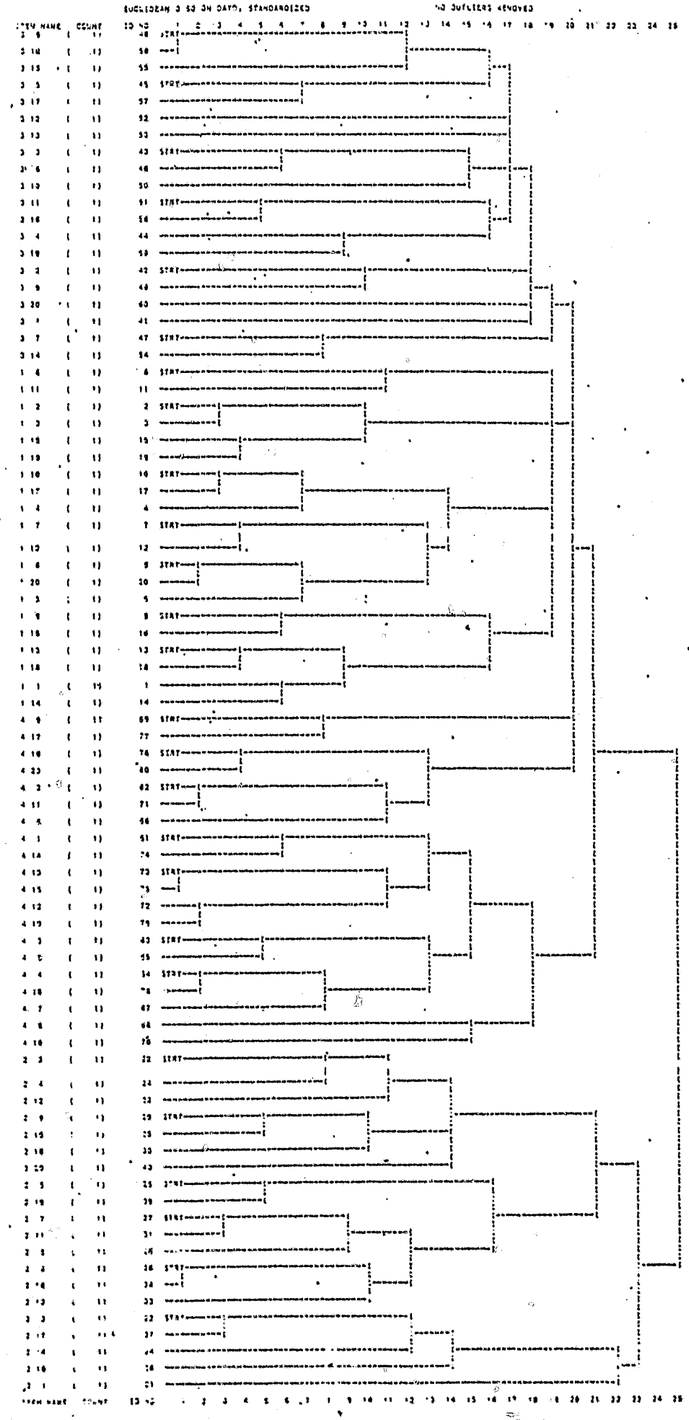


Figure 3D
TREE DIAGRAM FOR DATA SET D
WARD METHOD WITH EUCLIDEAN DISTANCE

level. The 5, 4, and 3-group partitions from the tree are shown on the left side of Table 2D, for comparison with the results of other methods. It can be seen that clusters 1 and 3 merge into a single group at a deceptively early stage of the WARD tree, while the highly elliptical and rather sparse cluster 2 does not come together into a single group until very late. It would be virtually impossible for a user to recognize clusters 1 and 3 as separate groups based upon this tree.

UPGMA and HMODE with Distance-Based Similarity Coefficients: The UPGMA and HMODE partitions on set D using d and Gower's coefficients are so similar that they will be discussed together. They produced identical partitions at the 4- and 3-group levels of their respective trees (which were the same as the MIKCA, trace W 4- and 3-group partitions). At the 5-group level, differences are minor; however they differ considerably from the MIKCA 5-group partitions.

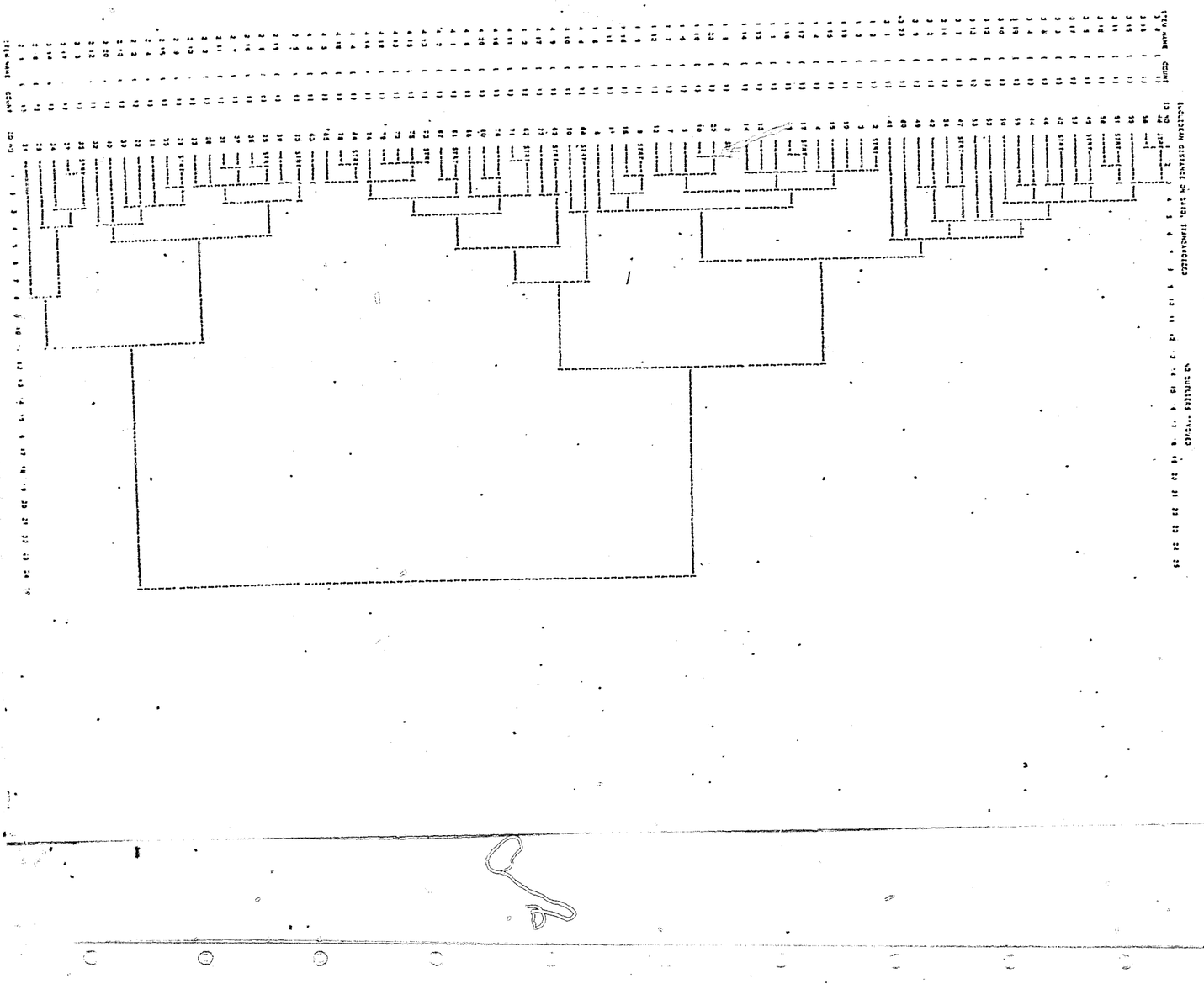
The trees for UPGMA-d and Gower, and HMODE-d and Gower are in Figures 4D-7D. The user looking for as much detail as possible would probably select the 5-group level of the UPGMA trees, thereby getting clusters 1, 3, and 4 nearly intact (except for a few outliers) and two groups representing the two ends of the ellipsoidal cluster 2. This partition is shown in Table 2D(b). It can be considered a good replication of the generated partition. However, the 4-group partition is not nearly as good because clusters 1 and 3 merge before the two segments of cluster 2 do.

The HMODE trees give the impression of either 4 or 3 groups, placing clusters 1 and 3 in the same group in either case. (Recall that HMODE trees show only the "secondary clusters" or modes, not the individual points.)

Of all the distance-based trees, the one from UPGMA-d (Figure 4D)

Figure 4D

TREE DIAGRAM FOR DATA SET D
GROUP AVERAGE METHOD WITH EUCLIDEAN DISTANCE



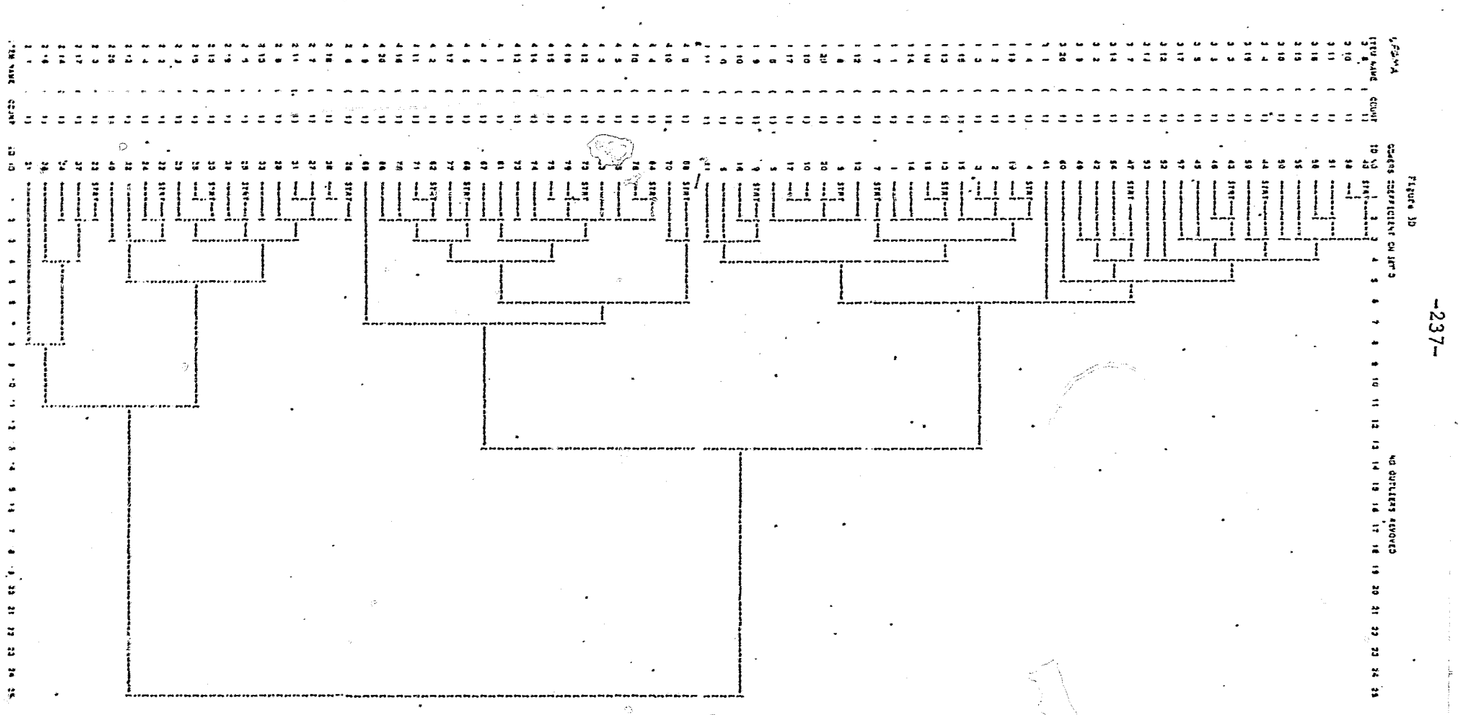
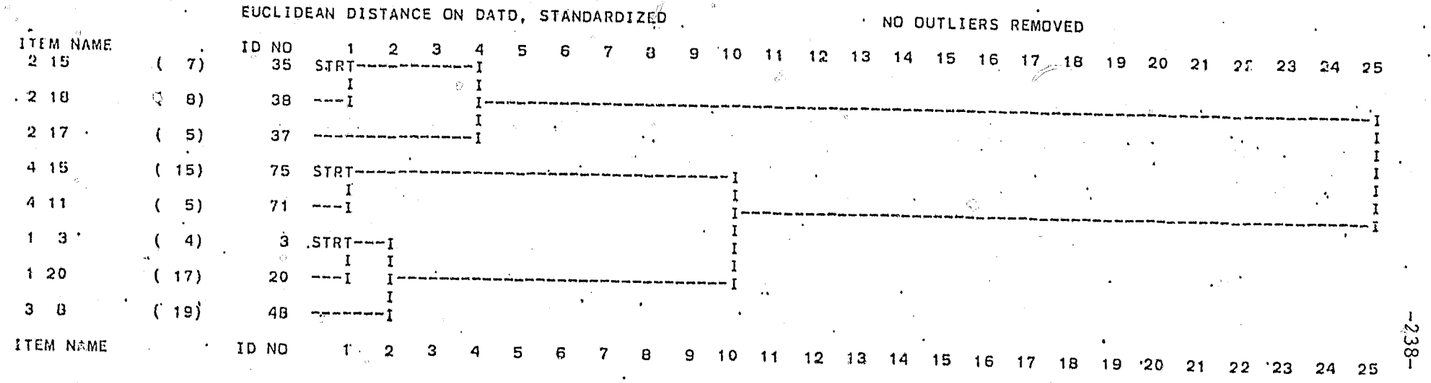


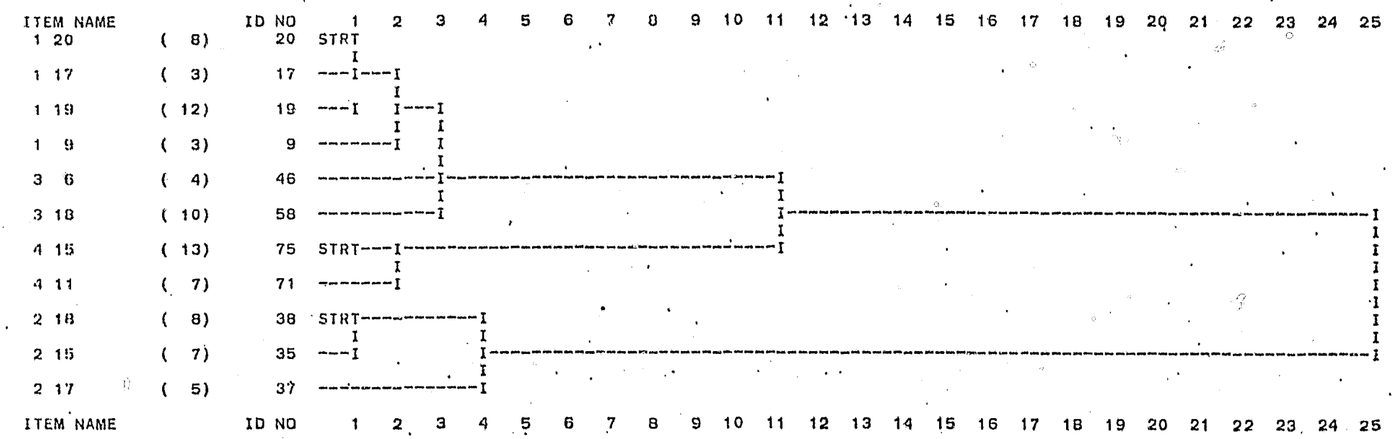
Figure 6D

TREE DIAGRAM FOR DATA SET D
HIERARCHICAL MODE SEARCH (HMODE) WITH EUCLIDEAN DISTANCE
(Tree shows merges of primary clusters only)



HMODE

GOWERS COEFFICIENT ON SET D — NO OUTLIERS REMOVED



certainly gives the most accurate impression of the structure of this data set. In Table 2D it can be seen that, at each level, the UPGMA and HMODE results are better than the WARD results. All three of these methods, though, had difficulty caused by the highly ellipsoidal shape of cluster 2: all split off the same 5 points of that cluster into a separate group. This, combined with the fact of relatively small separation between the other three clusters, produced the result that other clusters merged before cluster 2 came completely together.

UPGMA and HMODE with Angular Similarity Coefficients: The trees for UPGMA-cosine and r, and HMODE-cosine and r are given in Figures 8D-11D.. Again, the trees give somewhat different impressions as to the number of groups in the data. Likely choices from the separate trees would appear to be

Tree	# of groups (containing at least 3 points)
UPGMA-cosine	3 or 4 groups
HMODE-cosine	3 or 5 groups
UPGMA-r	3 or 4 groups
HMODE-r	2 or 4 groups

which may be compared with

UPGMA-d	5 groups
HMODE-d	3 or 4 groups

Thus comparing trees from different methods may give a rough idea, but not a precise one, of the number of populations involved in the data.

Of all the hierarchical techniques, only UPGMA-cosine produced a perfect 4-group solution (identical to the MIKCA, det W partition in Table 1D(b)). The 3-group cosine partitions from both UPGMA and HMODE were identical to the 3-group MIKCA partitions. In fact, all methods except WARD and those using the correlation coefficient produced the same 3-group partition.

Table 3D

Cluster Overlap Matrices for Data Set D
UPGMA and HMODE with Angular Similarity Coefficients

(a) UPGMA-cosine, 5 groups

UPGMA Partition	Generated Partition 1	2	3	4	Total
1	20	0	0	0	20
2	0	20	0	0	20
3	0	0	15	0	15
4	0	0	4	0	4
5	0	0	0	20	20
Outliers	—	—	1	—	1
Total	20	20	20	20	

(b) HMODE-cosine, 5 groups

HMODE Partition	Generated Partition 1	2	3	4	Total
1	20	0	1	0	21
2	0	20	0	0	20
3	0	0	8	0	8
4	0	0	11	0	11
5	0	0	0	20	20
Total	20	20	20	20	

(c) UPGMA-r, 4 groups

UPGMA Partition	Generated Partition 1	2	3	4	Total
1	20	0	0	0	20
2	0	20	0	0	20
3	0	0	6	19	25
4	0	0	13	0	13
Outliers	—	—	1	1	2
Total	20	20	20	20	

(d) HMODE-4, 4 groups

HMODE Partition	Generated Partition 1	2	3	4	Total
1	18	0	0	0	18
2	0	20	0	0	20
3	0	0	10	20	30
4	0	0	6	0	6
Outliers	2	—	4	—	6
Total	20	20	20	20	

(e) UPGMA-r, 3 groups

UPGMA Partition	Generated Partition 1	2	3	4	Total
1	20	0	6	19	45
2	0	20	0	0	20
3	0	0	13	1	14
Outliers	—	—	1	—	1
Total	20	20	20	20	

Figure 9D

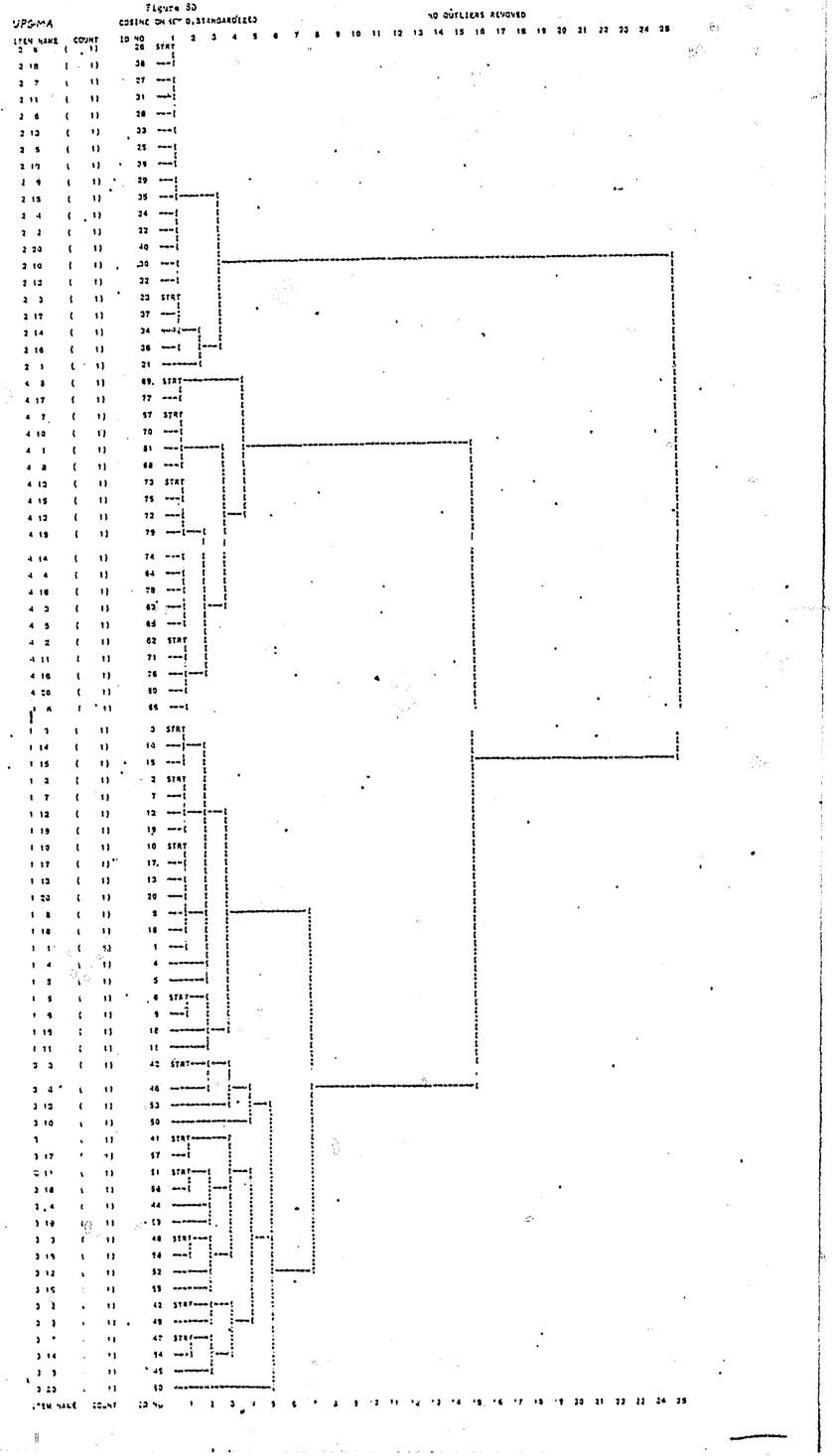
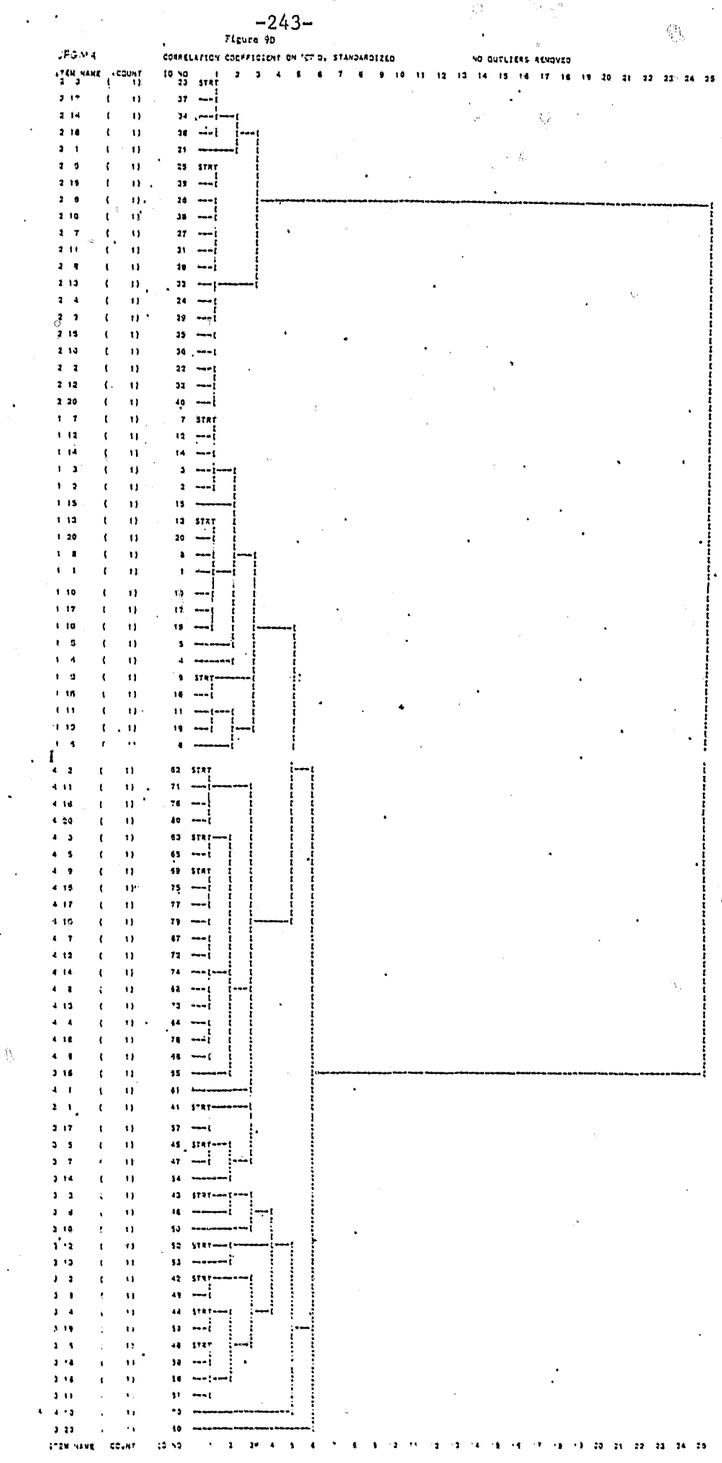


Figure 9D



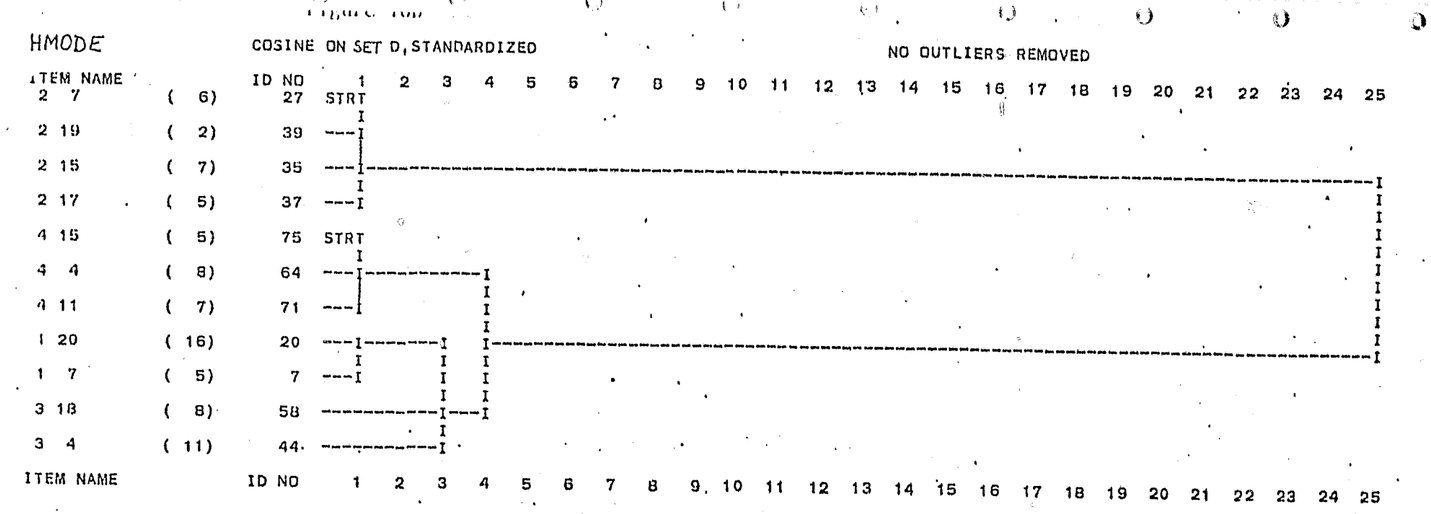
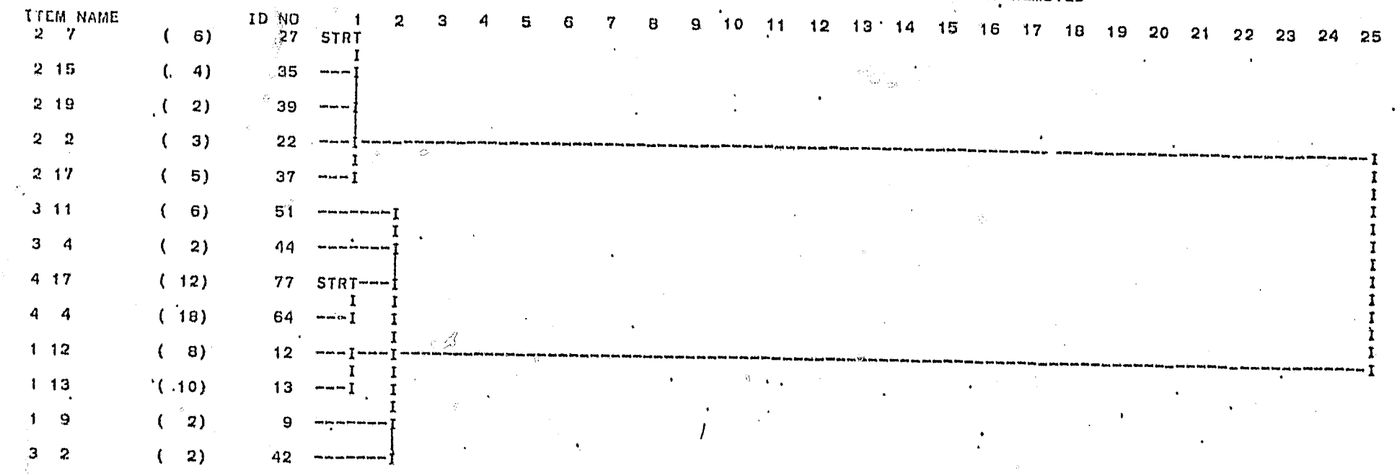

HMODE
COSINE ON SET D, STANDARDIZED
NO OUTLIERS REMOVED
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
2 7 (6) 27 STRT
2 19 (2) 39
2 15 (7) 35
2 17 (5) 37
4 15 (5) 75 STRT
4 4 (8) 64
4 11 (7) 71
1 20 (16) 20
1 7 (5) 7
3 18 (8) 58
3 4 (11) 44
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 11D

HMODE

CORRELATION COEFFICIENT ON SET D, STANDARDIZED — NO OUTLIERS REMOVED



CONTINUED

13 OF 9

Figure 12D

CLUSTER GRAPH

DATA SET D
UPGMA - R
X=VAR 1 Y=VAR 4

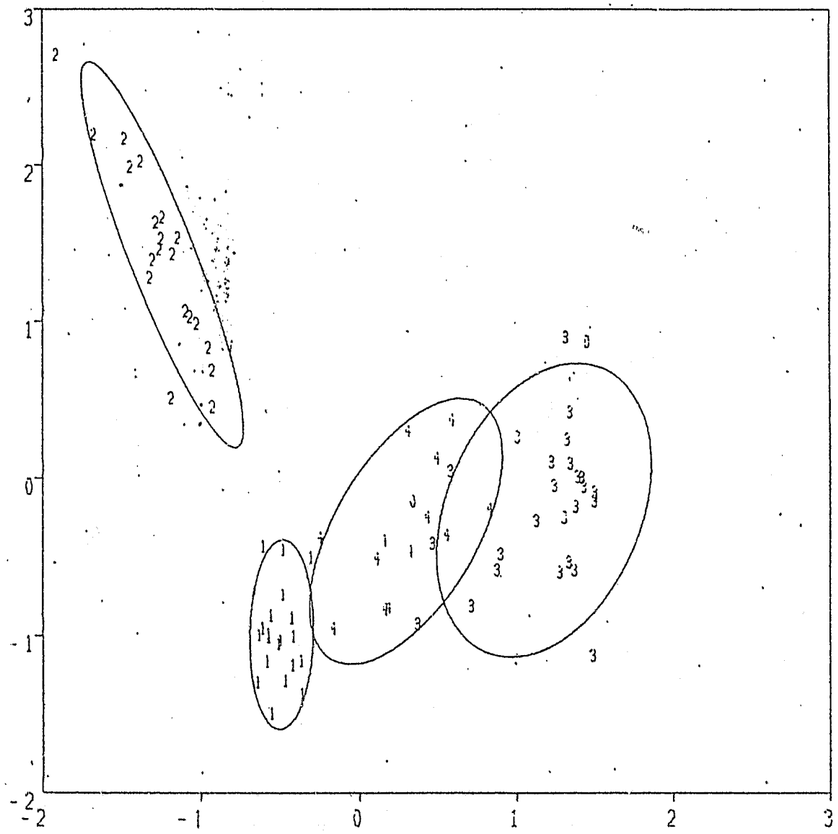


Figure 13D

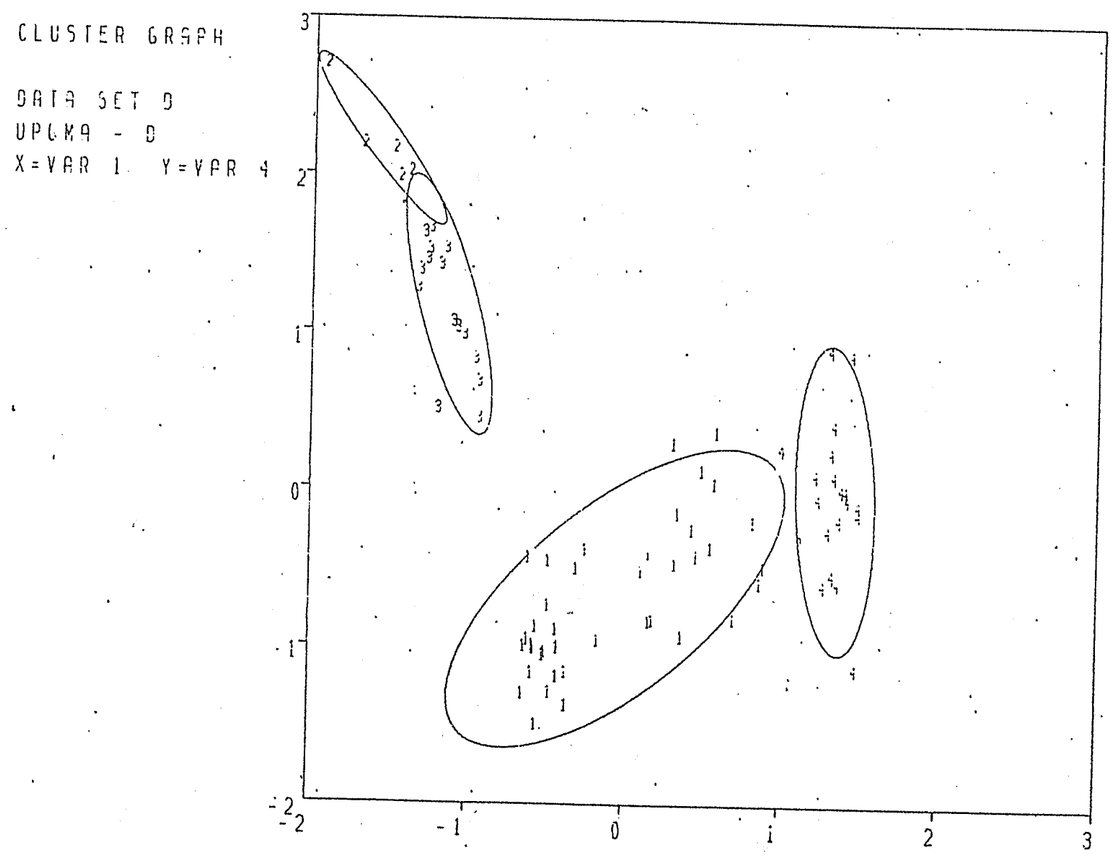


Table 3D shows a few of the partitions produced by r and cosine. The angular coefficients do not have the difficulty with the ellipsoidal cluster 2 that distance coefficients have, because cluster 2 does not subtend a large angle in most dimensions. In fact, cluster 2 is completed very early in the trees which involve r and cosine. On the other clusters, however, there is considerable difference in results produced by the two angular coefficients. Correlation tends to lump points from clusters 3 and 4 together, while the other coefficients do not. Its results at the 4-group level are reasonably good; but at the 3-group level, partitions obtained from r differ markedly from those of other methods.

Graphs showing the 4-group partitions produced by UPGMA-r and UPGMA-d are shown in Figures 12D and 13D respectively, to give a visual comparison of clustering results.

NMODE: In data set D, 3 modes were found--the same 3 groups identified by most other methods. The NMODE partition is the same as that in TAble 1D(a) with the exception that 4 points were left as outliers.

Comparison of the Partitions

The greatest agreement between the several clustering methods was obtained at the 3-group level. For some research purposes, this 3-group partition--shown in Table 1D(a)--might be satisfactory. This result indicates, however, that mere agreement between methods is not an adequate indication that all underlying populations have been found. Indeed, it is quite possible to find the "correct" 4 populations in this data, as evidenced by the success of the MIKCA, det W and UPGMA-cosine 4-group partitions. If research purposes suggest that too many clusters would be preferable over too few, then the UPGMA-d and HMODE-d 5-group partitions might also be considered better than the 3-group one.

An evaluation of the individual clusters obtained at the 3-, 4-, and 5-group levels will be undertaken in the next section. First, at the 4-group level, the clustering methods will be rank-ordered according to their success in duplicating the "correct" partition. The rank-ordering is shown in Table 4D, and is based primarily on the number of points misclassified by each method. (In this particular data set there was no difficulty in establishing a "best" match between the "correct" partition and each of the C.A. partitions --a problem which arises in other data sets to be discussed later.)

Table 4D includes values of the four coefficients of partition agreement which are described elsewhere. Each measures the agreement between the cluster analysis partition and the generated partition. All four coefficients rank the 4-group partitions in the same order that we did looking at number of points misclassified. Also included in Table 4D are coefficients of agreement for selected partitions not containing 4 groups. Within each k-level, the coefficients agree on the rank-ordering of the partitions (with one exception by Cramer's coef. at the 5-group level).

We are interested in judging the usefulness of these coefficients for comparing partitions containing different numbers of groups. The usefulness of Cramer's coef. must be questioned, since it gave a perfect agreement score of 1.0 to partitions containing 2, 3, 4, and 5 groups. This hardly fits our concept of a "perfect match" between partitions, which would reserve the top score of 1.0 for partitions that are exactly alike. The other three coefficients, although they differ in absolute magnitudes, produce very similar rankings of the partitions even when all of the k-levels are considered together. Furthermore, they agree with our assessment that the best of the 5-group partitions are better than any of the 3-group partitions.

Measures of Partition Agreement on Data Set D
with Success Ratings for the Four-Group Partitions

4-group partitions, in order of success in duplicating the clusters as generated	# of points misclassified (minimum)	Points left as outliers	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
1. UPGMA, cosine and MIKCA, det W, Mahal. D	0	0	1.000	1.000	1.000	1.000
2. MIKCA, det W, Euclidean d	1	0	.988	.950	.984	.983
3. UPGMA, r	6	2	.927	.736	.912	.870
4. HMODE, cosine	9	0	.912	.703	.891	.838
5. HMODE, r	10	6	.884	.620	.856	.782
UPGMA, both d and Gower						
6. HMODE, both d and Gower	25	0	.850	.591	.816	.750
MIKCA, trace W, d						
7. WARD, d^2	31	1	.731	.410	.720	.581
NMODE, d; no 4-group partition						

Other partitions, ordered by # of groups; within that, ranked according to Lambda	# of Groups	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
HMODE, r	2	.620	.388	1.000	.500
MIKCA, det W and trace W					
UPGMA with d, Gower, and cos					
HMODE with d, Gower, and cos	3	.873	.655	1.000	.800
NMODE, d					
UPGMA, r	3	.850	.593	.801	.750
WARD, d^2	3	.765	.464	.717	.642
	3	.597	.349	.707	.438
UPGMA, cosine					
HMODE, d	5	.975	.896	1.000	.958
UPGMA, d	5	.964	.854	.984	.941
MIKCA, det W, both D and d	5	.963	.846	.989	.933
HMODE, cosine	5	.961	.840	.985	.917
MIKCA, trace W, d	5	.960	.837	.984	.916
WARD, d^2	5	.821	.512	.816	.680
	5	.713	.371	.720	.505
HMODE, Gower	6	.899	.637	.920	.816

Those same 5-group partitions are also ranked higher than most of the 4-group partitions. This reflects the fact that, while a cluster had to be split at the 5-group level, there were generally fewer points misclassified at that level than at the 4-group level. Therefore, in this data set, we have no quarrel with the rankings produced by the Rand, Jaccard, and Lambda coefficients.

Evaluation of Individual Clusters

To assess the usefulness of several measures intended to evaluate individual partitions, see Tables 5D-8D. These contain, for 4 partitions, Huizinga's gamma coefficients for cluster separation, and measures of the homogeneity of the clusters on each of the variables. Table 5D gives these measures for the "correct" partition. The negative gammas for this partition result from the small separation between some of the generated clusters, and the large diameter of cluster 2.

Table 6D represents one of the poorer 4-group partitions, that produced by UPGMA and HMODE using d as the similarity measure. This partition was shown in Table 2D(d): clusters 1 and 3 are merged into a single group, while cluster 2 is split. We would consider it a poor representation of the structure of the data set, yet the gamma coefficients are higher for this partition than for the correct one because the maximum cluster diameter is smaller and the separations between groups are larger. Furthermore, Tables 7D and 8D show a 3-group and a 5-group partition, both of which gave higher gammas than the correct partition. This indicates that the gammas should not be used to compare two or more partitions, but only to assess compactness and separation after the "best" partition has been selected by other means.

The measures of within-cluster homogeneity of variables are also shown in Tables 5D-8D. They indicate very high homogeneity in all four partitions, however it can be seen that, quite predictably, overall homogeneity tends to decrease as the number of groups decreases.

An assessment of the usefulness of these evaluative measures, as well as the coefficients of partition agreement, will be taken up at the end of this chapter after other data sets have been considered.

Table 5D
Generated partition, 4 clusters
(Duplicated by UPGMA-cos and MIKCA, det W-D)

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	1053	1347	210	550

GAMMA -.2006

GAMMAB -.1225

GAMMAW -.4474

MAX. CLUSTER DIAMETER 5.3438
MIN. DIST. BTWN BOUNDARIES .8664

CLUSTER DIAMETERS

1	2	3	4
1.433	5.344	2.259	3.424

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4
1	.990	.938	.902	.985
2	.965	.753	.933	.828
3	.940	.732	.953	.857
4	.915	.644	.849	.773
5	.934	.532	.889	.823
6	.970	.844	.835	.967
7	.945	.812	.908	.967
8	.931	.775	.852	.976

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.954	.870	.870	.795	.794	.904	.908	.883

CLUSTER HOMOGENITIES

1	2	3	4
.949	.754	.890	.897

Table 6D
UPGMA-d and Gower, and
HMODE-d and Gower, 4 groups

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	1352	723	425	660

GAMMA .1247

GAMMAB .3031

GAMMAW -.2166

MAX. CLUSTER DIAMETER 3.4235
MIN. DIST. BTWN BOUNDARIES 1.1351

CLUSTER DIAMETERS

1	2	3	4
2.833	2.341	2.449	3.424

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4
1	.739	.953	.981	.985
2	.924	.848	.947	.828
3	.876	.921	.910	.857
4	.790	.917	.837	.773
5	.831	.569	.890	.823
6	.905	.916	.916	.967
7	.924	.954	.950	.967
8	.893	.849	.906	.976

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.859	.899	.880	.802	.824	.923	.941	.914

CLUSTER HOMOGENITIES

1	2	3	4
.860	.866	.917	.897

Table 7D

Most frequent 3-group partition
(Duplicated by MIKCA, UPGMA, and HMODE
using all coefficients except r)

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	1053	947	476	684
GAMMA -.0323				
GAMMAB .0530				
GAMMAW -.1793				

MAX. CLUSTER DIAMETER 5.3438
MIN. DIST. BTWN BOUNDARIES 1.2093

CLUSTER DIAMETERS

1	2	3
2.833	5.344	3.424

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3
1	.739	.938	.985
2	.924	.753	.828
3	.876	.732	.857
4	.790	.644	.773
5	.831	.532	.828
6	.905	.844	.967
7	.924	.812	.967
8	.893	.775	.976

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.850	.857	.835	.749	.754	.905	.906	.884

CLUSTER HOMOGENITIES

1	2	3
.860	.754	.897

Table 8D
HMODE-d, 5-group partition

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3; N4, ...	1352	1122	266	420

GAMMA .0241

GAMMAB .0930

GAMMAW -.2245

MAX. CLUSTER DIAMETER 3.4235
MIN. DIST. BTWN BOUNDARIES .9293

CLUSTER DIAMETERS

1	2	3	4	5
1.704	2.341	2.449	2.259	3.424

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5
1	.985	.953	.981	.915	.985
2	.944	.848	.947	.937	.828
3	.933	.921	.910	.951	.857
4	.919	.917	.837	.859	.773
5	.928	.569	.890	.884	.823
6	.966	.916	.916	.833	.967
7	.947	.954	.950	.903	.967
8	.927	.849	.906	.850	.976

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.966	.908	.913	.853	.862	.922	.913	.912

CLUSTER HOMOGENITIES

1	2	3	4	5
.944	.866	.917	.891	.897

DATA SET E:

DIFFERENTLY ORIENTED ELLIPSES, WITH MINOR CLUSTER OVERLAP

Description

Data set E originally contained 200 points measured on 3 variables. Of those points, 160 were drawn from 3 multi-variate normal subpopulations to represent 3 clusters of different sizes, while the other 40 were drawn from a uniform distribution over the ranges of the 3 variables to add "noise" to the data by providing some unclassifiable individuals. All 3 clusters are ellipsoidal in shape, with different amounts of within-cluster covariance.

The most isolated 10% of the points, based on their distance from all other points, were removed as "outliers" using program SIMOUT. Thus they were labelled as unclassifiable and were not included in any of the clustering attempts. Of the 20 removed, 19 came from the uniform "noise" distribution and 1 came from cluster 2. This indicates a particular kind of difficulty in the prior removal of unclassifiable individuals.

When subjected to cluster analysis, then, data set E contained 180 points: 3 clusters of 100, 29, and 30 points respectively, plus 21 "noise" points which may or may not fall within the territory of the clusters.

Parameters for the 3 subpopulations of the three variables are as follows:

Cluster 1 (n = 100)

Centroid: (0, 0, 0)

Standard deviations: $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 1$

No within-cluster correlation.

Cluster 2 (n = 29)

Centroid: (-6, 3, 6)

Standard deviations: $\sigma_1 = 2$, $\sigma_2 = 2$, $\sigma_3 = 1$

Correlation matrix:

1.0	- .8	- .1
-.8	1.8	0
-.1	0	1.0

Cluster 3 (n = 30)

Centroid: (6, 6, 4)

Standard deviation = 2 on each variable

Correlation matrix:

1.0	.8	.7
.8	1.0	.6
.7	.6	1.0

Figure 1E shows data set E in 2 dimensions, variable 1 by variable 2. In the third variable, cluster 2 is separated from cluster 1, while cluster 3 is not.

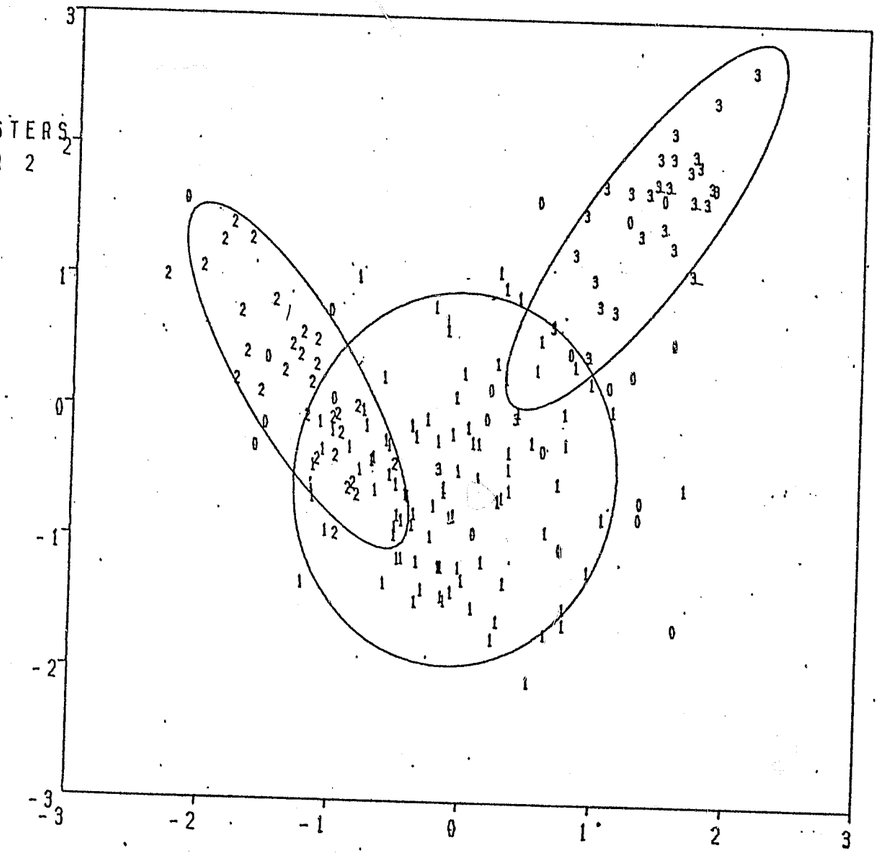
Shape, Orientation And Compactness: Cluster 1 is an ellipsoid whose axes are parallel to the major axes. It has a higher point density than the other two clusters.

Clusters 2 and 3 are oriented at different angles in space, with cluster 3 having more within-cluster covariance.

Separation: Cluster 1 and 3 overlap on all three variables, thus the boundary between them is fuzzy. Clusters 1 and 2 are separated on variable 3, clusters 2 and 3 are separated on variable 1.

CLUSTER GRAPH

DATA SET E
GENERATED CLUSTERS
X=VAR 1 Y=VAR 2



Discriminant analysis based on the three correct clusters was able to classify 98.7% of the cases correctly (not including the 21 unclassified "noise" points).

Point allocations are shown in the following table:

Predicted Group	Actual Group 1	2	3	Noise
1	100	0	2	7
2	0	29	0	8
3	0	0	28	6
Total	100	29	30	21

Most points (95.0%) were classified in their correct group with at least a .9 probability. All of six points which were assigned correctly but with less than a .9 probability were in the boundary region between populations 1 and 3.

A plot of this data on the first two discriminant functions is shown in Figure 2E. The fuzzy boundary between clusters 1 and 3 leaves membership in some doubt for, perhaps, 7 points. One point from population 3 clearly would be more appropriately assigned to cluster 1. (The 4's in the plot denote noise points.)

Clustering Results (Data Set E)

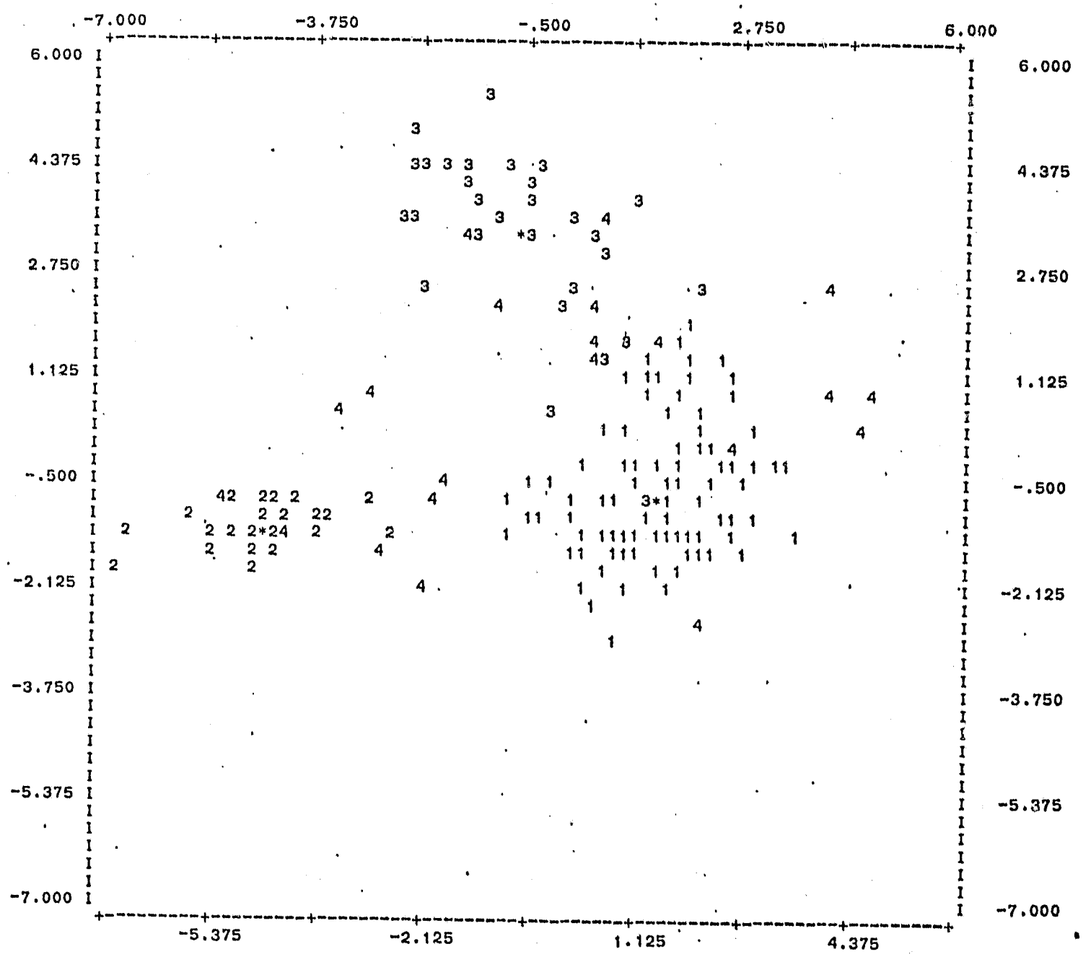
MIKCA, Minimizing Determinant W: Solutions containing 2, 3, and 4 groups were requested, using both Mahalonobis' D^2 and Euclidean distance as similarity measures. The match between these partitions and the correct one is shown in Table 1E (a)-(f).

DISCRIMINANT ANALYSIS ON SET E, AS GENERATED

Figure 2E

79/09/20. 14.19.12. PAGE 10

PLOT OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CENTROID.



MIKCA, Minimizing Trace W: The trace W criterion was tested on this particular data because of Everett's (1974) evidence that it would tend to split highly elliptical clusters. Solutions containing 3, 4, and 5 groups were requested, with Euclidean d as the similarity measure. The 3 and 4 group solutions are identical to the corresponding determinant W, Euclidean d solutions shown in Table 1E (d) and (f). The 5-group solution is shown in (g) of the table.

Observations: The 3-group solution using Euclidean d is excellent using either det W or trace W (Table 1E (d)). It allocates the classified points in the same way as discriminant analysis, with minor differences in the assignment of the noise points. The trace W criterion did not have trouble with the highly ellipsoidal clusters. Surprisingly, det W with Mahalanobis D^2 had more difficulty in the fuzzy boundary region between clusters 1 and 3. Its partition is unexpectedly similar to those described below in which cosine was used as the similarity coefficient.

In all of the 4-group solutions, cluster 1 is split while clusters 2 and 3 are nearly intact. Cluster 1 is, if anything, somewhat less ellipsoidal than the other two. It is possible that it was split because of its larger size (100 points). The highly ellipsoidal cluster 2 was not split until the 5-group level (g).

WARD: The minimum variance tree is shown in Figure 3E. Because short branch lines occur up to the last stages of the tree, the user would have difficulty deciding which cut to make. There is potential for solutions containing 2, 3, 6, 8, or 10 groups. The 3 and 6 group solutions will be examined here: they are matched with the correct partition in Table 2E (a) and (b).

Table 1E

Cluster Overlap Matrices for Data Set E
MIKCA: 2, 3, 4 and 5 cluster solutions

(a) 2 groups, det W, Mahalanolis D^2

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	90	29	2	14	135
2	10	0	28	7	45
Total	100	29	30	21	

(b) 2 groups, det W, Eudlidean d

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	100	29	4	17	150
2	0	0	26	4	30
Total	100	29	30	21	

(c) 3 groups, det W, Mahalanobis D^2

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	85	0	2	7	94
2	15	0	28	7	50
3	0	29	0	7	36
Total	100	29	30	21	

(d) 3 groups, Euclidean d
Identical solutions using det W & trace W

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	100	0	2	9	111
2	0	29	0	7	36
3	0	0	28	5	33
Total	100	29	30	21	

(e) 4 groups, det W, Mahalanobis D^2

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	43	0	0	5	48
2	57	0	4	5	66
3	0	29	0	7	36
4	0	0	26	4	30
Total	100	29	30	21	

(f) 4 groups, Euclidean d
Identical solutions using det W & trace W

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	53	0	0	5	58
2	47	0	4	5	56
3	0	29	0	7	36
4	0	0	26	4	30
Total	100	29	30	21	

(g) 5 groups, trace W, Euclidean d

MIKCA Partition	Generated partition 1	2	3	Noise	Total
1	72	0	1	3	76
2	28	0	4	6	38
3	0	17	0	4	21
4	0	12	0	4	16
5	0	0	25	4	29
Total	100	29	30	21	

Table 2E

Cluster Overlap Matrices for Data Set E
Hierarchical Methods with Distance-Based Similarity Measures

(a) WARD, Euclidean d^2, 6 groups

WARD Partition	Generated partition 1	2	3	Noise	Total
1	100	2	2	14	118
2	0	10	0	0	10
3	0	12	0	1	13
4	0	5	0	1	6
5	0	0	21	5	26
6	0	0	7	0	7
Total	100	29	30	21	

(b) WARD, Euclidean d^2, 3 groups

WARD Partition	Generated partition 1	2	3	Noise	Total
1	100	29	2	16	147
2	0	0	21	5	26
3	0	0	7	0	7
Total	100	29	30	21	

(c) UPGMA, Euclidean d, 6 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	10	0	0	4	14
2	72	0	1	1	74
3	15	0	8	4	27
4	0	29	0	5	34
5	0	0	20	2	22
6	0	0	1	4	5
Outliers	3	0	0	1	4
Total	100	29	30	21	

(d) UPGMA, Euclidean d, 4 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	100	0	9	10	119
2	0	29	0	5	34
3	0	0	20	2	22
4	0	0	1	4	5
Total	100	29	30	21	

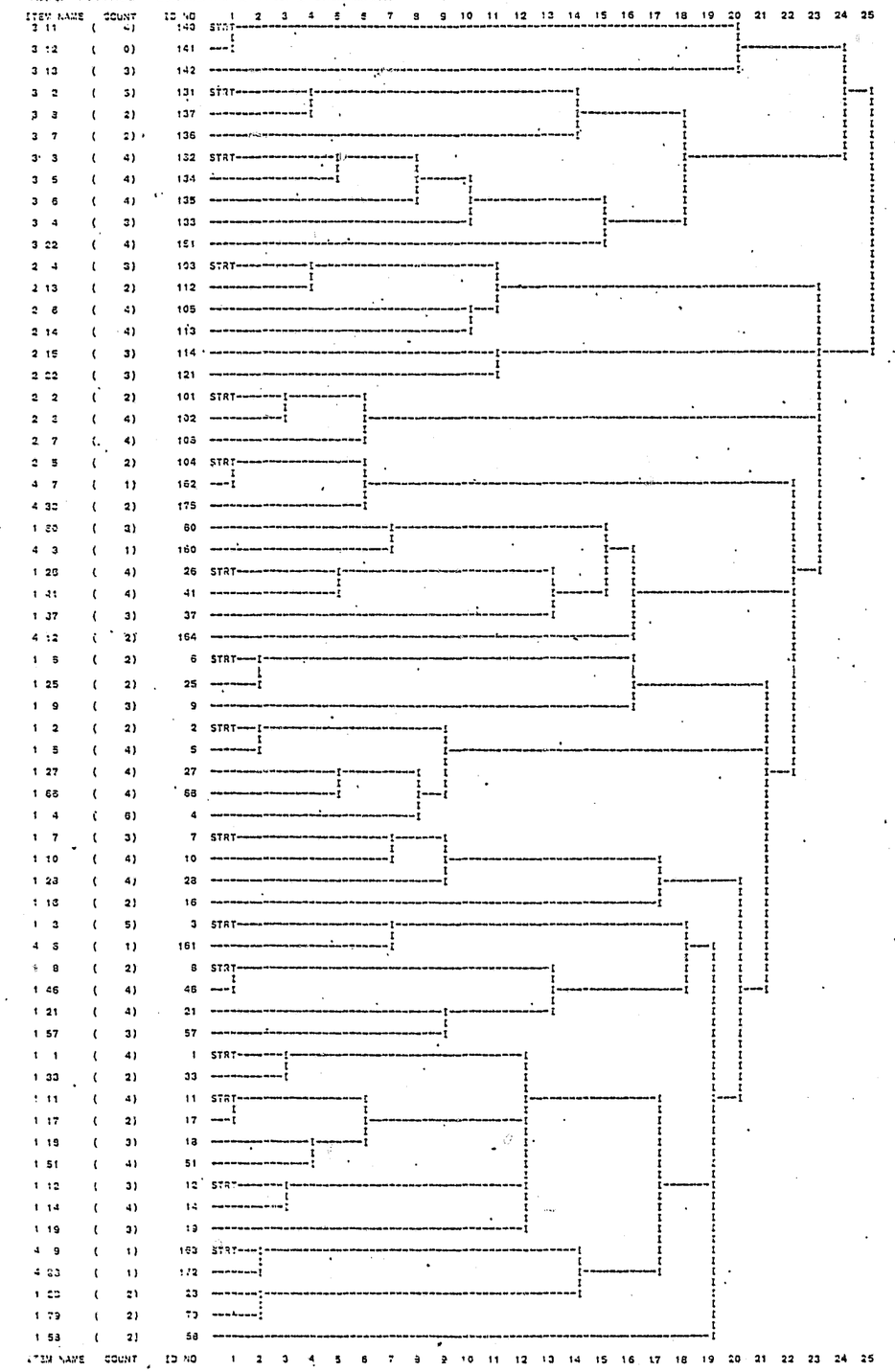
(e) UPGMA, Gower's coef., 5 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	83	0	1	5	89
2	17	0	8	5	30
3	0	29	0	5	34
4	0	0	20	2	22
5	0	0	1	4	5
Total	100	29	30	21	

(f) UPGMA, Gower's coef., 3 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	100	0	9	10	119
2	0	29	1	9	39
3	0	0	20	2	22
Total	100	29	30	21	

Figure 3E



The WARD tree differs from the other hierarchical trees to be discussed below in that cluster 1 comes together completely while the other two clusters are still in fragments--(a) in the table. Then pieces of cluster 2 are joined to cluster 1, one at a time, eventually producing the 3-group version (b). At no time does cluster 2 appear as a whole, distinct group. WARD was the only distance-based method that was unable to at least partially identify cluster 2 at the 3-group level. This could be due to either the ellipsoidal shape or the relatively low density of that cluster.

Because the Ward minimum variance method has been found by other researchers to perform well, we will look more closely at what happened in this data. The hierarchical progression is illustrated graphically in Figures 4E through 6E, which show the 10-group, 6-group, and 3-group WARD partitions. These may be compared with the correct clusters, shown in Figure 1E. In the graphs, the ellipse drawn for each group is based on a multiple of the mean Mahalanobis D^2 from the points to the group centroid, therefore the ellipse generally encloses most, but not all, of the points in the group. These graphs depict only two variables of the 3-dimensional space, so it must be remembered that population 2 is separated from population 1 in the third variable (Figure 1E).

Throughout the three WARD graphs, the two groups in the upper right-hand corner remain the same, containing pieces of population 3. The large group 1 is much more densely populated than any of the small groups. At each step of the hierarchy it swallows up one more of the small groups. Likely reason: When a very small group merges with an adjacent very large group there will be little increase in variance in the new slightly larger group. However, if

Figure 4E

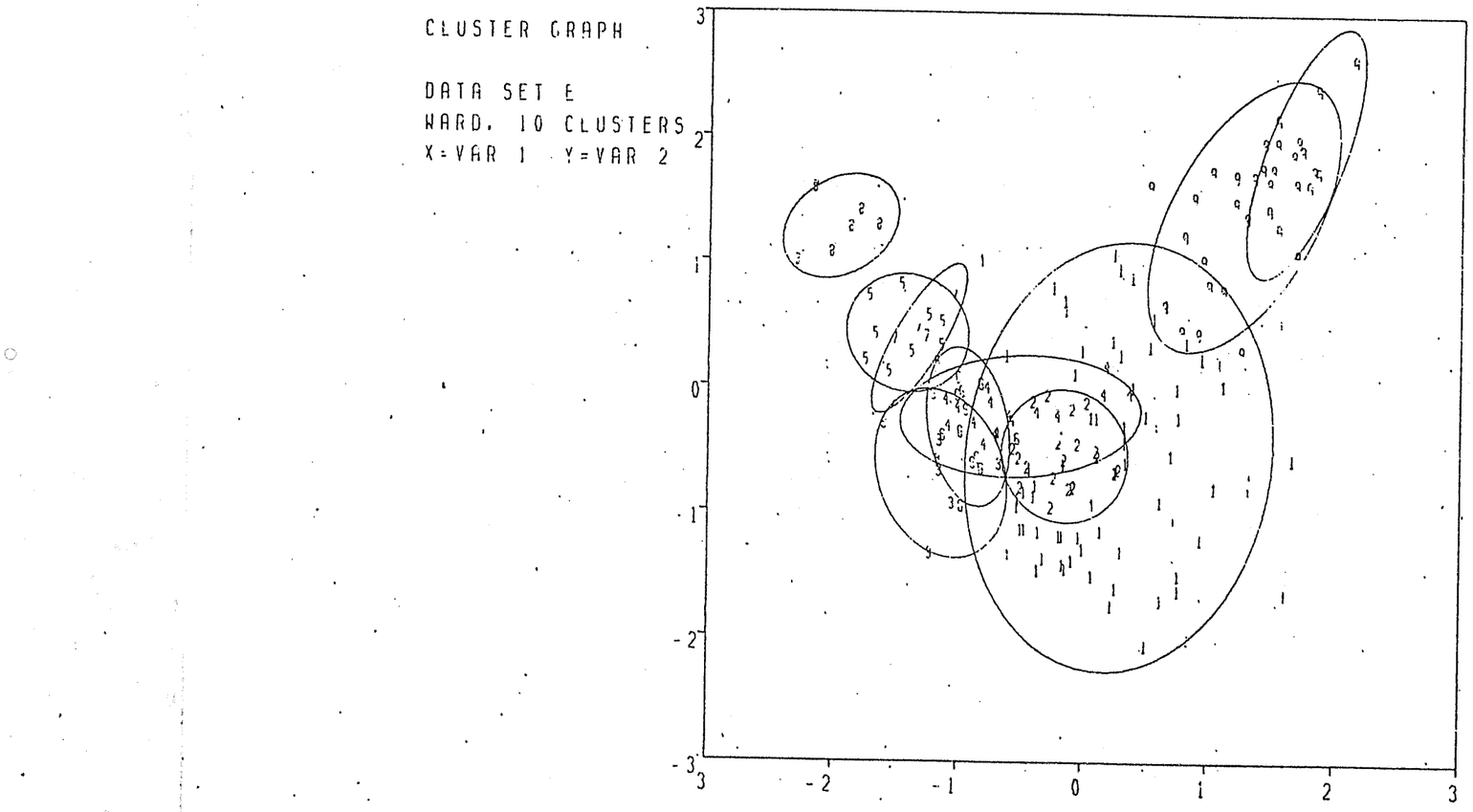

CLUSTER GRAPH
DATA SET E
WARD. 10 CLUSTERS
X=VAR 1 Y=VAR 2
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

Figure 5E

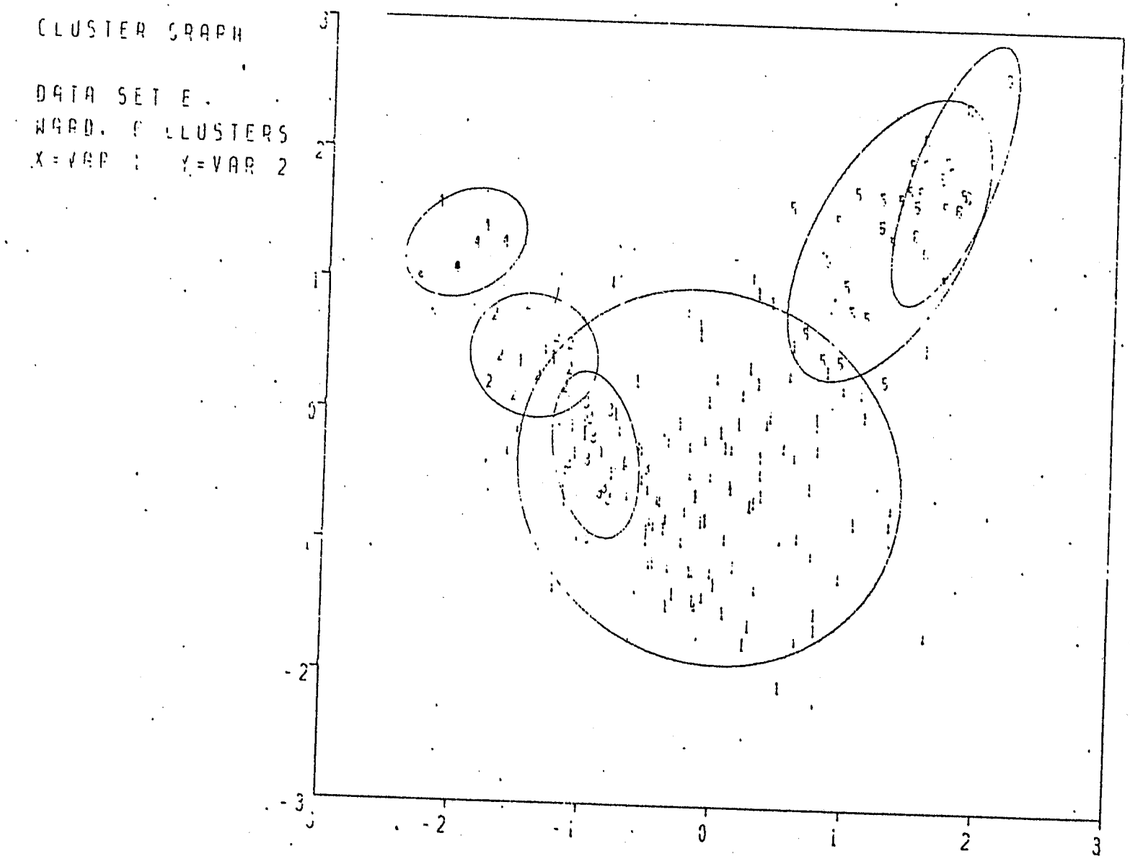
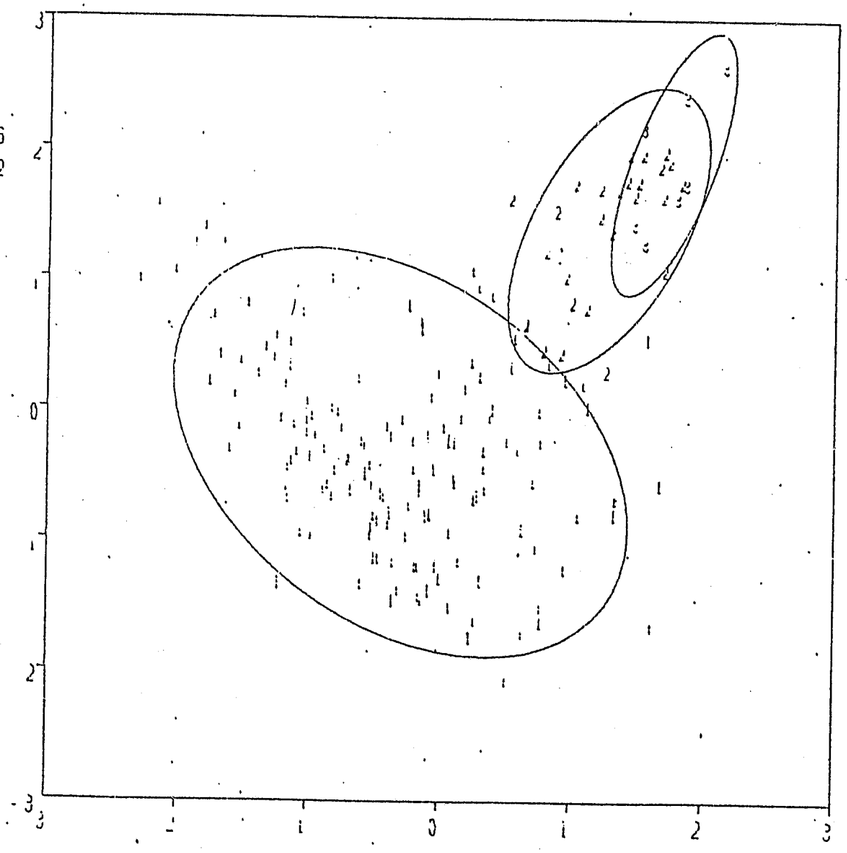

CLUSTER GRAPH
DATA SET E.
WARD. 6 CLUSTERS
X=VAR 1 Y=VAR 2
3
2
1
0
-2
-3
-2
-1
0
1
2
3

Figure 6E


CLUSTER GRAPH
DATA SET E
WARD. 3 CLUSTERS
X=VAR 1 : Y=VAR 2
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

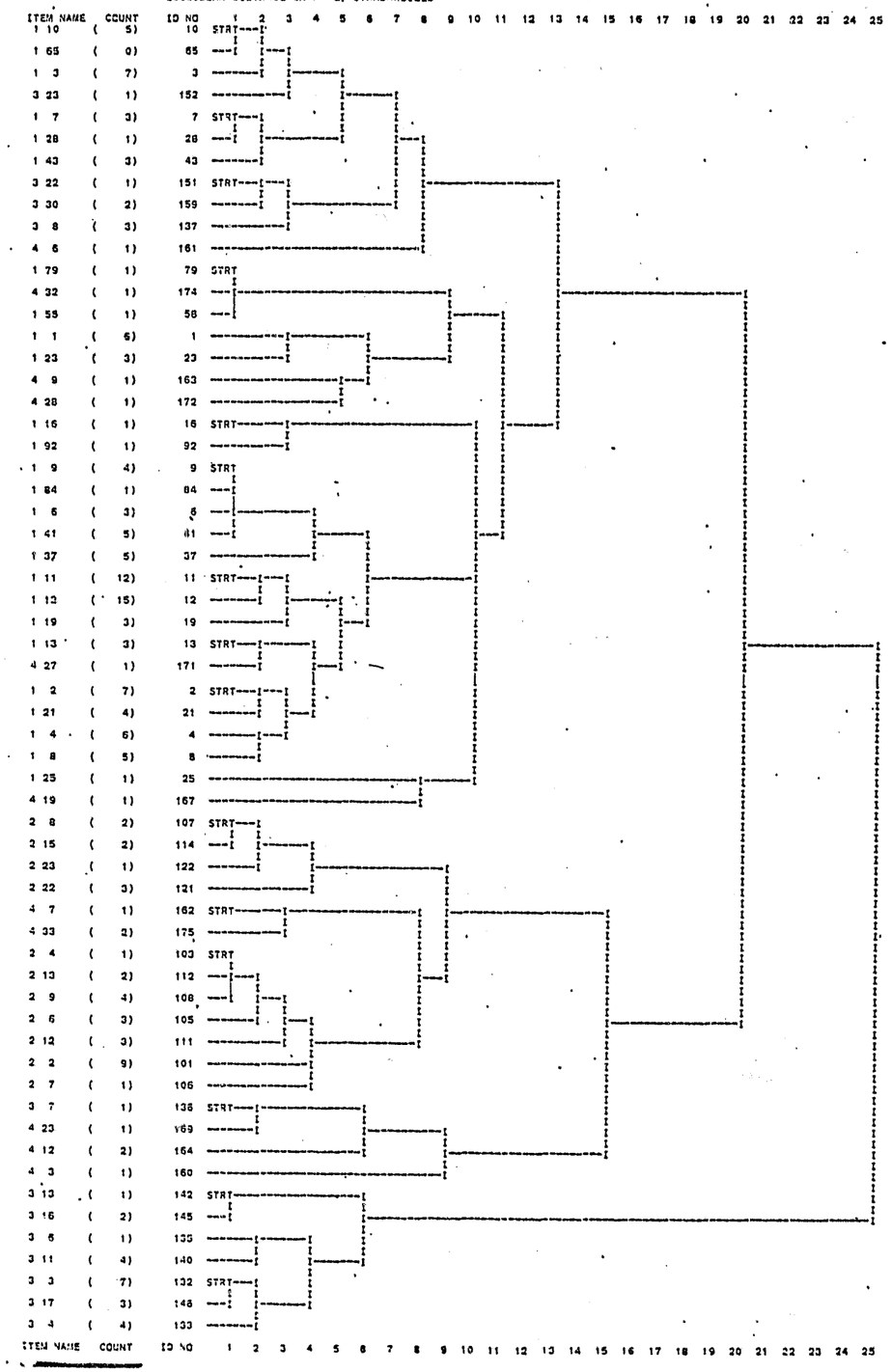
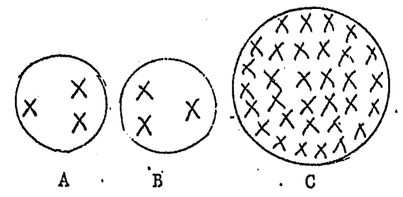


Figure 7E

that small group were to merge with an adjacent small group there would be a much greater impact on the total variance, even if the small groups are somewhat closer together. This can be depicted in a diagram:



Up to some threshhold separation based on the numbers of points in the clusters, group B will be joined to C rather than A under the minimum variance criterion.

Thus with little separation between clusters, the gradual "swallowing" of one by the other was caused by the fact that one cluster was both more compact and much larger than the other. The differences in size and compactness appear to be more important than the ellipsoidal shape of the clusters.

UPGMA With Distance-Based Similarity Coefficients: Figure 7E shows the tree for UPGMA with Euclidean d, and figure 8E shows the tree for UPGMA with Gower's coefficient. We have predicted (Chapter 7) that these two similarity coefficients would produce very similar classifications, which certainly holds true in this data set. Each tree was cut in two places which might reasonably be chosen by a user looking at the lengths of the branch lines. In order to be included in the partition, a group was required to contain at least 5 points. Points not in large enough groups were left as outliers. Table 2E (c)-(f) shows the match between each of these partitions and the correct one.

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ... 1 10374 3 5375

GAMMA -.9995

GAMMAB -.9998

GAMMAW -.9989

MAX. CLUSTER DIAMETER 5.6908
MIN. DIST. BTWN BOUNDARIES .0963

CLUSTER DIAMETERS

1	2	3
4.824	5.691	5.149

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3
1	.552	.155	.417
2	.397	.202	.505
3	.526	.176	.232

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3
.400	.374	.340

CLUSTER HOMOGENITIES

1	2	3
.492	.178	.385

(3) Table 7E
HMODE-r 3-group partition

NUMBER OF CLUSTERS 3

CLUSTER CENTERS

	1	2	3
1	.368	-.647	-.516
2	.499	.040	-.171
3	-.984	.156	.883

INTER-CLUSTER CENTROID DISTANCES

	1	2	3
1	0	1.533	2.105
2	1.533	0	1.943
3	2.105	1.943	0

COSINE OF THE ANGLE (AT THE ORIGIN)* BETWEEN CLUSTER CENTROIDS

	1	2	3
1	1.000	-.302	-.762
2	-.302	1.000	-.387
3	-.762	-.387	1.000

DISTANCE BETWEEN CLUSTER BOUNDARIES

	1	2	3
1	0	.096	.177
2	.096	0	.327
3	.177	.327	0

CLUSTER SQUARED DEVIATION SCORE AND NUMBER OF ELEMENTS

1	112.842	75
2	120.878	50
3	95.987	53

TOTAL OF SQUARED DEVIATION SCORES= 329.706

Figure 8E

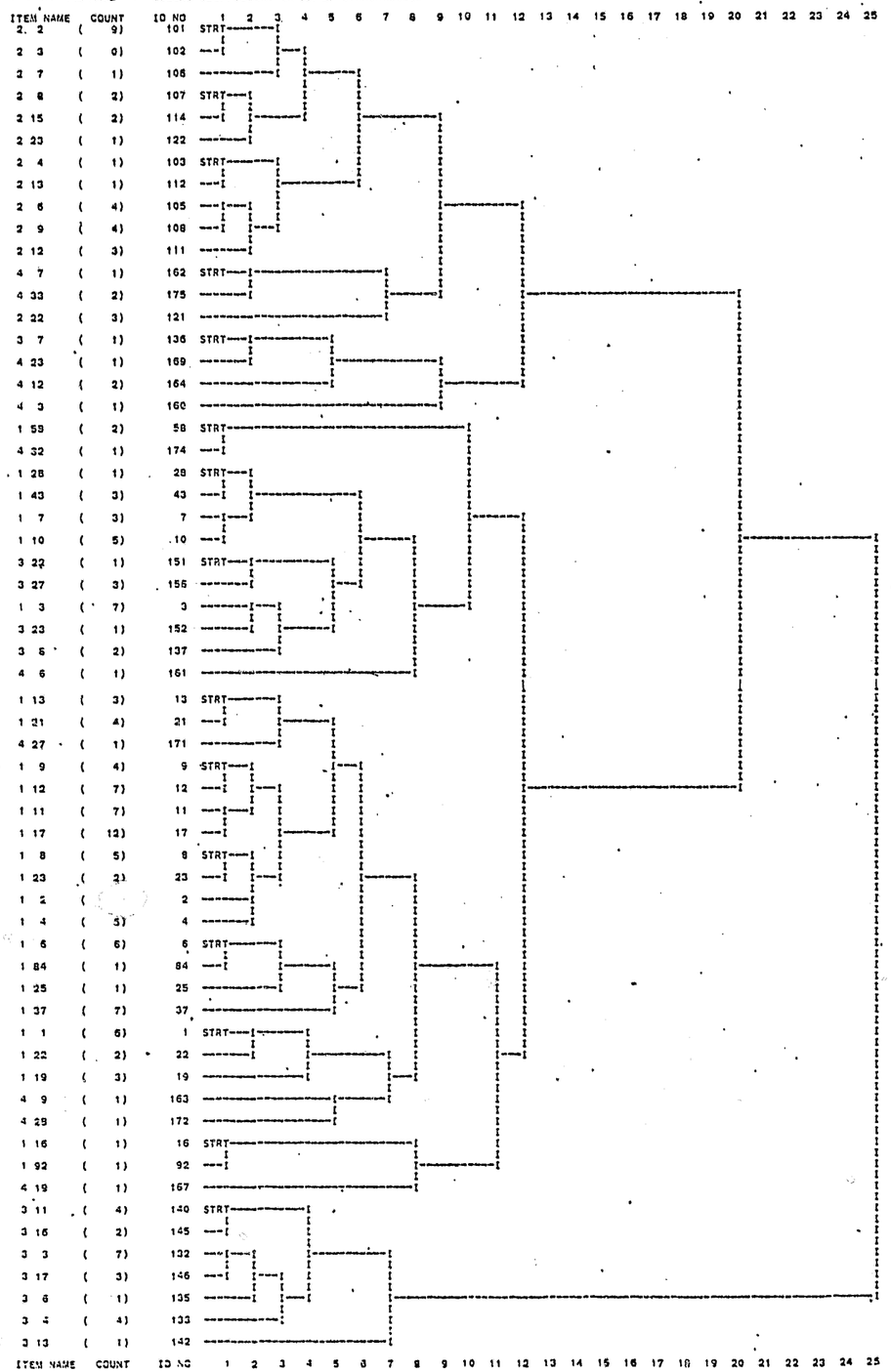


Figure 9E

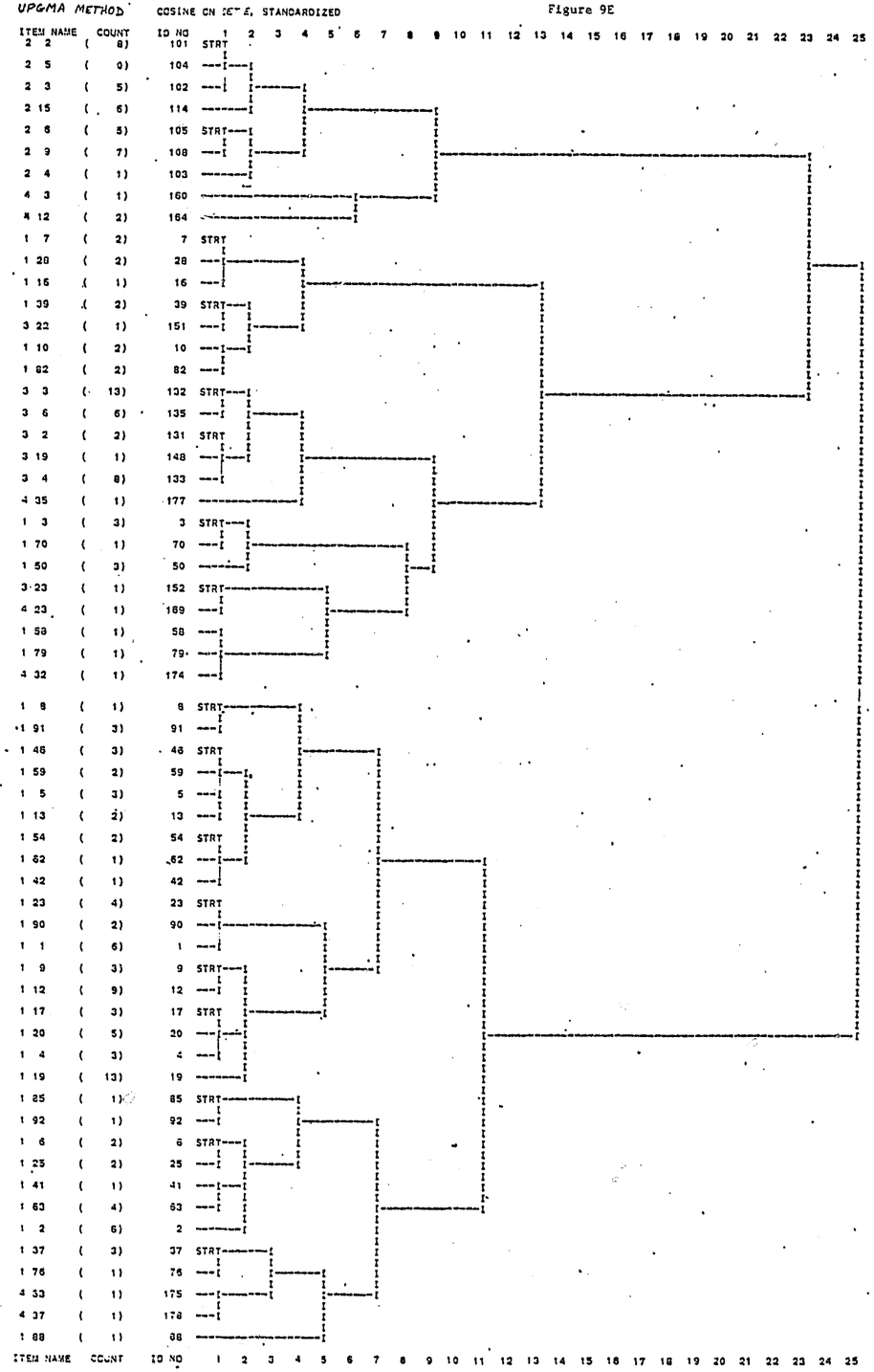
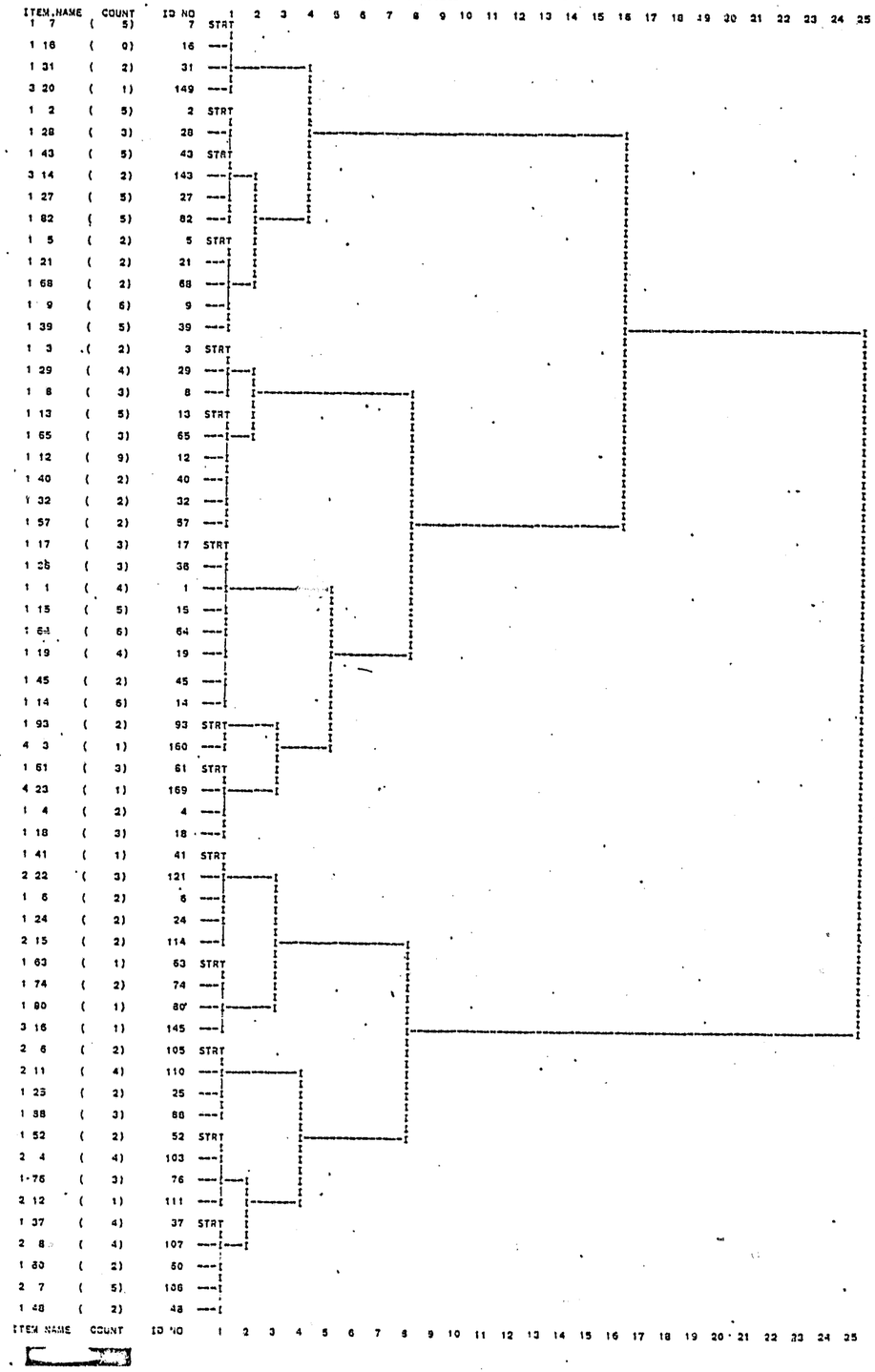


Figure 10E



The 6- and 5-group solutions both show the emergence of a group containing a mixture of points from clusters 1 and 3 in the fuzzy boundary region between those two clusters (group 3 in (c), group 2 in (e)). There may be the appearance of a mode where two clusters overlap. The hybrid group eventually merges with the group corresponding to cluster 1.

In (d) of Table 2E, the "extra" fourth group contains mostly noise points. The same group appears using both Gower's and distance, but maintains its separateness longer in the distance tree. At the 3-cluster level these two trees produce identical partitions.

UPGMA With Angular Similarity Coefficients: The UPGMA trees which resulted from use of cosine and Q-correlation, r, are shown in Figures 9E and 10E. Again, trees were cut where the branch lines were long, keeping in mind that we are most interested in partitions containing approximately 3 groups. The resulting groups are compared with the correct partition in Table 3E (a), (b) and (c). HMODE results using cosine and r are also included in Table 3E because they are strikingly similar to the UPGMA results.

The characteristics of the similarity coefficients had a marked effect on the clustering of this data. The grand centroid of the data set (and therefore the origin in Euclidean space after standardization) falls very near the boundary of cluster 1, and that cluster subtends a large angle in space (measured at the origin). Thus in early stages of the hierarchical clustering process using cosine, cluster 1 is fragmented into small pieces. Some of those pieces are closer, in an angular sense, to cluster 3 than to the main body of cluster 1. Since there is no empty space to provide a boundary between clusters 1 and 3, some of the fragments are assigned to the group

Table 3E

Cluster Overlap Matrices for Data Set E
UPGMA and HMODE with angular similarity coefficients

(a) UPGMA, cosine, 5 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	61	0	1	4	66
2	21	0	0	3	24
3	7	0	28	8	43
4	11	0	1	0	12
5	0	29	0	6	35
Total	100	29	30	21	

(b) UPGMA, cosine, 3 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	82	0	1	7	90
2	18	0	29	8	55
3	0	29	0	6	35
Total	100	29	30	21	

(c) HMODE, cosine, 6 groups

HMODE Partition	Generated partition 1	2	3	Noise	Total
1	76	0	1	5	82
2	3	0	27	6	36
3	7	0	0	0	7
4	4	0	0	1	5
5	4	0	1	1	6
6	0	29	0	4	33
Outliers	6		1	4	11
Total	100	29	30	21	

(d) HMODE, cosine, 3 groups

HMODE Partition	Generated partition 1	2	3	Noise	Total
1	80	0	1	6	87
2	14	0	28	7	49
3	0	29	0	4	33
Outliers	6		1	4	11
Total	100	29	30	21	

(e) UPGMA, r, 3 groups

UPGMA Partition	Generated partition 1	2	3	Noise	Total
1	59	0	8	10	77
2	27	0	19	4	50
3	14	29	3	7	53
Total	100	29	30	21	

(f) HMODE, r, 3 groups

HMODE Partition	Generated partition 1	2	3	Noise	Total
1	58	0	8	9	75
2	27	0	19	4	50
3	14	29	3	7	53
Outliers	1			1	2
Total	100	29	30	21	

containing cluster 3. The same pattern occurs when HMODE clusters this data using cosine, shown in Table 3E (c) and (d). The two clustering methods produce very similar 3-group solutions.

Q-correlation fragments cluster 1 even further, because the points of that cluster which are nearest the line x = y = z have "shapes" that are more-or-less randomly distributed around a flat profile. (The danger of using Q-correlation with nearly-flat profiles is discussed earlier). In data like this, clustering based on shape cannot be expected to duplicate clustering based on position in space. However there is remarkable agreement between UPGMA and HMODE on what the three shape-clusters should be: (e) and (f) in TAble 3E show identical partitions except for the two outliers left by HMODE.

Clearly, the characteristics of the similarity coefficients far outweighed any differences between the clustering methods in determining the clusters that were produced here using cosine and r. This highlights the importance of selecting the similarity coefficient which most nearly measures the aspects of similarity that are important in a particular research project. No single similarity measure should be considered the proper choice for all purposes.

HMODE With Angular Similarity Coefficients: The trees from HMODE using cosine and r are shown in Figures 11E and 12E. The resulting partitions have been discussed in the preceding section.

A note on the interpretation of HMODE trees: These trees only show linkages between secondary clusters, or modes. The "Item Name" at left is the label from the most-dense point in that secondary cluster. Linkages between individual points are not shown. The point count in parentheses gives the number of points in that secondary cluster at the time the clustering process

Figure 11E

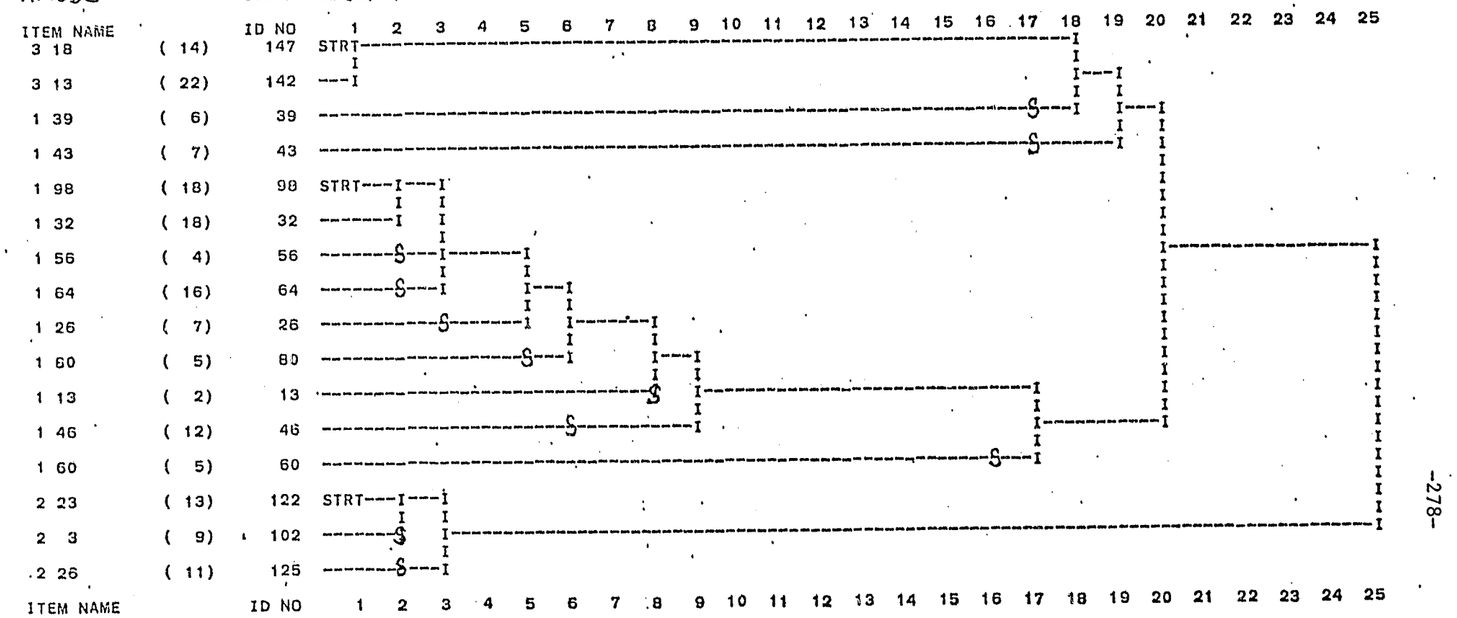

HMODE
COSINE ON SET E, STANDARDIZED
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
3 18 (14) 147 STRT
3 13 (22) 142
1 39 (6) 39
1 43 (7) 43
1 98 (18) 98 STRT
1 32 (18) 32
1 56 (4) 56
1 64 (16) 64
1 26 (7) 26
1 60 (5) 80
1 13 (2) 13
1 46 (12) 46
1 60 (5) 60
2 23 (13) 122 STRT
2 3 (9) 102
2 26 (11) 125

Figure 12E

HMODE

CORRELATION COEFFICIENT ON SET E, STANDARDIZED

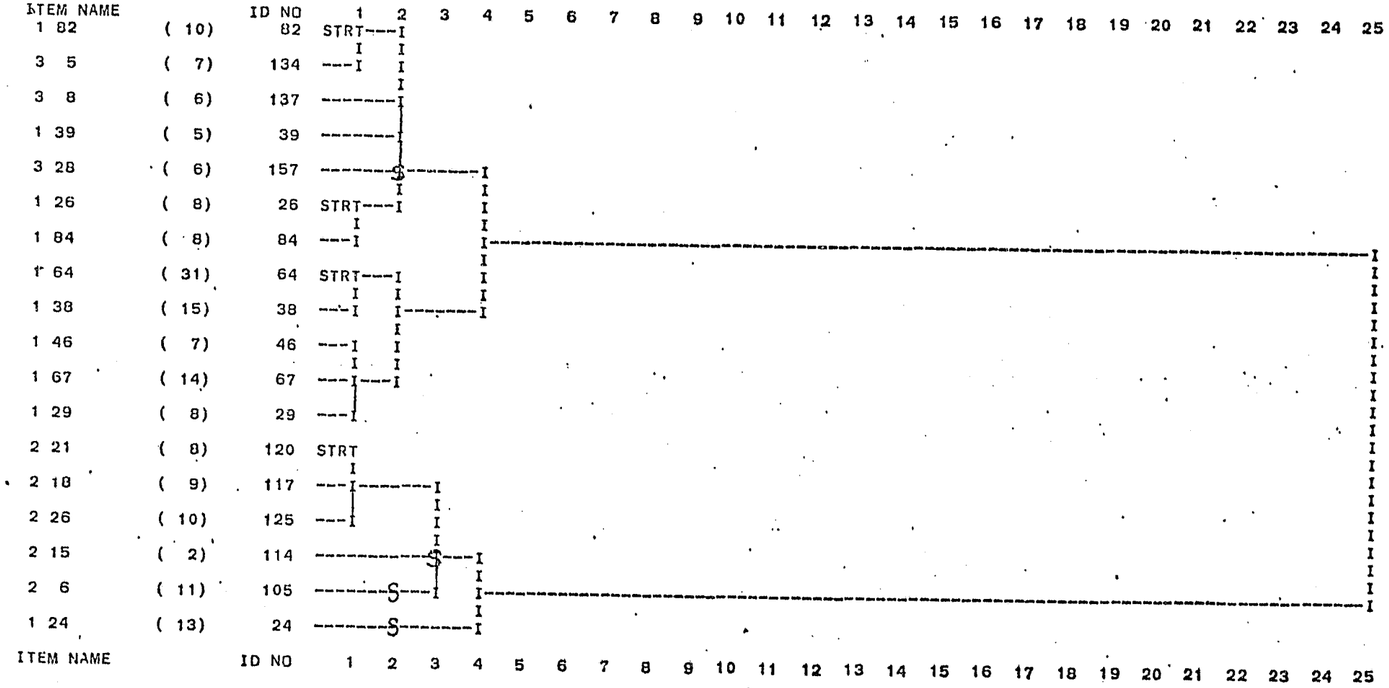
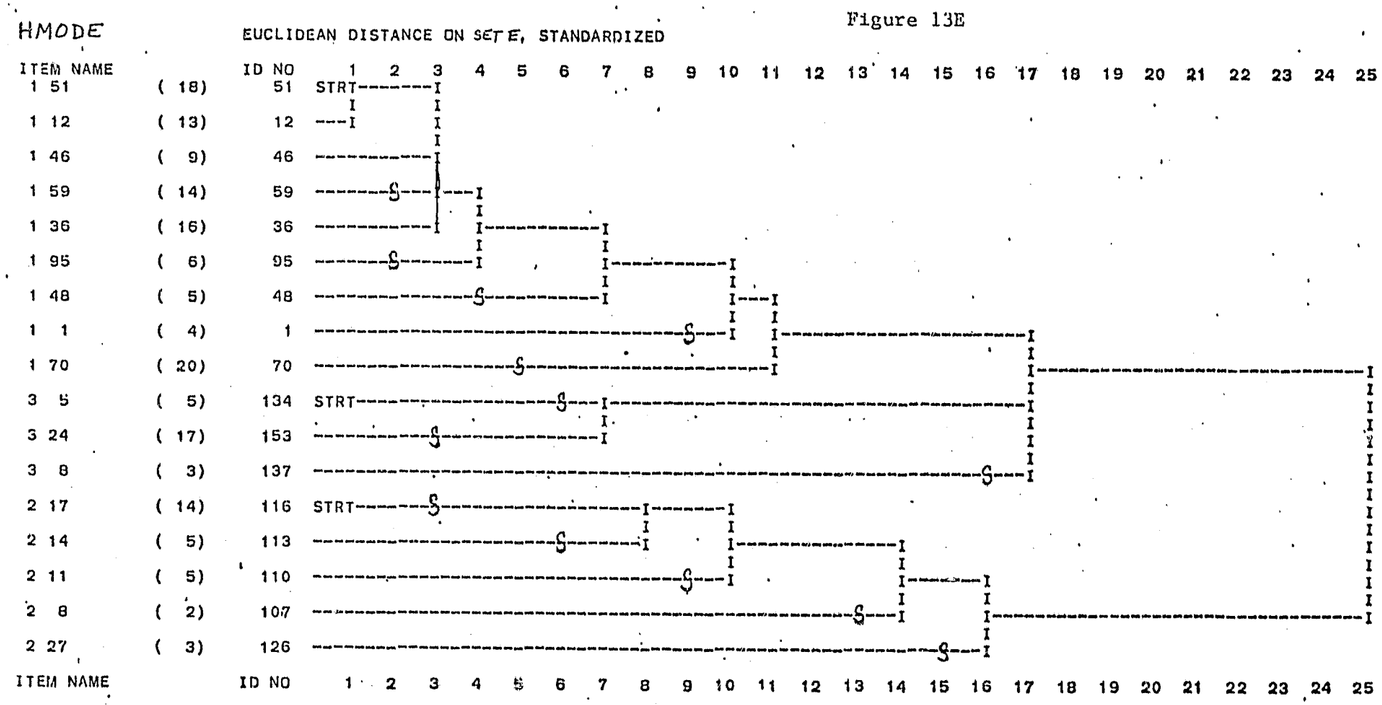


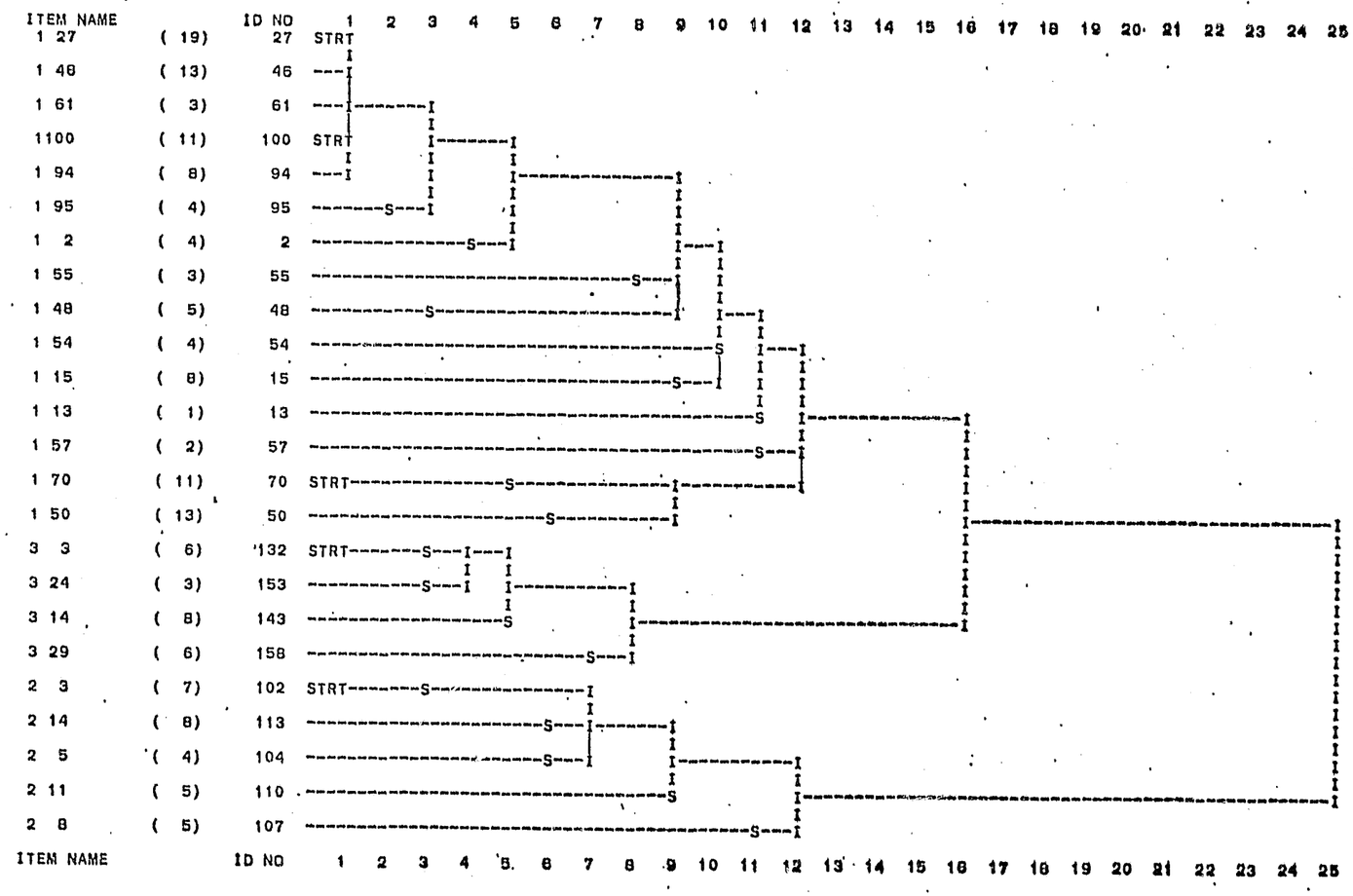
Figure 13E


HMODE
EUCLIDEAN DISTANCE ON SETE, STANDARDIZED
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
1 51 (18) 51 STRT
1 12 (13) 12
1 46 (9) 46
1 59 (14) 59
1 36 (16) 36
1 95 (6) 95
1 48 (5) 48
1 1 (4) 1
1 70 (20) 70
3 5 (5) 134 STRT
3 24 (17) 153
3 8 (3) 137
2 17 (14) 116 STRT
2 14 (5) 113
2 11 (5) 110
2 8 (2) 107
2 27 (3) 126
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

HMODE

GOWERS COEFFICIENT ON SET E, STANDARDIZED

Figure 14E



was terminated. The letter S printed on the tree shows the similarity level at which a secondary cluster was initialized (not printed for clusters initialized at level 1).

HMODE With Distance-Based Similarity Coefficients: HMODE trees using Euclidean d and Gower's coefficient are in Figures 13E and 14E. The 3-cluster partitions from those trees may be examined in Table 4E (a) and (b). The large number of outliers is governed by a parameter set within the HMODE program. The assignment of points to secondary clusters is terminated whenever all of the modes have been linked into a single primary cluster, provided that a specified proportion of the points has been classified. That proportion was set at .8 for all HMODE runs on data set E. Points that remain unclassified become outliers. Note in Table 4E that many of the outliers are noise points in the generated partition.

The HMODE results in Table 4E should be compared with the UPGMA tables showing d and Gower's coef., Table 2E (d) and (f). Except for the existence of outliers in the HMODE partition, the UPGMA and HMODE results are quite similar. In the HMODE, d, partition a point-by-point investigation of the 6 "outliers" that had been generated in cluster 3 shows that only one of those points was classified correctly into group 3 of the UPGMA partition. Perhaps it is best that questionable points are left in an outlier category, rather than being misclassified. Therefore the HMODE partition could be considered slightly preferable to the UPGMA partition in this case.

NMODE: Three modes were detected using the default value for the dense-point radius R. (See Table 4E (c).) Surprisingly, only half of cluster 2 was identified as a separate group, while most of the other half was assigned with the points from cluster 1. Thus NMODE, like WARD, was unable to

Table 4E

Cluster Overlap Matrices for Data Set E
Mode-Search Methods with Distance-Based Similarity Measures

(a) HMODE, Euclidean d, 3 groups

HMODE Partition	Generated partition 1	2	3	Noise	Total
1	98	0	4	3	105
2	0	27	0	2	29
3	0	0	20	2	22
Outliers	2	2	6	14	24
Total	100	29	30	21	

(b) HMODE, Gower's coefficient, 3 groups

HMODE Partition	Generated partition 1	2	3	Noise	Total
1	98	0	6	3	107
2	0	26	0	2	28
3	0	0	20	2	22
Outliers	2	3	4	14	23
Total	100	29	30	21	

(c) NMODE, Euclidean d, default R (3 groups)

NMODE Partition	Generated partition 1	2	3	Noise	Total
1	98	12	7	5	122
2	0	15	0	1	16
3	0	0	22	2	24
Outliers	2	2	1	13	18
Total	100	29	30	21	

create the correct boundary between populations 1 and 2. The NMODE result may be considered superior to the WARD result, however, because the correct number of modes is specified and the three modes do correspond to the three populations (albeit with 19 points misclassified).

Comparison of the Partitions

From the preceding tables and discussion, the 3-group partition from each method can be evaluated by considering numbers of points misclassified, along with some of the preferences discussed above. To determine the number of points misclassified it was necessary to match each of the groups in a partition with one of the populations. This was done in the way that would maximize agreement and minimize the number of misclassifications in each case. Since many of the clustering methods leave some points unclassified, the number of outliers must be considered along with the number of misclassifications. On the judgment that it is generally worse to misclassify a point than to leave it unclassified, a partition containing outliers was considered better than a similar one with no outliers if at least half of those outliers were misclassified by the latter partition.

In table 5E, the 3-group partitions are rank-ordered according to their success in duplicating the generated clusters. Partitions grouped together in brackets are quite similar. The five top-ranking partitions can all be considered good based on a comparison with discriminant analysis, which suggested that some misclassifications are to be expected in the fuzzy boundary region between clusters 1 and 3. The remaining 7 partitions show considerably more blurring of cluster boundaries.

The four coefficients of partition agreement are also shown in table 5E, in each case comparing the C.A. partition with the correct partition. For the

Table 5E

Measures of Partition Agreement on Data Set E
with Success Ratings for the Three-Group Partitions

3-group partitions, in order of success in duplicating the clusters as generated	# of points misclassified (minimum)	cluster pts. left as outliers	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
1. MIKDA, Euclidean d (both det W and Trace W)	2	0	.897	.778	.904	.752
2. HMODE, d	4	10	.904	.778	.787	.729
3. HMODE, Gower's	6	9	.891	.754	.777	.712
4. UPGMA, d - (4th grp. (considered outliers)	9	1	.843	.690	.722	.660
5. UPGMA, Gower's	10	0	.835	.680	.825	.631
6. MIKCA, Mahalanobis D^2	17	0	.803	.587	.800	.584
7. HMODE, cosine	15	7	.795	.554	.679	.578
8. UPGMA, cosine	19	0	.790	.562	.798	.576
9. NMODE, d	19	5	.789	.608	.702	.536
10. WARD, d^2	38	0	.650	.503	.635	.416
11. UPGMA, r	52	0	.636	.326	.537	.301
12. HMODE, r	52	1	.637	.321	.445	.297

Other partitions, ordered by # of groups; within that, ranked according to Lambda	# of groups	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
MIKCA, Euclidean d, det W	2	.630	.495	.855	.436
MIKCA, Mahalanobis D^2, det W	2	.619	.446	.735	.352
MIKCA, Mahalanobis D^2, det W	4	.766	.465	.725	.557
MIKCA, Euclidean d (both det & trace)	4	.766	.462	.725	.554
UPGMA, Gower's	5	.829	.604	.743	.591
UPGMA, cosine	5	.747	.420	.695	.567
MIKCA, d, trace W	5	.793	.511	.723	.533
HMODE, cosine	6	.802	.537	.731	.612
WARD, d^2	6	.840	.677	.743	.592
UPGMA, d	6	.800	.510	.756	.565

purpose of computing the Rand and Jaccard coefficients, outliers were treated as a single category; that is, all outliers are in the same group. These measures tend to rank-order the 3-group partitions in very much the same way that we did; in fact, lambda orders them exactly as we did. Where reversals occur, they are not consistent over the four coefficients. Note that the values of all four coefficients are generally lower than they were for data set D. That is because of the extra "noise" category, whose points tend to be distributed over all three groups. Only the methods which provide for outliers (mainly HMODE and NMODE) could provide a category which might correspond to the noise group.

The coefficients of agreement are shown also for some of the partitions that did not contain 3 groups, again comparing each partition with the correct 3-cluster one. Discussion of these results will be taken up at the end of this chapter in an evaluation of the four coefficients. Two brief observations: Rand's coefficient shows a tendency to increase as the number of clusters in the partition increases, and ranks a 6-group partition as highly as some of the "good" 3-group partitions. Cramer's V, on the other hand, shows a preference for the 2-group partitions, ranking one of them better than all but one of the 3-group partitions.

Evaluation Of Individual Clusters

Two 3-group partitions were selected for analysis of individual clusters: One good partition (HMODE with Euclidean d) and one poor partition (HMODE with correlation, r). These partitions are depicted graphically in Figures 15E and 16E, and may be compared with the correct partition in Figure 1E. Zeroes in these plots denote outliers or noise points.

Figure 15E

CLUSTER GRAPH

DATA SET E
HMODE - 0
X=VAR 1 Y=VAR 2

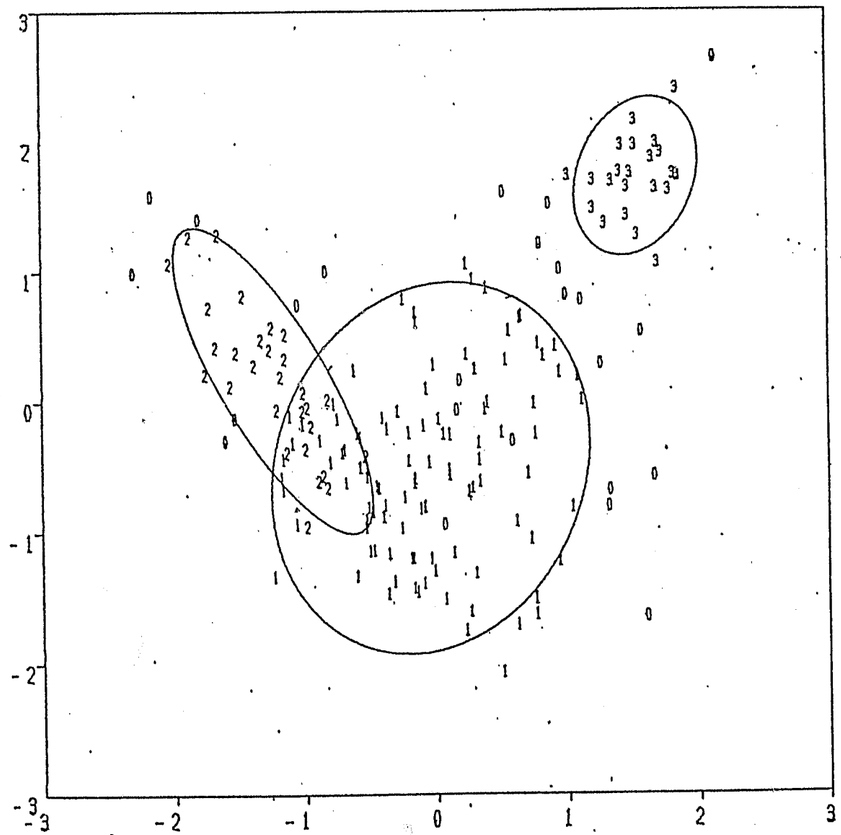


Figure 16E

CLUSTER GRAPH

DATA SET E
HMODE - R
X=VAR 1 Y=VAR 2

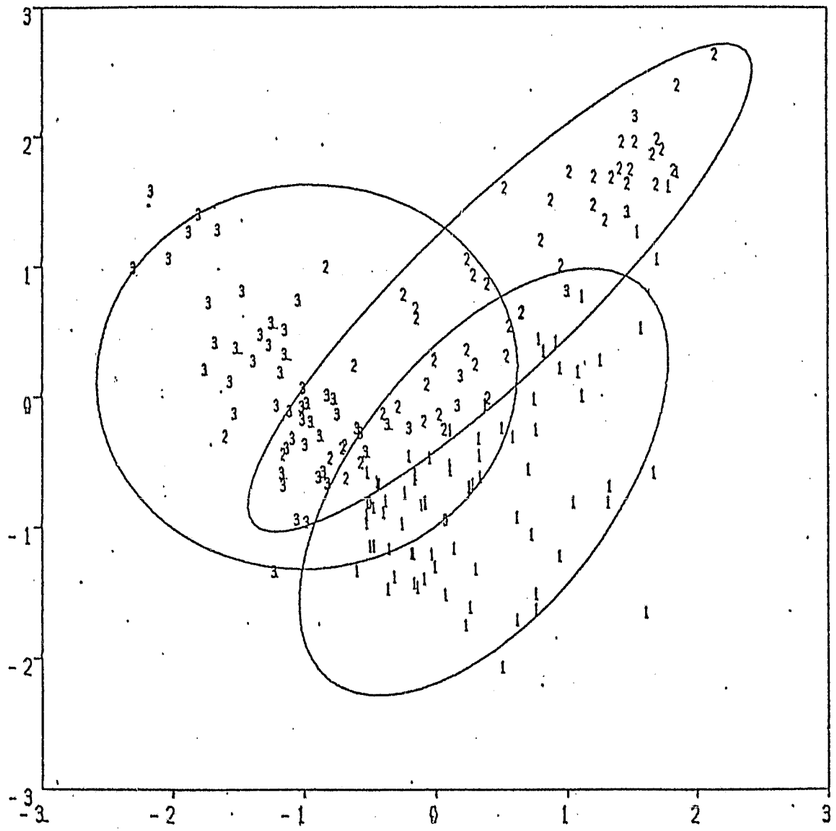


Table 6E shows measures of cluster separation, compactness, and homogeneity for the HMODE-d partition; table 7E shows those measures for the HMODE-r partition. Because the three original populations actually overlap slightly, the gamma measures are predictably low even for the "good" HMODE-d partition. However, the gammas are nearly -1 for the HMODE-r partition, indicating that the clusters have virtually 0 separation (GAMMA B) and that at least one cluster has a diameter nearly equal to the diameter of the entire data set, hence is anything but compact (GAMMA W). The truth of this can be seen by examining the HMODE-r graph (Figure 16e). A comparison of cluster diameters and distances between boundaries for the two partitions gives similar information.

The homogeneity of the variables offers another basis for comparison. It too is relatively high (over .6 in most cases) for the HMODE-d partition, and low (under .5 in most cases) for the HMODE-r partition.

It should be noted that the tests for cluster separation, compactness, and homogeneity were all designed with a points-in-space model of similarity in mind. If the similarity coefficient r had been chosen because the researcher was interested only in profile shape then these tests would not be appropriate; indeed, some measure of separation and compactness in shape should be devised.

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ... 2556 3437 1291 4806

GAMMA -.3636

GAMMAB -.1470

GAMMAW -.5765

MAX. CLUSTER DIAMETER 3.1425
MIN. DIST. BTWN BOUNDARIES .7444

CLUSTER DIAMETERS

1	2	3
3.142	2.413	1.980

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3
1	.668	.867	.948
2	.545	.681	.913
3	.859	.871	.727

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3
.745	.622	.842

CLUSTER HOMOGENITIES

1	2	3
.691	.806	.863

Table 6E
HMODE-d 3-group partition

NUMBER OF CLUSTERS 3

CLUSTER CENTERS

	1	2	3
1	-.040	-.508	-.654
2	-1.237	.163	1.395
3	1.542	1.717	1.084

INTER-CLUSTER CENTROID DISTANCES

	1	2	3
1	0	2.466	3.236
2	2.466	0	3.199
3	3.236	3.199	0

COSINE OF THE ANGLE (AT THE ORIGIN)* BETWEEN CLUSTER CENTROIDS

	1	2	3
1	1.000	-.610	-.777
2	-.610	1.000	-.024
3	-.777	-.024	1.000

DISTANCE BETWEEN CLUSTER BOUNDARIES

	1	2	3
1	0	.744	1.150
2	.744	0	2.536
3	1.150	2.536	0

CLUSTER SQUARED DEVIATION SCORE AND NUMBER OF ELEMENTS

1	96.478	105
2	16.278	29
3	8.661	22

TOTAL OF SQUARED DEVIATION SCORES= 121.416

DATA SET F:

OVERLAPPING CLUSTERS OF DIFFERENT COMPACTNESS

Description

Data set F, like set E, started with 200 points, 40 of which were uniformly distributed "noise". Again, the most isolated 10% were removed as unclassifiable "outliers". All 20 points removed came from the 40 original noise points, while the clusters remained intact.

The part of data set F that was subjected to cluster analysis, then, contained 180 points measured on 4 variables. Twenty points represent "noise" from the uniform distribution, while 160 points were drawn from 5 multivariate normal subpopulations to represent 5 clusters. A compact central cluster contains 80 points and the peripheral clusters contain 20 points each. All of the clusters are hyper-spherical in shape.

Parameters for the 5 subpopulations are:

Cluster 1 (n = 80)

Centroid: (10, 12, 10, 12)

Standard deviation = 1 on each variable

Cluster 2 (n = 20)

Centroid: (8.5, 10.5, 8.5, 10.5)

Standard deviation = 1 on each variable

Cluster 3 (n = 20)

Centroid: (12, 14, 12, 14)

Standard deviation = 1 on each variable

Cluster 4 (n = 20)

Centroid: (13, 15, 7, 9)

Standard deviation = 2 on each variable

Cluster 5 (n = 20)

Centroid: (7, 9, 13, 15)

Standard deviation = 3 on each variable

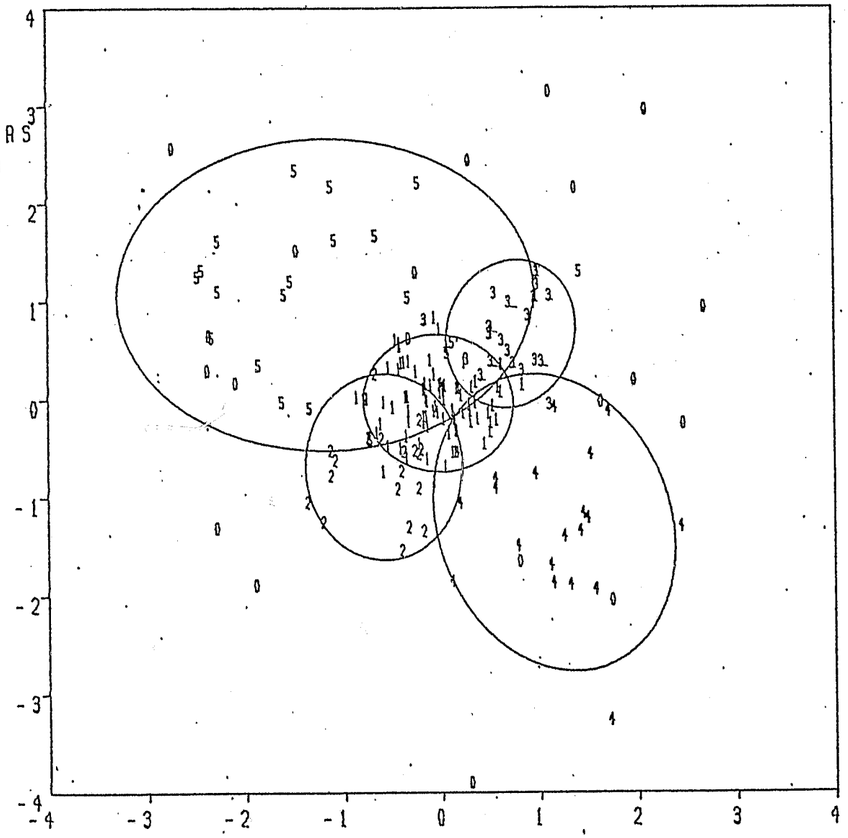
Figure 1 is a plot of data set F in two dimensions that show the greatest separation between clusters. The ellipses indicate the "territories" occupied by each of the clusters.

Size and compactness: Cluster 1 is much more dense than the others. While clusters 2 and 3 cover approximately the same sized territory as cluster 1, they contain only 1/4 as many points. Clusters 4 and 5 cover much larger territories with only 20 points each, hence they are much less compact than the first 3 clusters.

Separation: Each of the four surrounding clusters overlaps slightly into the "territory" of cluster 1, with cluster 2 having the most overlap. It will be a particular challenge to the clustering algorithms to identify 3 separate modes in clusters 1, 2, and 3, because the regions between those modes contain large numbers of points with only a reduced density to indicate a fuzzy boundary. From visual inspection of the 2-dimensional plots, it is not obvious that there are 5 modes in this data.

CLUSTER GRAPH

DATA SET F
GENERATED CLUSTERS
X=VAR 1 Y=VAR 3



Discriminant analysis based on the five correct clusters was able to classify 90.6% of the cases correctly (not including the 20 unclassified "noise" points). Point allocations are shown in the following table:

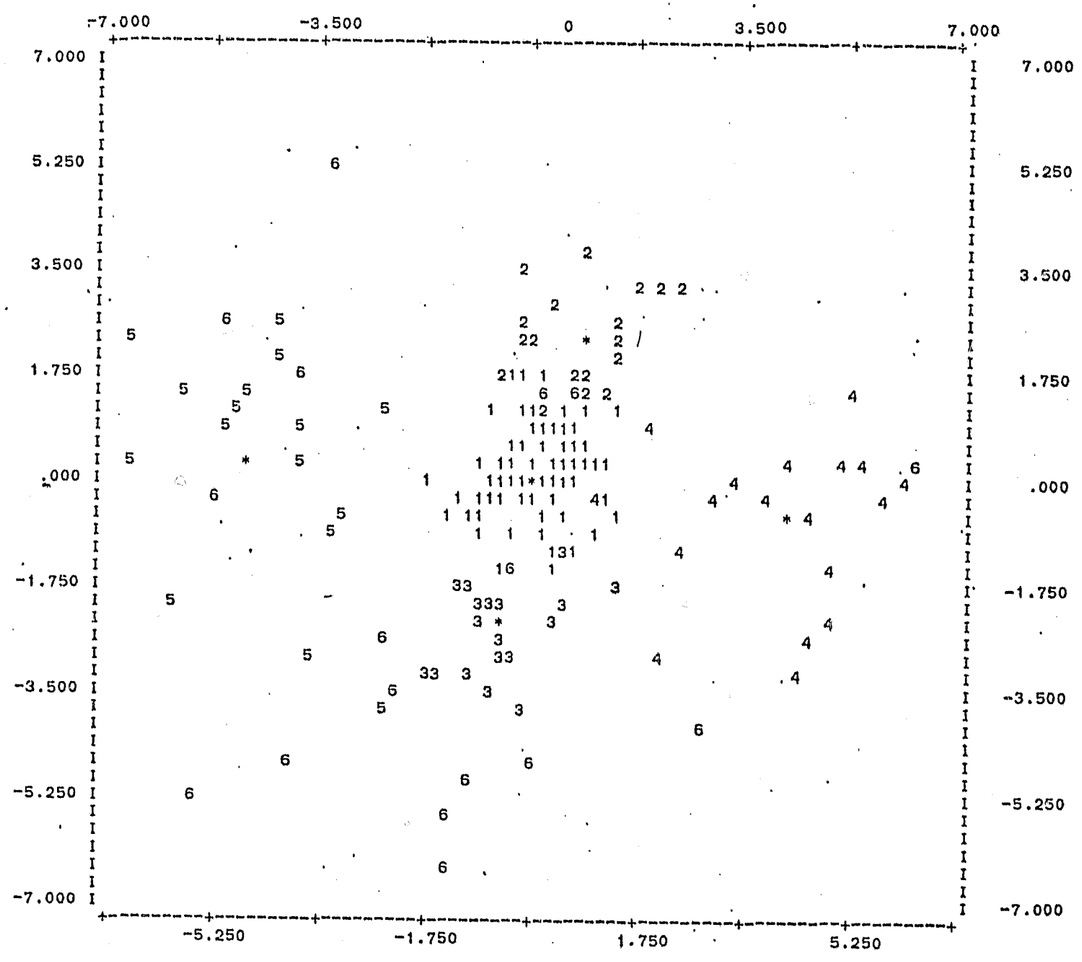
Predicated Group	Actual Group 1	2	3	4	5	Noise
1	73	1	1	1	0	0
2	5	19	0	1	0	4
3	2	0	19	1	3	10
4	0	0	0	17	0	2
5	0	0	0	0	17	4
Total	80	20	20	20	20	

Thus every cluster overlaps slightly with some other cluster. The fuzziness of the boundaries is even more obvious when correct classification with at least a .9 probability is required.

Cluster	Proportion of points classified correctly with at least .9 probability
1	17/80 = 21%
2	11/20 = 55%
3	17/20 = 85%
4	17/20 = 85%
5	16/20 = 80%

DISCRIMINANT ANALYSIS ON *SET F*, AS GENERATED Figure 2F 79/09/20. 14.18.59. PAGE 10

PLOT OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CENTROID.



The boundaries between clusters 1 and 2 and between clusters 1 and 3 are particularly fuzzy.

A plot of data set F in the space of the first two discriminant functions is shown in Figure 2F. (The 6's in the plot are noise points.)

Clustering Results (Data Set F)

In this 180-point data set we require that a group in a partition must contain at least 5 points. All other points will be designated "outliers" in that partition.

MIKCA, Minimizing Determinant W: Partitions containing 3, 4, and 5 groups were requested, using both Mahalanobis' D and Euclidean d. The matches between these partitions and the correct one are shown in Table 1F. Neither of the 5-group results are very satisfactory. The Mahalanobis D 5-group partition (Table 1F(c)) is probably the worst of the two, containing one group which combines nearly equal numbers of points from four of the original populations, and linking substantial parts of populations 2 and 4 into a single group. The Euclidean d 5-group solution (Table 1F(f)) is easier to understand, with populations 3, 4 and 5 fairly well identified but a wrong division is made between populations 1 and 2.

The best MIKCA results were obtained in the 4-group partitions, if one accepts the idea that it is reasonable to see clusters 1 and 2 as a single group. Furthermore, the two distance measures, D and d, produced very similar results at the 4-group level: populations 1 and 2 were merged, while clusters 3, 4, and 5 were clearly identified.

At the 3-group level, D and d again disagreed in the partitions produced. Mahalanobis D gave the result that an observer might expect looking at the

Table 1F

Cluster Overlap Matrices for Data Set F
MIKCA with Mahalanobis D: 3, 4, and 5 Group Partitions

(a) det W, D, 3 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	80	20	20	4	2	8	134
2	0	0	0	16	0	4	20
3	0	0	0	0	18	8	26
Total	80	20	20	20	20	20	

(b) det W, D, 4 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	77	20	0	2	0	3	102
2	3	0	20	0	3	9	35
3	0	0	0	18	0	3	21
4	0	0	0	0	17	5	22
Total	80	20	20	20	20	20	

(c) det W, D, 5 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	73	4	1	3	1	0	82
2	4	0	8	5	3	3	23
3	3	16	0	10	0	4	33
4	0		11	2	1	7	21
5	0	0	0	0	15	6	21
Total	80	20	20	20	20	20	

Table 1F(continued)

Cluster Overlap Matrices for Data Set F
MIKCA with Euclidean d: 3, 4, and 5 Group Partitions

(d) det W, d, 3 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	76	20	4	2	0	5	107
2	4	0	16	0	20	12	52
3	0	0	0	18	0	3	21
Total	80	20	20	20	20	20	

(e) det W, d, 4 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	78	20	1	2	0	3	104
2	2	0	19	1	3	9	34
3	0	0	0	17	0	3	20
4	0	0	0	0	17	5	22
Total	80	20	20	20	20	20	

(f) det W, d, 5 groups

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	0	0	0	0	15	5	20
2	19	0	5	2	4	2	32
3	61	20	0	1	0	2	84
4	0	0	0	16	0	2	18
5	0	0	15	1	1	9	26
Total	80	20	20	20	20	20	

discriminant plot: populations 1, 2 and 3 merged into a single group, and populations 4 and 5 identified separately. Euclidean d, however, formed one group from populations 1 and 2, another group combining populations 3 and 5.

Hoping to see the best possible results that could be achieved by a clustering method on set F and wishing also to consider the problem of local minima in MIKCA, one further experiment was tried. Usually, MIKCA begins by selecting 3 possible sets of starting centroids from among the data points, using a Monte-Carlo technique, then chooses that set which gives the best criterion value. To give MIKCA the best possible start, in this test we gave it the correct centroids of the 5 populations to use as starting centroids. The results were, indeed, much better than any of the other clustering attempts on this data. The partitions obtained using both Mahalanobis D and Euclidean d are shown in Table 2F. They are remarkably similar to the results of discriminant analysis. Mahalanobis D produced slightly fewer misclassified points than did Euclidean d.

The clustering criterion was minimization of the determinant of the within-group covariance matrix. Criterion values for the 5-group solutions were

MIKCA with Mahalanobis D

Random start: $\det W = .376 \times 10^6$

Centroids given: $\det W = .280 \times 10^6$

MIKCA with Euclidean D

Random start: $\det W = .314 \times 10^6$

Centroids given: $\det W = .286 \times 10^6$

Table 2F

Cluster Overlap Matrices for Data Set F
MIKCA with Correct Centroids Given

(a) det W, D, 5 groups (started with correct centroids)

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	72	1	1	2	0	1	77
2	8	19	0	0	0	3	30
3	0	0	19	0	3	8	30
4	0	0	0	18	0	3	21
5	0	0	0	0	17	5	22
Total	80	20	20	20	20	20	

(b) det W, d, 5 groups (started with correct centroids)

MIKCA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	71	0	2	2	0	2	77
2	9	20	0	0	0	2	31
3	0	0	18	1	3	8	30
4	0	0	0	17	0	3	20
5	0	0	0	0	17	5	22
Total	80	20	20	20	20	20	

Clearly both of the random starts produced criterion values that were merely local minima.

In many cases nothing better than a random start is available. But if approximate centroids can be obtained either theoretically or by preliminary clustering using some other algorithm (perhaps on a subset of the data) they should certainly be tried as starting centroids. The criterion value can then be compared with that obtained using a random start.

WARD: The WARD tree from set F shown in Figure 3F looks remarkably like its counterpart for set U, the uniform distribution. A few small groups do begin to form at an early stage of the set F tree, but they look like mere minor variations in density which are swallowed up one at a time by the large central group. This demonstrates to an extreme degree the tendency that we noted in the discussion of data set E above: When there is one large, compact cluster, the minimum variance criterion allows that one to dominate and swallow up any nearby smaller clusters.

No reasonable 5-group partition could be selected from this tree; indeed, the most reasonable conclusion to be drawn from the tree would be that the data has no significant structure and can be seen as a single cluster. For purposes of comparison with other methods, however, the 3-group partition is shown in Table 3F (a). At that level, all of the populations have been absorbed into the large group, with the two small groups containing mostly fragments of population 1.

The results of WARD on sets E and F throw serious doubts on its usefulness for clustering criminological data, which frequently contains a large central cluster and smaller "deviant" groups that must be differentiated.

Figure 3F

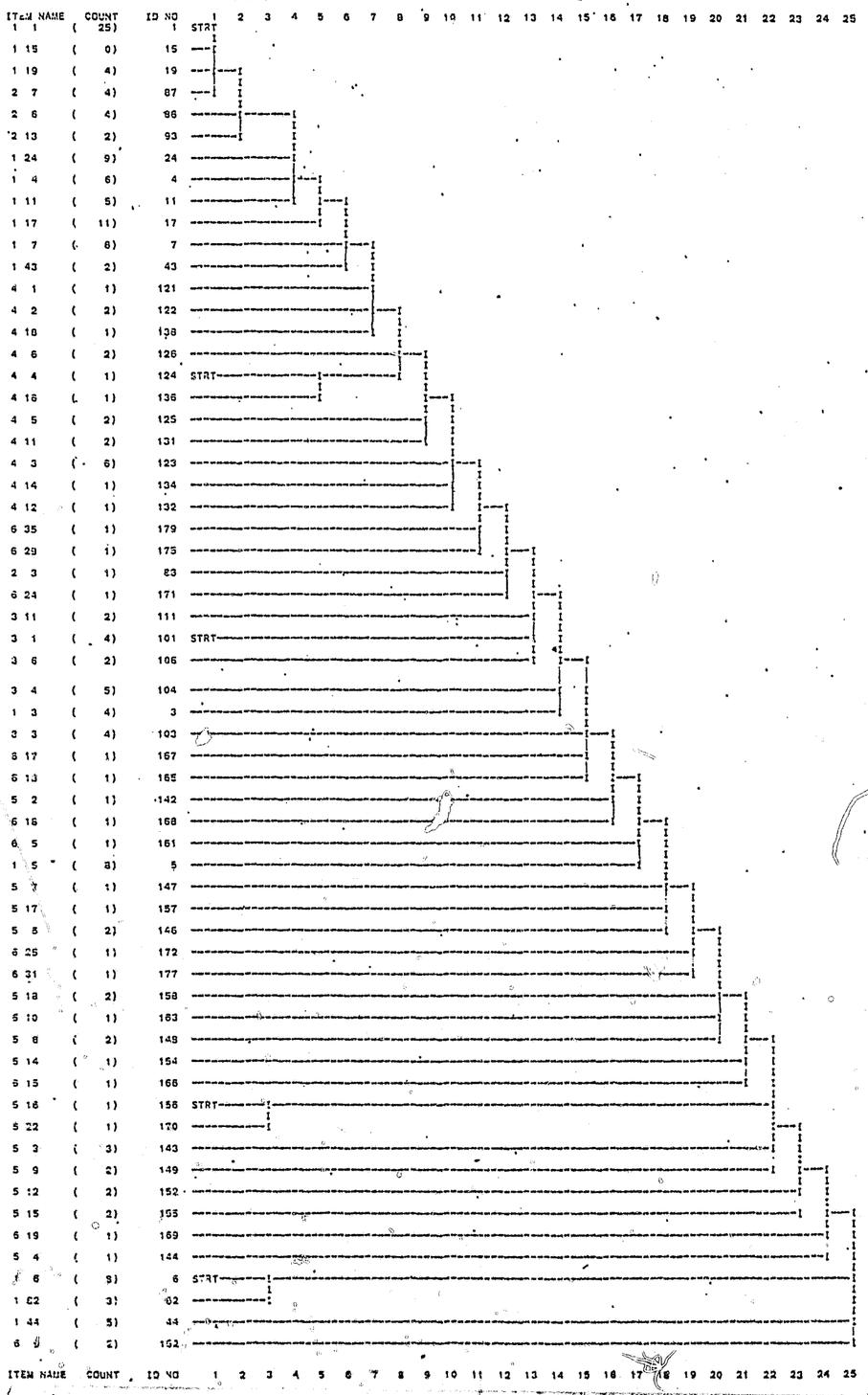


Table 3F

Cluster Overlap Matrices for Data Set F
Hierarchical Methods with Distance-Based Similarity Coefficients

(a) WARD - d^2, 3 groups

WARD Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	67	20	17	20	20	18	162
2	8	0	3	0	0	0	11
3	5	0	0	0	0	0	5
Outliers						2	2
Total	80	20	20	20	20	20	

(b) UPGMA - Gower's, 5 groups

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	73	10	1	1	0	0	85
2	7	0	19	1	0	1	28
3	0	9	0	0	0	0	9
4	0	0	0	8	0	0	8
5	0	0	0	0	7	0	7
Outliers		1		10	13	19	43
Total	80	20	20	20	20	20	

(c) UPGMA - Gower's, 3 groups

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	80	20	20	2	2	1	125
2	0	0	0	18	0	2	20
3	0	0	0	0	14	2	16
Outliers					4	15	19
Total	80	20	20	20	20	20	

Table 3F (cont.)

Cluster Overlap Matrices for Data set F
Hierarchical Methods with Distance-Based Similarity Coefficients

(d) UPGMA-d, 5 groups from selected tree branches

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	76	11	3	2	0	0	92
2	3	0	17	0	1	1	22
3	1	9	0	0	0	0	10
4	0	0	0	17	0	2	19
5	0	0	0	0	15	2	17
Outliers				1	4	15	20
Total	80	20	20	20	20	20	

(e) UPGMA-d, 5 groups from straight cut of tree branches

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	77	20	3	2	0	0	102
2	3	0	17	0	1	1	22
3	0	0	0	5	0	0	5
4	0	0	0	8	0	1	9
5	0	0	0	0	8	1	9
Outliers				5	11	17	33
Total	80	20	20	20	20	20	

(f) UPGMA-d, 3 groups

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	80	20	20	2	1	1	124
2	0	0	0	17	0	2	19
3	0	0	0	0	15	2	17
Outliers				1	4	15	20
Total	80	20	20	20	20	20	

UPGMA With Distance-Based Similarity Coefficients: The UPGMA trees obtained using Euclidean d and Gower's coef. are in Figures 4F and 5F. With some relief, we observe that they do show the existence of structure in set F and do not look like the corresponding trees from the uniform distribution. The two similarity coefficients produced very similar trees, therefore our discussion will focus on just the Euclidean d result, Figure 4F.

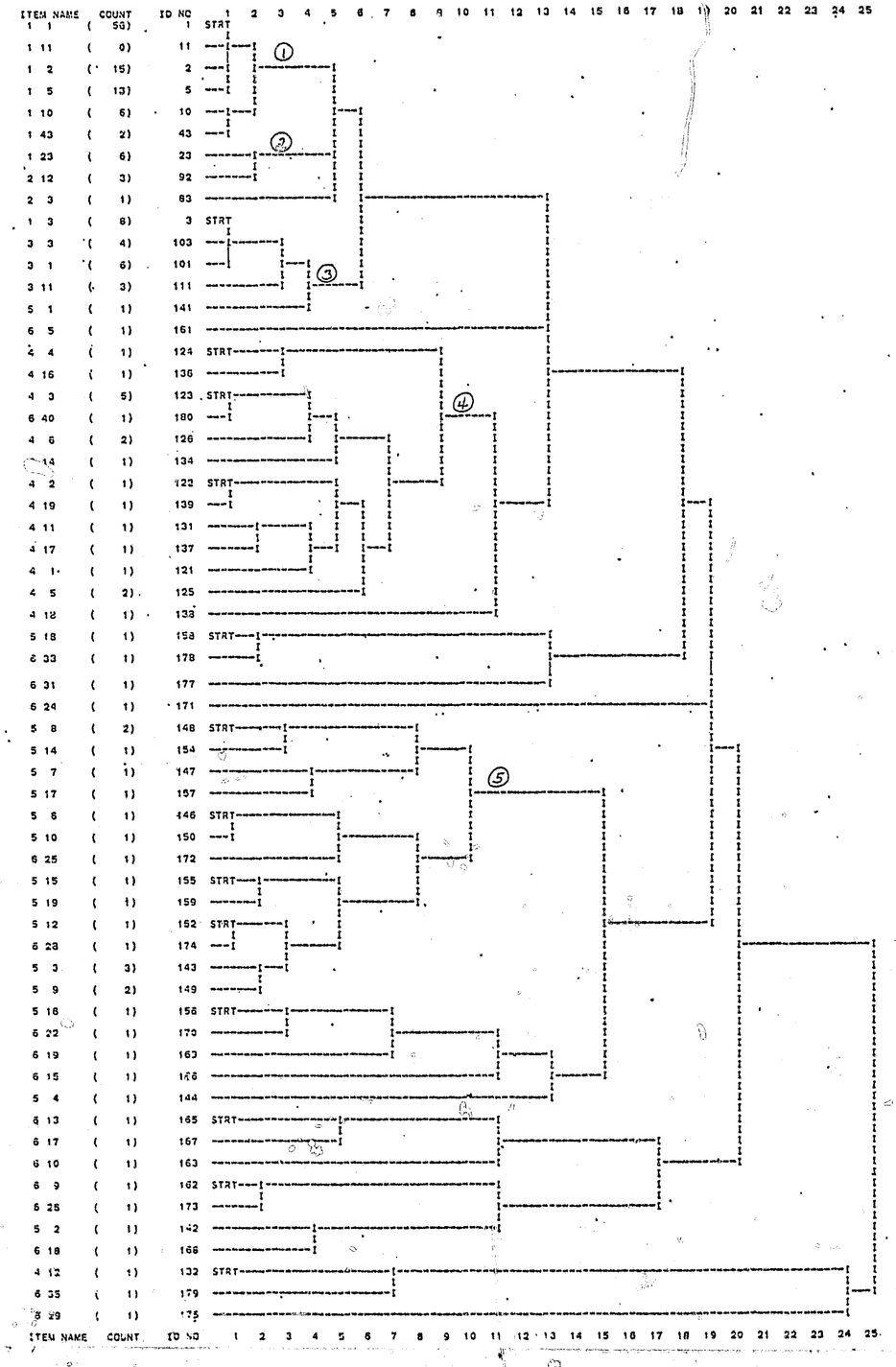
The trees as shown are incomplete, only showing the last 60 branch lines to allow a reasonably compact display of results. The number of points that had already been merged into each branch line before printing began are shown in parentheses to the left of the branch line. (Note: The count is taken one step too far, so that the 56 on the first branch of Figure 4F actually represents the sum of the points on the first two branch lines.) Many of the early merges, before printing of the tree began, took place within population 1. The first long branch line at the top of Figure 4F combines 6 small branches which represent a total of 92 points (group 1 in Table 3F (d)).

Figure 4F is a good example of a tree containing clusters of quite different compactness. Branch lines for the groups that best represent the original 5 populations are marked. Visual inspection of the tree clearly shows the development of groups corresponding to populations 4 and 5, but these groups are not fully formed until after the points from populations 1, 2, and 3 have merged into a single group. Although it may not be possible, without prior knowledge of the generated partition, to be confident of the separateness of population 2, it is certainly possible to recognize groups corresponding to populations 1 and 3. Therefore it should be possible to select groups based on branch lines from different levels of the tree, providing that those branch lines meet certain criteria. More work needs to

UPGMA

EUCLIDEAN DISTANCE ON SET F, STANDARDIZED

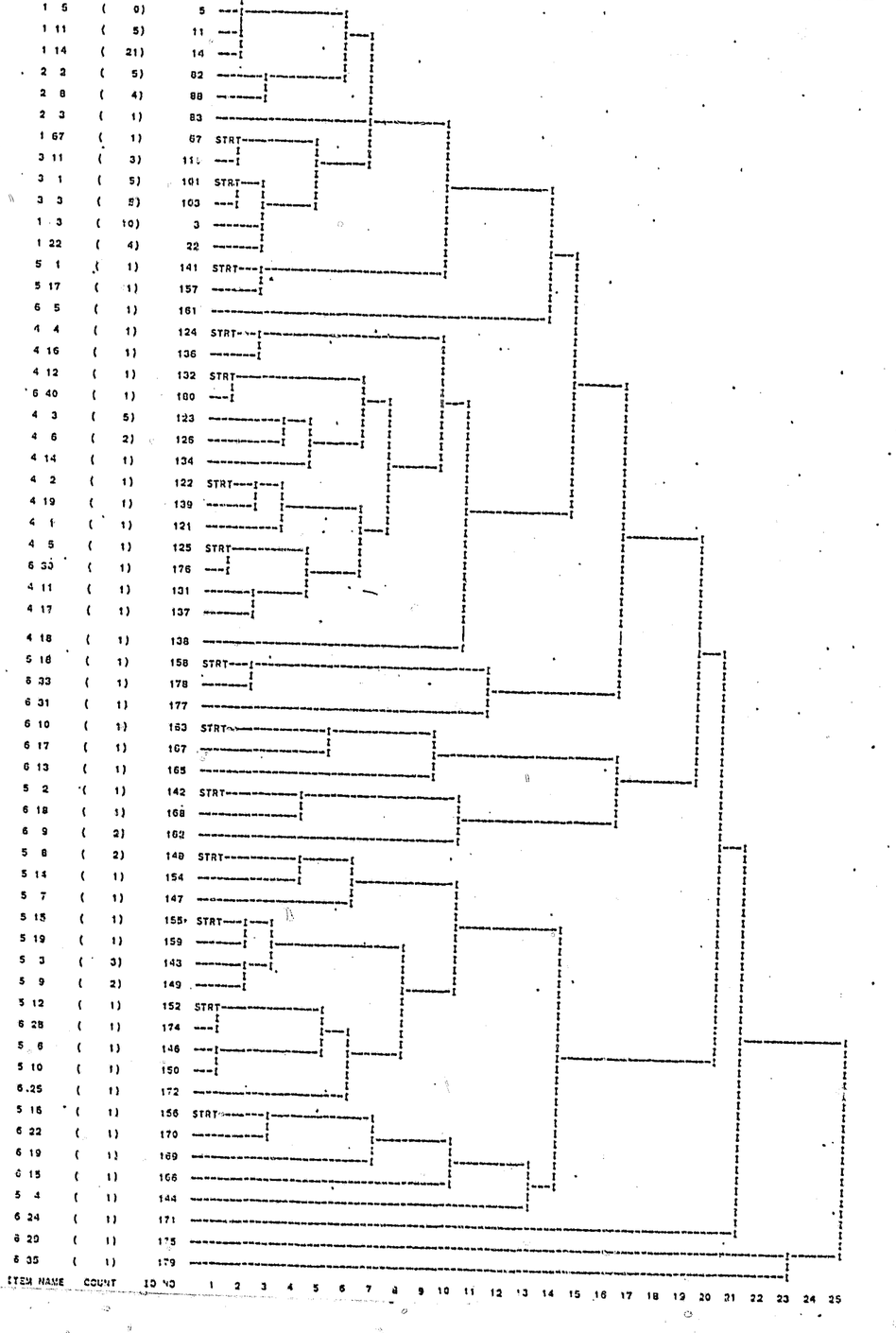
Figure 4F



UPGMA

COWERS COEFFICIENT ON SET F, STANDARDIZED

Figure 5F



be done to determine those criteria, however they would probably be based primarily on the relative lengths of branch lines and the number of points in the groups.

A 5-group partition selected in that way is shown in Table 3F (d). It provides the best match with the generated clusters of any of the methods tested. However, it should be viewed with some scepticism, since the selection was made with prior knowledge of the structure of the data.

Partitions resulting from the more conventional approach, using only branches at the same level of the tree, are shown in Table 3F also: (e) and (f) are from the UPGMA-d tree, while (b) and (c) are from the UPGMA-Gower's tree. In the 5-group partitions, populations 1 and 3 are separated while most of populations 4 and 5 remains unclassified or split into small groups. The 3-group partitions show populations 4 and 5 reasonably well, but the distinction between populations 1 and 3 is lost. Clearly some provision for modes of different density is needed.

UPGMA And HMODE With Angular Similarity Coefficients: The tree diagrams from UPGMA and HMODE using cosine and correlation as similarity measures are in Figures 6F-9F. As in data set E, the partitions obtained are influenced more by the similarity coefficient than the clustering method. Therefore the 5-group partitions created using cosine are shown together in Table 4F, while the partitions created by r are shown in table 5F.

After standardization of variables, the centroid of population 1 is quite near the origin. Cluster 1 subtends a 360° angle, therefore the angular coefficients must split it into many fragments. Under both UPGMA and HMODE, the cosine identifies modes for each of the other four populations and assigns points from cluster 1 to each of those modes, as well as to extra small

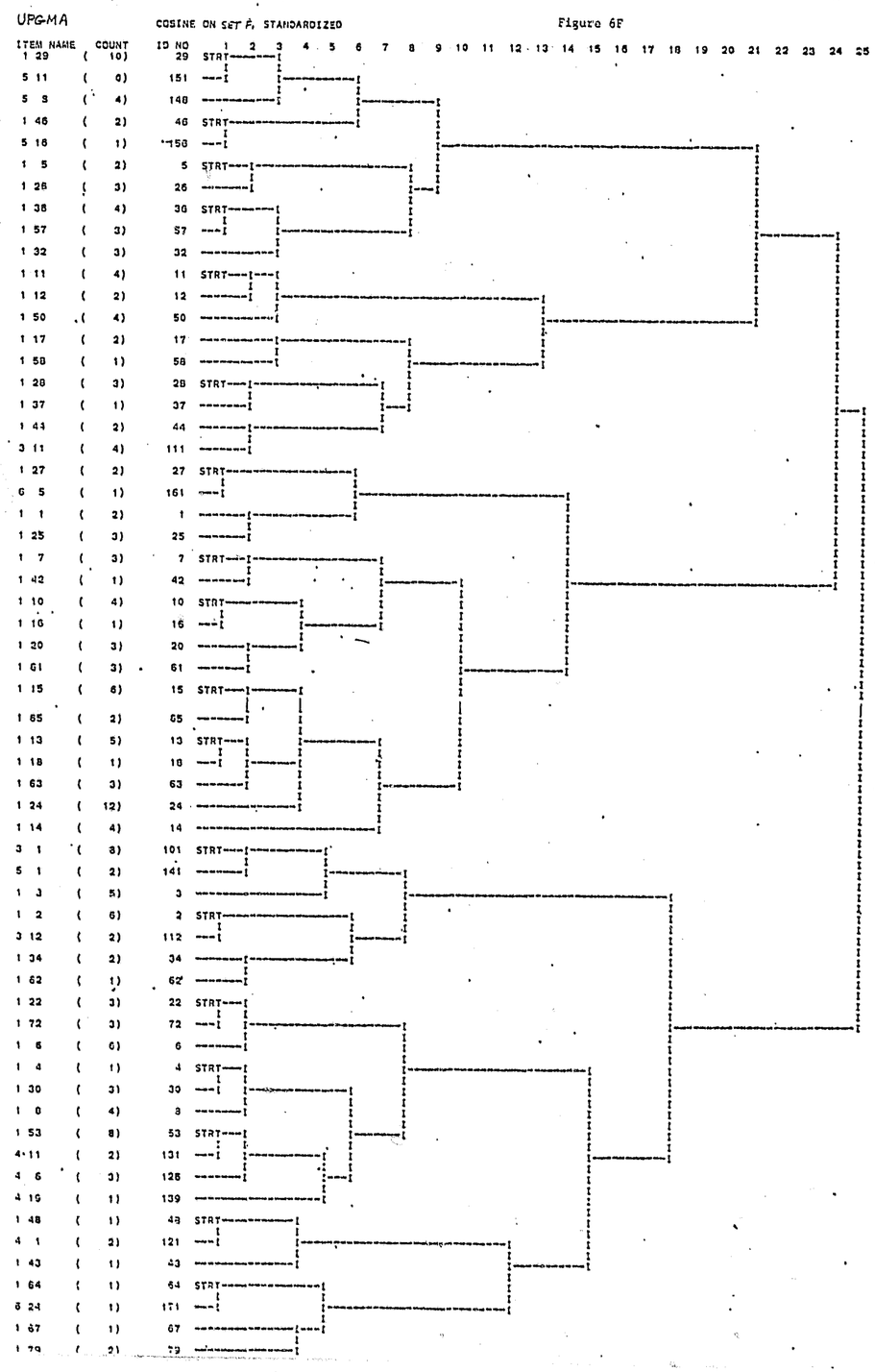
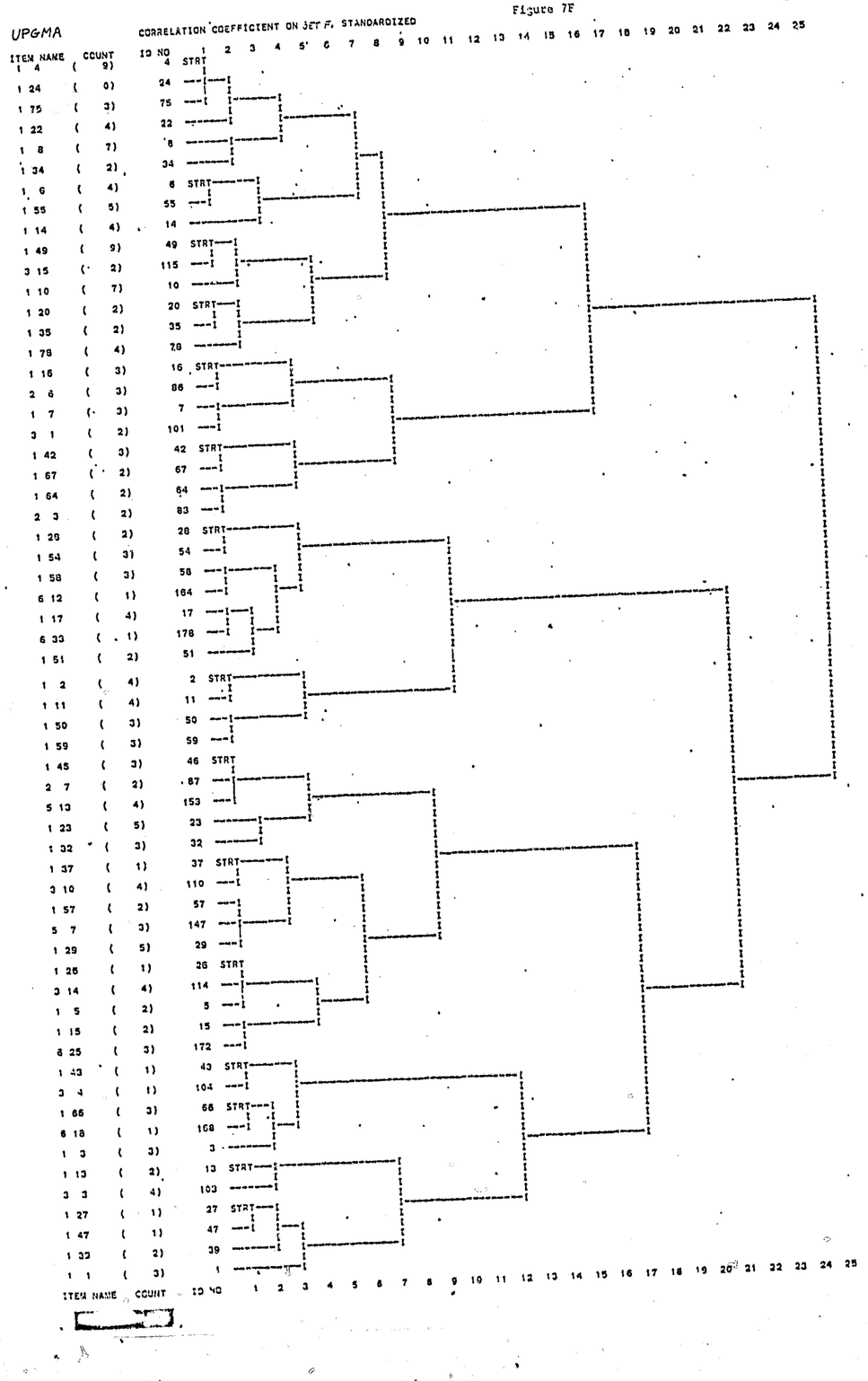


Figure 7F


UPGMA
CORRELATION COEFFICIENT ON SET F, STANDARDIZED
ITEM NAME (COUNT
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
1 24 (0)
1 75 (3)
1 22 (4)
1 8 (7)
1 34 (2)
1 6 (4)
1 55 (5)
1 14 (4)
1 49 (9)
3 15 (2)
1 10 (7)
1 20 (2)
1 35 (2)
1 78 (4)
1 16 (3)
2 6 (3)
1 7 (3)
3 1 (2)
1 42 (3)
1 67 (2)
1 64 (2)
2 3 (2)
1 28 (2)
1 54 (3)
1 58 (3)
6 12 (1)
1 17 (4)
6 33 (1)
1 51 (2)
1 2 (4)
1 11 (4)
1 50 (3)
1 59 (3)
1 45 (3)
2 7 (2)
5 13 (4)
1 23 (5)
1 32 (3)
1 37 (1)
3 10 (4)
1 57 (2)
5 7 (3)
1 29 (5)
1 26 (1)
3 14 (4)
1 5 (2)
1 15 (2)
8 25 (3)
1 43 (1)
3 4 (1)
1 66 (3)
6 18 (1)
1 3 (3)
1 13 (2)
3 3 (4)
1 27 (1)
1 47 (1)
1 33 (2)
1 1 (3)
4 STRT
24
75
22
8
34
6 STRT
55
14
49 STRT
115
10
20 STRT
35
78
16 STRT
86
7
101
42 STRT
67
64
83
28 STRT
54
58
164
17
178
51
2 STRT
11
50
59
46 STRT
87
153
23
32
37 STRT
110
57
147
29
26 STRT
114
5
15
172
43 STRT
104
66 STRT
108
3
13 STRT
103
27 STRT
47
39
1
ITEM NAME COUNT
ID NO

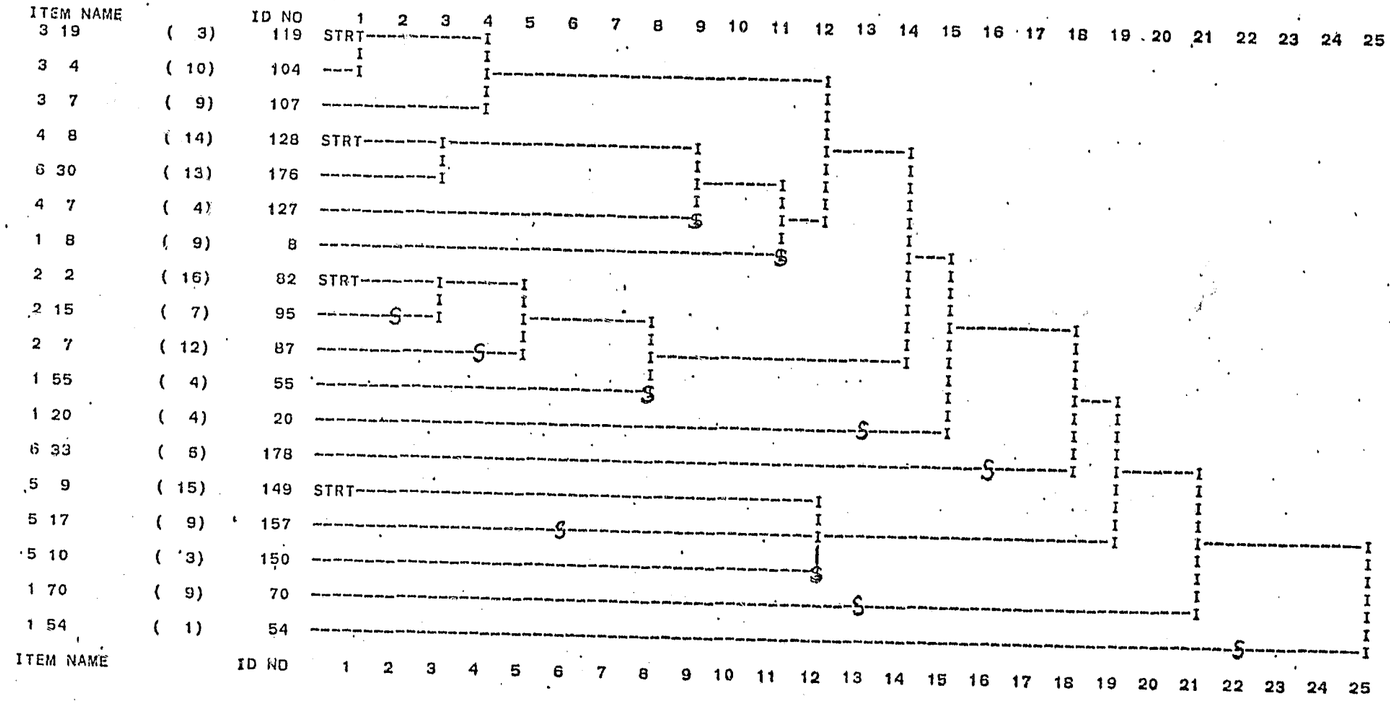
CONTINUED

4 OF 9

HMODE

COSINE ON SET F, STANDARDIZED

Figure 8F



HMODE

CORRELATION COEFFICIENT ON SET F, STANDARDIZED

Figure 9F

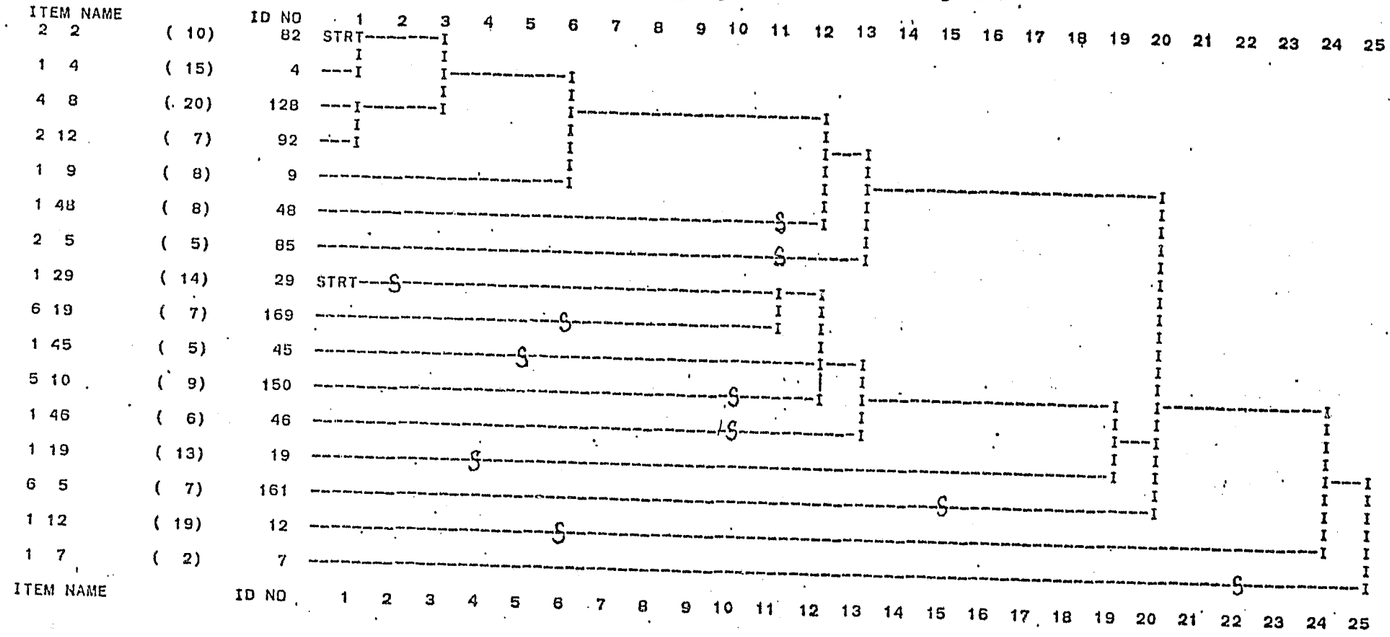


Table 4F

Cluster Overlap Matrices for Data Set F
Partitions using Cosine as Similarity Measure

(a) UPGMA-cos, 5 groups

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	32	19	0	2	0	3	56
2	5	0	14	0	1	6	26
3	17	0	5	18	0	3	43
4	9	0	0	0	17	6	32
5	17	1	1	0	2	2	23
Total	80	20	20	20	20	20	

(b) HMODE-cos, 5 groups

HMODE Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	18	0	19	17	0	8	62
2	19	18	0	0	0	2	39
3	7	1	0	0	1	0	9
4	6	0	0	0	16	5	27
5	2	0	1	0	1	2	6
Outliers	28	1		3	2	3	37
Total	80	20	20	20	20	20	

Table 5F

Cluster Overlap Matrices for Data set F
Partitions Using Correlation as Similarity Measure

(a) UPGMA-r, 5 groups

UPGMA Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	12	0	5	0	0	5	22
2	20	3	2	0	2	3	30
3	27	10	6	17	0	4	64
4	12	3	4	0	18	7	44
5	9	4	3	3	0	1	20
Total	80	20	20	20	20	20	

(b) HMODE-r, 5 groups

HMODE Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	11	1	4	0	1	2	19
2	4	0	0	0	0	3	7
3	29	12	8	20	0	4	73
4	8	1	1	0	1	2	13
5	10	2	4	0	18	7	41
Outliers	18	4	3			2	27
Total	80	20	20	20	20	20	

fragments which contain mostly cluster 1 points. (Although the HMODE-cos 5-group partition lumps clusters 3 and 4 together, they existed as separate modes throughout much of the tree.) The major difference between UPGMA and HMODE is in the large number of outliers and "cluster 1 fragments" left by HMODE, while UPGMA assigns all points to cluster-sized groups at an earlier stage.

While the cosine will only fragment clusters whose centroid is near the origin, r will fragment any cluster that is along the line $x_1 = x_2 = \ldots = x_m$: these are the points with nearly flat profiles. In Table 5F it can be seen that clusters 2 and 3, as well as cluster 1, are spread over several of the 5 groups for the above reason. Clusters 4 and 5, which do not fall on the line of equal coordinates after standardization, were recognized as separate modes but were then mixed with many points from the three fragmented clusters. As a method of clustering points in space, this is clearly a disastrous failure. If the correlation, r, is to be used to find clusters of points in space it would seem essential to first remove all points whose profile-shape is nearly flat.

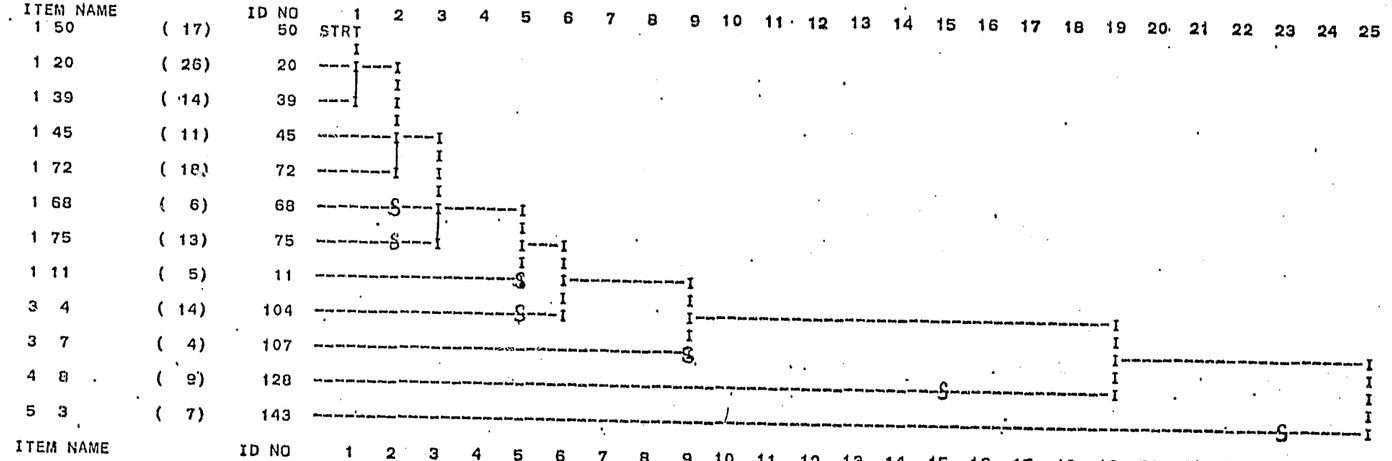
Similarly, if points are to be clustered using the cosine it would be wise to also cluster using distance, to detect and perhaps eliminate a group centered near the origin.

HMODE With Distance-Based Similarity Coefficients: The HMODE trees using Euclidean d and Gower's coefficient are in Figures 10F and 11F, with the resulting 5-group partitions summarized in Table 6F. The large number of outliers in the partitions is governed by the setting of a parameter in the HMODE program (described in our discussion of the HMODE results in data set E). The allowable proportion of outliers was set at .8 for both sets E and

HMODE

EUCLIDEAN DISTANCE ON SET F, STANDARDIZED

Figure 10F



HMODE

GOWERS COEFFICIENT ON SET F, STANDARDIZED

Figure 11F

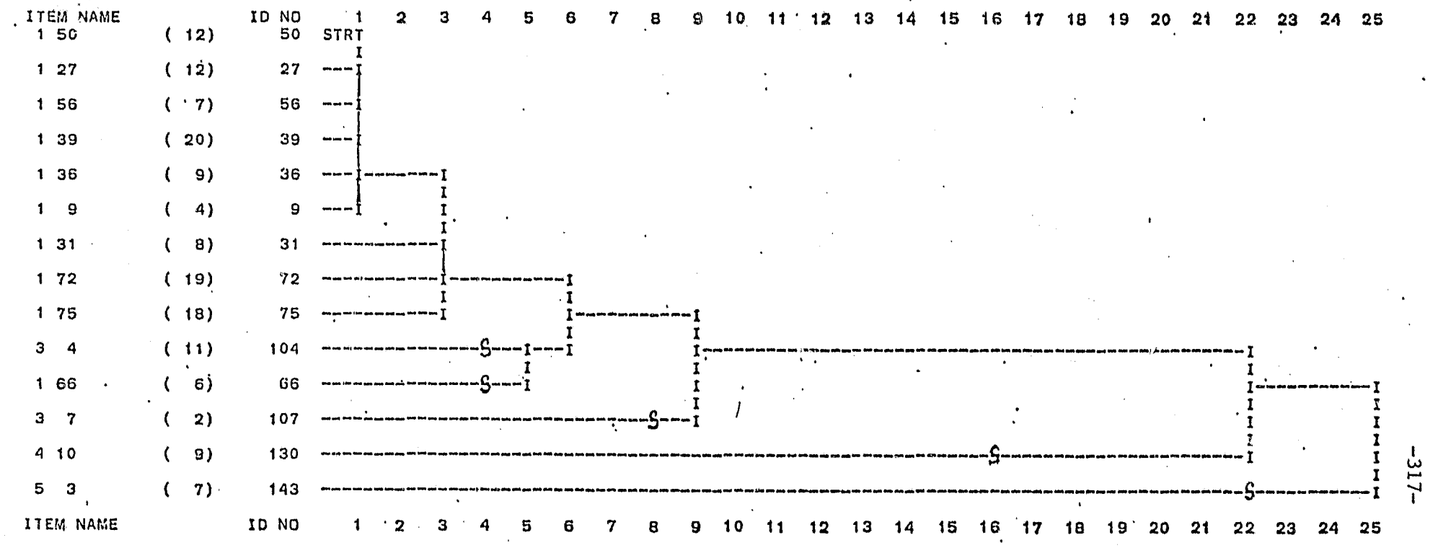


Table 6F

Cluster Overlap Matrices for Data Set F
HMODE with Distance-Based Similarity Coefficient

(a) HMODE-d, 5 clusters

HMODE Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	74	19	5	5	1	1	105
2	2	0	11	0	1	0	14
3	4	1	0	0	0	0	5
4	0	0	0	8	0	1	9
5	0	0	0	0	6	1	7
Outliers			4	7	12	17	40
Total	80	20	20	20	20	20	

(b) HMODE - Gower's, 5 clusters

HMODE Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	77	20	7	3	1	1	109
2	3	0	1	2	0	0	6
3	0	0	10	0	1	0	11
4	0	0	0	8	0	1	9
5	0	0	0	0	6	1	7
Outliers			2	7	12	17	38
Total	80	20	20	20	20	20	

(c) HMODE-d, 4 clusters

HMODE Partition	Generated Partition 1	2	3	4	5	Noise	Total
1	78	20	5	5	1	1	110
2	2	0	11	0	1	0	14
3	0	0	0	8	0	1	9
4	0	0	0	0	6	1	7
Outliers			4	7	12	17	40
Total	80	20	20	20	20	20	

F. Since approximately 10% of the points in each of these sets is added noise, this setting allows some of the points from the generated clusters to also be left as outliers.

The result in set F is that many of the points in the sparsely populated clusters 4 and 5 are relegated to outlier status. Small modes were detected among the points of clusters 4 and 5, but such small groups are often unstable, not as convincing evidence of the existence of separate populations as we would like. This points up a danger in allowing too many outliers: If clusters differ greatly in compactness, then the "outlier" category may contain large segments of the least compact clusters. Clearly in set F the assignment of points to groups was terminated too soon. (The UPGMA 5-group partitions obtained when all branches were cut at the same level also left much of clusters 4 and 5 in the outlier category.)

Using both d and Gower's coef., HMODE was unable to detect a separate mode in the vicinity of cluster 2. The points of cluster 2 were joined to secondary clusters dominated by points of cluster 1. It appears that the two modes are too close together to be recognized as separate populations under the HMODE criterion. A distinct mode was detected within cluster 3, however. Since clusters 1, 2, and 3 were all drawn from populations with equal variances, $\sigma^2 = 1$, this may give some indication of how far apart populations must be to be distinguished by HMODE. We compute the taxometric distance, $T=d/\sqrt{m}$ where m = number of variables, between the centroids of these populations:

$$T^2_{12} = \frac{d^2}{m} = \frac{4(1.5)^2}{4} = (1.5)^2; \quad T_{12} = 1.5$$

$$T^2_{13} = \frac{4(2)^2}{4} = (2)^2; \quad T_{13} = 2$$

In this very simple case, then, we see an indication that centroids of two populations must be separated by at least 2 standard deviations in each variable in order to be detected as separate modes by HMODE. This agrees with Wishart's (1973) theoretical discussion of the one-variable case, in which he points out that a mixture of two normal distributions is unimodal when the means are separated by less than 2 standard deviations.

Because HMODE distinguished only 4 of the 5 populations, the 4-group partition from the HMODE-d tree is included in Table 6F and should be compared with the MIKCA 4-group partitions in Table 1F. Although a mode was detected corresponding to a part of cluster 3, HMODE misclassified more of the boundary points between clusters 1 and 3; therefore its results may be considered inferior to both the UPGMA 5-group partitions and the MIKCA 4-group partitions.

NMODE: Using the default dense-point radius, R, only two modes were found, one combining virtually all of clusters 1, 2, and 3, the other containing just two points of cluster 4. There were 58 outliers, including most of cluster 4 and all of cluster 5. It appears that NMODE defines modes more strictly than HMODE does, requiring greater separation between them and allowing less difference in compactness.

Both larger and smaller values of the dense-point radius were tried. However, at both 1.5 R and 0.5 R only one mode could be found.

If it is important to the researcher to detect overlapping modal clusters, NMODE should certainly not be the method chosen.

Comparison of the Partitions

Data set F posed a difficult challenge for all of the clustering methods. None of the methods were particularly successful in separating clusters 1 and 2 (except when correct centroids were fed into MIKCA, which virtually constitutes giving away the right answer). The hint of a mode containing cluster 2 points can be seen in the UPGMA-d and UPGMA-Gower trees. However, the decision to call it a mode must be a tentative and subjective one on the part of the user, particularly because of the relatively small number of points it contains.

Our ranking of success of the 5-group partitions is shown in Table 7F, based on the above discussion as well as on the smallest estimate of numbers of points misclassified and the number of outliers. The relative positions of the two MIKCA random start solutions (#6 and #7) should be justified, since they do not correspond to the number of points misclassified. Here the type of misclassification was considered more important than the number of points misclassified. The partition from MIKCA-d is a much better reflection of the underlying populations than is the partition resulting from Mahalanobis D.

The four measures of partition agreement again correspond reasonably well with our ranking of the partitions. Our rankings were established without reference to the agreement coefficients, so it is interesting to note that Lambda and Cramer's coefficient agree with our relative positions of the two MIKCA random start solutions, while the Rand and Jaccard coefficients do not. Furthermore, only lambda agrees completely with our rankings in positions #5, 6, and 7. The differences in coefficient values are large, indicating that they may reflect a basic difference in the interpretation of "partition agreement." All of the coefficients reverse our placement of the HMODE-d and

Table 7F

Measures of Partition Agreement on Data Set F
with Success Ratings for the Five-Group Partitions

5-group partitions, in order of success in duplicating the clusters as generated	# of points misclassified (minimum)	cluster pts. left as outliers	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
1. MIKCA, D(centroids given)	15	0	.887	.642	.825	.695
2. MIKCA, d(centroids given)	17	0	.877	.616	.811	.675
3. UPGMA, d(branch lines individually selected)	21	5	.864	.611	.772	.681
4. UPGMA, Gower	20	23	.841	.559	.700	.585
5. UPGMA, d(all branches cut at same level)	34	16	.811	.536	.667	.562
6. MIKCA, d(random start)	53	0	.777	.419	.699	.500
7. MIKCA, D(random start	40	0	.844	.549	.647	.470
8. HMODE, d	37	23	.748	.443	.563	.434
9. HMODE, Gower	38	21	.757	.469	.579	.444
10. UPGMA, cosine	75	0	.702	.228	.557	.304
HMODE, cosine	*	34	.685	.212	.479	.252
UPGMA, r	*	0	.649	.167	.362	.125
HMODE, r	*	25	.639	.169	.347	.140
WARD, d^2: No reasonable 5-group partition						
NMODE, d: No 5-group partition						

Other partitions, ordered by # of groups; within that, ranked according to Lambda	# of groups	Rand's coef.	Jaccard's coef.	Cramer's coef.	Symmetric Lambda
UPGMA, d	3	.724	.467	.821	.577
UPGMA, Gower	3	.718	.462	.825	.574
MIKCA, Euclidean d, det W	3	.744	.466	.851	.514
MIKCA, Mahalanobis D, det W	3	.642	.402	.823	.425
WARD, d^2	3	.318	.221	.246	.017
MIKCA, Mahalanobis D, det W	4	.814	.550	.844	.624
MIKCA, Euclidean d, det W	4	.810	.547	.828	.608
HMODE, d	4	.755	.473	.629	.447

* In these partitions, no meaningful match with the generated clusters can be defined.

HMODE-Gower partitions, however the difference between these two partitions is minor.

Table 7F also lists the 3- and 4-group partitions that have been discussed, with measures of their agreement with the correct 5-group partition. Cramer's coefficient rates most of the 3-group partitions as good as or better than the best 5-group partitions. A quick comparison of some of the cluster overlap table demonstrates that the ratings by Cramer's coef. do not coincide with the intuitive idea of which match is better. (For example, Cramer's rates Table 1F (d) higher than Table 2F (a)). Cramer's coef. appears to incorporate a strong bias in favor of the partition with the smaller number of groups. This was noted also in the discussion of data set E.

On the other hand, each of the 4-group partitions listed is rated higher than its 5-group counterpart by all four of the agreement coefficients (with one exception: Rand's coefficient on MIKCA-D). This near-unanimity substantiates our own judgment that the 4-group partitions are superior, indicating that 4 clusters are the best one can reasonably expect to find in this data, with populations 1 and 2 seen as a single cluster.

That Rand's coefficient provided the one exception is probably no coincidence. We noted in data set E that Rand's coefficient tends to increase as the number of groups in the partition increases. Thus in this case there may be a bias in favor of the 5-group partition.

Evaluation of Individual Clusters

Three of the better partitions will be examined, involving different levels of coverage of the data set:

Figure 12F

CLUSTER GRAPH

DATA SET F
UPGMA-D, STRAIGHT
X=VAR 1 Y=VAR 3

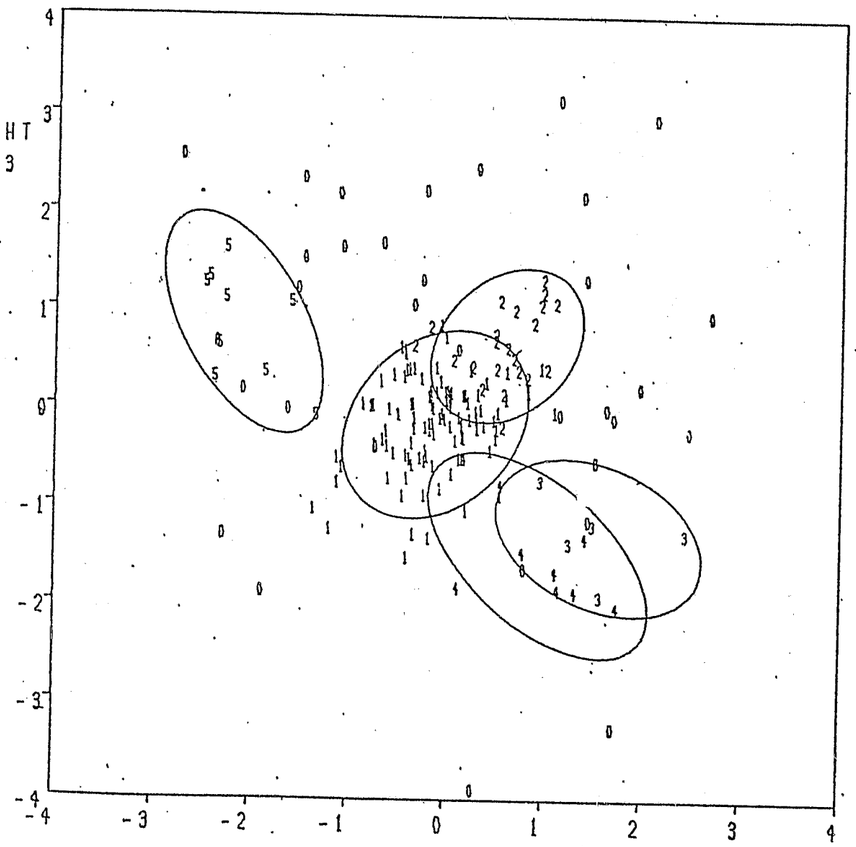


Figure 13F

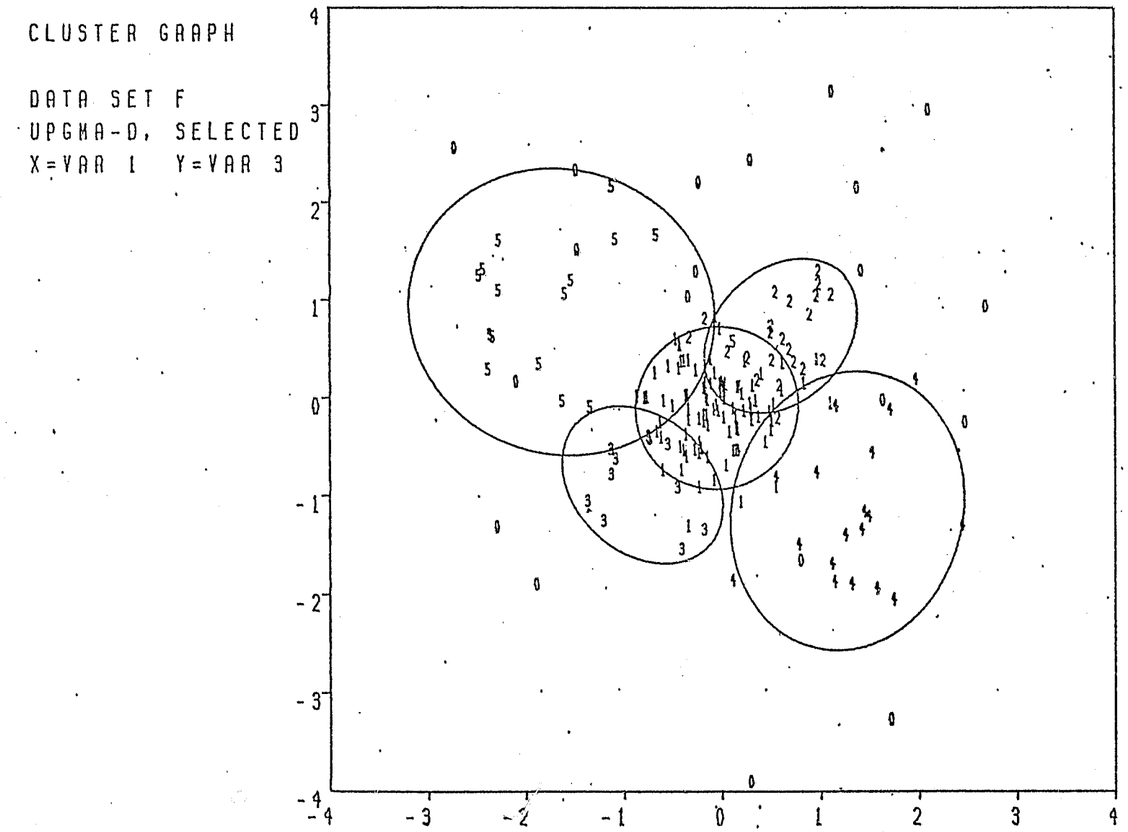
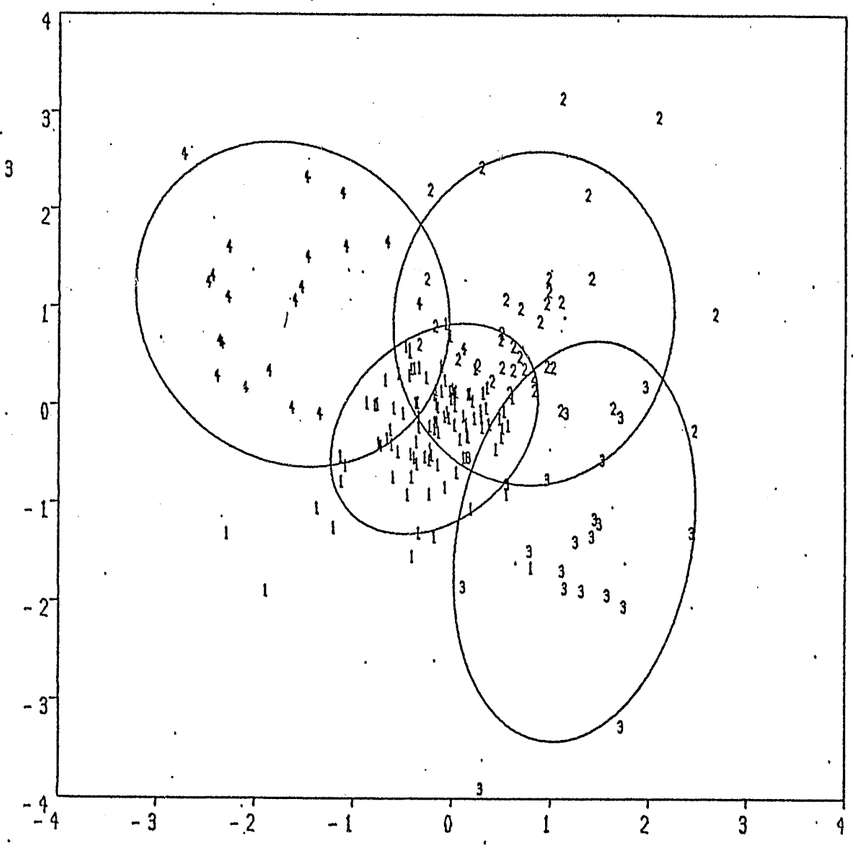

CLUSTER GRAPH
DATA SET F
UPGMA-D, SELECTED
X=VAR 1 Y=VAR 3
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4

Figure 14F

CLUSTER GRAPH

DATA SET F
MIKCA-MAHAL. D
X=VAR 1 Y=VAR 3



	# groups	# outliers	coverage
UPGMA-d, straight cut of branches	5	33	82%
UPGMA-d, selected branch lines	5	20	89%
MIKCA-Mahalanobis D	4	0	100%

Plots of these partitions are shown in Figures 12F-14F, and may be compared with the generated clusters in Figure 1F.

Measures of cluster separation, compactness, and homogeneity for the three partitions are in Tables 8F-10F. It has been determined in the preceding discussion of clustering results that the UPGMA-d 5-group partition involving selected branch lines is superior to the partition obtained from a straight cut of branches at one level. Comparing Tables 8F and 9F, though, one might come to the opposite conclusion. The gamma values show more separation between clusters in the "straight-cut" partition, and the cluster homogeneities in that partition are high. In contrast, the "selected" partition reflects the existence of two voluminous but sparsely populated clusters, showing two groups that are not particularly homogeneous - groups 4 and 5. However, the greater homogeneity in the straight cut partition is primarily due to the lower coverage of the data set; the thirteen more points classified in the selected partition greatly enlarge the spatial territory of the clusters.

Table 10F describes a partition involving complete coverage: the MIKCA-D, 4-group partition. The incorporation of "noise" points into the groups greatly enlarges their territories, so cluster homogeneity decreases tremendously. Unless a high degree of homogeneity is required for theoretical reasons, comparisons of homogeneity between partitions should probably be confined to those that involve the same amount of coverage.

UPGMA-d
Straight cut, 5 groups
(33 outliers)

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	1479	3788	139	5325

GAMMA -.6984

GAMMAB -.4384

GAMMAW -.9491

MAX. CLUSTER DIAMETER 3.0260
MIN. DIST. BTWN BOUNDARIES .4063

CLUSTER DIAMETERS

1	2	3	4	5
3.026	2.186	2.079	2.246	2.497

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5
1	.794	.858	.692	.681	.833
2	.839	.744	.491	.852	.688
3	.790	.854	.822	.727	.680
4	.820	.825	.689	.868	.698

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4
.795	.804	.790	.812

CLUSTER HOMOGENITIES

1	2	3	4	5
.811	.820	.673	.782	.725

NUMBER OF CLUSTERS 5

CLUSTER CENTERS

	1	2	3	4
1	-.124	-.101	-.183	-.300
2	.595	.523	.630	.595
3	1.554	.883	-1.320	-.280
4	.939	1.497	-1.504	-1.406
5	-2.099	-1.024	.834	1.273

INTER-CLUSTER CENTROID DISTANCES

	1	2	3	4	5
1	0	1.539	2.254	2.579	2.874
2	1.539	0	2.370	3.102	3.186
3	2.254	2.370	0	1.434	4.903
4	2.579	3.102	1.434	0	5.313
5	2.874	3.186	4.903	5.313	0

COSINE OF THE ANGLE (AT THE ORIGIN)* BETWEEN CLUSTER CENTROIDS

	1	2	3	4	5
1	1.000	-.927	.050	.410	-.158
2	-.927	1.000	.148	-.139	-.153
3	.050	.148	1.000	.849	-.901
4	.410	-.139	.849	1.000	-.866
5	-.158	-.153	-.901	-.866	1.000

DISTANCE BETWEEN CLUSTER BOUNDARIES

	1	2	3	4	5
1	0	.406	.821	1.526	1.038
2	.406	0	1.566	1.971	1.768
3	.821	1.566	0	.884	3.142
4	1.526	1.971	.884	0	3.604
5	1.038	1.768	3.142	3.604	0

CLUSTER SQUARED DEVIATION SCORE AND NUMBER OF ELEMENTS

1	76.419	102
2	15.102	22
3	5.225	5
4	6.977	9
5	8.810	9

TOTAL OF SQUARED DEVIATION SCORES= 112.533

Table 9F
UPGMA-d
Selected branches, 5 groups
(0 outliers)

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	1147	6804	139	4630
GAMMA	-.7978			
GAMMAB	-.7115			
GAMMAW	-.9417			

MAX. CLUSTER DIAMETER 3.7159
MIN. DIST. BTWN BOUNDARIES .4083

CLUSTER DIAMETERS

1	2	3	4	5
2.456	2.188	2.277	3.629	3.718

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5
1	.846	.858	.832	.667	.417
2	.882	.744	.814	.379	.224
3	.844	.854	.841	.527	.489
4	.841	.845	.767	.399	.554

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4
.780	.729	.770	.751

CLUSTER HOMOGENITIES

1	2	3	4	5
.853	.820	.814	.493	.421

NUMBER OF CLUSTERS 5

CLUSTER CENTERS

	1	2	3	4
1	-.050	-.036	-.106	-.258
2	.595	.523	.630	.595
3	-.812	-.694	-.889	-.684
4	1.267	1.367	-1.151	-1.054
5	-1.636	-1.053	.879	1.689

INTER-CLUSTER CENTROID DISTANCES

	1	2	3	4	5
1	0	1.413	1.345	2.329	2.883
2	1.413	0	2.721	2.655	2.953
3	1.345	2.721	0	2.962	3.092
4	2.329	2.655	2.962	0	5.091
5	2.883	2.953	3.092	5.091	0

COSINE OF THE ANGLE (AT THE ORIGIN)* BETWEEN CLUSTER CENTROIDS

	1	2	3	4	5
1	1.000	-.801	.759	.406	-.527
2	-.801	1.000	-.997	.041	.011
3	.759	-.997	1.000	-.062	.029
4	.406	.041	-.062	1.000	-.953
5	-.527	.011	.029	-.953	1.000

DISTANCE BETWEEN CLUSTER BOUNDARIES

	1	2	3	4	5
1	0	.406	.511	.720	1.038
2	.406	0	1.393	1.378	1.473
3	.511	1.393	0	1.983	1.425
4	.720	1.378	1.983	0	3.016
5	1.038	1.473	1.425	3.016	0

CLUSTER SQUARED DEVIATION SCORE AND NUMBER OF ELEMENTS

1	53.389	92
2	15.102	22
3	8.711	10
4	36.510	19
5	37.042	17

TOTAL OF SQUARED DEVIATION SCORES* 148.754

Table 10F
MIKCA-D
4 groups
(0 outliers)

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	355	9568	33	6154

GAMMA -.9518

GAMMAB -.9284

GAMMAW -.9893

MAX. CLUSTER DIAMETER 5.6528
MIN. DIST. BTWN BOUNDARIES .2710

CLUSTER DIAMETERS

1	2	3	4
5.024	5.653	4.802	4.929

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4
1	.751	.528	.644	.406
2	.768	.284	.433	-.266
3	.739	.331	.028	.353
4	.706	.212	.389	.088

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4
.653	.508	.530	.497

CLUSTER HOMOGENITIES

1	2	3	4
.741	.339	.373	.145

NUMBER OF CLUSTERS 4

CLUSTER CENTERS

	1	2	3	4
1	-.187	-.131	-.241	-.299
2	.823	.476	.889	.704
3	1.243	1.348	-1.383	-1.111
4	-1.630	-1.436	1.025	1.327

INTER-CLUSTER CENTROID DISTANCES

	1	2	3	4
1	0	1.916	2.489	2.834
2	1.916	0	3.064	3.175
3	2.489	3.064	0	5.268
4	2.834	3.175	5.268	0

COSINE OF THE ANGLE (AT THE ORIGIN)* BETWEEN CLUSTER CENTROIDS

	1	2	3	4
1	1.000	-.970	.225	-.123
2	-.970	1.000	-.091	-.045
3	.225	-.091	1.000	-.979
4	-.123	-.045	-.979	1.000

DISTANCE BETWEEN CLUSTER BOUNDARIES

	1	2	3	4
1	0	.271	.821	1.038
2	.271	0	.720	1.545
3	.821	.720	0	3.016
4	1.038	1.545	3.016	0

CLUSTER SQUARED DEVIATION SCORE AND NUMBER OF ELEMENTS

1	104.726	102
2	89.896	35
3	50.136	21
4	71.812	22

TOTAL OF SQUARED DEVIATION SCORES= 316.569

DATA SET U:

A UNIFORM DISTRIBUTION

The preceding data, set F, involved populations which overlapped so much that a question was raised as to whether two of the populations should be identified as separate modes at all. Yet most of the clustering methods partitioned the data into groups. Can they tell us that a data set should be regarded as a single large group, with no significant modes?

Data set U contains 150 points, randomly sampled from a single population which is uniformly distributed on 8 variables, each having mean = 0 and standard deviation = 1. Variations in density are random and small in magnitude. Thus the "correct" number of groups in this data is 1.

Clustering Results

MIKCA, Minimizing Determinant W: Partitions containing 4 groups and 6 groups were requested, using both Mahalanobis D and Euclidean d. MIKCA obligingly created them, giving no clue that the data is uniformly distributed. Group sizes were reasonable, and the partitions produced by D and d at the 6-group level were strikingly similar. The cluster overlap matrices in Table 1U (a) and (b) show the matches between three MIKCA partitions. The matches are certainly not bad enough to discourage an intrepid cluster-seeker.

WARD: The WARD tree, shown in Figure 1, clearly depicts a single population, with one or two points being added to the single large group at each level of the tree. Unfortunately, WARD produced a very similar tree for data set F: multiple modes that were not well-separated appeared also to be a single population. We conclude that this stairstep pattern in a WARD tree may

Table 1U

Cluster Overlap Matrices for Data Set U
Matches between MIKCA and UPGMA-d Partitions

(a) MIKCA, det W, 6 group partitions using D and d

MIKCA-d 6 groups	MIKCA-D, 6 groups 1	2	3	4	5	6	Total
1	22	1	2	0	0	0	25
2	1	15	0	0	0	0	16
3	0	1	32	0	0	0	33
4	0	0	2	34	0	0	36
5	1	0	2	2	1	14	20
6	1	0	1	0	18	0	20
Total	25	17	39	36	19	14	

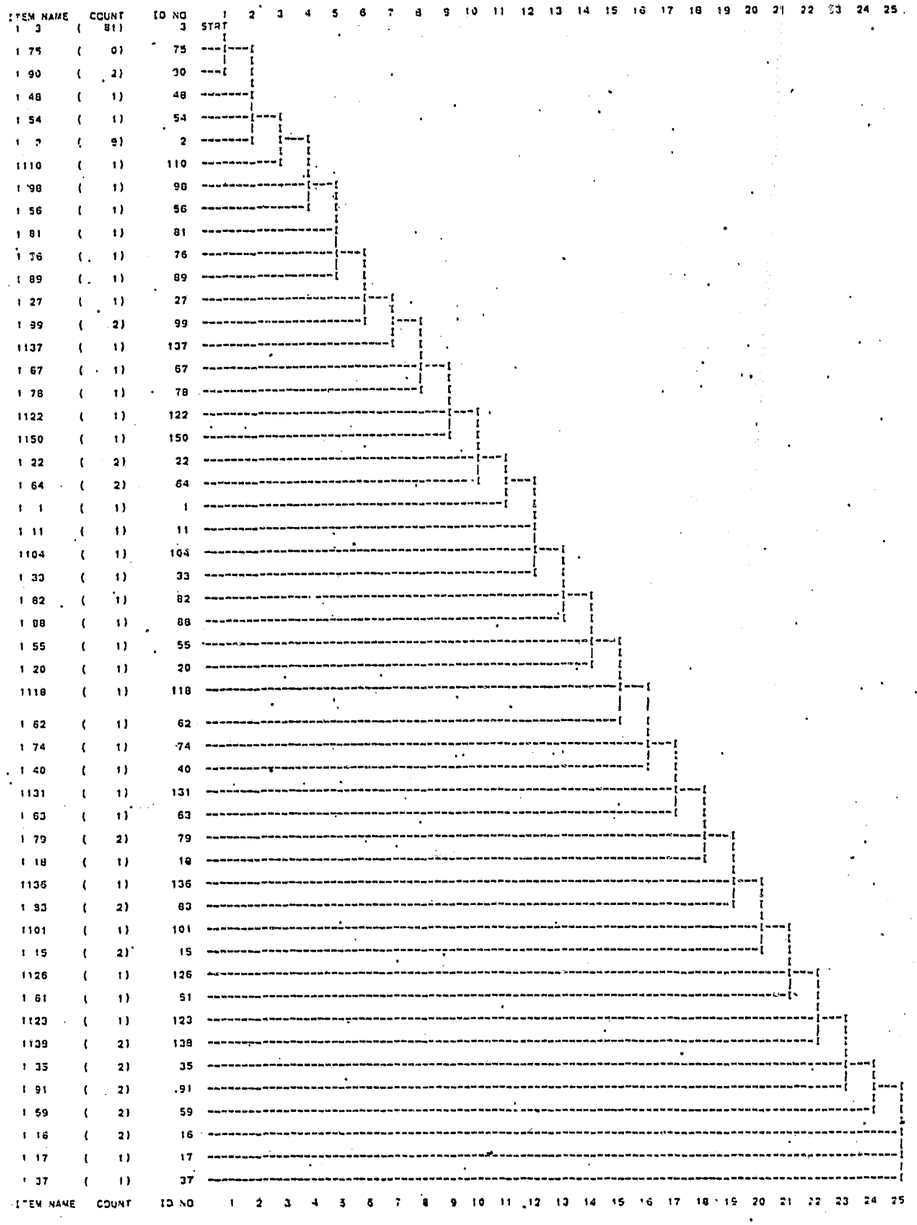
(b) MIKCA, det W, D: 4 and 6-group partitions

MIKCA-D 4 groups	MIKCA-D, 6 groups 1	2	3	4	5	6	Total
1	17	2	3	0	0	6	28
2	1	11	0	18	2	3	35
3	0	3	34	4	17	3	61
4	7	1	2	14	0	2	26
Total	25	17	39	36	19	14	

(c) MIKCA-D vs. UPGMA-d, 6-group partitions

UPGMA-d 6 groups	MIKCA-D, 6 groups 1	2	3	4	5	6	Total
1	5	0	2	0	1	1	9
2	5	1	8	30	1	11	56
3	2	0	3	4	0	0	9
4	1	1	26	0	1	0	29
5	12	14	0	1	0	1	28
6	0	1	0	1	16	1	19
Total	25	17	39	36	19	14	

Figure 1U



TREE DIAGRAM FOR DATA SET U
WARD METHOD WITH EUCLIDEAN DISTANCE
(Last 50 levels of the tree only)

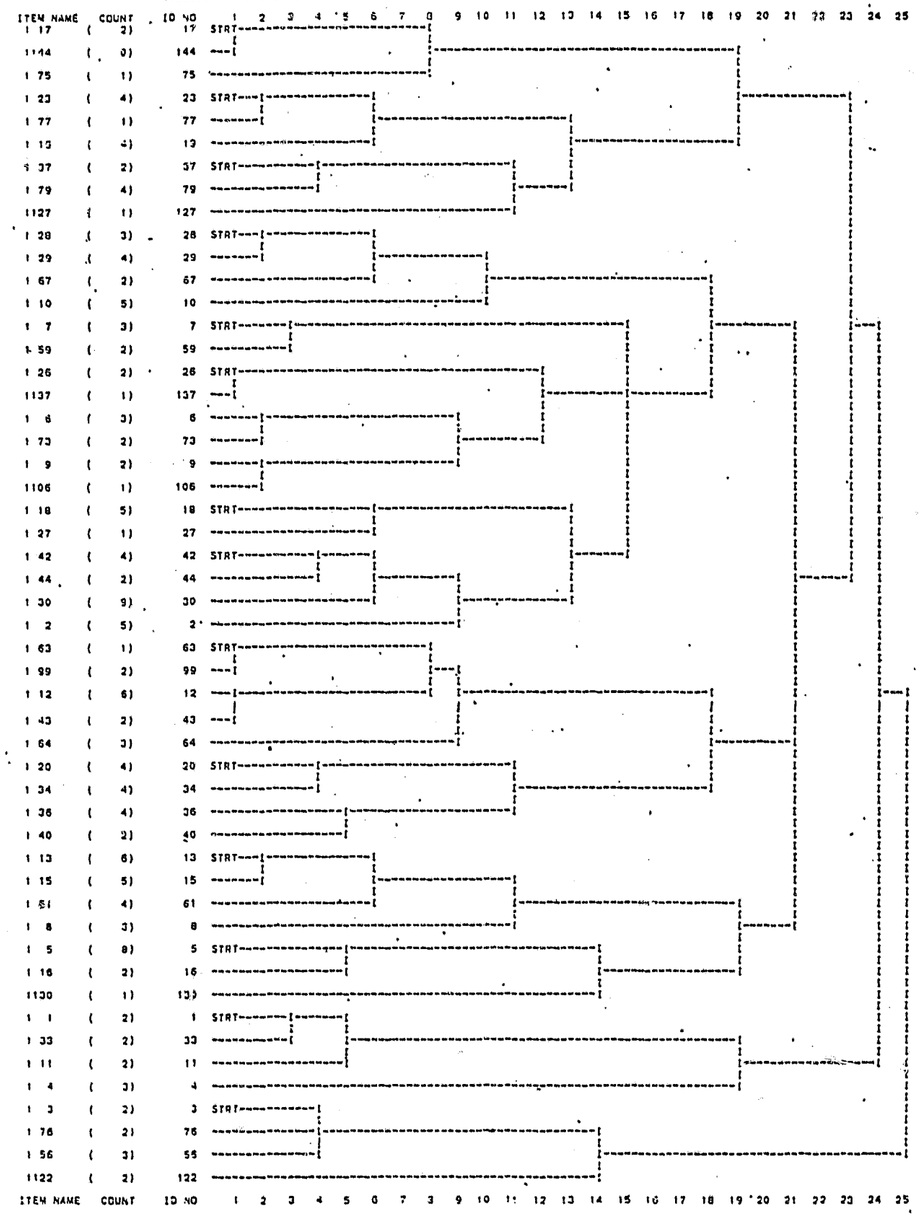
indicate that there are no substantial empty spaces between clusters, but it does not preclude the existence of multiple modes. If well-separated clusters are desired, then this tree is excellent proof of their absence. However, if data contains heavy background noise and/or closely spaced modes a more sensitive approach is needed.

UPGMA (all similarity coefficients): The UPGMA trees have a surprisingly similar appearance for all four similarity measures on this data. Figures 2U-5U show the trees using d, Gower, cosine, and r, respectively. Each figure depicts only the last one-third of the tree; that is 99 merges have occurred before printing of the tree began. At the left of each branch line, the number of points already merged on that branch is given in parentheses. In each tree, the early merges occurred fairly uniformly so that points are distributed more evenly over the branch lines than in UPGMA trees for our other data sets. The many small clusters come together at regular intervals along the tree. Thus branches are rather evenly spaced, making it difficult to single out a particular k-level as the "right" number of clusters. Most importantly, the last six to eight branches merge within a fairly small similarity range, indicating that there is no great distinction between them.

There is clearly a difference between these UPGMA trees and the ones produced for the other, more structured, data sets. But the difference is subtle and could easily be ignored. There is a danger that this method would be used to partition a uniform distribution into approximately 6 clusters just as MIKCA did.

In Table 1U (c), the UPGMA-d 6-group partition is matched with the MIKCA-D 6-group partition. There is a good match between the two methods on 4 of the 6 groups. This is not altogether surprising, since any arbitrarily drawn

Figure 2U



TREE DIAGRAM FOR DATA SET U
GROUP AVERAGE METHOD WITH EUCLIDEAN DISTANCE
(Last 50 levels of the tree only)

Figure 3U

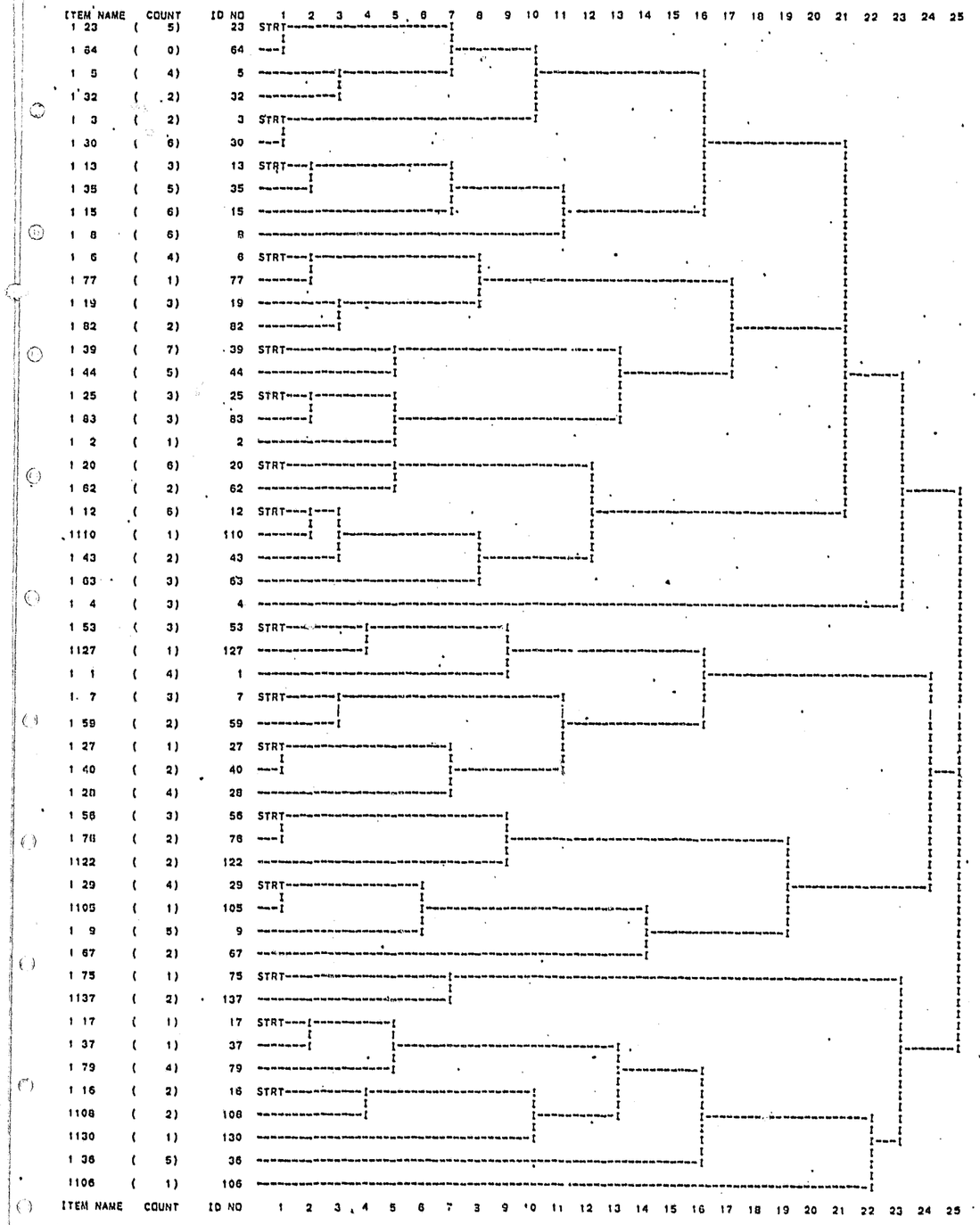
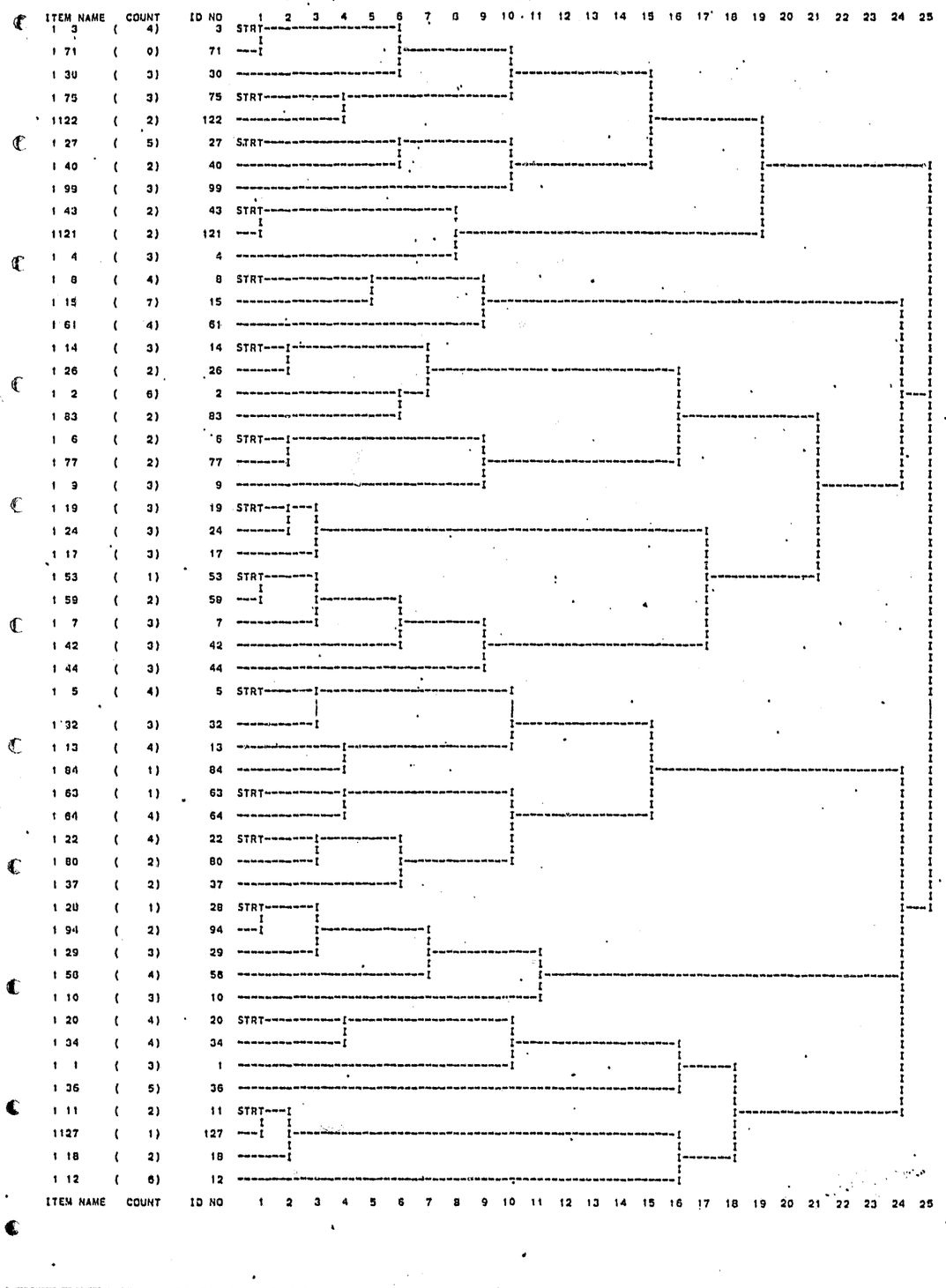
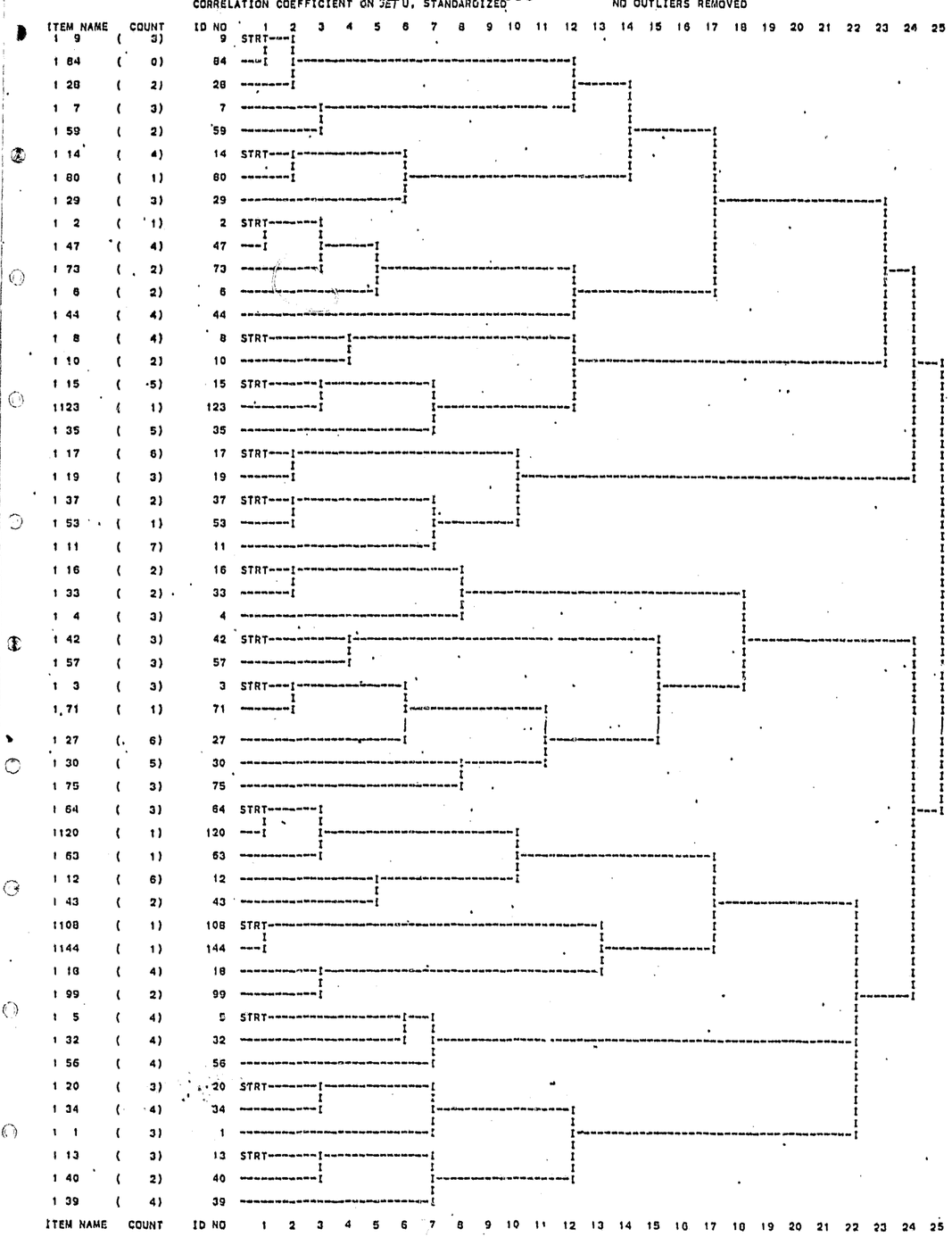


Figure 4

COSINE ON SET U, STANDARDIZED NO OUTLIERS REMOVED



CORRELATION COEFFICIENT ON SET U, STANDARDIZED NO OUTLIERS REMOVED



cartets (boxes) in space will group near neighbors together. However, it does illustrate that a moderate agreement between clustering methods does not guarantee the existence of "natural" boundaries in the data.

HMODE (All Similarity Coefficients): The HMODE trees produced using the four similarity measures are in Figures 6U-9U. Like the UPGMA trees, they could easily be interpreted as if there were clusters in the data. Again, there is a subtle difference between the trees for set U and those for our other data sets--not in the visual appearance of the trees but in the similarity ranges covered by the trees. The following table shows the ranges within which all secondary modes merge into a single group, for each of our artificial data sets. (Distance has been divided by $\sqrt{m}$ in order to compare data sets of different dimensions.)

	$d/\sqrt{m}$		Gower		Cosine		Correlation	
	First Merge	Last Merge	First Merge	Last Merge	First Merge	Last Merge	First Merge	Last Merge
Set D	.22	1.65	.96	.60	.99	-.04	.99	-.24
Set E	.16	.43	.97	.92	1.00	.91	.99	.95
Set F	.16	.63	.98	.92	.95	.87	.98	.92
Set U	.65	.78	.85	.82	.73	.66	.76	.70

It can be seen that merges of secondary modes in set U begin at a much lower similarity level (or higher distance) than in the other data sets, yet the merges are completed within a very narrow similarity range. The similarities given are those between nearest neighbors. Those for set U indicate even spacing between the secondary modes with no regions of high density.

Figure 6U

TREE DIAGRAM FOR DATA SET U
HIERARCHICAL MODE SEARCH (HMODE) WITH EUCLIDEAN DISTANCE
(Tree shows merges of primary clusters only)

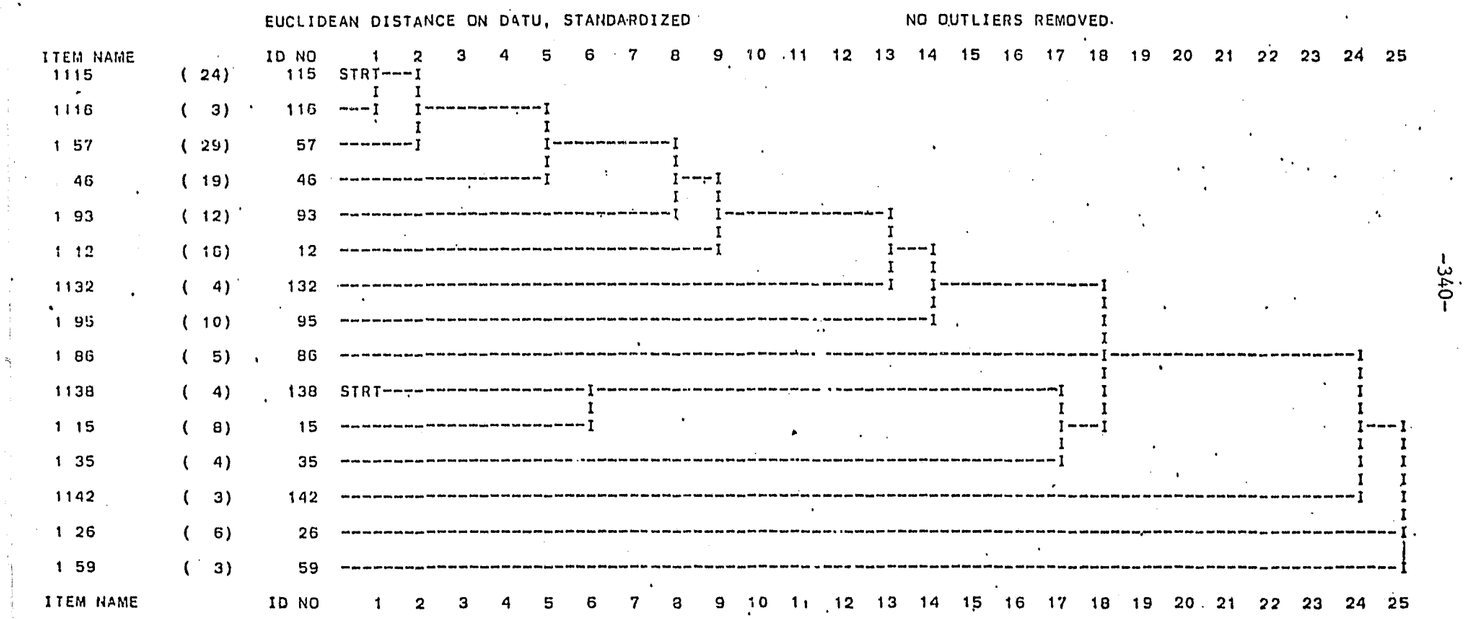


Figure 7U

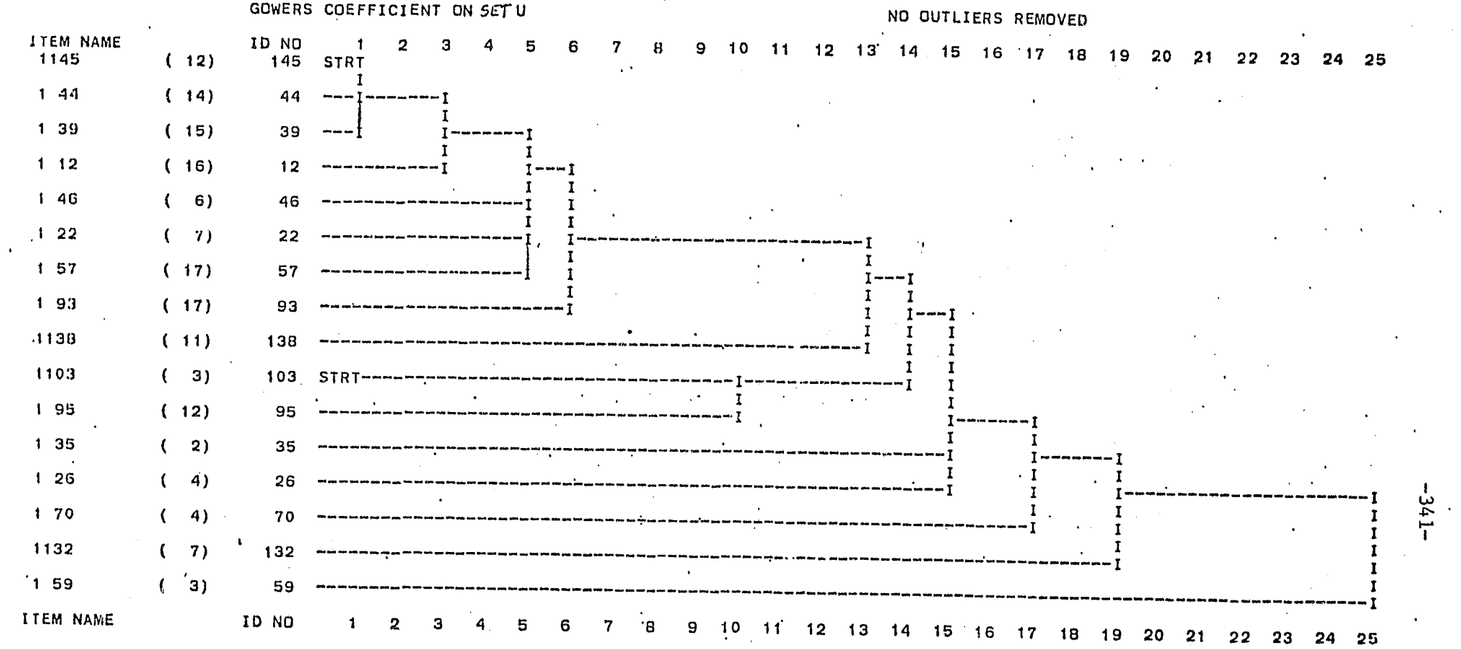

GOWERS COEFFICIENT ON SET U
NO OUTLIERS REMOVED
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
1145 (12) 145 STRT
1 44 (14) 44
1 39 (15) 39
1 12 (16) 12
1 46 (6) 46
1 22 (7) 22
1 57 (17) 57
1 93 (17) 93
1138 (11) 138
1103 (3) 103 STRT
1 95 (12) 95
1 35 (2) 35
1 26 (4) 26
1 70 (4) 70
1132 (7) 132
1 59 (3) 59
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 80

COSINE ON SET U, STANDARDIZED — NO OUTLIERS REMOVED

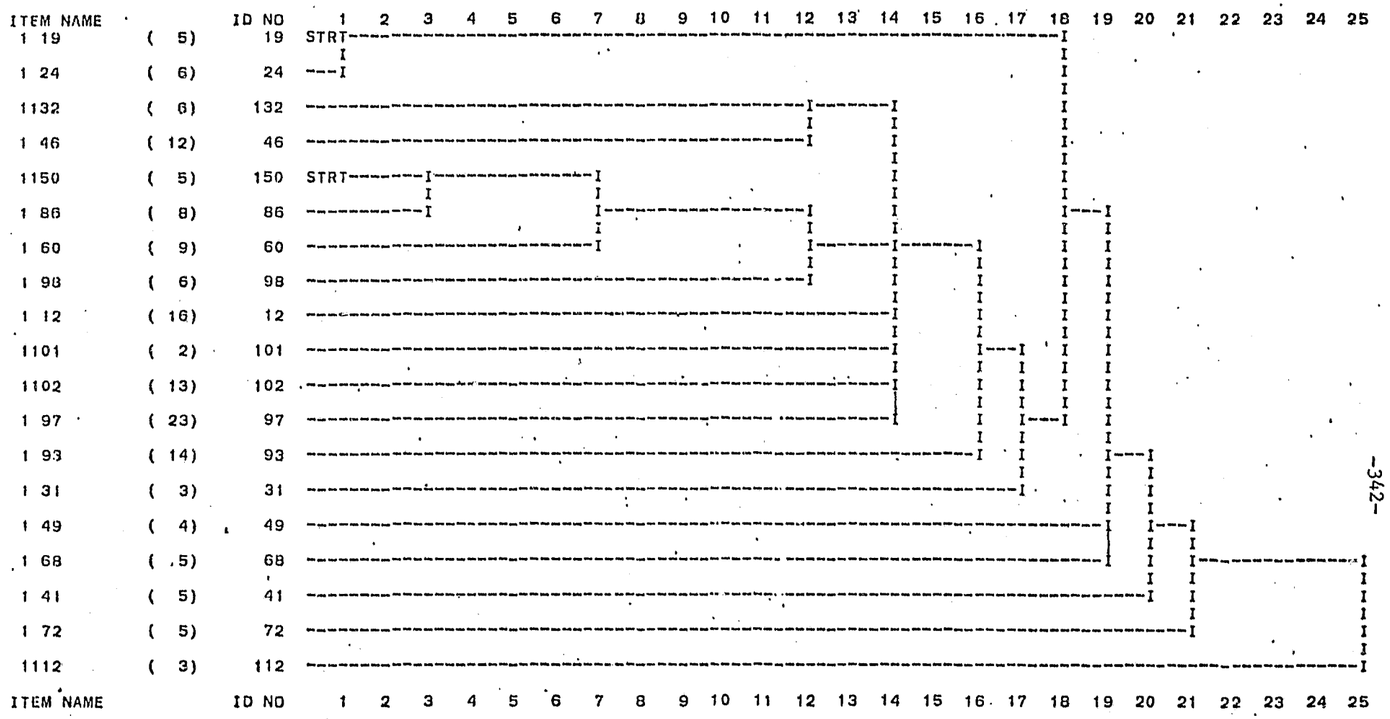


Figure 9U

CORRELATION COEFFICIENT ON SET U, STANDARDIZED NO OUTLIERS REMOVED

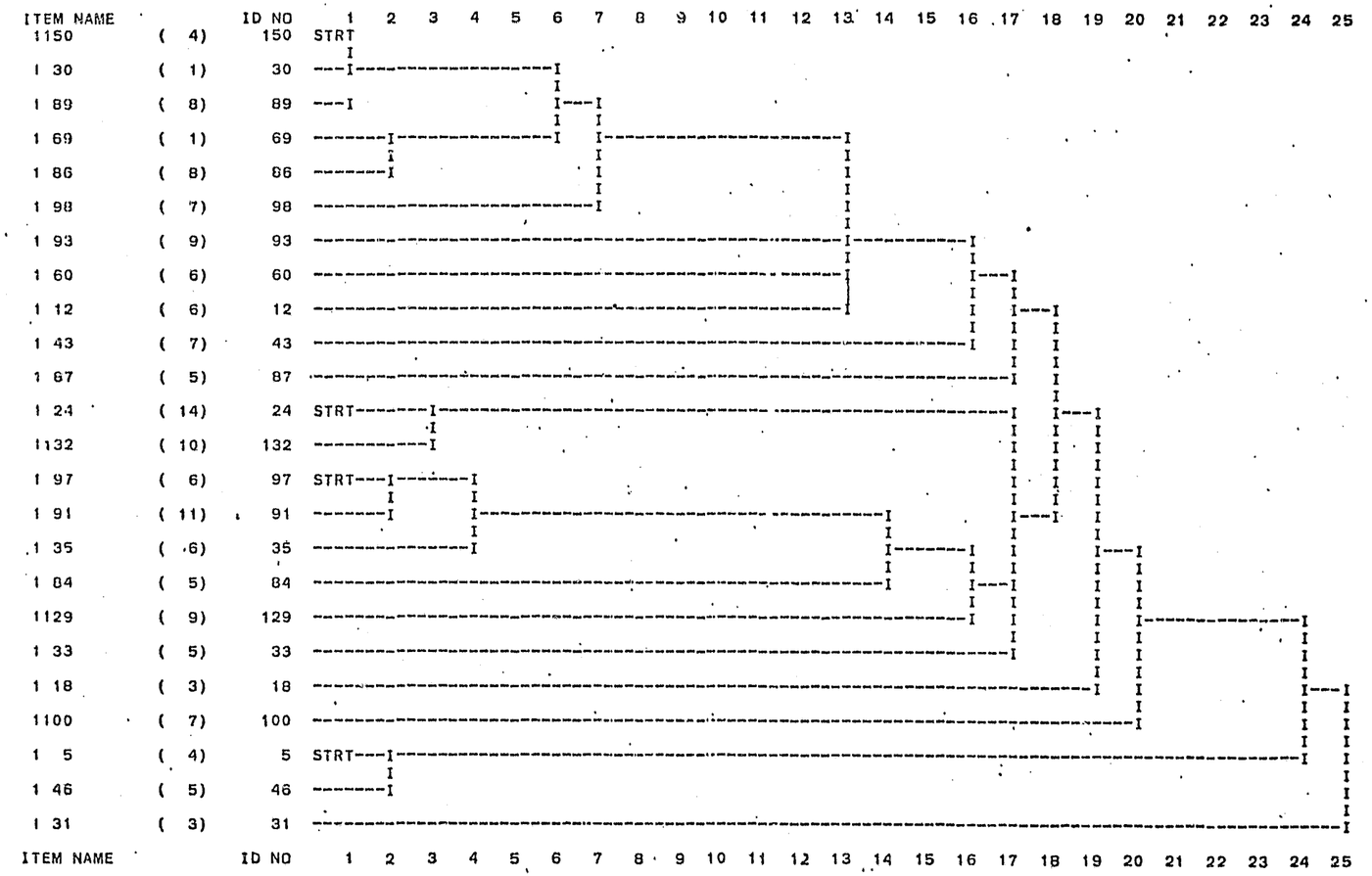


Table 2U

Cluster Overlap Matrices for Data Set U
Matches between HMODE and UPGMA Partitions

(a) UPGMA vs. HMODE, both with Euclidean d

HMODE-d	UPGMA-d, 7 groups								
7 groups	1	2	3	4	5	6	7	Outliers	Total
1	4	0	0	0	0	9	0	3	16
2	0	5	0	0	0	0	0		5
3	2	35	5	11	2	9	9	2	75
4	0	0	0	0	12	0	0		12
5	0	1	1	0	0	10	0		12
6	0	5	0	0	0	0	0	1	6
7	0	7	3	0	0	0	0		10
Outliers		3			4		7		14
Total	6	56	9	11	18	28	16	6	

(b) UPGMA, d vs. cosine

UPGMA-cos	UPGMA-d, 7 groups								
7 groups	1	2	3	4	5	6	7	Outliers	Total
1	5	3	0	0	0	18	1		27
2	0	19	0	0	0	0	1		20
3	0	9	8	1	0	6	0	5	29
4	0	1	0	10	3	4	7		25
5	1	12	0	0	0	0	7	1	21
6	0	0	0	0	15	0	0		15
7	0	12	1	0	0	0	0		13
Total	6	56	9	11	18	28	16	6	

(c) HMODE, d vs. cosine (with min cluster size = 6)

HMODE-d	HMODE-cosine, 7 groups								
6 groups	1	2	3	4	5	6	7	Outliers	Total
1	3	0	0	1	0	8	0	4	16
2	1	11	15	7	7	8	8	18	75
3	0	0	0	0	12	0	0		12
4	10	0	0	0	0	0	2		12
5	0	6	0	0	0	0	0		6
6	0	5	0	0	0	0	3	2	10
Outliers		6	3	3	4			3	19
Total	14	28	18	11	23	16	13	27	

If this rather sketchy evidence is ignored and partitions are elected from these trees, partitions from several methods might be compared. Table 2U shows the matches between several HMODE and UPGMA partitions based upon distance and cosine. These matches are much poorer than those in Table 1U, indicating that boundaries between the "groups" are quite unstable--an excellent clue that there may be no true boundaries at all.

This suggests that partition matching can be a useful tool in evaluating partitions. While agreement between two methods is not sufficient evidence that "natural" clusters have been found (as in Table 1U), the inability of two similar methods to find similar partitions may be good evidence that natural boundaries do not exist. This would only hold, however, when the clustering criterion and similarity measure used in the two methods are comparable. (Distance and cosine cannot be expected to agree on a cluster centered at the origin, and the correlation coefficient is even more likely to produce different results.)

NMODE: This method produced the strongest evidence that the data is uniformly distributed. It was unable to find any dense points, therefore it refused to search for clusters.

Comparison Of The Clustering Methods

Both NMODE and WARD provide clear indications that this data contains only 1 group. In contrast, the results of MIKCA, UPGMA, and HMODE, regardless of which similarity coefficient is used, are apt to be extremely misleading, encouraging an arbitrary partitioning of data in which no natural boundaries are present.

We recommend the use of NMODE for preliminary screening of a data set, because it so clearly establishes the lack of any dense regions in a uniform

distribution. Unfortunately, though, it defines dense points and modes very strictly, tending to underestimate the number of populations in a data set (e.g. sets D and F). Thus the search for modes should probably be done by a more sensitive method.

The use of partition matching as a cross-check on the partitions produced by several clustering methods was discussed above. For any cluster overlap matrix, the partition matching coefficients can be computed and may provide a useful measure of how well two C.A. partitions agree with each other. Table 3U gives the Rand, Jaccard, Cramer, and Lambda measures of partition agreement for the cluster overlap matrices shown in Tables 1U and 2U. For purpose of comparison, it also shows these measures of agreement for some analogous partitions in data sets E and F. Even though both sets E and F are quite noisy, and set F in particular has fuzzy boundaries between populations, still their measures of partition agreement do not drop nearly as low as those for set U. Unfortunately, sampling distributions for the agreement coefficients are unknown, so we cannot tell whether the differences we see here are statistically significant. They can only be treated as one clue to whether there are natural boundaries in a data set.

Evaluation Of Individual Clusters

Measures of the compactness, separation, and homogeneity of groups for three different partitions of set U are shown in Tables 4U-6U. Two of the partitions are based upon Euclidean distance, the third is based on cosine as the similarity measure. In all three partitions, the cluster evaluation statistics show very poorly-defined "clusters". The gamma measures are all below -.95, group diameters are large, while separation between groups is relatively small. Furthermore, the variables show low homogeneity within

Table 3U

Measures of Partition Agreement for Pairs of Partitions Produced by Cluster Analysis on the Same Data

Partition 1	Partition 2	# of groups p. 1	# of groups p. 2	# of outliers p. 1	# of outliers p. 2	Rand's coef.	Jacc. coef.	Cramer's coef.	Symmetric Lambda
Data Set U									
MIKCA-D	MIKCA-d	6	6	0	0	.933	.688	.880	.867
MIKCA-D	MIKCA-D	6	4	0	0	.748	.297	.598	.430
MIKCA-D	UPGMA-d	6	6	0	0	.791	.330	.603	.532
HMODE-d	UPGMA-d	7	7	14	6	.681	.213	.492	.284
UPGMA-cos	UPGMA-d	7	7	13	6	.782	.238	.613	.451
HMODE-d	HMODE-cos	6	7	19	27	.687	.150	.524	.259
Data Set E									
MIKCA-D	UPGMA-d	3	4	0	0	.911	.826	.893	.800
HMODE-d	UPGMA-d	3	4	24	0	.885	.770	.806	.765
UPGMA-cos	UPGMA-d	3	4	0	0	.761	.566	.774	.576
HMODE-d	HMODE-cos	3	3	24	11	.783	.545	.680	.530
Data Set F									
MIKCA-D	UPGMA-d	3	3	0	20	.884	.808	.867	.637
MIKCA-D	UPGMA-d	4	5	0	33	.875	.717	.796	.564
HMODE-d	UPGMA-d	5	5	40	33	.886	.743	.741	.732

(cosine and d partitions not comparable because of cluster centered at origin)

Table 4U
UPGMA-d, 6 groups, no outliers
Partition of Uniformly Distributed Data

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	8	8600	34	2533

GAMMA -.9925

GAMMAB -.9981

GAMMAW -.9735

MAX. CLUSTER DIAMETER 6.5719
MIN. DIST. BTWN BOUNDARIES 1.5892

CLUSTER DIAMETERS

1	2	3	4	5	6
5.273	6.572	4.397	5.142	5.681	5.410

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5	6
1	-.701	.034	.521	.261	.278	.179
2	.499	.141	.814	.254	.697	.434
3	.601	.337	-.349	.268	.260	.694
4	.686	.179	.720	-.147	.120	-.132
5	-.167	.356	-.168	.516	.102	.588
6	.416	.014	.432	.145	.244	.122
7	.577	.386	.605	.643	.258	.644
8	.364	.213	.789	.276	.135	-.182

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.127	.366	.329	.141	.307	.145	.469	.204

CLUSTER HOMOGENITIES

	1	2	3	4	5	6
	.309	.208	.420	.277	.262	.29
Size	9	56	9	29	28	19

Table 5U
HMODE-d, 7 groups, 14 outliers
Partition of Uniformly Distributed Data

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	1	6082	40	3057
GAMMA	-.9911			
GAMMAB	-.9997			
GAMMAW	-.9742			

MAX. CLUSTER DIAMETER 6.9006
MIN. DIST. BTWN BOUNDARIES 1.6662

CLUSTER DIAMETERS

1	2	3	4	5	6	7
5.691	3.005	6.901	4.299	5.952	4.120	4.536

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5	6	7
1	-.206	.865	.182	.256	.818	.920	-.058
2	.561	.539	-.089	.568	.034	.575	.747
3	.658	.849	.111	.099	.880	.440	.713
4	.685	.858	.112	.773	.525	-.440	-.104
5	-.072	.841	.160	.812	-.016	.692	.403
6	.174	.845	.059	.470	.136	.798	.682
7	.079	.832	-.059	.720	.570	.755	.944
8	-.068	.838	.100	.592	-.142	.031	-.062

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.239	.170	.328	.243	.241	.223	.205	.107

CLUSTER HOMOGENITIES

	1	2	3	4	5	6	7
	.226	.658	.072	.536	.351	.471	.408
Size	16	5	75	12	12	6	10

Table 6U
UPGMA-cosine, 7 groups, no outliers
Partition of Uniformly Distributed Data

--- ADEQUACY OF THE PARTITION---

TESTS FOR COMPACT-SEPARATED CLUSTERS

N1, N2, N3, N4, ...	177	9358	16	1624

GAMMA -.9655

GAMMAB -.9629

GAMMAW -.9805

MAX. CLUSTER DIAMETER 5.6913
MIN. DIST. BTWN BOUNDARIES 1.4397

CLUSTER DIAMETERS

1	2	3	4	5	6	7
5.691	5.370	5.523	5.339	5.056	4.887	4.142

VARIABLE X CLUSTER HOMOGENEITIES.

	1	2	3	4	5	6	7
1	.366	.710	.597	.222	.382	.330	.816
2	.705	.016	.181	.454	.280	.400	.867
3	.191	.596	.146	.219	.839	.129	.342
4	.254	.224	.041	.649	.755	.635	.257
5	.095	.284	.063	-.429	.318	.801	.425
6	-.260	.064	.255	.215	.174	.784	.170
7	.345	.279	.300	.726	-.522	.603	.204
8	.195	.679	.512	.535	.394	.261	.500

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8
.470	.394	.338	.383	.157	.164	.283	.438

CLUSTER HOMOGENITIES

	1	2	3	4	5	6	7
	.236	.356	.262	.324	.328	.493	.448
Size	27	20	29	25	21	15	13

groups. Knowing that these groups are not "natural" modes, but merely segments of a uniform distribution, we can establish some guidelines for interpreting the degree of homogeneity of clusters in other data sets.

In most cases, the groups with the highest overall homogeneities have the smallest number of points: In these three partitions, the only groups with cluster homogeneity above .5 contain 5 and 12 points respectively. All groups containing 20 or more points have cluster homogeneity below .4. Most of the larger groups are fairly homogeneous (over .5) on one or two variables. The partitioning process itself creates a certain amount of homogeneity. Therefore, homogeneity on one or two variables does not necessarily imply that the group is a "natural" mode.

It will be instructive to compare the cluster homogeneities for data set F with those for set U. Set F contains a very compact cluster as well as two spread-out, sparsely-populated clusters. Obtaining cluster homogeneity figures from tables 9F and 10F, we have:

From Table 9F (11% outliers)

Group	Cluster Homogeneity
1	.853
2	.820
3	.814
4	.493
5	.421

From Table 10F (no outliers)

Group	Cluster Homogeneity
1	.741
2	.339
3	.373
4	.145

These measures illustrate that cluster homogeneity measures are relative to the overall homogeneity of the data set. Inclusion of noise points lowers the homogeneity of all clusters. Also, the existence of a large, highly homogeneous cluster will lower the homogeneities of the sparser clusters.

Because the homogeneity measures will vary with the number of groups and the relative sizes of the groups, no cut-off points for "acceptable" homogeneity levels can be established. For descriptive purposes, though, homogeneity can be a useful concept. We would suggest that homogeneity below .4 should be considered relatively low, and that homogeneity above .6 might be considered relatively high. Furthermore, if all groups in a distance-based partition have low homogeneity one may suspect that the data is uniformly distributed.

CONCLUSIONS

The Use Of Partition Comparison

The contingency table, or cluster overlap matrix, showing the match between partitions is extremely useful. It gives a much more complete picture of the match than can be obtained from any single coefficient of agreement.

Frequently some clusters match well and others do not. A single coefficient cannot differentiate that situation from one in which none of the clusters match well. It can be clearly seen in the contingency table, however, and may indicate that some small groups are simply unstable collections of outliers, or may direct attention to partition differences caused by the use of different similarity coefficients in clustering. Thus information from the contingency table can lead to refinement of a typology produced by one clustering method based upon comparison with the results of another clustering method.

Agreement between partitions obtained from different methods should not be given too much weight in determining the correct number of clusters, however. In set D, the greatest agreement between partitions was obtained at the 3-cluster level despite the fact that the data was drawn from 4 populations. Thus agreement between clustering methods is not an adequate indication that all underlying populations have been recognized. Furthermore, in the uniform data, set U, substantial agreement existed between the UPGMA^{-d} and MIKCA 6-group partitions even though the data contains no natural boundaries. Thus agreement between methods might be regarded as a "good sign" but certainly not as convincing evidence of natural boundaries in the data. The absence of

agreement between methods which use similar clustering criteria and, in particular, comparable similarity measures, is perhaps a stronger indication of the absence of natural boundaries in the data. For example, in set U the agreement between UPGMA^{-d} and HMODE-d was very low, while on the other three data sets, agreement between these methods was quite high.

Coefficients of Partition Agreement

For our three structured artificial data sets, four measures of similarity between the cluster analysis partitions and the correct classification were computed: the Rand, Jaccard, Cramer, and lambda coefficients. Ostensibly, each coefficient should tell us which partitions are "most similar" to the correct one. Yet each uses a somewhat different criterion for determining similarity, therefore some disagreement is to be expected. Rankings by these coefficients were then compared with our own in-depth analysis of how well the cluster analysis partitions re-created the correct one.

The Rand, Jaccard, and lambda coefficients all agreed quite well with our rankings of the success of the cluster analysis partitions, even when partitions containing different numbers of classes were compared. (Numbers of classes did not differ by more than 2 in our comparisons, however.) Lambda agreed with our rankings slightly more often than the other two. Rand's coefficient showed a tendency to increase as the number of groups in the partition increased - a possible defect which should be taken into account when using that coefficient to compare partitions of different sizes.

Cramer's coefficient was judged unsatisfactory because it can attain its maximum value of 1.0 even when the two partitions are not identical. It also showed a decided tendency to give a higher ranking to the partition with the smaller number of clusters, whether warranted or not. Even in comparing

partitions of the same size, it deviated from our rankings more than did the other three coefficients.

It should be cautioned that while rankings produced by the Rand, Jaccard, and lambda coefficients can be compared, their actual values cannot. Usually, for a given pair of partitions the Rand value is higher than the Jaccard and lambda values. It is quite possible for lambda and Jaccard coefficients to take on a value of zero, thus they have a true range of 0 to 1. But it is virtually impossible for Rand's coefficient to drop to zero since that would require a = d = 0, i.e., no pairs of points placed together in both partitions and no pairs of points placed apart in both partitions. In our tests so far, the minimum value attained by Rand's coefficient is .318 and seldom has it dropped below .6 even when a match is poor. Thus its effective range is small.

We cautiously advance some very tentative guidelines for judging the level of agreement between two partitions on the basis of agreement coefficients. It should be recognized that these are derived from a study of artificial data sets that contain relatively small numbers of variables and 3 to 5 clusters plus a set of uniformly distributed data. Whether the guidelines are transferable to real data remains to be seen. Their tentative nature will be emphasized by the fact that it is quite possible, for example, to have "excellent" agreement according to Rand and only "moderate" agreement according to Jaccard and lambda.

Tentative Guidelines for Judging Partition Agreement
Based on Artificial Data Sets Containing 1-5 Clusters

Agreement Level	Rand's Coefficient	Jaccard's Coefficient	Symmetric Lambda
Excellent	Above .9	Above .75	Above .75
Moderate	.8 - .9	.6 - .75	.6 - .75
Questionable	.7 - .8	.5 - .6	.5 - .6
Poor	Below .7	Below .5	Below .5

Two factors should be taken into account in using these guidelines. First, Rand's coefficient tends to increase as the number of groups in the partitions increases, therefore the Rand guidelines should be raised for partitions containing more than 5 groups. Second, the existence of a large number of outliers may affect coefficient values, raising them if the two partitions have similar outlier groups, lowering them if one partition has a large outlier category while the other assigns most of the outliers to clusters.

The cluster overlap matrix (crosstabulation) for the two partitions will still be an important resource, revealing outlier effects and other systematic differences between the partitions which can be explained and perhaps lead to adjustments in one or both partitions.

Cluster Evaluation Coefficients

For comparing partitions to select the one that best reflects the structure of a data set, neither the gamma coefficients nor the homogeneity measures are appropriate. Higher scores on compactness will be achieved by partitions that break up a large sparse cluster into pieces; higher scores on separation will be received by a partition that merges two closely-spaced modes into one group (as in data set D).

For descriptive purposes, on the other hand, these measures can be useful. The gammas measure overall compactness and separation of groups in a partition. They are not entirely satisfactory because they rely on maximum cluster diameter and minimum cluster separation, possibly being overly influenced by one cluster of large diameter or one closely-spaced pair of clusters. These measures were designed with an ideal of compact, well-separated clusters in mind--an ideal that is seldom even approached by the criminological data within our experience.

In particular, when separation between clusters is very small, gamma within groups (GAMMA W) approaches -1, yet it has little meaning because the low score is caused by a lack of separation between two of the clusters rather than a lack of compactness in any of the clusters. Thus we doubt that the gamma measures will be appropriate for most criminological data.

The homogeneity measures give a much more complete picture of cluster compactness, including the homogeneity of each cluster on each variable as well as overall homogeneity for each cluster and each variable. If compact clusters are required, the homogeneity measures could potentially be used to establish minimum acceptable homogeneity levels. Several factors should be taken into account in interpreting the homogeneities, however.

1. In any partition, homogeneity will tend to increase as the number of groups increases. Thus homogeneities should usually only be compared for partitions of the same size.
2. If outliers are allowed; homogeneity will vary depending on the proportion of the data set that is assigned to clusters, with homogeneity tending to decrease as coverage increases.
3. The partitioning process creates a certain amount of homogeneity. Even on uniformly distributed data most groups are homogeneous on one or two variables, therefore, such evidence is not adequate proof of the existence of a natural mode.
4. The overall measures of "cluster homogeneity" are computed relative to the homogeneity of the data set as a whole. Thus the existence of a heavily populated and highly homogeneous cluster will lower the homogeneity measure of a sparser cluster.
5. These homogeneity measures are intended for clusters defined in a points-in-space model of similarity. They are not appropriate for evaluating clusters based on profile-shape. Measures of compactness and separation for such shape-clusters need to be developed possibly using correlations between cluster centroids. This will be important in the classification of psychiatric data since similarity of profile shape as measured by the correlation coefficient is frequently used in classification based on such common psychiatric tests as the 16PF and MMPI.

Given all the above qualifications, we suggest very loose guidelines for the use of the homogeneity measures for descriptive purposes: Homogeneity below .4 can be considered relatively low, and homogeneity above .6 can be

considered relatively high. Furthermore, if all groups in a partition have low homogeneity (within a points-in-space model) then there is some reason to suspect that the data is uniformly distributed.

Evaluation Of The Clustering Methods

From the detailed discussion above of the performance of each clustering method on each artificial data set, some key characteristics of the clustering methods can be seen.

K-means (MIKCA)

1. Number of clusters K must be known in avance. MIKCA cannot adjust it, and gives no clue if the wrong number is selected.
2. When the correct number of clusters K is provided, clustering results are generally good. The determinant W criterion performed better than the trace W criterion on set D. Contrary to predictions both determinant W and trace W produced excellent results on ellipsoidal clusters at different orientations in space (set E).
3. MIKCA may converge on a local minimum rather than finding the best solution. If estimates of cluster centroids are available from some other source it is worthwhile to provide them as a start for the clusters, then compare the results with those produced by the random start procedure.
4. The choice between Mahalanobis D and Euclidean d for use with the determinant W criteria is not clear cut. In set E, Euclidean d did better, probably because of the different orientations of the ellipses. In other sets they did about equally well. McRae (1973) argues that Mahalanobis D is more appropriate theoretically with the det W criterion. The advantages of Mahalanobis D would be more likely to show up in a data set which had some highly correlated variables—not true of any of the sets tested here.

5. No provision for outliers. All points are assigned to clusters.

Minimum Variance (WARD)

1. Results on the three structured data sets were poor. This method is apparently unable to distinguish two closely-spaced clusters, particularly if thre is considerable difference in their densities (described under data set E). This makes it inappropriate for many criminological applications.
2. If research purposes require clearly separated clusters, the Ward method might be appropriate. Other researchers have found it useful for other applications. Furthermore, when there are no empty spaces between groups it produces a tree diagram with a distinctive stairstep pattern that clearly implies only one group.

Hierarchical Group Average (UPGMA)

1. Clustering results using Euclidean d and Gower's coefficient are generally good. Results using cosine as the similarity measure are also good, except on a cluster which contains the origin (the grand centroid in standardized data).
2. Although the precise number of clusters cannot always be determined from the tree, a good approximation can be obtained. The UPGMA tree gives a very complete picture of the structure of a data set. To use it effectively requires considerable experience at interpreting tree diagrams, however. Relatively long branch lines indicate separate clusters. Alternatively, evenly developing groups with short branch lines at the last stages of the tree indicate uniformly distributed data.
3. When clusters differ greatly in density, in order to find them all it may be necessary to select branch lines from different levels of the tree. (See discussion on data set F.)

4. The most distant outliers can be easily recognized as the singletons and very small groups remaining at the last stages of the tree. This tends to result in a fairly small number of outliers in our experience, however. In set E, most of the added noise points were assigned to clusters by UPGMA.

Wishart's Hierarchical Mode Search (HMODE)

1. Clustering results using d, Gower's coefficient and cosine are generally good, except when cosine is used on a cluster centered at the origin as mentioned above. HMODE and UPGMA usually produce very similar results when using the same similarity coefficient on a structured data set. Our observations indicate that HMODE tends to misclassify slightly more points than UPGMA and MIKCA in the fuzzy boundary region between closely-spaced modes, unless the outlier-threshhold allows such points to be left as outliers.
2. It is often difficult to determine the correct number of clusters from HMODE trees. The number of secondary clusters provides a maximum for the number of modes, but this tends to be much greater than the actual number of clusters. Evidence that a data set has no cluster structure is unclear because the tree looks similar to some obtained from structured data.
3. Two very closely-spaced modes may be placed in the same secondary cluster and thus be impossible to detect separately. (However this only happened with modes so close that most other methods could not separate them either.)
4. HMODE has a good, flexible provision for outliers, allowing the user to specify the minimum number of points that must be classified and allowing the program to continue assigning points beyond that time if the

hierarchical merging of secondary clusters is not complete. Very small secondary clusters that form later in the tree can also be designated as outliers. In the data sets containing added noise, most noise points were left as outliers. Care should be taken not to set the threshhold so as to allow too many outliers, since a large part of the sparsest cluster may then be left among the outliers. A requirement that 90% of points be classified before point assignments are allowed to stop seems safer than the 80% level used for sets E and F.

Huizinga's Mode Search (NMODE)

1. Clustering results are poor for closely-spaced clusters, but reasonably good for separated clusters. In effect, NMODE defines modes more strictly than HMODE, requiring more separation between them and allowing less difference in compactness. NMODE would not be appropriate for finding overlapping modal clusters.
2. NMODE tends to underestimate the number of clusters in a data set in two ways: two closely spaced clusters are often identified as one, and a very sparse cluster may go unrecognized and be left in the outlier category.
3. Often too many points are left as outliers, since only dense-points are assigned to clusters. Points in the sparse outer region of a cluster may not be assigned to it.
4. NMODE can be recommended for preliminary screening of a data set, because it appears to provide a lower bound for the number of clusters. It also provides valuable information about the lack of structure in a data set, when no dense points can be found. However, it is not accurate enough on closely-spaced clusters to provide a final clustering solution.

Importance Of The Similarity Coefficient

Four similarity coefficients, Euclidean d, Gower's coefficient, cosine, and correlation, were tried with both the UPGMA and the HMODE clustering programs. Thus joint comparison of similarity measures and clustering methods is possible.

For each structured data set, but not for the uniform one, clustering results using d and Gower's coefficient are very similar. This result was predicted by our study of similarity coefficients in Chapter 7, and helps to solidify those findings.

The cosine (or normalized vector product) measures similarity in angular position. It produced partitions very similar to those of d and Gower, except for one cluster in each of data sets E and F which contains the grand centroid of the data set within its boundaries. However, the "growth pattern" of the clusters to be seen in the cosine trees is clearly different from that in the d and Gower trees. This can be seen by comparing cosine and d partitions produced at several tree levels.

Correlation, which measures similarity in profile shape, differs much more from the other coefficients, particularly in sets E and F. This should not be surprising since those data sets contain the type of clusters that correlation cannot identify in one piece: random deviations from a nearly flat profile. In all of the data sets, the growth pattern of the correlation tree is quite different from that of any of the other trees.

Yet in each data set, including the uniform one, agreement between partitions produced by UPGMA and HMODE when both use correlation is relatively high. The same is true when both use cosine. And it is true for all data sets except the uniform one when both clustering methods use d or Gower.

Furthermore, comparison of cluster "growth patterns" usually shows similar cluster development within UPGMA and HMODE when the similarity measure is the same for both.

Thus when two similarity coefficients genuinely measure different aspects of similarity they can be expected to produce considerably different partitions on some data sets. The differences attributable to the similarity measures in such data will tend to overshadow any differences between the UPGMA and HMODE algorithms.

Determining The Number Of Clusters

No truly satisfactory solution has been found to the problem of determining the number of clusters in a set of data. Of the clustering methods considered here, UPGMA provides the most information about the structure of the data set, and its tree is probably the best guide in the choice of partition size. Several likely tree levels may be selected, to be evaluated in terms of the research purposes. Also, branch lines at different levels of the tree may be used, possibly representing clusters of different density.

The two mode-search methods appear to provide a bracket for the correct number of clusters, with NMODE giving a minimum and HMODE giving a maximum.

One cannot expect to distinguish separate modes beyond a certain level of population overlap. Set F represents an initial attempt to test the minimum level of cluster separation needed in a very simply structured data set. For many research purposes it may be quite satisfactory to treat two closely-spaced modes as one cluster.

Outliers

In most cases it would be preferable to set a subject aside as an

"outlier" rather than misclassify it. The outliers, then, would have to be studied as individuals. Thus identifying too few outliers may lead to misassignments, while identifying too many outliers may lead to an unduly large number of individual analyses if a sparse cluster is overlooked.

The UPGMA and HMODE methods allow considerable user control of the number of outliers. NMODE tends to leave too many subjects in the outlier category, while MIKCA makes no provision for outliers at all but assigns all points to clusters. In the "noisy" data encountered in criminology it will generally be advisable to remove a few of the most isolated points before using MIKCA, since their inclusion can distort the optimization process that MIKCA uses in clustering.

Recommendations

As a result of this study, we recommend the UPGMA clustering method. It provides the user with the most information about the structure of the data set, once some experience in interpreting the tree diagrams has been gained. Methods like NMODE and HMODE which were designed to tell the user the correct number of clusters do not, in practice, do so very reliably. Furthermore, they provide less information to aid the user in making that decision.

For a points-in-space model of similarity, Euclidean d or Mahalanobis D will generally be the most appropriate similarity coefficient because all features of the data are considered, whereas cosine and correlation systematically ignore some information. Distance and cosine partitions can sometimes be fruitfully compared when the angular position of clusters is of interest.

It is advisable to try two different clustering algorithms on the same data, using the same similarity coefficient, as a tentative check on the

stability of the groups that are found. UPGMA and HMODE are a good combination because both allow for a variety of similarity coefficients and because on uniformly distributed data HMODE tended to produce less stable partitions than the other methods--a clue to lack of structure in the data.

Unfortunately, our UPGMA and HMODE programs are limited to a maximum of 200 subjects. For larger data sets, UPGMA clustering might be done on a randomly sampled subset to find the approximate number of clusters and even estimates of the cluster centroids. Then clustering of the entire data set can be done using MIKCA, which produces good results when the number of clusters is known.

BIBLIOGRAPHY

Anderberg, M.R.
1973 Cluster analysis for applications. Academic Press: New York.

Baer, D.J.
1970 "Taxonomic classification of male delinquents from auto-biographical data and subsequent recidivism." Journal of Psychology, 76(1), 27-31.

Ball, G.H.
1970 Classification analysis. Technical Report. Stanford Research Institute: Menlo Park, California.

Bartko, J.J., J.S. Strauss, and W.T.J. Carpenter
1971 "An evaluation of taxometric techniques for psychiatric data." Classification Society Bulletin, 2(3), 1-29.

Bezdek, J.
1973 "Numerical taxonomy with fuzzy sets." Journal of Mathematical Biology, 1(1), 57-71.

Blashfield, R.K.
1976 "Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods." Psychological Bulletin, 83, 377-88.

Cunningham, K.M., and J.C. Ogilvie
1972 "Evaluation of hierarchical grouping techniques - a preliminary study." Computer Journal, 15(3), 209.

Delabre M., A. Bianchi, and M. Veron
1973 "Classification of aquatic bacteria." Annals of Microbiology, 124(4), 489-506.

Edelbrock, C.
1978 "Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody." National Institute of Mental Health, American Classification Society Conference.

Edelbrock, C.
1979 "Comparing the accuracy of hierarchical clustering algorithms: The problem of classifying everybody." Multivariate Behavioral Research, 14, 367-84.

Everitt, B.
1974 Cluster analysis. Halsted Press: New York.

Final Report

MULTIVARIATE TAXONOMIC CLASSIFICATION FOR CRIMINAL JUSTICE RESEARCH

VOLUME 3

OCTOBER, 1980

79121

Prepared For:
National Institute of Justice
Washington, D.C.

NCJRS
JUN 26 1981
ACQUISITIONS

FINAL REPORT

MULTIVARIATE TAXONOMIC CLASSIFICATION FOR CRIMINAL JUSTICE RESEARCH

(Project No. 78-NI-AX-0065)

Volume 3: APPLICATIONS OF NUMERICAL CLASSIFICATION METHODS IN CRIMINOLOGY AND GENERAL CONCLUSIONS

Principal Investigator: Tim Brennan Ph.D.*

Project Staff: Mary Downton
Randy Blair
Dick Williams Ph.D.

Behavioral Research Institute
Boulder, Colorado

October, 1980

Prepared for:

National Institute of Justice
Office of Research and Evaluation Methods
Department of Justice
Washington, D.C. 20531

*Now at Human Systems Institute, Inc., 1906 13th St., Suite 304, Boulder, Colorado, 80302.

ACKNOWLEDGEMENTS

This project has benefitted from the input of many people. We thank particularly George Silberman of the National Institute of Justice (OREM) for his continued support, encouragement, and understanding of this project. Richard L. Linster (OREM), has also been extremely supportive of our work. We are indebted to Judge Horace B. Holmes and the staff members of the Boulder County Juvenile Probation department for allowing us access to selected data. Similarly, we acknowledge the support of Dr. Delbert S. Elliott in allowing us access to sub-samples of his larger national study of juvenile delinquency. In our analysis of prisoner classification, the staff at the Reception and Diagnostic Center at the Canon City Correctional facility was extremely helpful and kind. We thank particularly Lou Hesse, and Dr. Gideon Weiss, as well as Dr. Tom Crago. These persons were consistently supportive in our studies of prisoner classification. Dr. Herb Eber also provided invaluable help in regard to our analysis and understanding of the prisoner data.

Finally, I would like to acknowledge my appreciation to the project staff: Mary Downton, Dick Williams and Randy Blair. These people were a delight to work with and I am indebted to them and their intelligence, motivation and effort.

- T.B.

TABLE OF CONTENTS

Volume 1: AN EVALUATIVE OVERVIEW OF CLASSIFICATION IN CRIMINAL JUSTICE

Page

Acknowledgements

Chapter 1. Evaluative Criteria for Criminological Classification 1

- Introduction 2
- Structural Aspects of Classification Systems 4
- The Attribute (Measurement) Space of Classifications 5
- The Vertical Dimension of Classification 7
- The Horizontal Dimension of Classification 8
- Pragmatic Criteria for Classification Systems 15
- Conclusion 26
- Bibliography 27

Chapter 2. The Uses of Classification Within Criminology and Criminal Justice Systems

- Introduction 31
- Uses of Classification in the Field of Research Criminology 32
- Classification for Treatment 38
- Classification in the Criminal Justice System 40
- Conclusion 43
- Bibliography 44

Chapter 3. The Failure of Classification in Criminology: Some Results of Inadequate Classification 50

- Introduction 51
- The Relation Between Inadequate Classification Research and the Failure of Classification in Practical Contexts 52
- Ethical Issues in Applied Criminological Classification 61
- Conclusion 69
- Bibliography 70

Chapter 4. A Brief Examination of Decision Making and Classification in the Criminal Justice System

- Introduction 74
- Conclusions 91
- Bibliography 94

Chapter 5. Bibliographic Materials on the Use of Classification in Criminological Research and Its Use Within Criminal Justice Systems 96

- Introduction 97
- Bibliographic Materials 98

Volume 2: EVALUATING STATISTICAL CLASSIFICATION METHODS (MONTE CARLO STUDIES)

Chapter 6. Concepts of Similarity as Used in Criminological Studies 112

- Some General Problems Regarding Similarity as Used in Prior Criminological Research 113
- Explanatory Theory as a Guide to the Specification of Similarity 122
- Similarity in Criminological Studies Using Quantitative Classification Techniques 125
- Conclusions 128
- Bibliography 130

Chapter 7. Some Theoretical Observations on Distance Coefficients Versus Correlation Coefficients for Classification Studies 135

- Introduction 136
- Mahalanobis D^2: Orthogonality or Implicit Weighting of Attributes 137
- Euclidean Distance vs. Correlation: Some Strong Opinions 141
- Resolving the Paradox: A Brief Look at the Components of Similarity 142
- Some Comments on Correlation 146
- Distance Versus Correlation Coefficient: The 'Points-in-Space' Model Versus Profile Shapes 152
- Other Ways of Measuring "Shape" 152
- Bibliography 154

Chapter 8. Empirical Comparisons of Several Similarity Coefficients 156

- Purpose 157
- Method 157
- Conclusions 169
- Bibliography 178

Chapter 9. Evaluating Classifications: A Comparison of Five Coefficients for Partition Agreement 180

- Method 185
- Results 191
- Conclusions 206
- Bibliography 208

Chapter 10. Evaluation of Several Cluster Analysis Methods Using Artifical Data Sets 209

Introduction 210
Evaluation of Approaches 223
Results - Data Set D: Clusters of Different Shape and Separation 226
Data Set E: Differently Oriented Ellipses, With Minor Cluster Overlap 256
Data Set F: Overlapping Clusters of Different Compactness 291
Data Set U: A Uniform Distribution 331
Conclusions 353
Bibliography 367

Volume 3: APPLICATIONS OF NUMERICAL CLASSIFICATION METHODS IN CRIMINOLOGY AND GENERAL CONCLUSIONS

Chapter 11. Classification Analysis for the Epidemiologic Description of Crime Patterns 369

Introduction 370
Classification Approaches in the Epidemiology of Delinquency 377
A Behavioral Classification of Juvenile Violence 382
A Behavioral Classification of Juvenile Theft 400
A Behavioral Classification of Juvenile Drug Consumption Patterns 416
Discussion 426
Bibliography 430

Chapter 12. The I-Level Classification: An Evaluation by Taxometric Methods 432

Introduction 433
Methods 438
Data 439
Results 444
The Overlap Between Natural Clusters 458
Description of the Core Clusters 474
Cluster 1 474
Cluster 2 477
Cluster 3 478
Cluster 4 479
Cluster 5 480
Cluster 6 482
Conclusions 497
Bibliography 501

Chapter 13. A Classification of Probation Youth 503

Research Sequence 513
Results 515
Evaluation of ES Clustering Variables 517
Clustering Results for the HSPQ 523
Discussion 530
Bibliography 532

Chapter 14. Classification by Personality of a Prison Sample 533

Methods of Analysis and Data 537
Results 545
Prisoner Personality Types Based on the CAQ 569
Conclusions 595
Bibliography 598

Chapter 15. Making Predictions from Empirical Classifications Using Class Membership as a Dummy Variable 599

Method 601
Criteria 602
Predicting Violence 603
Validating the Predictions 609
Conclusions 611
Bibliography 614

Chapter 16. Concluding Remarks and Recommendations 615

The Diffusion and Utilization of Numerical Classification into the Criminal Justice System 616
Bibliography 625

CHAPTER 11

CLASSIFICATION ANALYSIS FOR THE EPIDEMIOLOGIC DESCRIPTION OF CRIME PATTERNS

INTRODUCTION

Epidemiology requires an accurate mapping or description of criminal behavior. Univariate description has been criticized for its fragmentation of patterns into separate atomized bits of criminal behavior. Most studies of incidence and epidemiology of crime have utilized simple frequency counts within this atomistic univariate approach. This practice represents a serious distortion in the description and enumeration of criminal and delinquent behavior since it is known that offenders generally do not constrain their antisocial behavior to one type of offense (Hood & Sparks, 1970, and others). The same offender may commit a number of different offenses.

Multi-dimensional approaches may help to avoid the above type of distortion. If patterns of offender behavior exist, then they ought to be discovered and described accurately by the epidemiological analysis. Univariate, and even most of the more complex cross-classification approaches are generally inadequate for the discovery and description of offense behavior. Therefore, we propose to examine the utilization of taxometric methods in the discovery, description, and enumeration of patterns of criminal and delinquent behavior. These three particular subgoals (discovery, description, and enumeration) are clearly the main tasks of general epidemiology. We do NOT mean to imply that univariate frequency approaches should be omitted or replaced by this multi-dimensional approach. Rather, we suggest that if the multi-dimensional taxonomic approach is used to augment the univariate approach, then a much clearer, more complete, and less misleading epidemiology of criminal behavior will be available.

In our present report we examine the epidemiology of juvenile crime. This task involves the use of self-report data assessed on a national sample of

delinquents. The self-report instruments focused on delinquent behavior. Consequently, we provide an evaluation of reliability and validity and potential advantages and weaknesses of self-report data for criminal behavior epidemiology.

Brief Considerations in the Use of Self-Report Data

As Bellack and Hersen (1977) cogently point out, use of self-reported data as the basis of criminological research has undergone numerous fluctuations of popularity. Critics are quick to decry its supposedly devastating lack of objectivity (i.e., Nettler, 1974), while its adherents claim it covers a range of data not accessible to external observation and not revealed by official records.

Despite such evidence of controversy, within the realm of research into criminal behavior it is generally admitted that the use of self-report data is necessary to provide a more complete epidemiology of crime (Hindelang, Hirschi, and Weiss, 1975; Swanson and Mobley, 1976). Official statistics gathered from police and court records generally exhibit biases which result from selective procedures of arrest, dubious record-keeping, purposes inherent in the criminal justice system, etc. (Swanson & Mobley, 1976; Empey, 1978; Ageton & Elliott, 1978; Gould, 1968; Erickson, 1972). Police, court, and prison records are not generated to provide a complete description of criminal behavior. If the methodology underlying the compilation of official statistics is in question, however, the collection and use of self-reported criminal data can be just as ambiguous. For one thing, as Nettler (1974) says, it is always "ticklish to ask people to recall their 'bad' behavior." Nonetheless, problematic though it may be, we assert that careful use of self-report data can lead to a remarkably improved picture of the description and enumeration of criminal behavior.

Validity of Self-Report Behavioral Data

One of the initial considerations which must be taken into account in the construction or use of any behavioral measurements for epidemiology is the question of validity. In this case, the relevant concern is to determine to what extent the use of self-reported behavior is valid and if there is an improvement in validity over official statistics.

The validity of self-report data (e.g., self-report as an accurate reflection of criminal behavior) has been addressed in a number of studies, but cannot be unquestionably established. As Ageton and Elliott (1978) point out, most studies concerning self-report validity have been inconclusive. Several types of validation have been suggested in this connection (Ageton & Elliott, 1978; Nettler, 1974):

1. Comparison of self-report data with external measures of behavior, such as official statistics. Studies along these lines have been less than convincing. Erickson (1972) and Gold and Williams (1969), among others, report high response validity, meaning, in these cases, that self-report data and official statistics more or less agree. Others, however (Empey, 1978), point out the divergence of self-report and official statistics. All in all, this method of validation is not particularly useful. For one thing, use of one scale of dubious validity to check the validity of another questionable scale is problematic (Nettler, 1974). Nettler (1974) points out that a criterion external to both measures is necessary. Several writers (Erickson, 1972; Gould, 1968) stress that an explicit theory of the relation between self-report measures and external statistics is needed before comparisons can be made. As Gould (1968) and Nettler (1974) write, self-report scales and official reports do not necessarily measure the same thing; they do not cover a "common base of behaviors."

2. Comparison of self-report data with offenses reported by other informants. This provides a much better method of dealing with the problem of concealment. Most studies (Gold, 1966; Goodstein & Russel, 1977) report a relatively high correlation level, though Jessor et al. (1968) report only moderate correlations. Voss (1963) and Gibson (1970) reported extremely high correlations between self-reports and the reports of informants; however, their samples only included "official" delinquents who most likely had little to conceal (Donovan, 1972).

3. The use of "known-group" differences; that is, the expectation that a known delinquent group would be apt to score higher on a self-report delinquency scale. This has been a more successful index of validity. Short and Nye (1957), Erickson and Empey (1963), and Gold (1966) all report that adolescents apprehended by the police score higher on self-report scles than others.

4. Ageton and Elliott (1978) report the use of "lie scales", bogus items, and social desirability scales to test self-report validity. Petzel, Johnson, and McKnip (1973), for instance, used a bogus drug item on a scale and found that under four percent reported using the drug. Clark and Tifft (1966) found that subjects extensively changed their answers on a self-report scale when retested with a polygraph. They found, in fact, that most items were under-reported. They concluded that self-report data is generally accurate over a wide range, but that there is differential validity on specific items. Bellack and Hersen (1977), however, stress that answers to individual questions might be valid while "summative scores and grouped data" are not.

It is evident that, as mentioned above, tests relating to the validity of self-report scales are not totally conclusive. Ageton and Elliott (1978)

write, "There is simply no truly satisfactory independent criterion measure against which SRD (self-report delinquency) can be validated." However, the tests referred to above tend more toward supporting the validity of SRD than not. Bellack and Hersen (1977) suggest that any variance between SR and actual behavior may result from misused assessment procedures, not from the lack of validity in self-report data. This suggests the need for more stringent criteria of validity.

Reliability

Another major question area regarding self-report data is reliability, although this is considerably less problematic than the issues surrounding validity. Reliability, to paraphrase Nettler (1974), entails judging the consistency of the measuring tool utilized. This has basically been posited in terms of internal consistency and retest consistency. Ageton and Elliott report that studies based on an internal consistency approach (Elliott et al., 1976; Elliott & Knowles, 1978) show high levels of reliability.

Studies utilizing test-retest criteria for reliability are less conclusive, but still reveal a fairly high level of reliability (Clark and Tifft, 1966; Kulik, Stein, & Sarbin, 1968). Clark and Tifft (1966) found higher reliability in a retest situation, after subjects were threatened with the use of a lie detector. Nettler (1974) points out that scaling techniques promote consistency in that they weed out questions which "don't hang together."

Problems Affecting Validity and Reliability

The issues of validity and reliability in self-report data raise a number of difficulties, the majority of which are often commented upon but seldom actually remedied:

1. The problem of lying: Each respondent, for a multitude of related reasons, may choose to conceal or exaggerate certain information. Gold (1970) points out that there seems to be little difference in levels of concealment among races or social classes, but there are some differences between the sexes.

2. The problem of context, or relativity of response: Each respondent will answer according to his or her interpretation of the question and the context in which it is given. This includes attention to vocal intonation (in an interview) or the relative positions of questions (in a questionnaire) (Nettler, 1974), as well as questions which may be culture or class-bound (Hindelang, Hirschi, and Weiss). This category also entails differential kinds of administration of the self-report scale (interview, questionnaire, individual or group, etc.).

3. The problem of inaccurate memory, particularly with questions that cover a long range of time (Ageton & Elliott, 1978; Nettler, 1974).

4. The problem of anonymity: Will the subjects respond better to an anonymous measure than to one in which they are identified. Hindelang et al. considers this an important item, while the study of Kulik, Stein, and Sarbin (1968) suggests that it actually makes little difference.

5. The problem which arises when normative response categories ("fairly often," "mostly," etc.) are used. These terms are subject to a wide range of differential interpretations (Hindelang, Hirschi, & Weiss; Ageton & Elliott, 1978).

6. The problem regarding representativeness of the items on a response scale; that is, the extent to which these items accurately cover the desired realm of (delinquent) behavior (Empey, 1978; Hindelang, Hirschi, & Weiss).

Gould (1968), for instance, repeats the criticism that self-report scales tend to concentrate on merely trivial offenses.

7. The problem of including more than one item on a response scale which would cover the same event, and the related problem of "inclusive items" which subsume a number of possibly different events under a single heading (Hindelang, Hirschi, & Weiss; Ageton & Elliott, 1978).

Potential Advantages of Self-Report

Though there are a number of difficulties involved in using self-report data, as partially enumerated above, the advantages of this approach tend to be decisive, particularly when compared with the use of official crime statistics.

1. Coverage. Most writers agree that self-report has the potential of covering a much wider range of the "deviant domain" than is provided by other forms of data. Hindelang, Hirschi, and Weiss stress that self-report data can provide a superior estimate of the incidence and pattern of crime. Swanson and Mobley (1976) claim that SR provides more complete data and does not over-simplify the heterogeneous nature of delinquency as official statistics often do.

2. Effects of Labeling. The use of self-report can eradicate or avoid negative labeling effects from the data, and also measure deviance as a continuous variable rather than creating the false dichotomy of "delinquent" and "nondelinquent."

3. Self-report as a Reflection of Actual Behavior. Several writers agree that self-reported behavior is a more complete behavioral picture than that provided by official statistics. By using descriptive measure of deviancy, self-report data is closer to looking at behavior in terms of what is actually

done rather than what law is violated. Official statistics, on the other hand, tend to reflect official response to behavior rather than the behavior itself (Empey, 1978). Clark and Haurek write that official statistics are "indices of negative social response to behavior and not necessarily indices of the actual quality and quantity of juvenile behavior . . ." By examining the nexus of actual deviant behavior, self-report can overcome certain biases inherent in official statistics. Empey (1978), for instance, shows that most "criminal" behavior remains unrepresented in the official records: "If all, or even a significant part of all, law violations became a part of the official record, the result would be unprecedented: a large majority of all adolescents would be official delinquents."

4. Contexts of Behavior. Self-report data can also provide a more complete contextual representation of deviant behavior. Bellack and Herson (1977) claim that one of self-report's major advantages resides in its ability to reveal aspects of the "individual's subjective experiences"; that is, the personal context of his or her behavior. Nettler (1974) stresses the extent to which self-reported behavior gives a clearer picture of the "social location of the crime."

CLASSIFICATION APPROACHES IN THE EPIDEMIOLOGY OF DELINQUENCY

Purposes

In the following studies we will examine the classification approach to the clarification and simplification of epidemiological self-report data on juvenile deviant behavior. Again, our primary interest focuses upon the manner in which taxometric methods provide structures for the description and interpretation of data. Since we are dealing with a large national sample of

adolescents, assessed on a comprehensive set of delinquent behaviors, our primary purpose is description. We will examine the degree to which the different taxometric methods provide results which lead to the same basic descriptive structures. This leads naturally into the issue of discovery of "real types."

Description of the Sample

The present study utilizes the data from the second year of a National Longitudinal Youth Survey.[1] The parent study is a five-wave panel design with a national probability sample of 1726 adolescents, aged 11 through 17, in 1976. The overall sample was selected and interviewed initially between January and March, 1977. The second the third surveys were completed between January and March, 1978 and 1979. The fourth and fifth surveys will be conducted between January and March of 1980 and 1981. Whereas the original NIMH funded study involved follow-up on only the odd age cohorts, an LEAA grant allowed for follow-up of the even age cohorts of the original sample. The present report uses the LEAA component of the overall total sample. The data base for the present study consists of those youth who were 13, 15, and 17 during the 1977 calendar year. Taxometric methods are strictly limited in regard to sample size capacity. Therefore, to conform to the requirements of the available methods, we selected a random sample of 200 cases from the overall LEAA data file as our analysis sample for clustering.

---

1 The National Youth Survey is funded by The Center for Studies of Crime and Delinquency, NIMH (MH27552) and the National Institute for Juvenile Justice and Delinquency Prevention, LEAA (78-JN-AX-0003).

CONTINUED

5 OF 9

The youth panel was drawn from a sample of households in the continental United States based on a multistage cluster sampling design (see Huizinga, 1978). The sample contained 2375 eligible youth aged 11 through 17 in 1976. Of these youth, 73% agreed to participate in the study and completed interviews. A comparison of the age, sex, and ethnicity of eligible youth not interviewed with participating youth indicated that the loss rate from any particular age, sex, or ethnic group appeared to be proportional to that group's representation in the population. Furthermore, a comparison with U.S. Census estimates suggests that participating youth are representative of the total 11 through 17 year old youth population in the United States with respect to these characteristics (see Huizinga, 1978).

Instruments and Administration

The national study of delinquency and drug behavior used a very extensive self-reported delinquency instrument (Elliott and Ageton, 1980). This instrument attempted to be comprehensive in its coverage of various kinds of delinquent behavior. In our present analysis we will concentrate on three subdomains of delinquency: theft, violence, and drug-taking. Each specific behavioral item had certain common response categories. For our present taxonomic purposes we use the frequency with which the respondent had committed a particular delinquent act, or consumed a particular drug during the prior year. Two measures of this were available: a straight frequency measure, and a categorical response; i.e., 1) once a month, 2) once every 2-3 weeks, 3) once a week, 4) 2-3 times a week, 5) once a day, and 6) 2-3 times a day. This latter categorical response was used in the following studies of theft and violent behavior. In the drug classification study the categorical response was identical except for two extra categories at the low frequency end of the

response format. These are a) once or twice, and b) once every two or three months.

The interviews were conducted in person with the youth, at the youth's home. Each interview lasted about one and a half hours. A payment was made to the respondent.

The particular items used for our three taxonomic studies are as follows.

Theft Items. The following six offenses make up the attribute space for theft:

1. Stolen (or tried to steal) a motor vehicle, such as a car or motorcycle.
2. Stolen (or tried to steal) something worth more than $50.
3. Stolen (or tried to steal) things worth $5 or less.
4. Taken a vehicle for a ride (drive) without the owner's permission.
5. Stolen (or tried to steal) things worth between $5 and $50.
6. Stolen (or tried to steal) something at school such as someone's coat from a classroom or cafeteria, or a book from the library.

Violence Items. The following ten behaviors were used in the attribute space for violence:

1. Carried a hidden weapon other than a plain pocket knife.
2. Attacked someone with the idea of seriously hurting or killing him/her.
3. Been involved in gang fights.
4. Hit (or threatened to hit) a teacher or other adult at school.
5. Hit (or threatened to hit) one of your parents.
6. Hit (or threatened to hit) other students.
7. Had (or tried to have) sexual relations with someone against their will.
8. Used force (strong-arm methods) to get money or things from other students.

9. Used force (strong-arm methods) to get money or things from a teacher or other adult at school.

10. Used force (strong-arm methods) to get money or things from other people.

Drug Items. In our classification of drug profiles the frequency categories for the following 13 drugs were assessed:

Beer	Amphetamines
Wine	Barbiturates
Hard Liquor	Codeine
Tobacco	Cocaine
Marijuana	Inhalants
Psychedelics	Angel Dust
Tranquilizers	

Taxometric Methods

In examining the above epidemiological data on self-reported delinquency we have made use of two cluster analytic methods which embody two different definitions of a "class." First, the UPGMA method uses a "average-linkage" approach. Second, the HMODE method attempts to find "natural clusters" (see Hartigan, 1975; Everitt, 1974; and others).

Two similarity coefficients, cosine (cos) and Euclidean distance (d) are used to create the similarity matrices which are used as input to the above two cluster methods.

A BEHAVIORAL CLASSIFICATION OF JUVENILE VIOLENCE

Hierarchical Structure of Violence Classifications

The diagrams presented below indicate the taxometric structures that have emerged from the analysis of 120 youth who had indicated at least one violent behavior. In each case we give the hierarchical structure that emerged directly from the cluster analytic method, either UPGMA or HMODE, and secondly, we give the related schematic diagram which indicates the behavioral profiles that are embedded within the taxometric dendrograms. The UPGMA analysis with Euclidean distance gives a particularly interesting tree diagram which suggests the stair-step pattern that is usually associated with a relatively unclustered or unstructured data set. This particular diagram gives no clear indication as to where and whether clusters actually exist in the data set. A very large number of isolated points fuse together early in forming the relatively non-violent or "minimal violence" cluster. All of the singleton points at the base of this tree diagram were regarded as outlying "singletons." The related schematic diagram (Figure 2B) indicates the behavioral sense that emerged from this particular analysis. The vast number of points in this analysis formed into a very large cluster of 89 cases, which indicated only occasional or minimal violence. Four small clusters showing various increasing levels of violence were discovered also in the tree. These are indicated in the schematic diagram (Figure 2B). It would have to be concluded that the distance coefficient is less successful in the discovery and description of well separated groups of profiles than is the cosine similarity coefficient. This latter coefficient is exemplified in Figure 1A and the related schematic 1B.

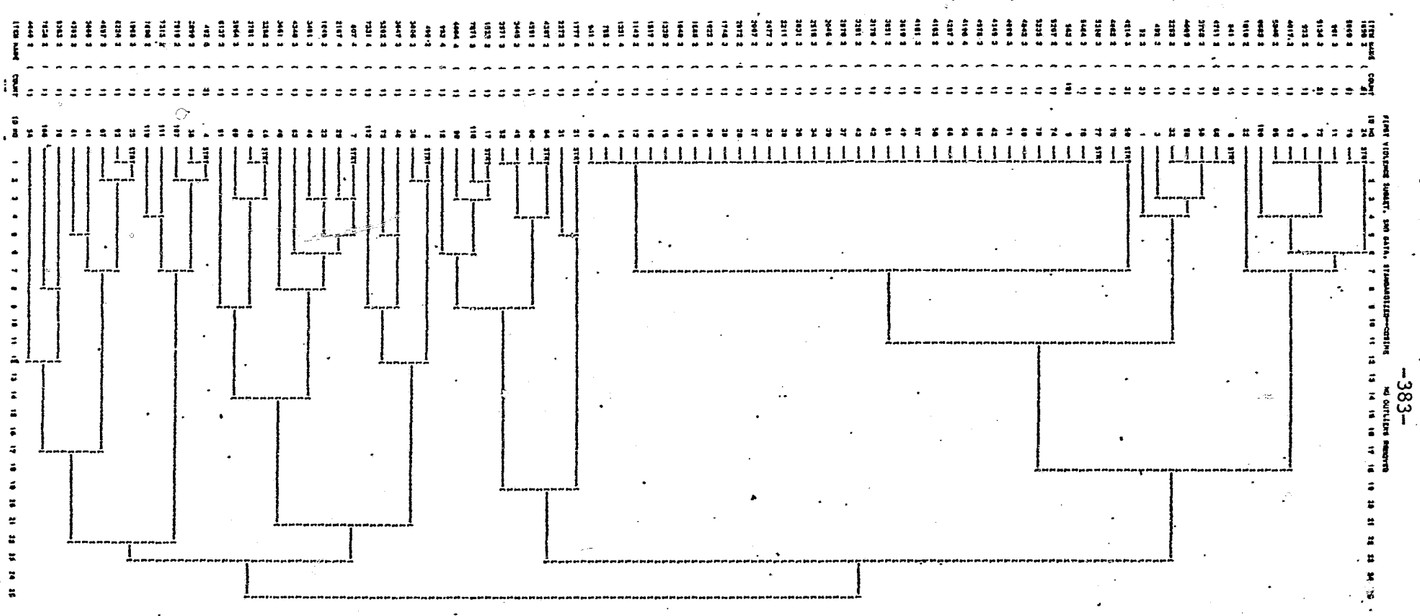


Figure 1A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Violence: Method - UPGMA with Cosine

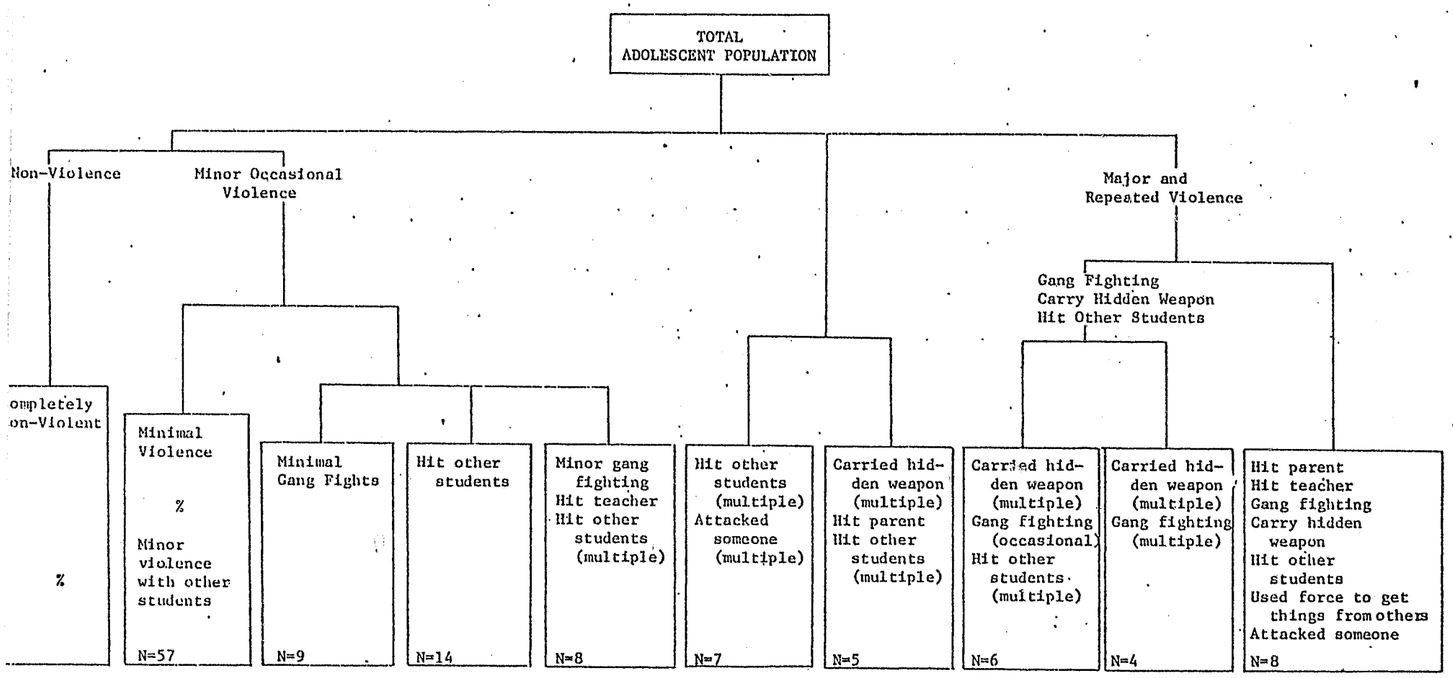


FIGURE 1B: SCHEMATIC DIAGRAM OF PATTERNS OF VIOLENT BEHAVIOR WITHIN A NATIONAL SAMPLE OF ADOLESCENTS
Taxonomy generated by UPGMA Clustering using Cosine as a Similarity Coefficient

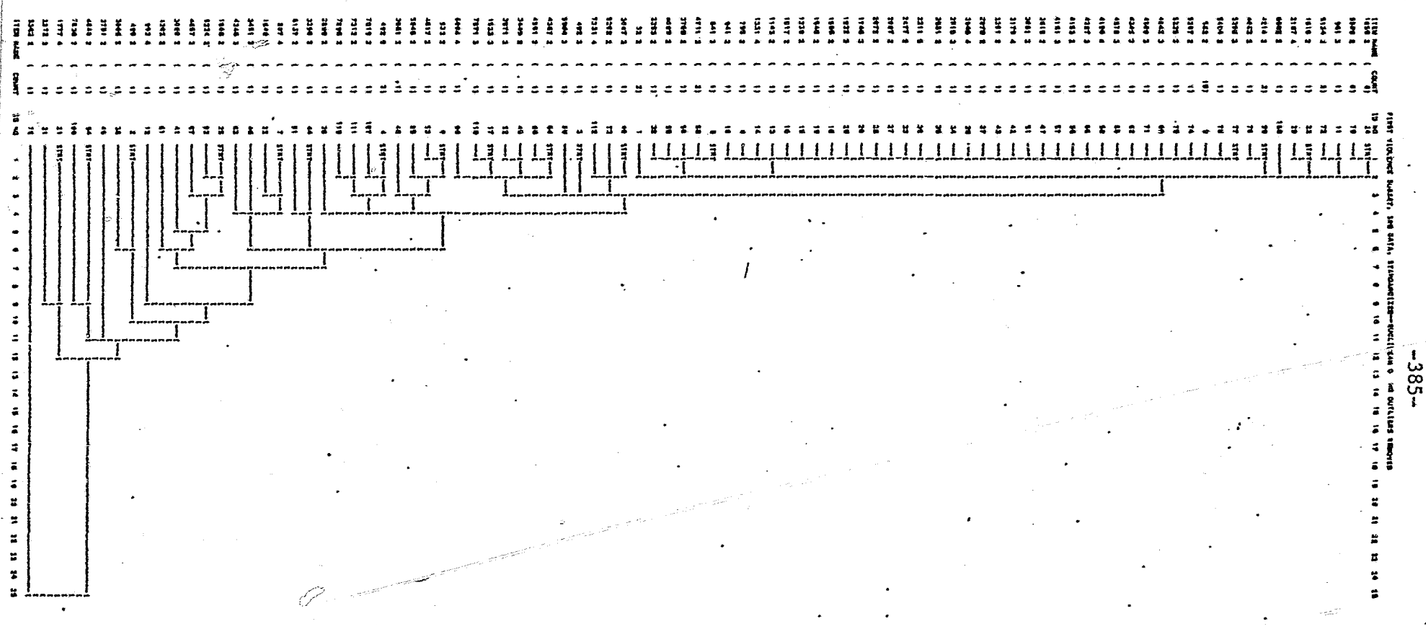


Figure 2A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Violence: Method - UPGMA Analysis with Euclidean D

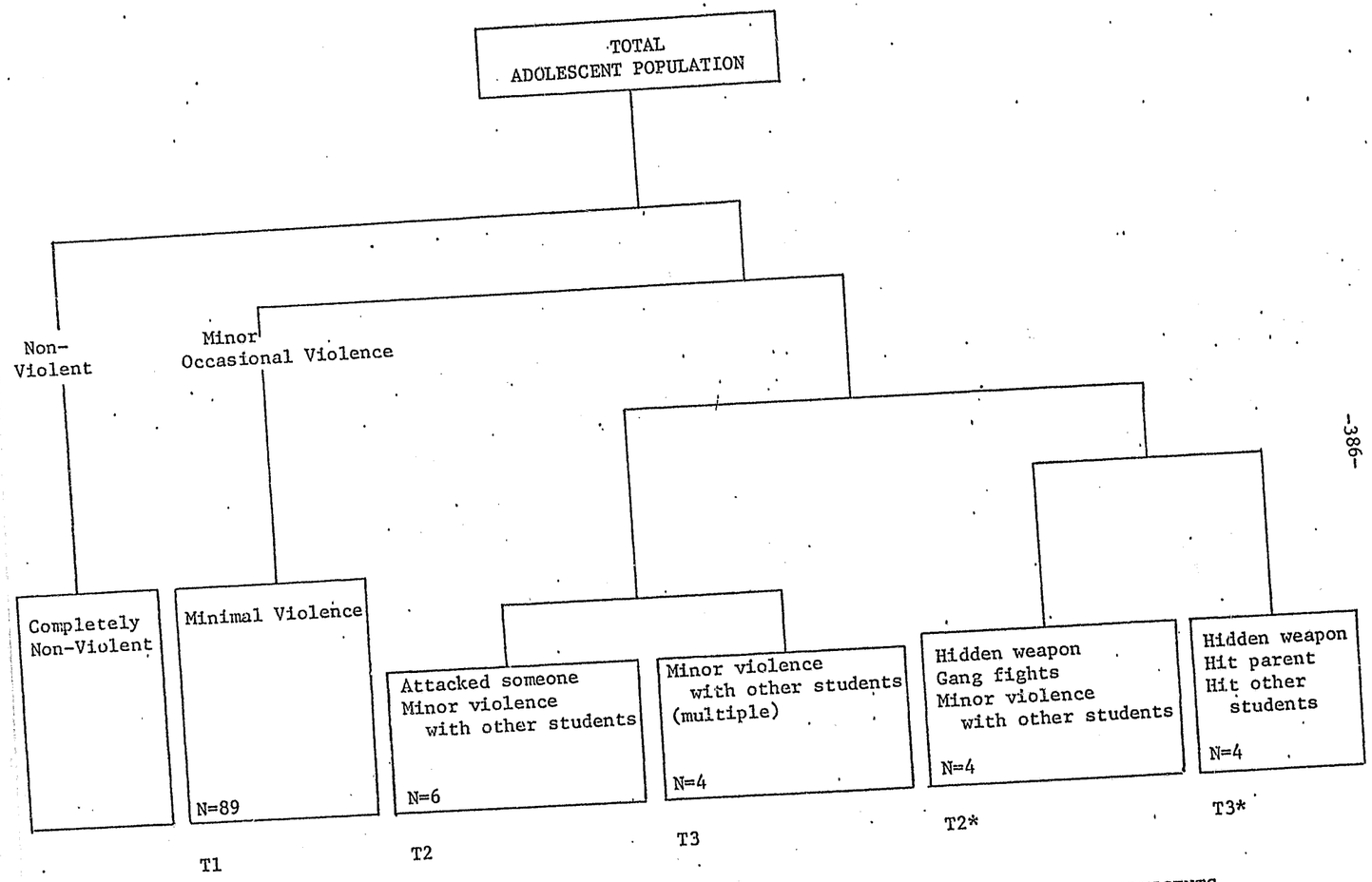


FIGURE 2B:SCHEMATIC DIAGRAM OF PATTERNS OF VIOLENT BEHAVIOR WITHIN A NATIONAL SAMPLE OF ADOLESCENTS
Taxonomy generated by UPGMA Clustering and Euclidean Distance

The schematic 1A indicates a number of likely candidates for well separated clusters in this data set. Again, a very large number of minimal violence cases are clustered together at the earliest stages of the analysis. This indicates the possibility of a well matched cluster with the UPGMA Euclidean distance analysis.

One section of this hierarchical classification includes groups of youth who are involved in major and repeated violence. Another component of the tree contains youth who are involved in minimal violence, very intermittent or occasional gang fights, or minor infractions with other students.

Turning to the HMODE analysis, we again find that the cosine dimilarity coefficient is much more sensitive to variations of pattern than is the distance coefficient. Many cases are classified as outliers by the distance coefficient and it seems also to collapse groups together which were recognized separately by the cosine similarity coefficient. Examining the schematic figures from the HMODE analyses, however, also reveals the same basic patterns that were found with the UPGMA clustering method. One set of branches in the dendrograms deals with major and repeated violence. A second deals with intermittent violence and a third deals with minor or occasional violence. It might be noted that seven or eight clusters in each case are found by the cosine method and fewer clusters are discovered by the distance method. This, again, suggests that the choice of a similarity coefficient is more important than the choice of a clustering method.

Descriptions of Replicated Violence Profiles

Minimal Violence. This cluster is picked up by all of the cluster methods and the different similarity coefficients. The vast majority of these youth fall into this particular cluster. These youth have perpetuated virtually no

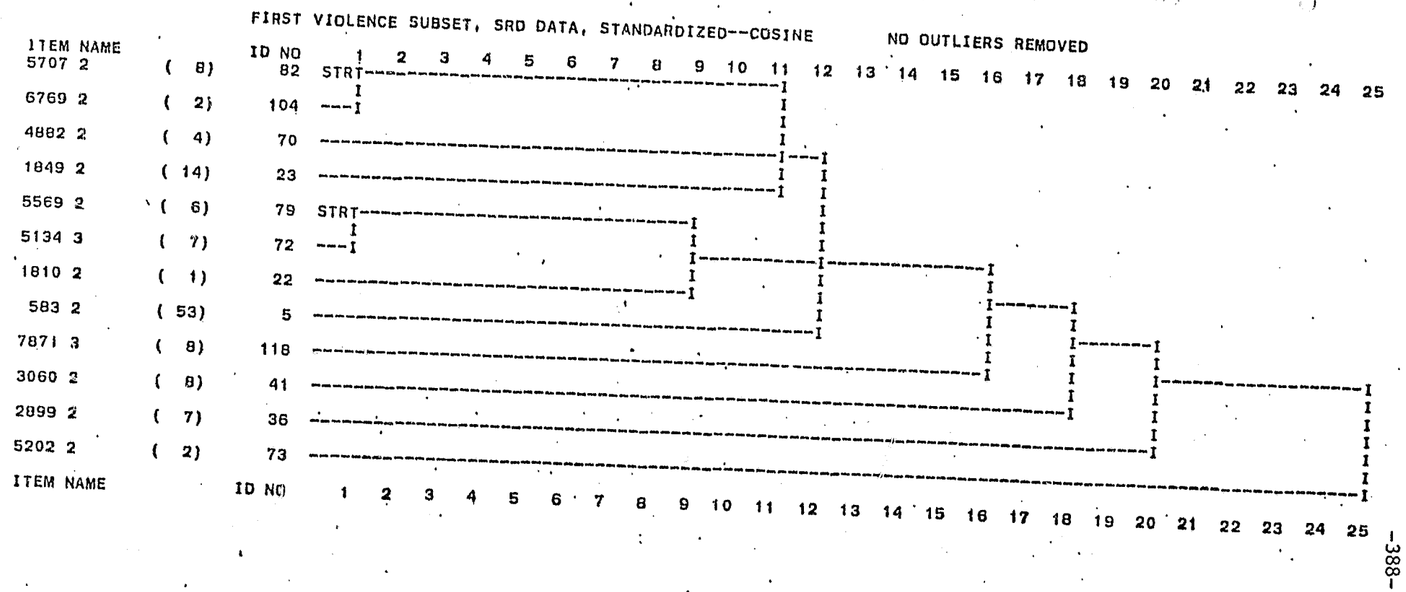


Figure 3A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Violence: Method - HMode Analysis with Cosine

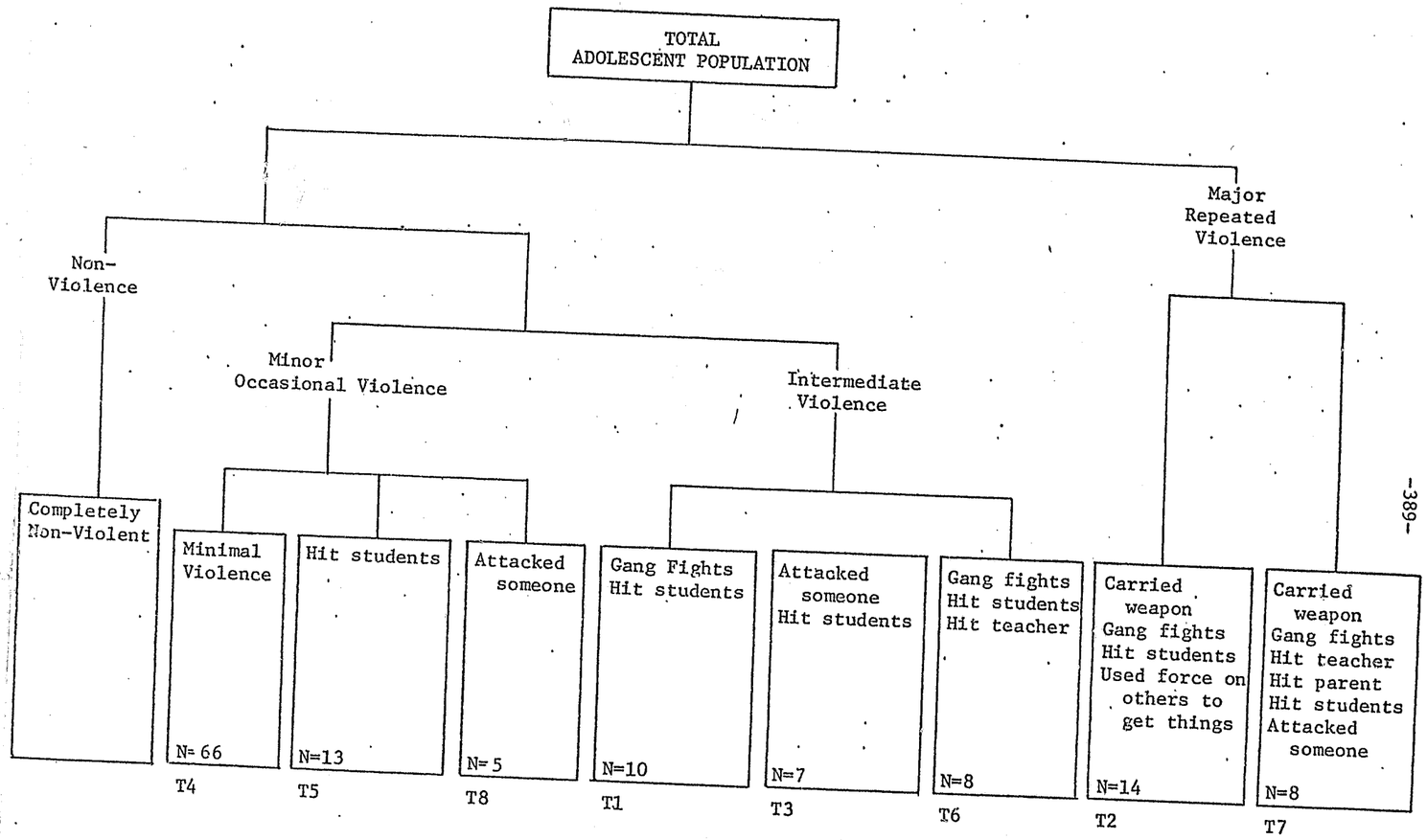


FIGURE 3B: SCHEMATIC DIAGRAM OF PATTERNS OF VIOLENT BEHAVIOR WITHIN A NATIONAL SAMPLE OF ADOLESCENTS
Taxonomy generation by Hierarchical Mode Analysis with Cosine as a Profile Similarity Coefficient

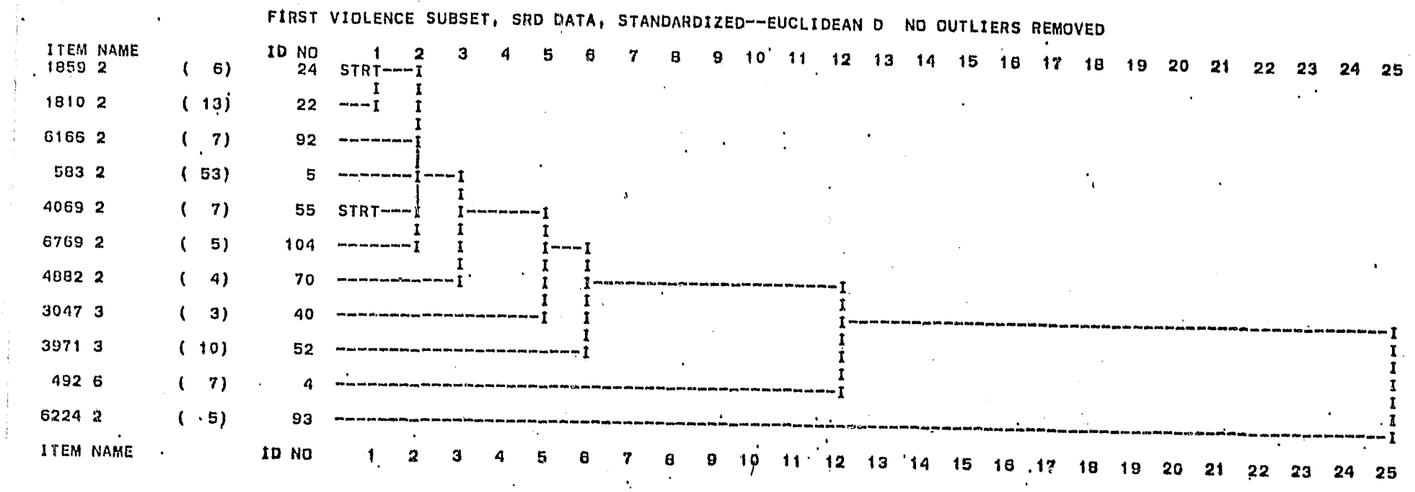


Figure 4A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Violence: Method - HMode with Euclidean D

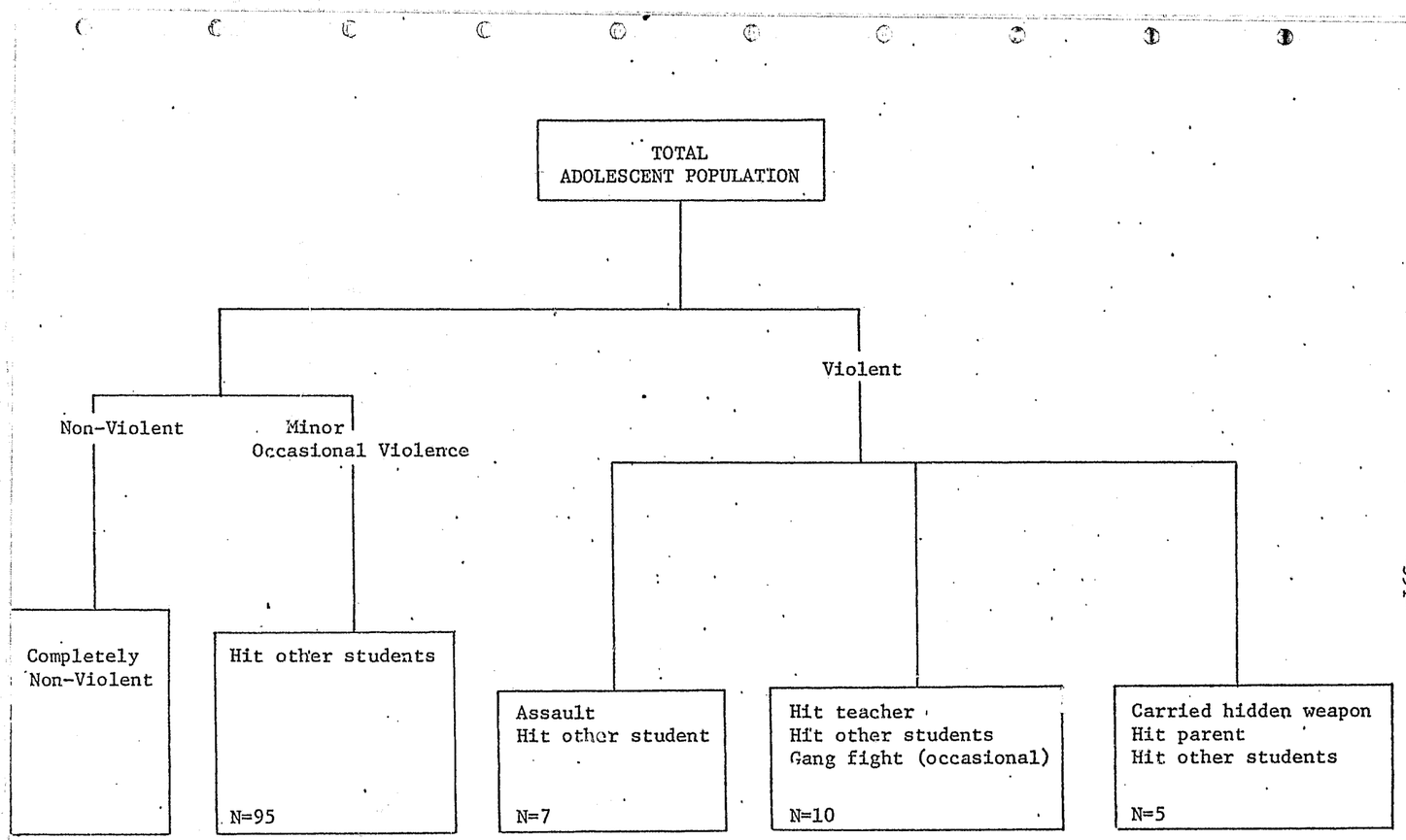


FIGURE 4B: SCHEMATIC DIAGRAM OF PATTERNS OF VIOLENT BEHAVIOR IN A NATIONAL SAMPLE OF ADOLESCENTS
Taxonomy generated by Hierarchical Mode Analysis and Euclidean Distance.

violent behaviors. The only violent behavior scores indicated are perhaps an occasional threat to a teacher or another student. This single act is either the threat of violence or an actual violent interaction, usually with another student. It might be noted that in the two analyses involving the cosine coefficient close to 60% of the youth fall into this cluster, whereas in the two analyses involving Euclidean distance, about 80% of the students fall into this category. An examination of the homogeneity coefficient for this cluster indicates that it is highly homogeneous in every single one of the four analyses. However, it may be noted that the homogeneity in the Euclidean distance analyses is much lower than that in the cosine analyses. This indicates that the larger numbers of persons being classified into this cluster in the distance-based analyses are bringing about a clearly less homogeneous cluster. This suggests that the cosine results are not leading to an artificial paring down of this particular cluster.

Intermediate and Replicated Violence. All of the analyses are picking up clusters which illustrate two or three violent behaviors. The most common of these profiles is for youth to show a violent interaction with a teacher, with other students, and to be perhaps involved in occasional gang fighting. Some of these clusters indicate two of these three behaviors; e.g., being involved in gang fights and being involved in violent interactions with other students but not hitting a teacher. A much smaller proportion of these youth fall into this type of cluster. Of the 120 students involved in these analyses, we estimate that about 20 youth, or close to 18%, are involved in this level of intermediate or low violence.

Major and Repeated Violence. The two analyses involving the cosine method clearly picked up profiles in which a number of different violent acts were

recorded and committed a multiple number of times by the youth. The violent behaviors within these clusters subsume the earlier, more moderate, violence and add a number of new forms of violent behavior. These clusters, in addition to violence against teachers, parents, and other youth, are characterized by gang fighting, carrying a hidden weapon, using force to get money or things from other people, and attacking someone with the idea of seriously hurting or killing them. Again, of the 120 youth involved in these taxometric analyses, about 20 youth are involved in this kind of profile. Thus, we estimate about 18% of this analysis sample is involved in major and repeated violence. It was not really possible to accurately delineate the boundaries between subtypes within this repeated violence syndrome. The UPGMA and HMODE methods, when using the cosine similarity coefficient, both indicated the presence of this multiple violence profile but drew the boundaries between the subtypes in slightly different ways. Nevertheless, the main core of violent behaviors against teachers, parents, and other students, together with the carrying of hidden weapons, was present in all of the various subtypes that were found.

A Note on Reliability

We have not formally classified these violence classifications against each other. Therefore, formal reliability testing is not conducted upon these classification systems. The replication of the various clusters, however, can be seen in Tables 1A through 3A. Only certain of the profiles are clearly replicated, while others are obviously unreliable. The most clear replication occurs at the level of nonviolent, minor occasional violence, and repeated, multiple violence. These basic structures recur in all of the tree diagrams. However, the more refined subdivisions within these classes, particularly the

TABLE 1A; PROFILE INTERPRETATION DATA FOR ADOLESCENT VIOLENCE TAXONOMY
HMode with Euclidean D as a Similarity Coefficient

	Cluster/Type Number							
	1		2		3		4	
Cluster Size	95		7		10		4	
Average Squared Deviation	7.712		1.302		13.280		4.063	
	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score
Specific Offenses								
1. Carried a hidden weapon other than a plain pocket knife	.000	1.642	-.126	1.429	-.320	1.100	.802	1.819
2. Attacked someone with the idea of seriously hurting or killing him/her	-.130	1.063	2.132	2.143	-.262	1.000	.157	1.200
3. Been involved in gang fights	.013	1.432	-.240	1.143	.161	1.600	-.190	1.200
4. Hit (or threatened to hit) a teacher or other adult at school	-.129	1.095	-.267	1.000	1.482	2.200	.024	1.200
5. Hit (or threatened to hit) one of your parents	-.158	1.021	-.248	1.000	-.248	1.000	4.001	2.000
6. Hit (or threatened to hit) other students	-.029	2.463	-.170	2.286	.317	2.900	.396	3.000
7. Had (or tried to have) sexual relations with someone against their will	.042	1.032	-.159	1.000	-.159	1.000	-.159	1.000
8. Used force (strong-arm methods) to get money or things from other students	-.011	1.158	-.210	1.000	.042	1.200	-.210	1.000
9. Used force (strong-arm methods) to get money or things from a teacher or other adult at school	0	1.000	0	1.000	0	1.000	0	1.000
10. Used force (strong-arm methods) to get money or things from other people	-.123	1.000	-.123	1.000	1.352	1.300	-.123	1.000
Chi-square from uniform Distribution Index								

--- ADEQUACY OF VARIABLES---

Table 1B

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4
1	-.052	.784	.965	-2.310
2	.178	.373	1.000	.123
3	-.169	.890	.284	.847
4	.137	1.000	-.323	.575
5	.624	1.000	1.000	1.000
6	-.090	.433	-.037	.687
7	-.257	1.000	1.000	1.000
8	-.164	1.000	.363	1.000
9	1.000	1.000	1.000	1.000
10	1.000	1.000	-10.007	1.000

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8	9	10
-.011	.257	-.024	.168	.695	-.021	-.021	.001	1.000	.059

CLUSTER HOMOGENITIES

1	2	3	4
.221	.848	-.476	.492

TABLE 2A: PROFILE INTERPRETATION DATA FOR ADOLESCENT VIOLENCE TAXONOMY
HMode with Cosine as a Similarity Coefficient

	Cluster/Type Number															
	1		2		3		4		5		6		7		8	
Cluster Size	10		14		7		53		13		8		8		4	
Average Squared Deviation	1.171		19.096		1.302		.000		.773		2.819		43.186		0	
Specific Offenses	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score
1. Carried a hidden weapon other than a plain pocket knife	-.320	1.100	1.645	4.429	-.126	1.429	-.379	1.000	-.379	1.000	-.305	1.125	1.023	3.375	.211	2.000
2. Attacked someone with the idea of seriously hurting or killing him/her	-.052	1.100	-.262	1.000	2.132	2.143	-.262	1.000	-.262	1.000	-.262	1.000	1.309	1.750	-.262	1.000
3. Been involved in gang fights	.861	2.400	.699	2.214	-.240	1.143	-.365	1.000	-.298	1.077	.182	1.625	.839	2.375	-.365	1.000
4. Hit (or threatened to hit) a teacher or other adult at school	-.267	1.000	.149	1.286	-.267	1.000	-.267	1.000	-.267	1.000	1.555	2.250	1.191	2.000	-.267	1.000
5. Hit (or threatened to hit) one of your parents	-.248	1.000	-.248	1.000	-.248	1.000	-.248	1.000	-.248	1.000	-.248	1.000	3.470	1.875	-.248	1.000
6. Hit (or threatened to hit) other students	-.554	1.800	.565	3.214	-.170	2.286	-.396	2.000	1.126	3.923	.396	3.000	.890	3.625	-1.187	1.000
7. Had (or tried to have) sexual relations with someone against their will	-.159	1.000	-.159	1.000	-.159	1.000	-.159	1.000	-.159	1.000	-.159	1.000	2.232	1.375	-.159	1.000
8. Used force (strong-arm methods) to get money or things from other students	-.210	1.000	.872	1.857	-.210	1.000	-.210	1.000	-.210	1.000	-.210	1.000	.736	1.750	-.210	1.000
9. Used force (strong-arm methods) to get money or things from a teacher or other adult at school	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000
10. Used force (strong-arm methods) to get money or things from other people	-.123	1.000	.930	1.214	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000
Chi-square from uniform Distribution Index																

--- ADEQUACY OF VARIABLES---

Table 2B

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5	6	7	8
1	.965	-1.772	.784	1.000	1.000	.956	-2.278	1.000
2	.561	1.000	.373	1.000	1.000	1.000	-7.461	1.000
3	.284	-1.735	.890	1.000	.941	.137	-4.683	1.000
4	1.000	.206	1.000	1.000	1.000	-.063	-8.113	1.000
5	1.000	1.000	1.000	1.000	1.000	1.000	-1.257	1.000
6	.889	-1.522	.433	1.000	.221	-.253	-1.494	1.000
7	1.000	1.000	1.000	1.000	1.000	1.000	-9.897	1.000
8	1.000	-2.643	1.000	1.000	1.000	1.000	-6.172	1.000
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	1.000	-7.098	1.000	1.000	1.000	1.000	1.000	1.000

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8	9	10
.425	.346	.151	.209	.846	.312	.255	.074	1.000	.031

CLUSTER HOMOGENEITIES

1	2	3	4	5	6	7	8
.870	-1.056	.848	1.000	.916	.678	-3.936	1.000

TABLE 3A; PROFILE INTERPRETATION DATA FOR ADOLESCENT VIOLENCE TAXONOMY
UPGMA with Cosine as a Similarity Coefficient

	Cluster/Type Number													
	1		2		3		4		5		6		7	
Cluster Size	66		5		7		10		14		8		8	
Average Squared Deviation	.193		9.170		1.302		7.374		.777		2.819		43.186	
	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score
Specific Offenses														
1. Carried a hidden weapon other than a plain pocket knife	-.343	1.061	.211	2.000	-.125	1.429	2.218	5.400	-.337	1.071	-.305	1.125	1.023	3.375
2. Attacked someone with the idea of seriously hurting or killing him/her	-.262	1.000	-.262	1.000	2.132	2.143	-.052	1.100	-.262	1.000	-.262	1.000	1.309	1.750
3. Been involved in gang fights	-.232	1.152	-.015	1.400	-.240	1.143	1.299	2.900	-.302	1.071	.182	1.625	.839	2.375
4. Hit (or threatened to hit) a teacher or other adult at school	-.267	1.000	.024	1.200	-.267	1.000	-.122	1.100	-.267	1.000	1.555	2.250	1.191	2.000
5. Hit (or threatened to hit) one of your parents	-.248	1.000	-.248	1.000	-.248	1.000	-.248	1.000	-.248	1.000	-.248	1.000	3.470	1.875
6. Hit (or threatened to hit) other students	-.468	1.909	1.187	4.000	-.170	2.286	.079	2.600	1.074	3.857	.396	3.000	.890	3.625
7. Had (or tried to have) sexual relations with someone against their will	-.159	1.000	-.159	1.000	-.159	1.000	-.159	1.000	-.159	1.000	-.159	1.000	2.232	1.375
8. Used force (strong-arm methods) to get money or things from other students	-.210	1.000	2.567	3.200	-.210	1.000	-.084	1.100	-.210	1.000	-.210	1.000	.736	1.750
9. Used force (strong-arm methods) to get money or things from a teacher or other adult at school	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000	0	1.000
10. Used force (strong-arm methods) to get money or things from other people	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000	-.123	1.000
Chi-square from uniform Distribution Index														

--- ADEQUACY OF VARIABLES---

Table 3B

VARIABLE X CLUSTER HOMOGENEITIES

	1	2	3	4	5	6	7
1	.980	.303	.784	-1.415	.975	.956	-2.278
2	1.000	1.000	.373	.561	1.000	1.000	-7.461
3	.876	.770	.890	-2.826	.945	.137	-4.683
4	1.000	.575	1.000	.787	1.000	-.063	-8.113
5	1.000	1.000	1.000	1.000	1.000	1.000	-1.257
6	.947	-4.011	.433	-.141	.243	-.253	-1.494
7	1.000	1.000	1.000	1.000	1.000	1.000	-9.897
8	1.000	-4.100	1.000	.841	1.000	1.000	-6.172
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000

VARIABLE HOMOGENEITIES (UNIVARIATE CORRELATION RATIO)

1	2	3	4	5	6	7	8	9	10
.514	.352	.140	.274	.847	.284	.261	.284	1.000	1.000

CLUSTER HOMOGENITIES

1	2	3.	4	5	6	7
.980	-.146	.848	.181	.916	.678	-3.936

highly multiple violent class, are not obviously exact replicates of each other. They do, however, contain the basic pattern of a large number of offenses against parents, peers, teachers, gang fighting, and the carrying of a hidden weapon.

It should be noted that the one hundred and twenty youth on whom these classifications are conducted were identified as all of the youth who had conducted at least one violent act. The proportions therefore given in the above discriptions refer to this particular sample. These proportions therefore do not refer to general national population norms.

A BEHAVIORAL CLASSIFICATION OF JUVENILE THEFT

In this section we examine the classification of behavioral profiles which resulted from the application of the various cluster analytic methods to the attribute space consisting of six theft items. The sample for this particular study consisted of the selection from the full juvenile delinquency sample of all of those youth who indicated at least one theft behavior committed during the prior year. Thus, this is a purpose of sample and all of the youth who entered into the cluster analytic study had committed at least one theft act during the prior year. As in the above study of violence, therefore, the proportions attached to each behavioral profile class do not reflect the proportion of all youth falling into these classes but indicate the proportion of all theft behavioral profiles which would be accounted for by the particular type.

Hierarchical Structure of the Theft Classifications

The diagrams presented below illustrate the hierarchical tree structrues which emerged from the various analyses of this theft data.

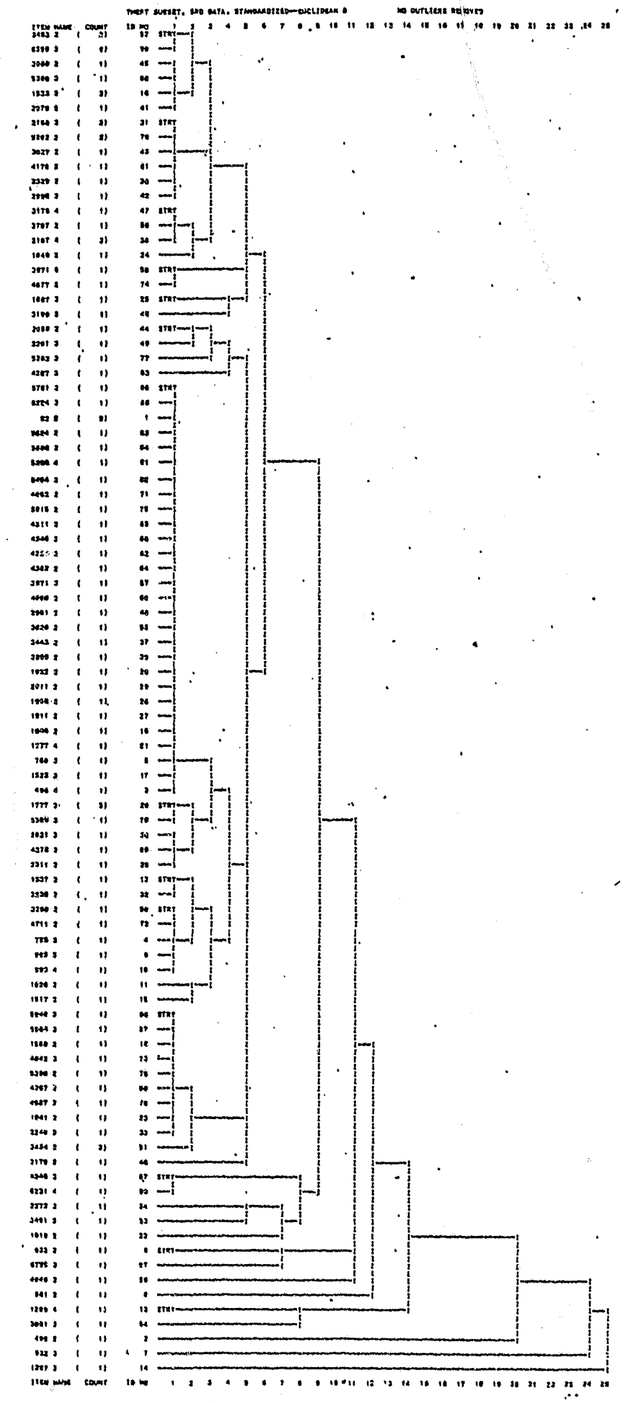


Figure 5A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Theft: Method - UPGMA Analysis with Euclidean D

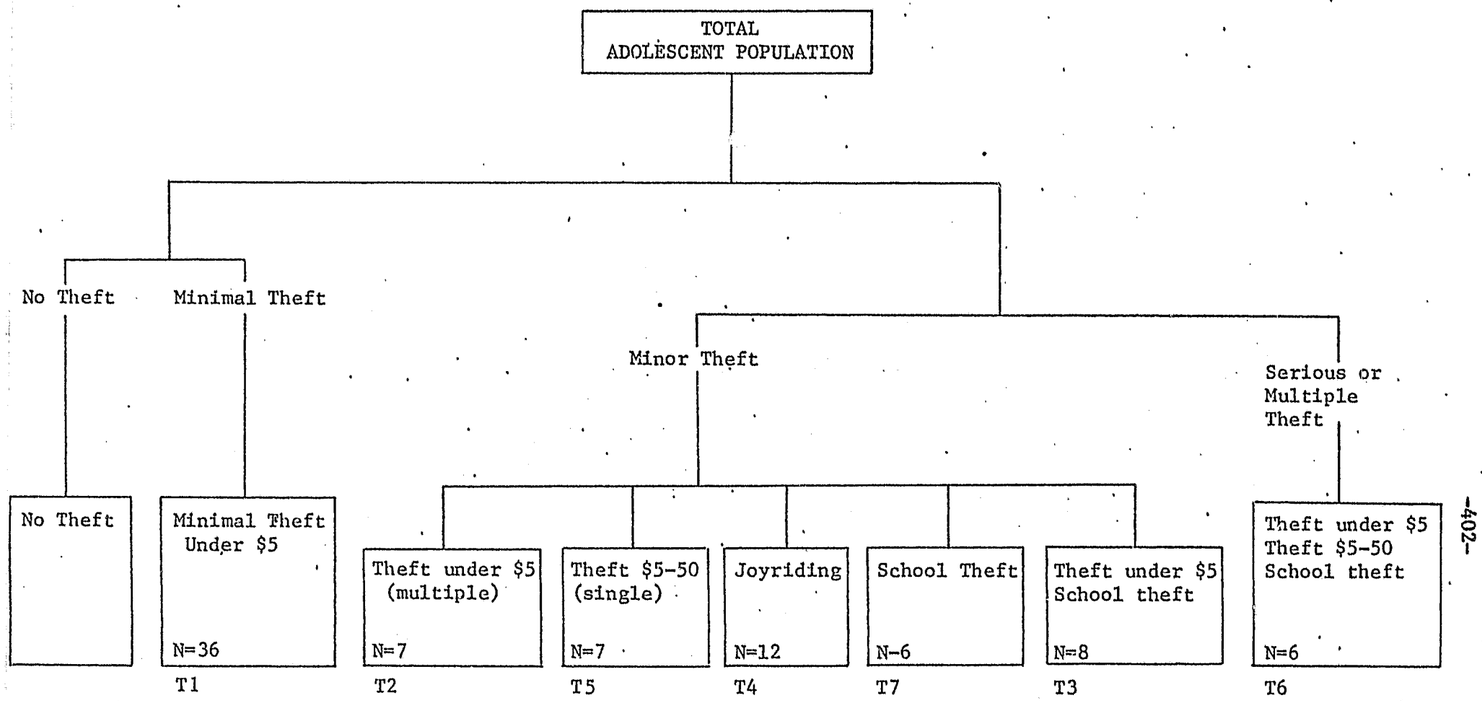


FIGURE 5B: SCHEMATIC DIAGRAM OF PATTERNS OF THEFT IN A NATIONAL SAMPLE OF ADOLESCENTS
Taxonomy generated by UPGMA and Euclidean Distance as a Profile Similarity Coefficient

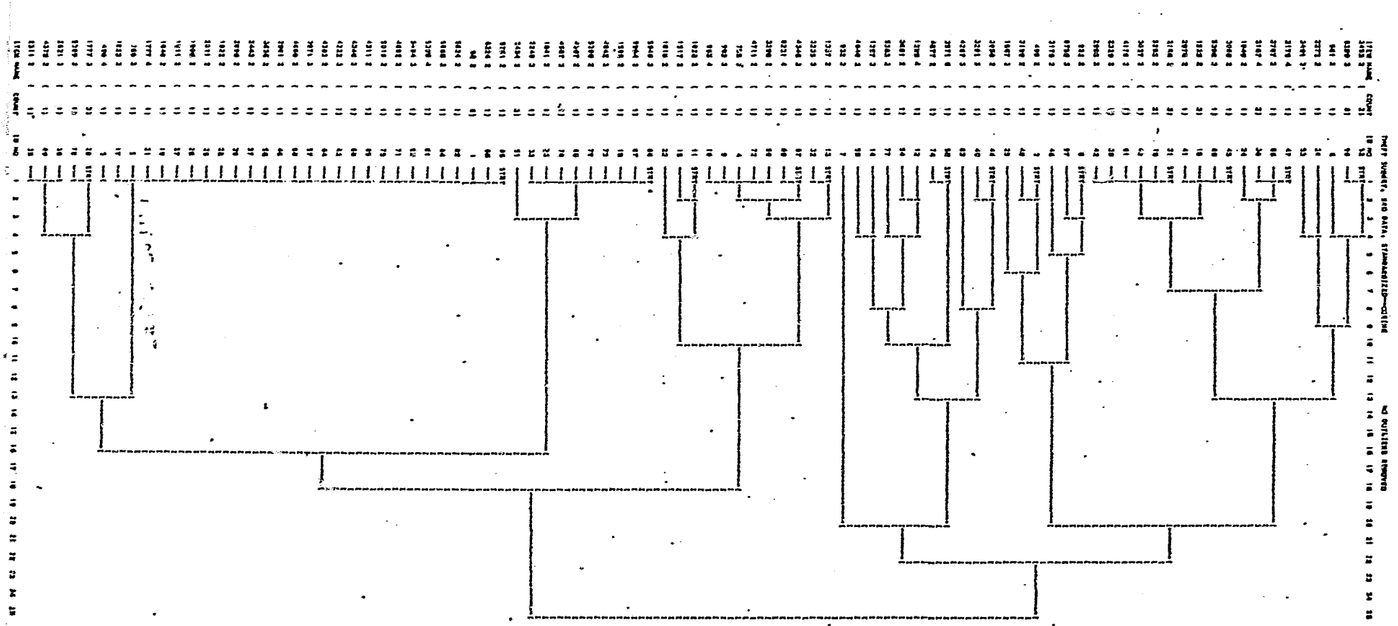


Figure 6A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Theft: Method – UPGMA Analysis with Cosine

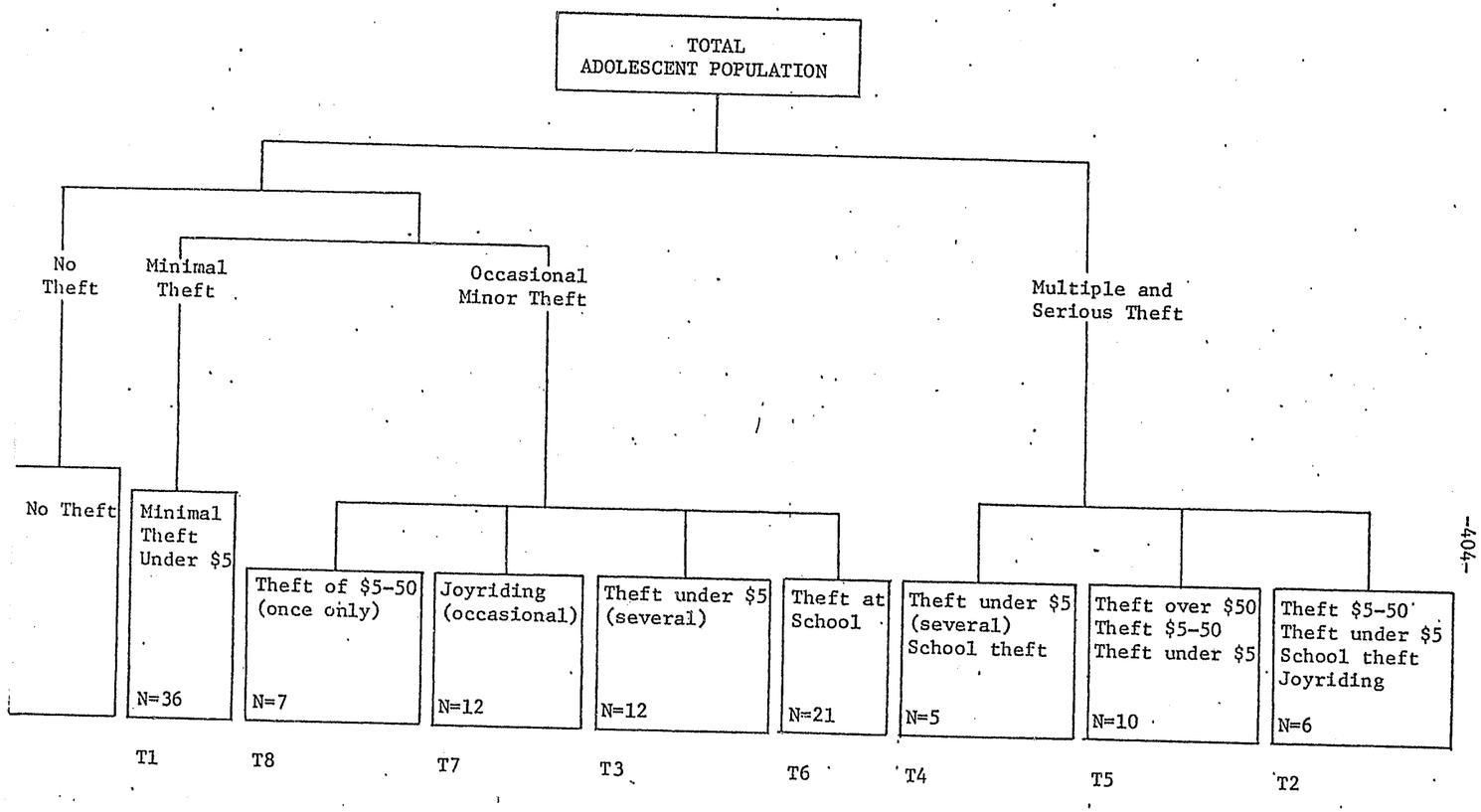


FIGURE 6B:SCHEMATIC DIAGRAM OF PATTERNS OF THEFT IN A NATIONAL SURVEY SAMPLE OF ADOLESCENTS
Taxonomy generated by UPGMA Clustering and Cosine as a Profile Similarity Coefficient

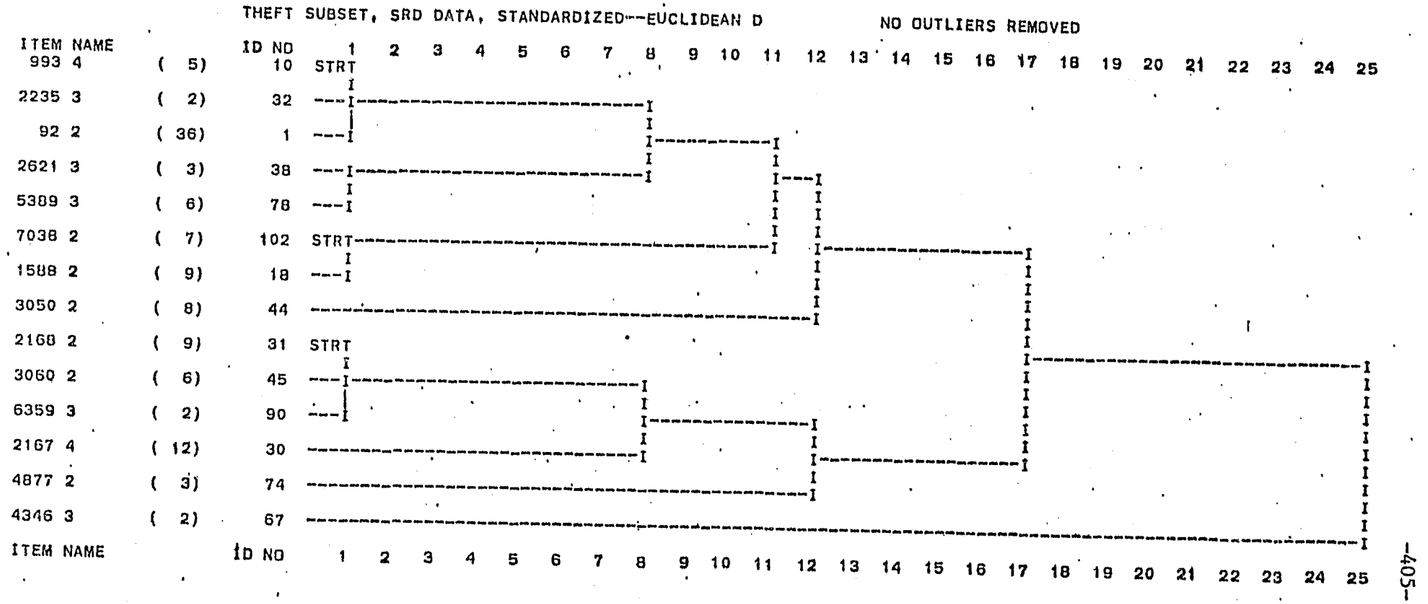


Figure 7A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Theft: Method - HMode Analysis with Euclidean D

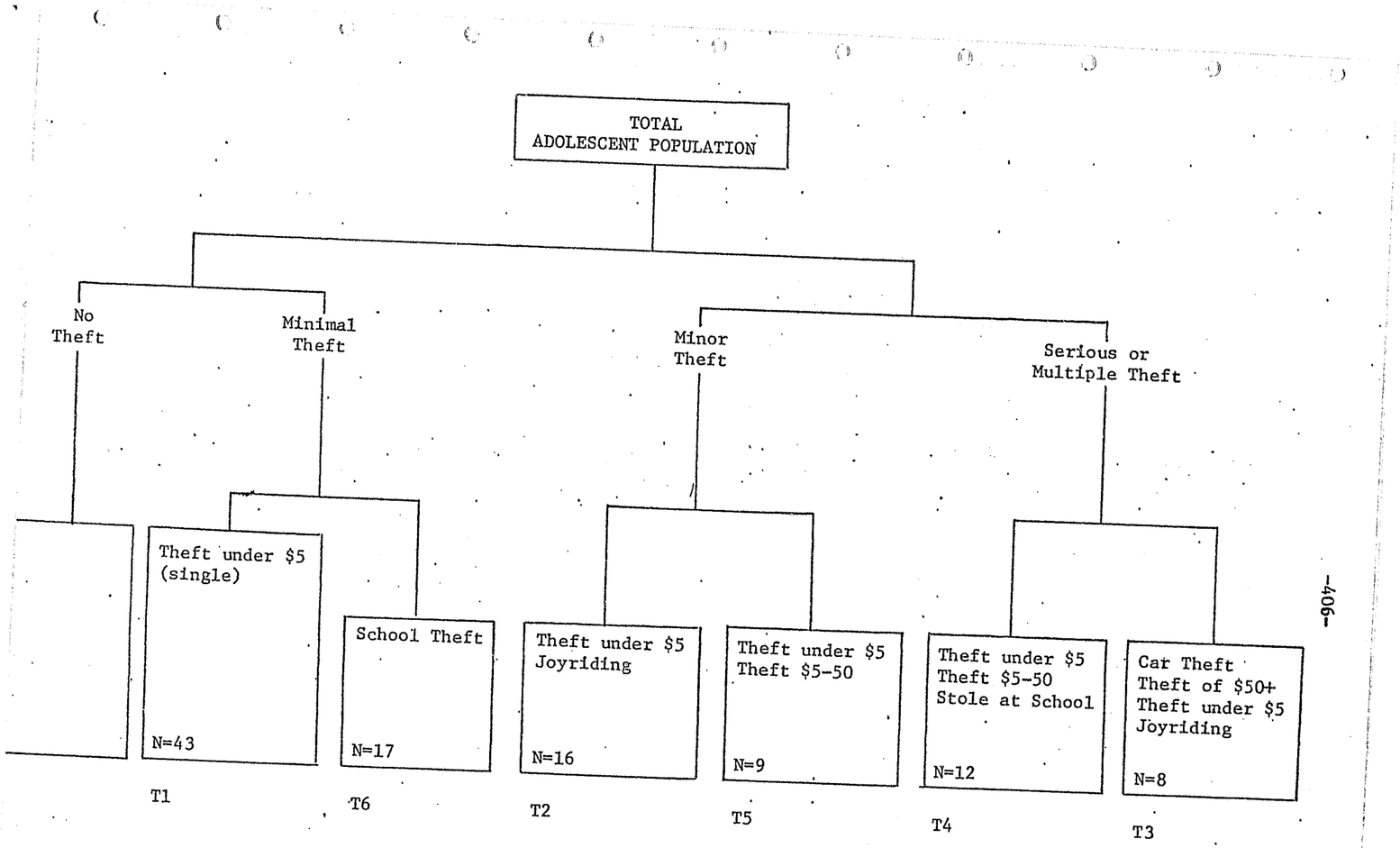


FIGURE 7B: SCHEMATIC TAXONOMIC TREE FOR ADOLESCENT THEFT PATTERNS
Generated by Hierarchical Mode Clustering with Euclidean
Distance Similarity Coefficient

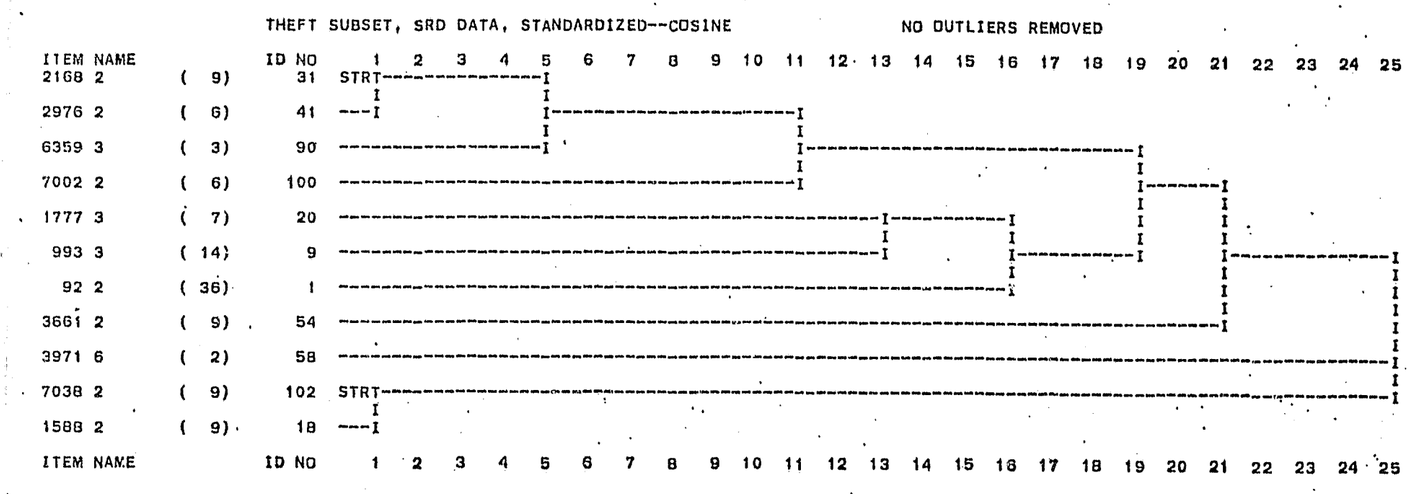


Figure 8A

Dendrogram to Indicate Hierarchical Classification of an Adolescent National Sample According to Profiles of Theft: Method - HMode Analysis with Cosine

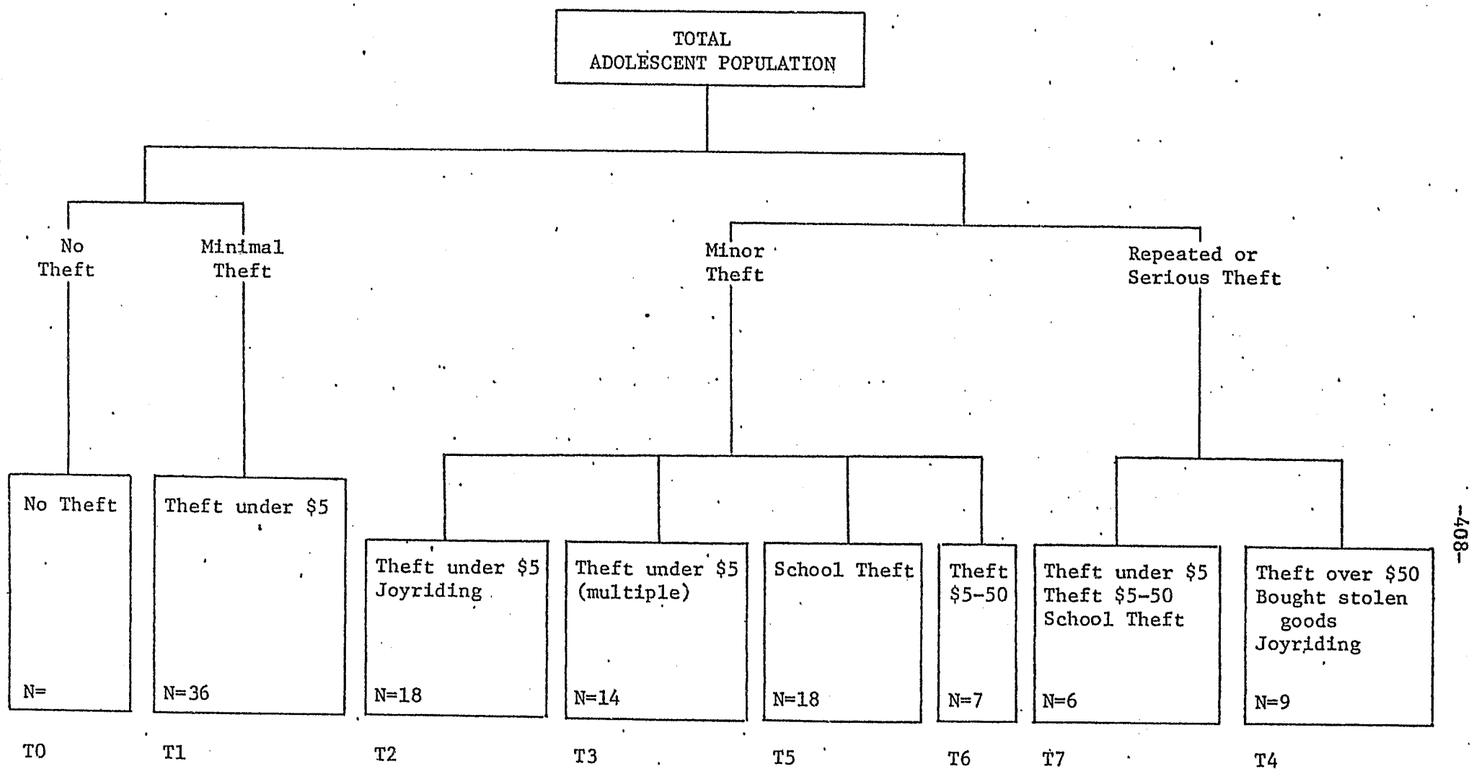


FIGURE 8B: SCHEMATIC DIAGRAM OF PATTERNS OF THEFT IN A NATIONAL SAMPLE OF ADOLESCENTS
Taxonomy generated by Hierarchical Mode Analysis with Cosine as a profile Similarity Coefficient

All of these hierarchical tree structures illustrate the same basic form of hierarchy. They indicate that there is a fairly dense closely-packed and very large cluster consisting of minimal theft behaviors. This large cluster accounting for about thirty percent of the samples in each case consists of youth who have perhaps committed only one theft of under five dollars, or who have committed only one school-related minor theft. This cluster can be seen in each of the tree structure diagrams as the largest of all the clusters. This is clearly the densest area of the measurement space in terms of numbers of youth. The size of this cluster is comparable in all of the different analyses.

A second major region or class within this set of hierarchies consists of classes of youth who have committed some minor theft beyond the minimal theft of the prior class. Each of the analyses identifies some overlapping clusters within this general area. For instance, repeated school theft consists of one such subclass of minor theft. Additionally, a group of youth indulge in occasional joy riding as a further class. A further class consists of youth who have stolen in the area of five to fifty dollars. The four different analyses define the minor theft subclasses around this narrow set of variables. There is an obvious overlap between many of the profiles identified and this section of the hierarchical structures. The schematic diagrams in particular illustrate the major behavior within each of these subclasses.

The final general class consists of multiple or serious theft. In this case, all of the analyses have deliniated certain clusters which I define by two, three or four of the different theft behaviors. These small classes of youth have indulged repeatedly in a larger number of separate theft offenses. It might be noted again that the Euclidean distance metric is less successful in discovering separate behavioral profiles than is the cosine method. Figures 5B

and 7B indicate the schematic tree diagrams emanating from the two analyses UPGMA and HMode when the Euclidian distance metric is used. One cluster discovered by all of the methods consists of youth who have indulged in school theft, theft under five dollars, and theft in the region five to fifty dollars. In each case, these offenses are perpetuated multiply by the youth. Other patterns are found in two of the analyses by the addition of joy riding and car theft to this basic pattern.

Descriptions of Replicated Profiles

Tables 4 through 8 indicate the standard score and the raw means score for the behavioral profiles emanating from the various cluster analytic approach. Each of these separate clusters has a distinctive behavioral profile across the six theft items. The raw mean score is probably the better indicator than the standardized distance score. A score of one indicates that the youth has not committed this particular act. A score of two indicates that they committed the offense once or twice, and so forth.

It is interesting to note the sizes of the homogeneity coefficients for these analyses of the theft clusters. Many of them are exceptionally high, i.e. in the range .8 and above. This basically indicates that virtually every member of the cluster has the identical profile. This occurs largely because of the small number of items in the present analyses (six only) and the fact that a great majority of the members in any particular cluster score one, indicating never, on each of the items. In those instances where the cluster homogeneities are low, they become exceedingly low as a result of the general high homogeneity of the other clusters in the analysis. It is also noteworthy that clusters which do not illustrate very high homogeneity are often those at the high multiple end of the theft spectrum. This is only to be expected if a group of youth are characterized by positive scores on

TABLE 4: PROFILE INTERPRETATION DATA FOR ADOLESCENT THEFT TAXONOMY
UPGMA with Euclidean D as a Similarity Coefficient

	Cluster/Type Number						
	1	2	3	4	5	6	7
Cluster Size	36	7	8	12	7	6	9
Average Squared Deviation	.000	.124	.114	.114	.149	.084	.000
	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score
Specific Offenses							
1. Stolen (or tried to steal) a motor vehicle, such as a car or motorcycle	-.137 1.000	-.137 1.000	-.137 1.000	-.137 1.000	-.137 1.000	-.137 1.000	-.137 1.000
2. Stolen (or tried to steal) something worth more than $50	-.279 1.000	-.279 1.000	-.279 1.000	-.279 1.000	-.279 1.000	-.279 1.000	-.279 1.000
3. Stolen (or tried to steal) things worth $5 or less	-.206 2.000	1.131 3.714	-.011 2.250	-.790 1.250	-.540 1.571	-.335 1.833	-.985 1.000
4. Taken a vehicle for a ride (drive) without the owner's permission	-.403 1.000	-.403 1.000	-.403 1.000	1.076 2.000	-.403 1.000	-.403 1.000	-.403 1.000
5. Stolen (or tried to steal) things worth between $5 and $50	-.459 1.000	-.459 1.000	-.459 1.000	-.459 1.000	.663 2.000	.663 2.000	-.459 1.000
6. Stolen (or tried to steal) something at school such as someone's coat from a classroom, or cafeteria, or a book from the library	-.602 1.000	-.602 1.000	1.237 2.000	-.602 1.000	-.602 1.000	1.237 2.000	1.237 2.000
Homogeneity Coefficient	1.00	.98	.98	.98	.97	.98	1.00

TABLE 5: PROFILE INTERPRETATION DATA FOR ADOLESCENT THEFT TAXONOMY
UPGMA with Euclidean D as a Similarity Coefficient

	Cluster/Type Number							
	1		2		3		4	
Cluster Size	52		23		12		4	
Average Squared Deviation	.498		.486		.114		.764	
	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score
Specific Offenses								
1. Stolen (or tried to steal) a motor vehicle, such as a car or motorcycle	-.137	1.000	-.137	1.000	-.137	1.000	.178	1.250
2. Stolen (or tried to steal) something worth more than $50	-.279	1.000	-.279	1.000	-.279	1.000	1.255	2.000
3. Stolen (or tried to steal) things worth $5 or less	-.026	2.231	-.477	1.652	-.790	1.250	-.595	1.500
4. Taken a vehicle for a ride (drive) without the owner's permission	-.403	1.000	-.403	1.000	1.076	2.000	-.403	1.000
5. Stolen (or tried to steal) things worth between $5 and $50	-.265	1.173	-.166	1.261	-.459	1.000	.102	1.500
6. Stolen (or tried to steal) something at school such as someone's coat from a classroom, or cafeteria, or a book from the library	-.602	1.000	1.237	2.000	-.602	1.000	-.602	1.000
Homogeneity Coefficient	.92		.92		.98		.83	

TABLE 6: PROFILE INTERPRETATION DATA FOR ADOLESCENT THEFT TAXONOMY
UPGMA with Cosine as a Similarity Coefficient

Cluster/Type Number

	1		2		3		4		5		6		7		8	
Cluster Size	36		6		12		5		10		21		12		7	
Average Squared Deviation	.000		8.831		1.961		2.841		12.429		.408		.114		.149	
	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score
Specific Offenses																
1. Stolen (or tried to steal) a motor vehicle, such as a car or motorcycle	-.137	1.000	-.137	1.000	-.137	1.000	-.137	1.000	.366	1.400	-.137	1.000	-.137	1.000	-.137	1.000
2. Stolen (or tried to steal) something worth more than $50	-.279	1.000	-.023	1.167	-.279	1.000	.028	1.200	2.481	2.800	-.279	1.000	-.279	1.000	-.279	1.000
3. Stolen (or tried to steal) things worth $5 or less	-.206	2.000	1.094	3.667	1.549	4.250	1.665	4.400	.028	2.300	-.577	1.524	-.790	1.250	-.540	1.571
4. Taken a vehicle for a ride (drive) without the owner's permission	-.403	1.000	2.555	3.000	-.403	1.000	-.403	1.000	.484	1.600	-.403	1.000	1.076	2.000	-.403	1.000
5. Stolen (or tried to steal) things worth between $5 and $50	-.459	1.000	.663	2.000	.008	1.417	.214	1.600	1.561	2.800	-.138	1.286	-.459	1.000	.663	2.000
6. Stolen (or tried to steal) something at school such as someone's coat from a classroom, or cafeteria, or a book from the library	-.602	1.000	1.237	2.000	-.602	1.000	1.605	2.200	-.050	1.300	1.237	2.000	-.602	1.000	-.602	1.000
Homogeneity Coefficient	1.00		-.77		.64		.41		-1.30		.93		.98		.97	

TABLE 7: PROFILE INTERPRETATION DATA FOR ADOLESCENT THEFT TAXONOMY
HMode with Cosine as a Similarity Coefficient

	Cluster/Type Number						
	1	2	3	4	5	6	7
Cluster Size	36	18	14	9	18	7	6
Average Squared Deviation	.000	5.341	2.450	22.886	1.196	.149	.084
Specific Offenses	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score	Stand. Dist. / Raw Mean Score
1. Stolen (or tried to steal) a motor vehicle, such as a car or motorcycle	-.137 1.000	-.137 1.000	-.137 1.000	1.542 2.333	-.137 1.000	-.137 1.000	-.137 1.000
2. Stolen (or tried to steal) something worth more than $50	-.279 1.000	-.194 1.056	-.279 1.000	2.447 2.778	-.194 1.056	-.279 1.000	-.279 1.000
3. Stolen (or tried to steal) things worth $5 or less	-.206 2.000	-.162 2.056	1.632 4.357	-.032 2.222	-.335 1.833	-.540 1.571	-.335 1.833
4. Taken a vehicle for a ride (drive) without the owner's permission	-.403 1.000	1.569 2.333	-.403 1.000	.583 1.667	-.403 1.000	-.403 1.000	-.403 1.000
5. Stolen (or tried to steal) things worth between $5 and $50	-.459 1.000	-.085 1.333	.182 1.571	1.536 2.778	-.459 1.000	.663 2.000	.663 2.000
6. Stolen (or tried to steal) something at school such as someone's coat from a classroom, or cafeteria, or a book from the library	-.602 1.000	.011 1.333	-.339 1.143	-.398 1.111	1.340 2.056	-.602 1.000	1.237 2.000
Homogeneity Coefficient	1.00	.01	.56	-3.31	.78	.97	.98

TABLE 8: PROFILE INTERPRETATION DATA FOR ADOLESCENT THEFT TAXONOMY
HMode with Euclidean D as a Similarity Coefficient

	Cluster/Type Number											
	1		2		3		4		5		6	
Cluster Size	43		16		8		12		9		17	
Average Squared Deviation	.264		5.682		24.671		3.292		.540		.290	
	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score	Stand. Dist.	Raw Mean Score
Specific Offenses												
1. Stolen (or tried to steal) a motor vehicle, such as a car or motorcycle	-.137	1.000	-.137	1.000	1.595	2.375	-.137	1.000	-.137	1.000	-.137	1.000
2. Stolen (or tried to steal) something worth more than $50	-.279	1.000	-.183	1.063	2.404	2.750	-.151	1.083	-.279	1.000	-.279	1.000
3. Stolen (or tried to steal) things worth $5 or less	.012	2.279	-.206	2.000	-.011	2.250	.574	3.000	-.206	2.000	-.527	1.588
4. Taken a vehicle for a ride (drive) without the owner's permission	-.403	1.000	1.630	2.375	.336	1.500	-.157	1.167	-.403	1.000	-.403	1.000
5. Stolen (or tried to steal) things worth between $5 and $50	-.459	1.000	-.108	1.313	1.645	2.875	.756	2.083	.663	2.000	-.459	1.000
6. Stolen (or tried to steal) something at school such as someone's coat from a classroom, or cafeteria, or a book from the library	-.602	1.000	-.142	1.250	-.602	1.000	1.237	2.000	-.602	1.000	1.237	2.000
Homogeneity Coefficient	.95		.01		-3.7		.40		.90		.95	

a larger number of theft behaviors. The cluster itself has a much higher possibility of being relatively spread out within this attribute space.. These analyses, nevertheless, are encouraging in that they indicate four general patterns of theft, i.e. no theft, minimal theft, minor theft patterns, and multiple or serious theft. Although we have not conducted formal cross-classification of the typologies against each other, an inspection of Tables 4 through 8 and of their related schematic hierarchical structures indicates that the different classifications have a great deal in common with each other and in many cases the same cluster profiles are delineated.

A BEHAVIORAL CLASSIFICATION OF JUVENILE DRUG CONSUMPTION PATTERNS

The taxonomic structures emerging from our analyses of drug patterns are fairly stable and reliable. The different grouping methods and similarity coefficients provided taxonomic structures of this adolescent population that are consistent with each other. The following table indicates this agreement.

Table 9
Agreement Coefficients Between Taxonomic Classifications
Produced By Different Clustering Methods At The 6-Cluster Level

Partitions Being Compared	Rand's Coefficient	Symmetric Lambda
UPGMA (Cos) vs. HMODE (Cos)	.958	.702
UPGMA (Cos) vs. MIKCA (D)	.946	.694
MIKCA (D, 5 clusters) vs. MIKCA (D,6)	.976	.923

These levels of agreement are encouraging when held against the fact that these methods approach the construction of taxonomic classes in quite different ways, and that different approaches were used in assessing similarity.

The Overall Taxonomic Structure of Adolescent Multiple Drug Use

Since the results of the different taxometric analyses are so similar we will utilize the output of the UPGMA with Cosine similarity coefficient to illustrate the general structure of the taxonomy. The schematic outline of the structure of this adolescent sample is provided by Figure 9. This schematic figure indicates that drug consumption classes range from no drug use through minor (once or twice) alcohol consumption to high frequency polydrug consumption (classes 4 and 7). These classes vary considerably in both number of users), profile of drugs used, and compactness. Tables 10 and 11 provide the numerical information of the relative sizes, drug consumption pattern, and the compactness of the clusters. Homogeneity scores are "average squared deviations" and in these tables indicate the level of compactness or dispersion of each cluster. The largest clusters are those with either no drug use (type 0) and only occasional alcohol use (type 1). The more complex patterns of drug use occur with far lower frequency - see Figure 9.

It should be noted that certain clusters are highly compact, i.e., tobacco only (type 3), occasional minor alcohol use (type 1), and regular alcohol consumers (types 5 and 6). Type 2 - consisting of users of alcohol, tobacco and marijuana, also presents a reasonably large and fairly compact type. Other clusters are much less compact. Much higher deviation from the control profile is found, as might be expected, among the youth who consume many drugs in various quantities. For example, types 4 and 7 are widely scattered in the measurement space yet both clustering methods UPGMA and MIKCA identified type 4 (high frequency polydrug use) and both methods indicate the high dispersion of its class members. Type 7 on the other hand is basically a collection of outliers characterized by diverse patterns of extermely heavy drug use (see descriptions below). It does not form a coherent "type" and the

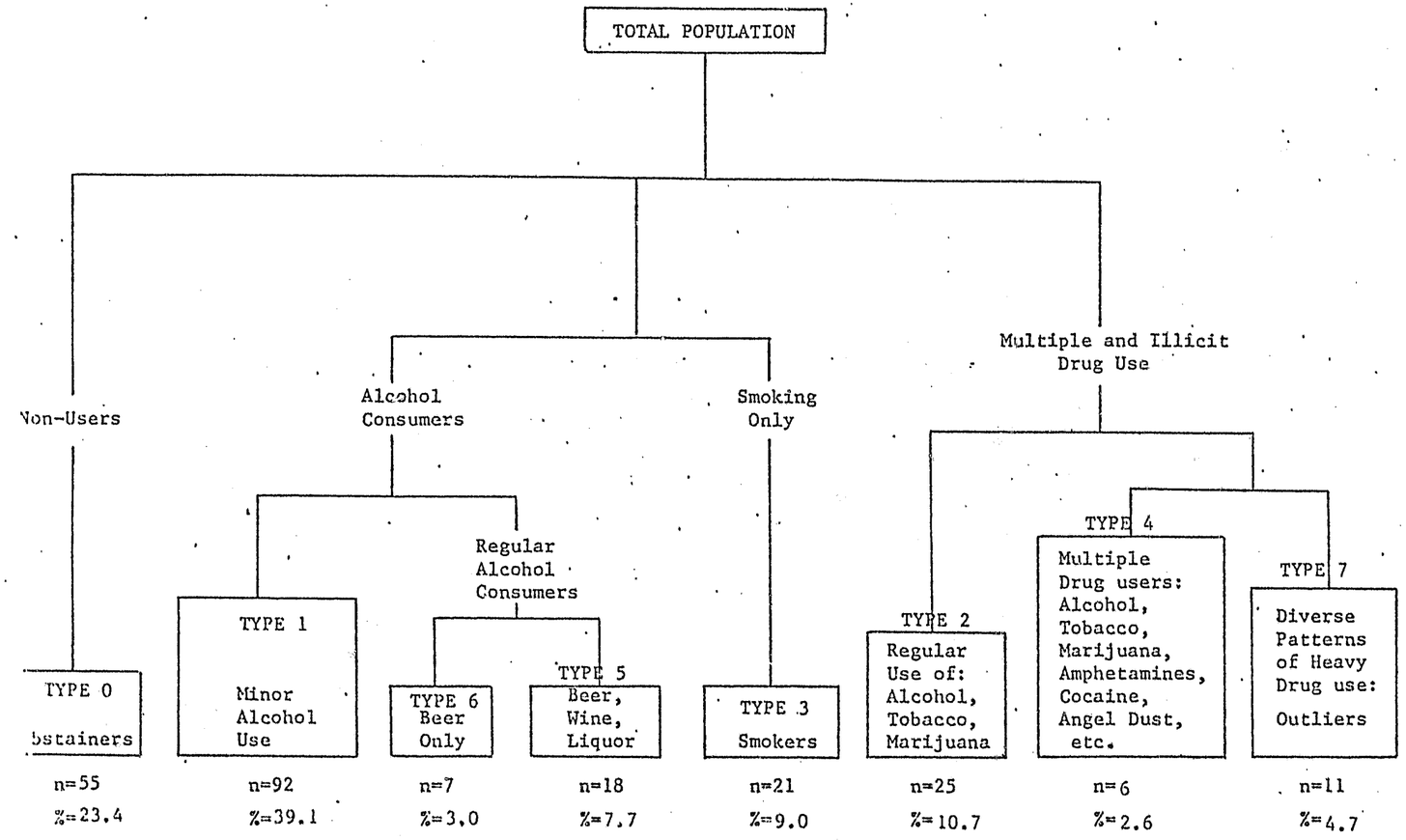


Figure 9

SCHEMATIC TAXONOMIC SYSTEM IMPLIED BY
UPGMA CLUSTER ANALYSIS WITH COSINE SIMILARITY COEFFICIENT

TABLE 10

POLYDRUG CONSUMPTION PATTERNS: STANDARDIZED AND RAW CLUSTER FREQUENCY SCORES:
MIKCA CLUSTER ANALYSIS WITH COSINE SIMILARITY COEFFICIENT

		Cluster Number 1		2		3		4		5		6		
al Sample portions	N=	95		16		31		7		20		9		
	%=													
UBSTANCES		Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Total Score Raw Mean
Beer		-.549	1.989	.919	5.062	-.205	2.710	1.777	6.857	1.391	6.050	.518	4.222	3.139
Wine		-.361	1.705	1.137	3.875	-.380	1.677	1.519	4.429	.706	3.250	.380	2.778	2.228
Hard Liquor		-.488	1.168	.941	3.375	-.033	1.871	2.178	5.286	.471	2.650	.986	3.444	1.922
Tobacco		-.633	1.158	1.590	8.812	1.447	8.323	1.312	7.857	-.679	1.000	-.486	1.667	3.339
Marijuana		-.494	1.095	1.884	6.438	-.364	1.387	2.389	7.571	-.158	1.850	1.738	6.111	2.206
Psychedelics		-.088	1.000	.900	1.375	-.088	1.000	-.088	1.000	-.088	1.000	-.088	1.000	1.033
Tranquilizers		-.116	1.021	.027	1.125	-.145	1.000	2.404	2.857	-.145	1.000	.160	1.222	1.106
Amphetamines		-.197	1.000	-.012	1.188	-.197	1.000	4.302	5.571	-.197	1.000	-.087	1.111	1.200
Barbiturates		-.150	1.000	-.009	1.063	-.150	1.000	3.071	2.429	-.150	1.000	.100	1.111	1.067
Codeine		-.083	1.000	-.083	1.000	-.030	1.032	-.083	1.000	-.083	1.000	-.083	1.000	1.050
Cocaine		-.107	1.000	-.107	1.000	-.107	1.000	2.420	2.571	-.107	1.000	.071	1.111	1.067
Inhalants		-.112	1.011	-.131	1.000	-.014	1.065	.388	1.286	-.040	1.050	-.131	1.000	1.072
Angel Dust		-.199	1.000	.310	1.313	-.199	1.000	3.295	3.143	-.118	1.050	-.018	1.111	1.122
geneity Scores		.947		-.314		.832		-5.735		.770		.476		
. Sq. Deviation		.687		16.011		2.117		75.048		2.838		6.059		

TABLE 11

POLYDRUG CONSUMPTION PATTERNS: STANDARDIZED AND RAW CLUSTER FREQUENCY SCORES:
UPGMA CLUSTER ANALYSIS WITH COSINE SIMILARITY COEFFICIENT

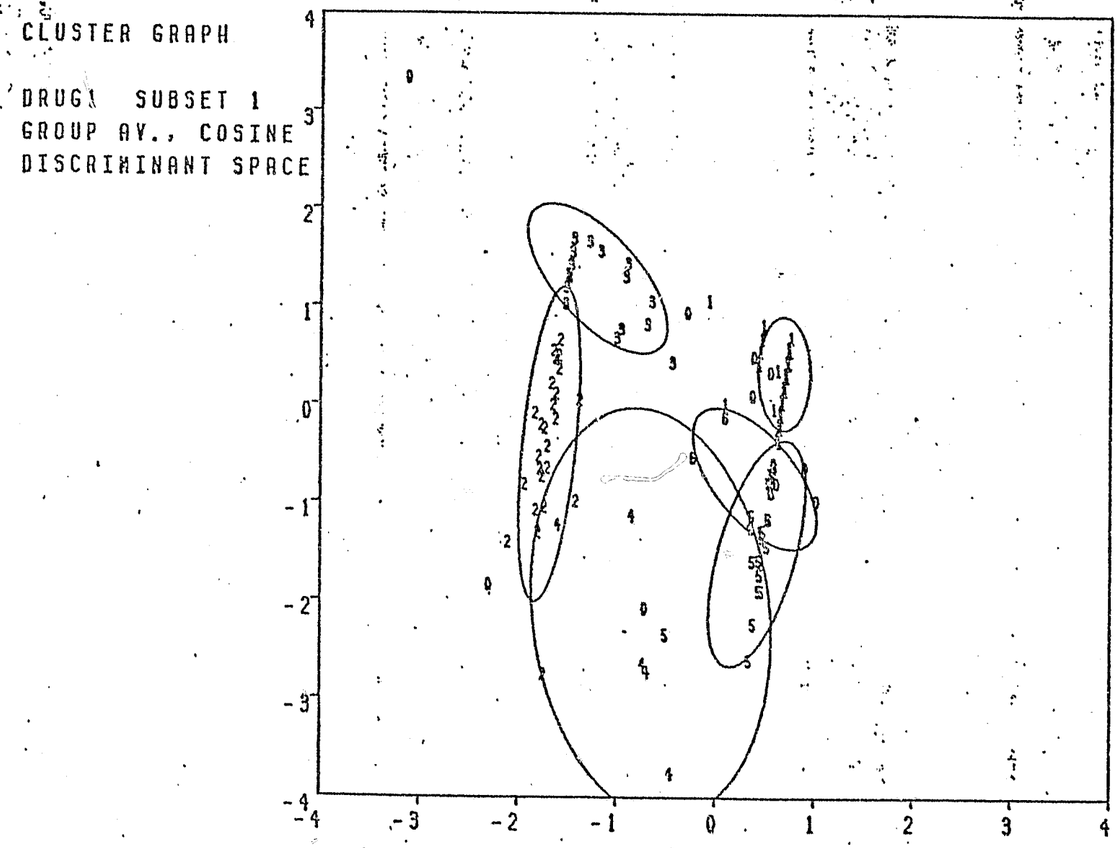
		Cluster Number												
		1		2		3		4		5		6		
Total Sample	N=	92		25		21		6		18		7		
Proportions	%=													
SUBSTANCES		Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Stand. Mean	Raw Mean	Total Score Raw Mean
1. Beer		-.560	1.967	.909	5.040	-.681	1.714	1.447	6.167	1.394	6.056	1.368	6.000	3.139
2. Wine		-.345	1.728	.699	3.240	-.552	1.429	.993	3.667	1.147	3.889	-.650	1.286	2.228
3. Hard Liquor		-.499	1.152	.905	3.320	-.412	1.286	1.777	4.667	.986	3.444	-.505	1.143	1.922
4. Tobacco		-.641	1.130	1.598	8.840	1.367	8.048	1.257	7.667	-.631	1.167	-.472	1.714	3.339
5. Marijuana		-.503	1.076	1.084	4.640	-.473	1.143	2.505	7.833	.230	2.722	-.028	2.143	2.206
6. Psychedelics		-.088	1.000	.018	1.040	-.088	1.000	-.088	1.000	-.088	1.000	-.088	1.000	1.033
7. Tranquilizers		-.115	1.022	-.145	1.000	-.145	1.000	3.516	3.667	-.145	1.000	-.145	1.000	1.106
8. Amphetamines		-.197	1.000	-.039	1.160	-.197	1.000	2.756	4.000	-.197	1.000	-.197	1.000	1.200
9. Barbiturates		-.150	1.000	-.150	1.000	-.150	1.000	3.984	2.833	-.150	1.000	-.150	1.000	1.067
10. Codeine		-.083	1.000	-.083	1.000	-.004	1.048	-.083	1.000	-.083	1.000	-.083	1.000	1.050
11. Cocaine		-.107	1.000	-.107	1.000	-.107	1.000	2.841	2.833	-.107	1.000	-.107	1.000	1.067
12. Inhalants		-.131	1.000	.014	1.080	-.131	1.000	.172	1.167	-.030	1.056	-.131	1.000	1.072
13. Angel Dust		-.199	1.000	.127	1.200	-.199	1.000	4.149	3.667	-.109	1.056	-.199	1.000	1.122
Homogeneity Scores		.952		.498		.950		-5.661		.736		.909		
Ave. Sq. Deviation		.623		6.262		.615		72.165		3.237		1.018		

only common feature is that all of these outliers have a pattern of excessive use of drugs of a particular kind.

To indicate the relativity of the dispersions of these various drug-using classes of youth, we have mapped the sample into the two-dimensional space of the first two discriminant functions. The input variables to the discriminant analysis are the frequency of use scores for the 13 drugs studied in this analysis. Although 5 significant discriminant functions are identified, we utilize only the first two in this graphical presentation. These two discriminant functions account for 89% of the total dispersion in the data. The behavioral meaning of the first two functions emerges from the mapping of the types in the discriminant space.

The horizontal axis (first discriminant function) is defined primarily by the frequency of tobacco use, and clearly serves to discriminate between youths who do not use tobacco (types 0, 1, 6, and 5) and those who do (types 3,2, 4 and 7). This underlines the importance of tobacco consumption as a major structural feature of this taxonomy. The vertical axis has a larger number of high discriminant coefficient loadings (frequencies of use of alcohol, tobacco, marijuana, amphetamines and barbiturates). Thus, it differentiates between those who have only occasional use of single drugs (types 1, 3, and 6) from those using multiple drugs with higher frequencies (types 5, 2 and 4). This axis indicates that the high frequency, multidrug user region of the discriminant space is at the bottom of Figure 10. We notice, here, the confirmation of the extreme dispersion of the subjects classified into type 4. It is also likely that the heavy drug using outliers (type 7) would also be dispersed widely in this region of the space.

FIGURE 10 DISCRIMINANT FUNCTION MAPPING OF THE MULTI-DRUG CLASSIFICATION SYSTEM.



Consumption Patterns of Multiple Drug Use Among the Youth Population

Tables 10 and 11 provide the standardized means and raw frequency scores for the use of the thirteen drugs[3] for the clusters identified by UPGMA and MIKCA methods. The homogeneity of each of these clusters is indicated by the average squared deviation scores and the homogeneity index H, (see Tryon and Bailey, 1970: 162). In each analysis certain recurrent groups appeared with highly similar polydrug use profiles. In addition to the similarity of the central polydrug profiles of these groups, it is also encouraging to find that the relative sizes (N's) and relative dispersions in the multivariate measurement space share the same consistency. The profiles are as follows:

Cluster 1: Minor Alcohol Use (39.1% of Sample). This is the largest group. It is characterized by occasional use of either beer or wine. There is no evidence of the consumption of any other drugs. The HMODE analysis indicated that this cluster may consist of two subtypes - one subtype showing only occasional beer consumption and the other subtype showing only occasional wine consumption. The majority of these youth have consumed beer or wine only once or twice during the last year.

Cluster 2: Alcohol, Tobacco, and Marijuana Use (10.7% of the Sample). The youth falling into this cluster show a complex pattern of consuming beer, wine, and hard liquor, in addition to frequent use of tobacco and marijuana. They show no evidence of consuming psychedelics, tranquilizers, or any of the other drugs in our study. They consume tobacco daily, beer two or three times a month, and wine or hard liquor only four or five times during the year. Marijuana is consumed every two or three weeks (Table 10).

[3]Heroin was omitted from the taxonomic analyses as a result of insufficient numbers of users.

Cluster 3: Smokers (9.0% of the sample). The only feature of this group is the daily use of tobacco. There is minimal evidence of the use of any other drug. The use of tobacco, therefore, stands out as the main defining feature of this class of youth.

Cluster 4: Multiple Drug Consumers (2.6% of the sample). This small cluster consists of youth who respond that they have used virtually all the drugs in the study. This includes frequent use of beer, wine, hard liquor, tranquilizers, amphetamines, barbiturates, angel dust, etc. This cluster is located towards the high and multiple drug use end of this measurement space (see Figure 10). A second important feature is to note that this cluster, although containing few members, is extremely scattered in the measurement space. The polydrug consumption patterns are diverse. This is indicated by the average squared deviation and the very low homogeneity score (see Tables 10 and 11). The sturctural features of this cluster are identical for both UPGMA and MIKCA cluster analysis. The most dominant consumption style is the daily use of marijuana and tobacco, with alcohol being consumed at least once a week. The other drugs are consumed less frequently.

It may be noted that this group contains virtually no use of psychedelics and no use of codeine, and only minor use of inhalants.

Cluster 5: Alcohol Users. This fairly compact cluster is characterized by multiple use of all three alcoholic drinks: beer, wine, and hard liquor, although a few members use a rather small amount of marijuana. It is differentiated from Cluster 2 by the fact that these youth do not smoke. Beer is consumed about once a week, while wine and hard liquor are drunk only occasionally. Similarly, marijuana has been consumed by some members only once every two or three months.

Cluster 6: Beer Drinkers. This small cluster is characterized by the weekly use of alcohol. It differs from groups 2 and 5 in that beer is the only alcoholic beverage consumed by these youth. They show extremely low scores for wine, hard liquor, tobacco, and marijuana. Therefore, there is virtually no use of any of the other hard drugs.

Cluster 7: Diffuse and Severe Drug Use - Multiple Patterns. The sample members falling into this class are characterized by heavy and diverse use of drugs. They are, however, extremely scattered within the high consumption region of the measurement space. No highly recurrent patterns emerged from this collection of respondents. They are dissimilar from each other in terms of the particular range of drugs used, and in the frequency of use. The following patterns are illustrative of this diversity:

1) Amphetamine use, coupled with daily use of alcohol, tobacco and marijuana.
2) Daily use of tobacco and marijuana and weekly use of psychedelics with less frequent use of alcohol and tobacco.
3) Daily use of codeine, with occasional use of alcohol.
4) Weekly use of both marijuana and hard liquor but minimal beer or wine use.

Each of these patterns represent particular styles of polydrug use which is not found with sufficient frequency in this sample to allow the formation of a class or sub type of drug users.

Demographic and Educational Correlates of the Polydrug Types

Establishing the validity of any typology involves examining its correlates with other variables which did not enter into its construction.

This is a large and extensive undertaking, and in the present paper we can only begin this particular task. We have initially correlated the above classes against age, sex, ethnicity, and grade point average. In regard to sex and age, no significant differences between the types were found. However, it was noticed from the contingency table that the large type 1 (minor or occasional use of alcohol) had a majority of younger youth aged 13 and 15. Similarly, there was no significant association between the various drug classes and the ethnic groups. A marginally significant relation was found in relation to grade point average (A,B,C,D) to the drug using classes. In this analysis it was noticed that the types 3 and 5 (i.e., the smokers, and the multiple drug users), had lower grade point averages than the other types. The small number of cases falling into the serious drug users and the outlying group makes any statement of these groups relatively unreliable. The age and grade point average relationships to this typology are consistent with general expectations regarding the age and educational performance of multiple drug users, although the finding with regard to smokers is somewhat surprising.

DISCUSSION

The present Chapter illustrates the use of multivariate taxonomy techniques for the purposes of descriptive exploratory analysis in the context of the epidemiology of crime. Three offense domains were chosen for illustrative purposes: theft, violence and drug consumption. In total, these domains included, respectively, 6, 19, and 13 separate behavioral items. This gives a total of 29 separate behavioral offense measures.

The summarization and 'data-compression' function of taxometric classification is well illustrated by these analyses. Each of the domains is summarized by a single, highly homogeneous (for the most part), and strongly

integrative taxonomy. The separate items are integrated with each other into a fairly well-organized taxonomy. Thus, twenty-nine separate behavioral items can be replaced (or summarized) by only three taxonomies. We concede that additional measures of the accuracy of the behavioral taxonomies could have been used. Our intention, however, in the present Chapter, was clearly exploratory, and we did not utilize the full range of evaluative techniques. These are examined in other chapters of this report. We would argue, however, that the organization of data and the descriptive summarization of complex multivariate domains, as illustrated in the present Chapter, represents a use of taxometric methods which may have extreme utility for epidemiological descriptive purposes.

Regarding the function of exploratory analysis to discover 'new' patterns, the present analysis has also been informative. The precise combinations of particular drugs, or theft or violence patterns, that adolescents follow appear to have been 'discovered' by the taxometric methods. Admittedly, the knowledgeable practitioners may make an educated guess at the prevailing patterns. On the basis of his/her experience, such a practitioner might come up with an intuitively appealing set of profiles. However, the present Chapter indicates that the taxonomic methods represent a useful alternative approach for this task of discovery.

In terms of ranking the various approaches utilized, we have relied on the work described in other chapters for the artificial data sets. In those analyses, there was, in fact, a 'known correct' solution. The various analytic approaches could be compared according to their kinds and levels of deviation from the correct solution. In the present instance, no such attempt at ranking

was conducted. We are more concerned with the commonality of results and the consistency of description. These results, particularly in the case of the drug-profile analysis, show a striking similarity to each other. The symmetric Lambda scores obtained for the drug patterns were highly encouraging. Yet, it is also clear that cluster analytic methods with different cluster-definitions similarity measures and linking-principles will tend to find different partitions in data. This is most clearly illustrated at the sub-class level within the present analyses. The boundaries between classes, and the memberships of cases within classes are not identical. However, an examination of the accumulated results indicates substantial agreement between the various analyses. Surprisingly, the impact of the different similarily coefficients was again probably more drastic than that of the grouping method.

Thus, the two prior considerations of the epidemiological specialist regarding the logical nature and definition of a 'class' and of 'similarity' are again of initial importance for the use of classification methods. Our present results have the effect of forcing us to the conclusion that the numerical taxometric methods can incorporate many different definitions of 'class' and 'similarity' and these will often recover different features of information from epidemiological data. It is essential that the definitions of these key concepts be consistent with those of the epidemiological specialist. Yet, in our review of criminological literature, no clear discussions of these concepts exists. The clarification of these concepts for use in epidemiology is therefore a critical task if taxometric methods are to be used with optimally good results and an optimal clarity of understanding of the results that are obtained. The emergence of numerical methods for classification may force the substantive specialist to re-examine and clarify the meanings of the

definitions of 'class' and 'similarity' if they wish to make the best use of these methods.

BIBLIOGRAPHY

Ageton, S.S., and D.S. Elliott
1978 "The effects of legal processing on delinquency proneness." Institute of Behavioral Science, University of Colorado: Boulder.

Clark, J.P., and E.W. Haurek
1966 "Age and sex-roles of adolescents and their involvement in misconduct: A reappraisal." Sociology and Social Research, 50, 495-508.

Clark, J.P., and L.L. Tifft
1966 "Socio-economic class and area and correlates of illegal behavior among juveniles." American Sociological Review, 31, 516-23.

Elliott, D.S., and B.A. Knowles
1978 "Social development and employment." In Employment statistics and youth. U.S. Government Printing Office: Washington, D.C.

Elliott, D.S., S.S. Ageton, M. Hunter and B.A. Knowles
1976 "Research handbook for community planning and feedback instruments." Behavioral Research Institute: Boulder, Colorado.

Empey, L.T.
1978 American delinquency. The Dorsey Press: Homewood, Illinois.

Erickson, M.L.
1972 "The changing relationship between official and self-reported measures of delinquency: An exploratory predictive study." Journal of Criminal Law, Criminology, and Police Science, 63(3), 388-95.

Erickson, M.L., and L.T. Empey
1963 "Court-records, undetected delinquency and decision-making." Journal of Criminal Law, Criminology, and Police Science, 54, 456-69.

Everitt, B.
1974 Cluster analysis. Halsted Press: New York.

Gold, M.
1966 "Undetected delinquent behavior." Journal of Research in Crime and Delinquency, 3, 27-46.

Gold, M.
1970. Delinquent behavior in an American city. Brooks/Cole Publishing Co.: Belmont, California.

Gold, M., and J.R. Williams
1969 "The national survey of youth - search of official records." Unpublished monograph. Institute for Social Research: Ann Arbor, Michigan.

Gould, L.C.
1969 "Who defines delinquency: A comparison of self-reported and officially-reported indices of delinquency for three racial groups." Social Problems, 16, 325-36.

Hartigan, J.A.
1975 Clustering algorithms. John Wiley and Sons: New York.

Hindelang, M.J., T. Hirschi, and J.G. Weiss
1975 "Self-reported delinquency: Method and substance. Proposal submitted to National Institute of Mental Health.

Hood, R., and R. Sparks
1970 Key issues in criminology. McGraw-Hill: New York.

Huizinga, D.
1978 "Are there any clusters?" Technical Report 78-2. Behavioral Research Institute: Boulder, Colorado.

Jessor, R., et al.
1968 "Satisfaction and drinking patterns of college students." Quarterly Journal of Studies on Alcohol, 29, 101-16.

Kulik, J.A., K.B. Stein, and T.R. Sarbin
1968 "Dimensions and patterns of adolescent anti-social behavior." Journal of Consulting and Clinical Psychology, 32, 375-82.

Nettler, G.
1974 Explaining crime. McGraw-Hill Book Co.: New York.

Short, J.F., Jr., and F.I. Nye
1957 "Reported behavior as a crierion of deviant behavior." Social Problems, 5, 207-13.

Swanson, R.M., and M.J. Mobley
1976 "Patterns of delinquent behavior: Dimensions of self-reported behaviors." LEAA: Washington, D.C.

Tryon, R.C., and D.E. Bailey
1970 Cluster analysis. McGraw-Hill Book Co.: New York.

Voss, H.L.
1963 "Ethnic differentials in delinquency in Honolulu." Journal of Criminal Law, Criminology, and Police Science, 54, 322-27.

Elliot, D.S., and S.S. Ageton
1980 "Reconciling race and class differences in self-reported and and official estimates of delinquency." American Sociological Review, 45(1), 95-110.

CHAPTER 12

THE I-LEVEL CLASSIFICATION: AN EVALUATION BY TAXOMETRIC METHODS

INTRODUCTION

The I-level classification system (Sullivan, Grant and Grant, 1957) is probably one of the most widely used, best-known, and controversial classification systems. It is used for the theoretical explanation of delinquent behavior, as well as the classification and treatment of delinquent youth. The system seems to be well-grounded in the theories of ego-development, personality structure (Sullivan, 1940), and socialization (Sarbin, 1952). Loevinger (1976) describes the I-level system as a developmental characterology. Each of the major types is seen as manifesting increased levels of psychological development in general, and interpersonal maturity in particular. Development proceeds in a direction of increasing involvement with others, greater perceptual and cognitive discrimination, perceptual accuracy, and personal effectiveness. Each stage exhibits, and is partially defined by, a "core" problem. Fixations at particular stages are seen as being possible. Extensive descriptions of the I-level stages and subtypes are provided in a variety of sources (see Loevinger, 1976; Pauker and Hood, 1979; etc.). An abbreviated description of the system is given below.

Basic I-level	Subtype Code	Subtype Name	Typical Responses to Frustration, Problems
I_2 1	Aa	Asocial, aggressive	Responds with active demands and open hostility when frustrated
2	Ap	Asocial, passive	Responds with whining, complaining, and withdrawing when frustrated
I_3 3	Cfm	Passive conformist	Responds with immediate compliance to authority or persons in power
4	Cfc	Cultural conformist	Conforms to specific reference group

Basic I-level	Subtype Code	Subtype Name	Typical Responses to Frustration, Problems
5	Mp	Manipulator	Tries to undermine the power of authority figures and/or usurp the power role for themselves
I_4 6	Na	Neurotic, acting out	Responds with condemnation of self and tries to outrun or avoid conscious anxiety
7	Nx	Neurotic, anxious	Responds with symptoms of emotional disturbance to conflicts stemming from feelings of inadequacy and guilt
8	Se	Situational, emotional reaction	Responds to crisis by acting out
9	Ci	Cultural identifier	Responds to identification with a deviant value system by living out his/her delinquent beliefs

Whereas the basic I-level stages were theoretically derived and specified, the subtype levels were empirically derived (Pauker and Hood, 1979). The following basic questions are addressed by the present work.

The Use of Cluster Analytic Methods for Theory Testing and Confirmatory Analysis

Model-testing has been mentioned as one of the major potential uses of taxometric methods (Ball, 1970; Brennan, 1972; and others). A well-defined typological model--defined in terms of class stereotype and class boundaries--can be compared with the natural structures emerging from the data. Various "goodness-of-fit' coefficients and procedures can be used to assess the adequacy of the proposed theoretical model in regard to its consistency with the real data. The present study follows this model of comparing a proposed or theoretical taxonomic model with real data.

The Jesness Instrument versus the Interview Method of Reaching an I-level Designation

The Jesness psychometric procedure and the interview method represent alternative procedures for reaching an I-level designation for a youth. There is some debate on the relative merits of the two procedures. The psychometric approach appears to be more reliable, more efficient, more objective, and cheaper (see the reviews in Pauker and Hood, 1979). It may be gradually displacing the interview method as the most commonly used approach in reaching an I-level designation. Palmer (1980 - personal communication) suggests that, by and large, the two approaches give comparable results, with a small tendency on the part of the psychometric procedure to give lower I-level designations. This would result in a small percentage of any sample of youth falling into lower I-level classes with the psychometric procedure.

It is to be stressed that our present examination pertains purely to the I-level designations of the psychometric procedure and not to those of the interview approach. Any generalizations from the present results to I-levels obtained from interviews would be critically dependent on the isomorphism between the psychometric and interview methods of obtaining the I-level classification.

Identification of the "Natural Typological Structure" within Early Adolescence as Assessed by the Jesness Dimensions

Although the Jesness dimensions are utilized to identify the I-level type designations of youth, it is quite possible that an alternative typological structure may lie embedded within this complex multi-dimensional personality assessment domain. Our most basic question in this research was whether such a structure exists, and, if so, how would it compare with the I-level designations that were generated for the same youth, from the same

dimensions. If the I-level classes are real, then we hypothesize that they should be discovered by the empirical classification methods. If they constitute a weak, nonexistent (artifactual), or purely theoretical kind of structure, then they may show only very weak (or nonexistent) boundaries, weak internal homogeneity, and may have no overlap with the more "natural" boundaries that are contained within the data. These questions are addressed in the present research.

The "Objective Reality" of the Types: Do the "Types" Exist?

Questions have been raised regarding whether the types and subtypes within the I-level system actually exist, or whether they simply represent theoretical concepts with little objective basis in reality. An early study raising this question was that of Butler and Adams (1966). A most interesting finding stemmed from the attempt by Butler and Adams to replicate the I-level classes on a sample of delinquent girls by conducting Q-factor analysis on the Jesness dimensions. Their empirical typology was compared to the I-levels (obtained by the interview method), with the surprising finding that no clear relationship existed. Similarly, Keck (1978) as reported by Pauker and Hood (1979) found no relationship between I-level (interview method) and the Jesness scale scores, or the MMPI scales. Other attempts to establish concurrent validity for the I-level system (assessed either by the Interview or the Jesness dimensions) by specifically cross-classifying it against other typological systems have met with very mixed success. Carbonell (1980) found no relation between Megargee's MMPI types and the I-level (Jesness method); Smith (1974) found only a weak relation between the I-level (Jesness method and Eysenck's personality typology. Werner's (1975) typology based on the CPI had little overlap with I-level. The implications of these findings are that

the I-level classes established by the Jesness method have minimal or weak homogeneity on many basic personality dimensions (i.e., the MMPI, Eysenck's dimensions, the Quay system). Another implication is that these I-level classes also may have relatively low overlap with I-level as established by the Interview method.

Cross-validation against relevant continuous variables, on the other hand, provides more encouraging support for the I-level. Reviews by Pauker and Hood (1979), Jesness (1974), Palmer (1977), and others indicate that the I-level has significant and meaningful correlations with many variables such as I.Q., internal-external control, socioeconomic status, race, etc.. While the set of cumulative research is encouraging in regard to I-level as a continuous "dimension" of interpersonal development, the earlier weak associations with other typologies raises questions regarding the appropriate "number" of types, the boundary conditions, and the levels of internal homogeneity of these types.

Boundaries and Internal Homogeneities:
How Well Structured is the I-level Typology?

Related to the above question of the "reality" of the I-level types is the question of establishing the strength of the boundary conditions and internal homogeneities of these types. The reality or existence of any set of types is largely governed by the presence of discontinuities, or strong boundaries, for the types. This is inevitably coupled with certain atypicalities on a subset of type-attributes on which each type is homogeneous. To date, there has not appeared in the literature on the I-level classification an explicit examination of the nature of the boundary conditions and internal homogeneities of the system. The present chapter will provide an initial examination of these issues.

The Relation of I-level to Delinquent Behavior: What is this Relationship?

Kornfeld et al. (1975) raise the question of the exact relation between I-level and delinquency. These reviewers criticize most of the prior research for using only samples of delinquents and incarcerated youth. They point out that very few studies were available which examined broader samples of both delinquent and nondelinquent youth. The present study uses both delinquent and nondelinquent youth and can be used as a vehicle to examine new aspects of this issue. Specifically we will examine:

1. How the I-level system (Jesness method relates to delinquent behavior;
2. How strong is the linear relation between the Jesness dimensions (taken separately, and then in a linear composite via multiple regression) and delinquent behavior; and
3. How strong is the relation between the natural typological structure of the Jesness data and delinquent behavior.

METHODS

In the initial stage of examining the typological structures within the Junior High School sample, seven different approaches to clustering were used. The two hierarchical methods UPGMA and HMODE were used with three separate similarity coefficients: Euclidean distance, cosine, and the correlation coefficient. In addition we also used the NMODE method, which is confined to the Euclidean distance metric. The first six of these analyses (i.e. the UPGMA and HMODE analyses) provided hierarchical trees of the taxonomic structure of the data.

In a second stage of our examination we utilized iterative relocation clustering (the MIKCA program), with minimum det W as the clustering

criterion. Variations in the use of the MIKCA program stemmed from the selection of different initial seed points (see Everitt 1974) and different k-levels (i.e. numbers of clusters). The particular starting solutions are indicated on the appropriate tables in the results section.

DATA

Fourteen hundred junior high school youth in California were tested with the Jesness inventory during 1978. This data was kindly made available to the present project by Dr. Carl Jesness. Each of the youth had already been classified into an I-level status using the discriminant function procedure (see Moloff & Jesness 1974).

Since the agglomerative clustering methods cannot deal with sample sizes greater than 200 we selected a 15% random sample from the total junior high school group. This sample contained approximately equal numbers of male and female youth (48% and 52% respectively). In regard to age, 53% of the youth were 13, 42% were 14, and the remaining few were 12 or 15 years old. The following table indicates the breakdown of the sample according to I-levels:

General I-level Stage	Sub-types Within I-level	
I-level 2	1. unsocialized aggressive	1.5%
	2. unsocialized passive	2%
I-level 3	3. immature conformist	7.5%
	4. cultural conformist	15.1%
	5. manipulator	9.5%
I-level 4	6. neurotic acting out	45.7%
	7. neurotic anxious	10.6%
	8. situational emotional	6.0%
	9. cultural identifier	2.0%

The above assignments to the stages and substages of the I-level indicate that the categories of cultural conformist (15%), neurotic acting out (45%) and neurotic anxious (10.6%) were the most prominant sub-types within this adolescent sample. To look at the more general I-level categories, we also formed a simplified version of the Jesness I-level assignation consisting of 4 general categories. This collapsing conformed to the more basic I-level groupings as indicated by the earlier literature (See Loevinger 1976). The first two categories (i.e. unsocialized aggressive and unsocialized passive) were collapsed into the unsocialized group i.e. the I-2 level. The I-3 maturity level was then formed by collapsing the immature conformists (3), the cultural conformists (4), and the manipulators (5), into one general class. Finally, we formed two sub-groups at the I-4 level, joining the neurotic acting out (6) and neurotic anxious (7) into one group, and collapsing the 'situational emotional' and 'cultural identifiers' into another. Thus, these four groups correspond to the levels I-2, I-3, I-4, levels with two sub-groups at the I-4 stage.

The data sample also contains a self-reported delinquency score. This scale represents the sum of fourteen delinquent behaviors committed during a one-year period. Each item is scored from 1 to 5 using simple frequency response format. The delinquent acts included such behaviors as: fighting, vandalism, threats, various types of theft, drug and alcohol use, truancy, and running away from home. The total delinquency scale scores within this data set, therefore, ranged from 14 to 70. On the basis of a histogram of these scores we established the following four classes of deliquency:

Delinquency Level	Delinquency Scale Score	Number of Cases in the Analysis Sample	Sample %
1 = Non-delinquents	14 - 20	97	48.5
2 = Minor delinquency	21 - 27	49	24.5
3 = Multiple delinquency	28 - 34	25	12.5
4 = Frequent and multiple delinquency	35 - 41	16	8.0
5 = Frequent and serious delinquency	42 - 70	12	6.0

It can be seen, therefore, that there are 97 youth in this subsample who exhibit virtually no delinquent behavior while there are 12 youth who are very highly delinquent.

The Measurement Space: The Jesness Inventory

The Jesness Inventory is a personality-attitude test which was developed by Carl F. Jesness primarily for the purpose of distinguishing delinquent from nondelinquent youths, and for differentiating among delinquents of various types. It consists of 155 true or false items from which, through the use of cluster and factor analysis, eleven scales were derived (including an indicator of delinquency proneness, the asocial index). The scales are as follows.

1. Social Maladjustment. This scale delineates a set of attitudes associated with "inadequate or disturbed socialization." It measures the extent to which a person shares the attitudes of those "who demonstrate inability to meet environmental demands." Use of this scale reveals that delinquents tend to show a negative self-concept, feel unhappy, feel worried and misunderstood, distrust authority, blame others for their problems, are bothered by feelings of hostility, "generously evaluate parents," are sensitive to criticism, and "accept much behavior regarded as antisocial."

2. Value Orientation. This measures the tendency to share the attitudes and opinions of "persons in the lower socioeconomic classes.". Items from this scale selected for use were those which separated social class groupings based on ratings for the father's occupation. Within the context of this scale, it seems that delinquents look for trouble, luck, and thrills, are characterized by fear and failure, a gang orientation, toughness, and a tendency to "view internal tension in terms of specific symptoms."

3. Immaturity. Here the emphasis is on a displayed tendency to possess attitudes which would be considered usual for persons of a younger age. Items were chosen that discriminated between different age groups in the non-delinquent sample. Items suggested that some delinquents are naive about expressing motivations, tend to suppress problems, lack insight, and express anxiety through somatic symptoms.

4. Autism. This scale measures the tendency to "distort reality" in order to accord with personal desires and needs. The self in this case does not seem to be clearly distinguished from the nonself or objective reality. Those who score high on this scale tend to see themselves as smart, good looking, and tough, to hear things, daydream, and feel something is wrong with their minds. They also prefer to be alone, are fearful, and "express many somatic complaints."

4. Alienation. This scale measures the degree of distrust and estrangement a person displays in attitudes toward others (especially others representing authority). It has been found that delinquents score higher on this scale than nondelinquents at virtually every age level. High scorers tend to have poor interpersonal relationships, to be critical and intolerant of others, project hostile feelings onto others, and deny personal problems.

CONTINUED

6 OF 9

6. Manifest Aggression. The focus of this scale is on the acute awareness of unpleasant feelings (particularly anger and frustration), the tendency to react quickly with emotion, and "perceived discomfort concerning the presence and control of these feelings." The high scorer is typically concerned with controlling feelings, expresses disappointment with others, and is frustrated in efforts to "understand and feel comfortable with himself."

7. Withdrawal-Depression. This scale measures a perceived lack of satisfaction with self and others, along with a tendency toward isolation. High scorers feel depressed, sad, misunderstood, prefer to be alone, feel that "fighting is bad," and are displeased by displays of aggressiveness.

8. Social Anxiety. This scale deals with a perceived emotional discomfort associated with interpersonal relationships. High scorers feel nervous tension, see themselves as sensitive to criticism, and are unduly shy.

9. Repression. Measures the suppression of feelings and emotion which, under normal circumstances, would be displayed, or the failure to label these emotions. High scorers do not admit the presence of negative feelings, and are generally not critical of themselves and others.

10. Denial. Measures the reluctance to acknowledge unpleasant events or aspects of reality "often encountered in daily living." This scale emerged as a separate cluster from the repression scale, and does not correlate highly with it. High scorers see their parents as without fault and admit no conflict with them, deny personal inadequacies or unhappiness, and are unwilling to criticize others. Low scores on this scale indicate a willingness to admit problems, conflicts, and inadequacies.

11. Asocial Index. This special scale measures the extent to which personal or social problems tend to be solved without regard for social customs or rules.

RESULTS

General Taxonomic Structure of the Sample

The natural mode-seeking method (NMODE) with a dense point radius of 2.5 detected 6 modes in the analysis sample. This was a clear suggestion of cluster-like structure in the data. It was discovered, however, that one of the modes was very large (112 points) while the other modes were small. It was also found that 40 points were left as outliers.

With a smaller dense point radius (R=1.66), NMODE detected only 2 small natural modes. Surprisingly, both of these fell within the large 112 point cluster that had been discovered in the first analysis. This suggested that the large dense cluster might fruitfully be broken down into at least 2 separate subtypes.

The hierarchical methods also provided evidence of typological structure. All of the dendrograms produced by the UPGMA analyses exhibited long branch lines in the later stages of the tree, again suggesting the existence of clusters. Prior to any interpretation of these clusters, however, we examined their stability in the data.

Table 1 examines the stability or overlap between the clusters which emerged at various levels of the tree diagrams for the various hierarchical methods. The agreement between UPGMA and HMODE (both using cosine) reaches an encouraging level (symmetric lambda = 0.66). When Euclidean distance is used lambda falls below 0.5. This agreement between the two methods when utilizing distance is, however, misleadingly low. This stems partially from the fact that there is a large discrepancy in numbers of outliers between the two analyses. This analysis, however, also provides confirmation that some relatively stable clusters exist within a 'points-in-space' model of

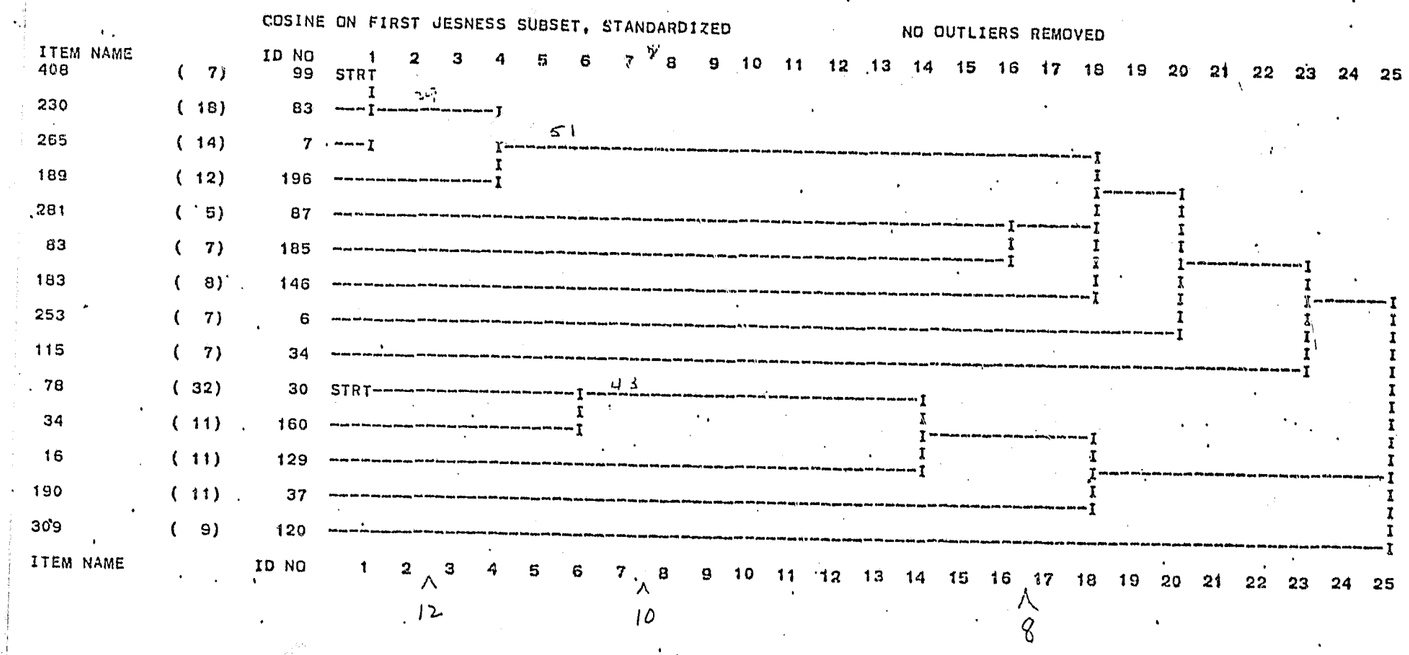
TABLE 1

Measures of Partition Agreement for JES1
Comparing UPGMA and HMODE Partitions

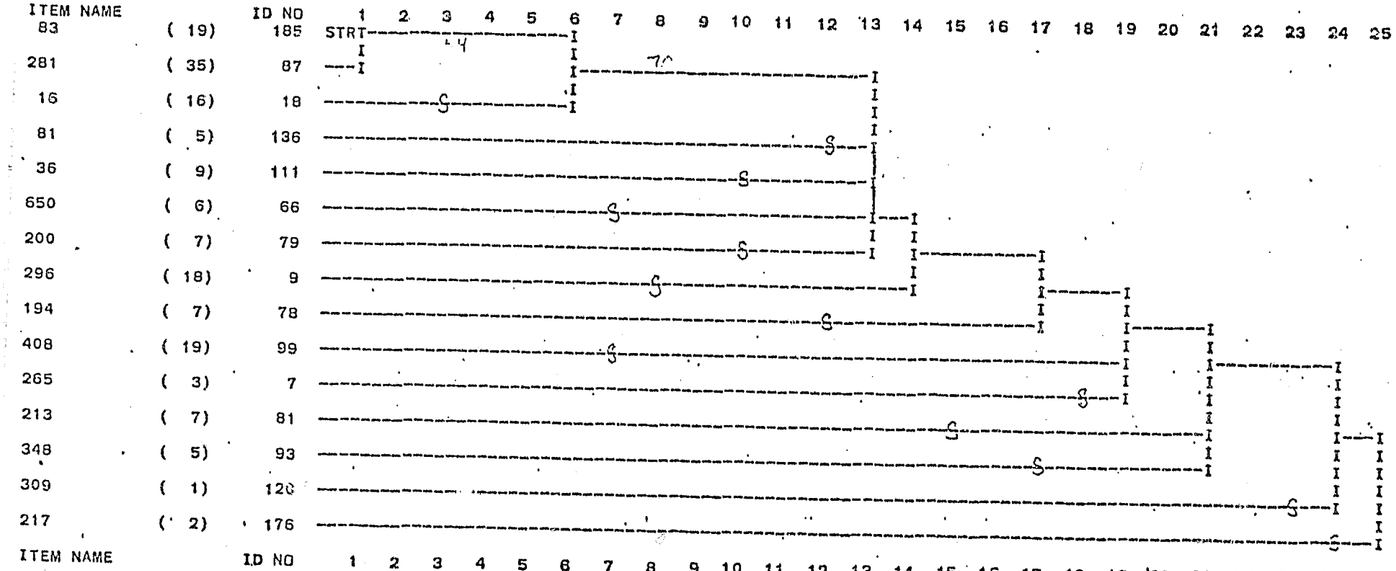
Partitions (same sim. coef.)	# Clusters*	# Outliers	Rand	Jaccard	Lambda
UPGMA-d (190) vs. HMODE-d (6)	7x6	4, 46	.787	.457	.463
UPGMA-d (190) vs. HMODE-d (6)	4x5	18, 51	.786	.468	.482
UPGMA-cos (190) vs. HMODE-cos (6)	8x8	3, 40	.882	.527	.664
UPGMA-cos (193) vs. HMODE-cos (6)	6x8	0, 40	.844	.476	.616
UPGMA-r (191) vs. HMODE-r (6)	8x6	0, 42	.742	.258	.385
Partitions (different sim. coef.)					
UPGMA-d (190) vs. UPGMA-cos (193)	4x5	18, 6	.731	.321	.466
UPGMA-cos (190) vs. UPGMA-r (192)	9x7	0, 0	.813	.321	.517

*The number of clusters at a particular tree-level depends in part on the minimum number of points required to make a cluster. Smaller groups are thrown into the "outlier" category. In this table, where the number of groups for a given partition changes it is because the minimum cluster size was changed. (Treatment of such small numbers of points has little impact on the agreement coefficients.)

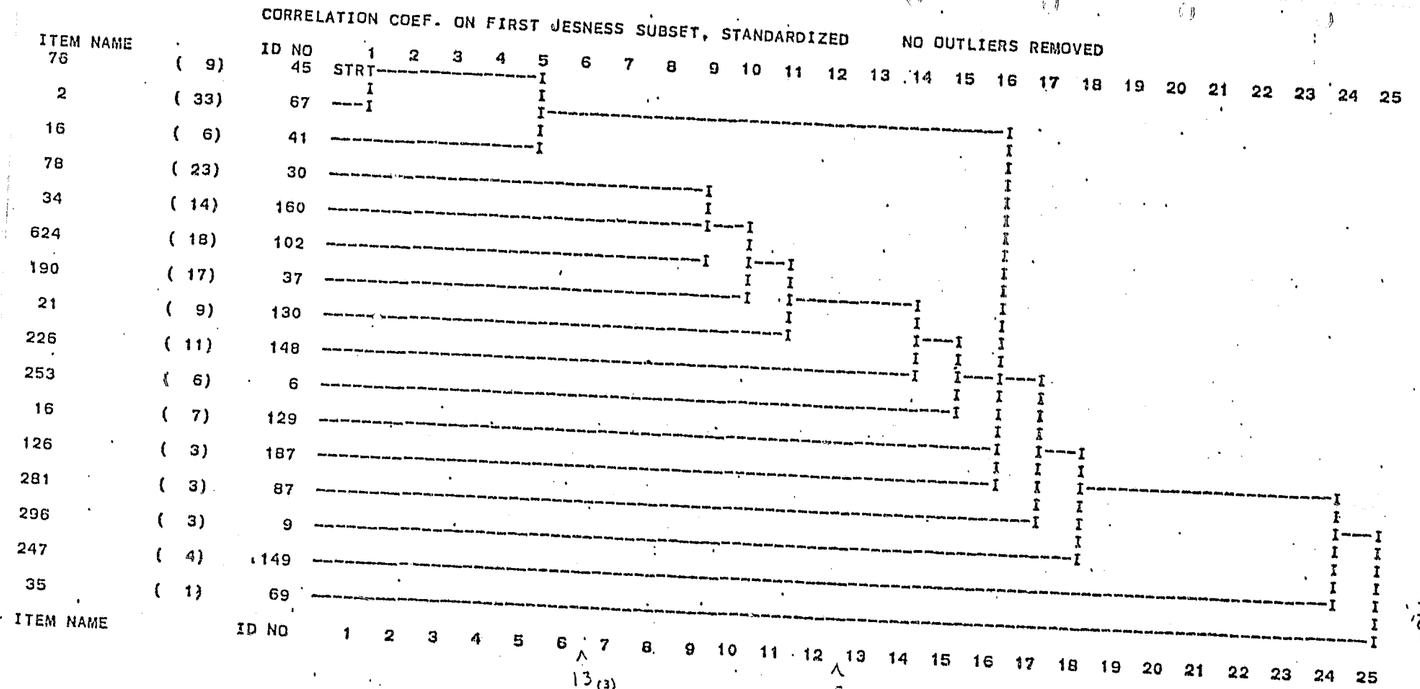
COSINE ON FIRST JESNESS SUBSET, STANDARDIZED — NO OUTLIERS REMOVED



EUCLIDEAN D ON FIRST JESNESS SUBSET, STANDARDIZED NO OUTLIERS REMOVED



CORRELATION COEF. ON FIRST JESNESS SUBSET, STANDARDIZED NO OUTLIERS REMOVED

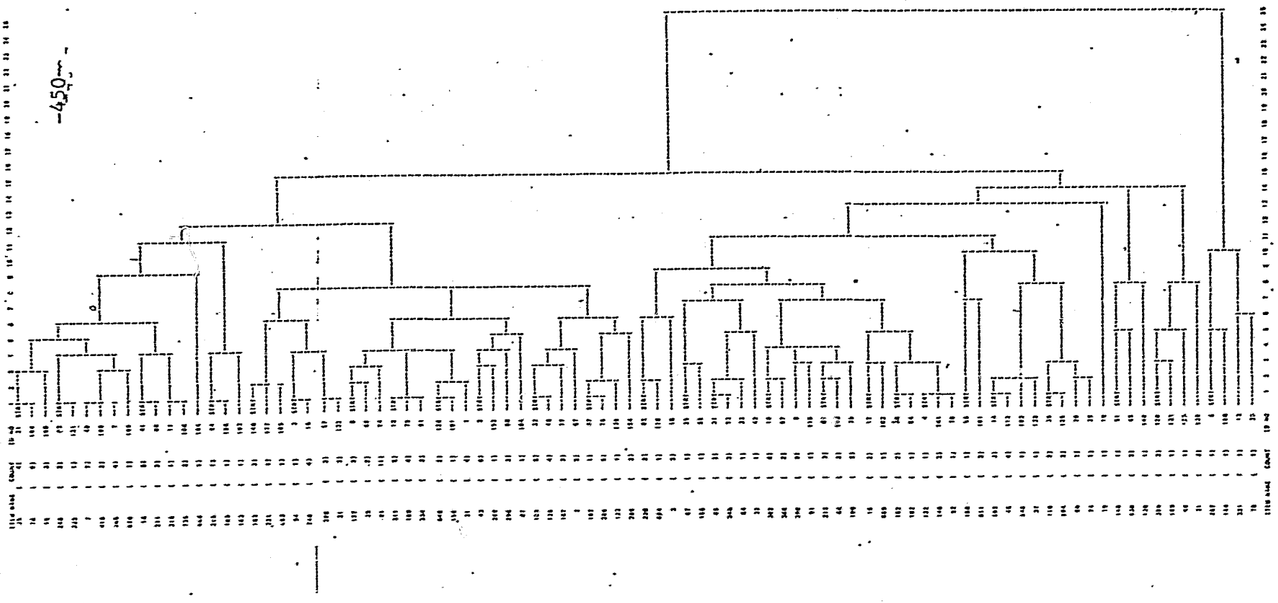


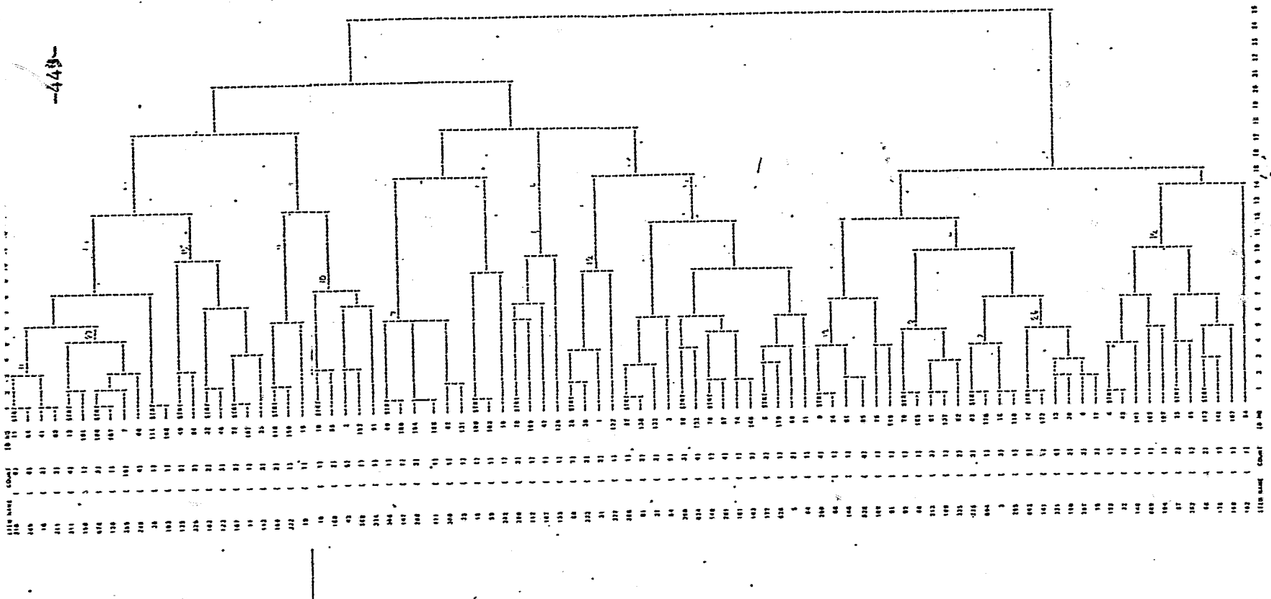
13 (3)

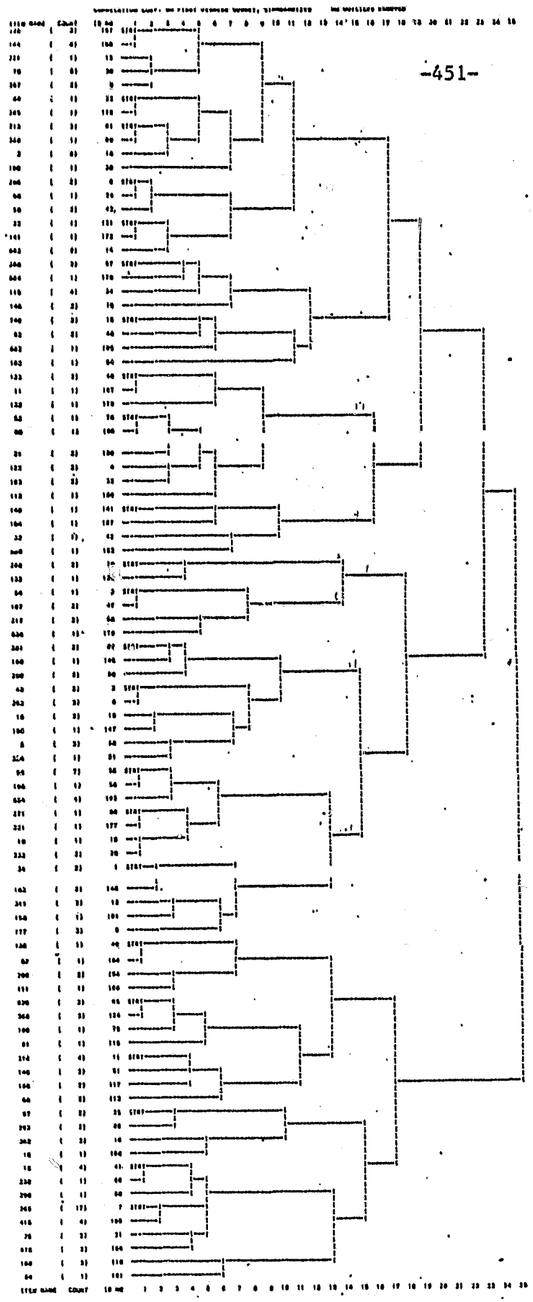
4 (5)

9 (3)

5 (5)


450


449



similarity. The worst stability however, is found when the correlation coefficient is used. When the Q-correlation is used with these two methods symetric Lambda is only .39. This suggests that the partitions based only on profile shape are unstable.

Alternative Classification Structures Using Iterative Relocation Methods

Additional analyses were then conducted on the same data using the K-means method (MIKCA) of clustering. We experimented with different starting 'seed points' in the MIKCA analysis. Specifically we utilized the cluster centroids which emanated from the following: UPGMA-d with 4 and 7 clusters, UPGMA-cos with 5 and 7 clusters. We then examined the partitions which resulted from utilizing random starts for the MIKCA analysis. This aspect of the study serves both as a methodological comparison between the hierarchical methods and the K-means method, and as an attempt to refine the hierarchical partitions. Basically, 11 separate partitions were created with the MIKCA method, ranging from 4 to 7 groups. To reduce the risk of obtaining local minimum solutions we utilized several sources of starting centroids. The starting configurations and the clustering criterion values for the resulting partitions are show in Table 1(a).

In evaluating this table it should be noted that the criterion values can only be compared for those partitions which contain the same number of classes. With a random start it might be noted that with more clusters (i.e. from 4 through 5, 6, and 7 clusters) there is a steady lowering (or improvement) of the criterion value.

A second point to notice is that in only one case did the imposed starting centroids produce a lower criterion value than the random starts (partition No. 3 at the 4 group level). It appears that the iterative relocation tends

Table 1(a)

MIKCA Analyses of Jesness Sample: Starting Solutions, K-levels, and Resulting Criterion Values

MIKCA Partition #	# of Groups	Source of Starting Centroids	Criterion (det W)
1	4	Random start	$.96 \times 10^{21}$
2	4	Centroids of first 4 SRD categories	$.99 \times 10^{21}$
3	4	Centroids of 4 I-level categories	$.95 \times 10^{21}$
4	4	Centroids from UPGMA-d level 190	1.01×10^{21}
5	5	Random start	$.75 \times 10^{21}$
6	5	Centroids of 5 SRD categories	$.83 \times 10^{21}$
7	5	Centroids from UPGMA-cos level 193	$.86 \times 10^{21}$
8	6	Random start	$.64 \times 10^{21}$
9	7	Random start	$.530 \times 10^{21}$
10	7	Centroids from UPGMA-d level 180	$.58 \times 10^{21}$
11	7	Centroids from UPGMA-cos	$.534 \times 10^{21}$

to converge upon a local minimum which is close to the input starting solution. The agreement level between a cluster analysis partition that serves as a source of starting centroids and the resulting MIKCA partition in these instances is relatively high. Examples of this can be seen in Table 1(c) with MIKCA partitions 4, 7. 10, and 11, all of which produced high values of the Lambda and Rand coefficients when compared with the partition that supplied their starting centroids.

By the MIKCA clustering criterion, then, the random start solutions were generally better than the others. The best of the MIKCA solutions are compared in Table 1(d) and show moderately high agreement levels between partitions. Thus we are encouraged to assume the existence of certain similarities of "membership" between these classifications. We emphasize,

Table 1(c)

Agreement between Source of Starting Centroids and Resulting MIKCA Partitions

MIKCA Partition #	Source of Starting Centroids	# groups	Rand	Jaccard	Lambda
2	Delinquency levels	4x5	.638	.233	.185
3	I-level Categories	4x4	.618	.283	.303
4	UPGMA-d level 190	4x4	.808	.477	.652
6	Delinquency levels	5x5	.666	.245	.165
7	UPGMA-cos level 193	5x5	.824	.445	.627
10	UPGMA-d level 180	7x7	.873	.410	.637
11	UPGMA-cos level 190	7x7	.840	.364	.596

Table 1(d)

Agreement Between the MIKCA Partitions which had the Best Criterion Values

MIKCA Partitions	# groups	Rand	Jaccard	Lambda
1 vs. 3	4x4	.789	.414	.626
3 vs. 5	4x5	.754	.307	.452
5 vs. 8	5x6	.818	.350	.563
5 vs. 9	5x7	.834	.364	.587
8 vs. 9	6x7	.822	.286	.508
9 vs. 11	7x7	.837	.293	.509

however, that these symmetric Lambda values, while indicating some general stabilities, do not by any means indicate that exactly the same clusters have emerged from different sets of analyses (the details of this can be examined in Table 2).

The Lambda agreement coefficients were much lower when comparing cluster analytically generated classifications with the I-level classifications provided by Jesness, or with the classification of delinquent behavior as provided by the Jesness data. This indicates that the cluster analytic centroids have converged upon structures that are much different from these external partitionings of the data. This is one of the major substantive findings of the present study.

Refinement of a "Best" Partition

The analyses so far suggest that between 5 and 7 core clusters exist in this data set (when the minimum cluster size is set at 7 points). We have chosen the 7 cluster partitions to be examined in more detail in order to avoid loss of information that would result from considering too few clusters. The following three 7-cluster partitions were selected as the best: (1) UPGMA Cosine (level 190) because of it's high stability (2) MIKCA analysis 9 (random start) and (3) MIKCA analysis 11 (UPGMA cosine starting solution, level 190).

The two MIKCA partitions have almost equally low criterion values. It might be noted that there is good agreement between MIKCA (analysis 11) and the UPGMA-cosine (level 190) with 7 clusters (Lambda=.60). This level of agreement while underlining that there is not a complete cluster overlap, indicates that the different methods have again converged upon a roughly similar partitioning of the data. Similarly, MIKCA analysis 9 and MIKCA

Table 2

Comparison of 7-Group Partitions of JES1

(a)

MIKCA #9 (Random Start)	MIKCA #11 (UPGMA-cos Start) Partition 1	2	3	4	5	6	7	Total
1	7	–	(14)	–	8	1	–	30
2	(20)	(18)	–	4	–	–	–	42
3	–	2	6	–	9	(17)	–	34
4	–	–	–	(28)	–	–	–	28
5	–	–	–	–	(14)	–	1	15
6	2	2	–	–	5	–	(12)	21
7	1	5	(13)	4	–	6	–	29
Total	30	27	33	36	36	24	13	

Sym. lambda .509 Rand .837 Jaccard .293

(b)

UPGMA-cos level 190	MIKCA #11 (UPGMA-cos Start) Partition 1	2	3	4	5	6	7	Total
1	2	–	1	–	–	–	(9)	12
2	–	(13)	–	–	–	–	1	14
3	(23)	4	5	–	1	–	1	34
4	–	–	2	–	–	(11)	–	13
5	2	8	9	(35)	–	1	–	55
6	–	2	4	–	(35)	(12)	2	55
7	–	–	7	–	–	–	–	7
Outliers	3	–	5	1	–	–	–	9
Total	30	27	33	36	36	24	13	

Sym. lambda .596 Rand .840 Jaccard .364

(c)

UPGMA-cos level 190	MIKCA #9 (Random Start) Partition 1	2	3	4	5	6	7	Total
1	2	–	–	–	–	(10)	–	12
2	–	(11)	1	–	–	2	–	14
3	10	(21)	–	–	–	3	–	34
4	1	–	7	–	–	–	5	13
5	–	9	–	(28)	–	–	(18)	55
6	10	1	(22)	–	(15)	6	1	55
7	1	–	3	–	–	–	3	7
Outliers	6	–	1	–	–	–	2	9
Total	30	42	34	28	15	21	29	

Sym. lambda .449 Rand .796 Jaccard .255

analysis 11 have a symmetric Lambda of .51 between them. Again this indicates a rough similarity in the general partition structure of this space.

Table 2 displays the actual overlap of clusters in the three best partitions. Table 2(a) indicates that at least 6 stable "core" clusters exist with rather fuzzy boundary regions between the cluster cores. The matches are circled: The seventh cluster in each of these partitions was created by splitting a single cluster of the other.

Table 2(b) compares the UPGMA cosine partition at the 7 cluster level with the MIKCA analysis 11, UPGMA cosine start (symmetric lambda = 0.60). Table 2(c) compares UPGMA cosine at level 7 with MIKCA analysis 9 which used a random start. As might be expected the symmetric lambda of 0.45 for the second analysis is much lower because of the different "seed" points.

Some differences between the cosine partition and the MIKCA partitions result from similarity coefficient differences. Several clusters in the two MIKCA partitions are very close to the origin; in fact, cluster 3 of the MIKCA 11 partition contains the origin within its boundary. A cluster containing the origin however cannot be found in one piece using cosine as a similarity measure. In Table 2(b) cluster 3 of the MIKCA 11 partition is distributed across nearly all of the cosine clusters. This cluster in the MIKCA-11 partition is the cluster which contained the origin. Clusters 1 and 2 of the MIKCA 9 partition are spread across a large number of UPGMA cosine clusters. These types of disagreement might be expected whenever cosine and distance partitions are compared. A generally consistent pattern of types emerges from all 3 partitions. The major differences occur only in the profiles of the clusters which are close to the origin.

THE OVERLAP BETWEEN NATURAL CLUSTERS AND I-LEVEL

1. Hierarchical Methods

In this section we will compare the cluster analytic classes that have emerged from this data set with the I-level classification of these youth reached by the Jesness procedure. Table 3 presents the results from our analyses utilizing UPGMA-cosine, and UPGMA-distance. This table indicates that although there is a highly significant relationship (as can be seen by the significant chi-square value) the strength of this association between the two partitions is very weak. The symmetric Lambda values indicate that neither partition has much ability to predict the other. The Lambda value is low (.13) when comparing the UPGMA-cosine analysis with I-level classification. On the other hand the UPGMA-d has an even poorer relationship to the I-level (symmetric Lambda is .12). Table 3 illustrates that at the mid ranges of these I-level classes (classes 3 through 7) there appears to be virtually no relationship between the I-level classes and the empirically generated cluster types. It is only at the extremely low and high I-level classes that any substantial overlap can be seen with the empirically generated clusters. At these extremes the relationship with the UPGMA-cosine partition is quite good. It may be noticed that all 7 members of the lowest I-level class (Jesness classification) fall into one large empirical cluster (i.e. cluster 6). Additionally, 15 of the 16 highest I-level designations fall into another large empirical cluster (i e. cluster 5). The UPGMA-distance partition on the other hand overlaps well only at the highest I-levels. Also, many of the lower I-level students are left as outliers. This is consistent with a general finding that the use of Euclidean distance leads to a larger number of outliers.

Table 3

Jesness' I-levels Compared with
Two UPGMA Partitions

(a)

UPGMA-cos level 190	I-levels 1	2	3	4	5	6	7	8	9	Total
1	-	-	-	-	1	4	7	-	-	12
2	-	-	1	1	-	6	5	1	-	14
3	-	-	-	2	1	28	3	-	-	34
4	-	-	2	6	3	2	-	-	-	13
5	-	-	6	8	8	14	4	11	4	55
6	3	4	5	10	4	27	2	-	-	55
7	-	-	-	1	-	5	-	-	-	6
8	-	-	-	2	1	4	-	-	-	7
Outliers	-	-	1	-	1	1	-	-	-	3
Total	3	4	15	30	19	91	21	12	4	

Chi-square = 148.4 (probability = .000)

L_1 = .181, L_2 = .065, L_s = .131

(b)

UPGMA-d level 180	I-levels 1	2	3	4	5	6	7	8	9	Total
1	-	-	3	1	2	26	11	1	1	45
2	-	-	-	2	1	11	1	-	-	15
3	-	1	5	9	7	13	-	-	-	35
4	-	-	2	1	4	10	-	10	3	30
5	-	-	-	1	-	6	2	-	-	9
6	1	1	2	2	-	3	-	-	-	9
7	-	-	-	-	-	2	4	-	-	6
8	-	-	2	6	4	10	1	-	-	23
Outliers	2	2	1	8	1	10	2	1	-	27
Total	3	4	15	30	19	91	21	12	4	

Chi-square = 218.56 (probability = .000)

L_1 = .195, L_2 = .019, L_s = .122

Table 3(c)

Inter-centroid Comparison between Jesness Method I-level
Classes and the UPGMA-cos Cluster Analysis Results

Inter-centroid Distances

	I-level Classes 1	2	3	4	5	6	7	8	9
1	5.064	4.691	3.490	3.493	3.461	2.423	1.541	5.414	4.363
2	5.976	4.350	2.011	3.183	2.463	1.981	1.765	3.628	2.603
3	6.302	5.675	2.971	3.064	2.416	1.355	2.154	3.899	2.895
4	4.017	3.115	1.669	1.463	1.632	2.494	2.838	4.245	3.856
5	6.707	5.765	2.463	3.262	2.118	2.891	3.465	1.295	1.410
6	2.802	2.490	2.946	2.084	2.944	2.672	3.088	5.898	5.167
7	5.978	4.765	1.904	2.635	1.717	2.221	2.905	3.368	3.212
8	4.842	4.916	2.897	2.212	2.250	1.499	2.854	4.340	3.425

Inter-centroid Cosines (at the Origin)

	I-level Classes 1	2	3	4	5	6	7	8	9
1	.193	.055	-.425	-.329	-.756	.435	.833	-.506	-.293
2	-.463	.054	.288	-.688	-.396	.081	.531	.233	.429
3	-.693	-.832	-.633	-.584	-.362	.800	.290	.077	.271
4	.602	.661	.514	.652	.495	-.768	-.234	-.133	-.352
5	-.621	-.536	.311	-.243	.560	-.408	-.312	.969	.862
6	.898	.803	-.028	.577	-.235	-.022	-.003	-.856	-.885
7	-.436	-.171	.399	-.105	.485	-.181	-.232	.373	.112
8	.191	-.339	-.649	.130	-.260	.556	-.319	-.232	-.088

The HMODE partitions are related to the Jesness I-level classes in much the same way as the corresponding UPGMA partitions. We utilized the UPGMA partitions for our further analysis rather than HMODE partitions, because the UPGMA had few outliers.

2. Similarity Relations between I-level Centroids and the Natural Clusters

I-2: I-level subclasses 1 and 2; i.e., the lowest asocial I-2's in the sample, are both classified into natural cluster 6 (socially maladjusted). Table 3(c) indicates the relations of similarity. The cosine values for these comparisons are 0.89 and 0.80 respectively, indicating an extremely high level of mutual similarity between these clusters.

I-3: The two conformist subclasses Cfm and Cfc of I-3 level both classify with the natural cluster 4 (repressed hostile youth). The cosine coefficients are 0.51 and 0.65 respectively, while the distance coefficients reinforce the closeness at 1.46 and 1.63 respectively. The homogeneity of the natural cluster 4 is higher than that of the two conformist subclasses (0.39 versus 0.18 and 0.36). This suggests that the subdivision into two subtypes of conformist may be an artificial separation. The low H-coefficient of the Cfm group further undermines the idea that the passive conformist class can be regarded as a viable cluster.

The third I-3 subclass--the manipulator (Mp)--is more strongly affiliated with the large natural cluster 5 (mature youth). It is also not a particularly homogeneous class and its members are spread across a number of different natural clusters. This I-level subclass, although close to natural cluster 5 (cos = 0.56, D = 2.1), is also similar to the repressed hostile youth of cluster 4 (cos = 0.49 and D = 1.63). It occupies a position on the boundary between the two natural clusters. Again, this is consistent with its low homogeneity (H = 0.25).

I-4: The two neurotic subclasses, Neurotic Acting-Out Na. and Neurotic-anxious Nx, also occupy a similar region of the measurement space. However, each seems to have only a single primary affiliation among the natural clusters. Na is similar to natural cluster 3 (insightful youth). The similarity is high (cos = 0.80, D = 1.3).

Nx is similar to natural cluster 1 (withdrawn neurotic youth). Again, the similarity is high (cos = 0.83, D = 1.5).

The clarity of these strong affiliations suggests that the division of neurotic youth into the two subclasses is correct. However, the homogeneity scores for these two classes are terrible (0.13 and 0.26 for the Na and Nx respectively). Furthermore, the evidence from the intercentroid similarity matrices indicates that Na is also similar to natural class 1, while Nx is very close to natural class 2 (shy, nervous youth). The data indicates that these two classes straddle the natural boundaries. This leads to their low homogeneities, and also to a blurring of the differences between them. This "blurring" results from the absence of strong boundaries. However, in both cases they are, in fact, strongly affiliated with natural clusters. This provides partial evidence of their "reality" status.

The situational reaction youth (Se) are totally subsumed within the large natural cluster 5 (mature youth). The similarity levels are cos = 0.97 and D = 1.3. However, this large natural cluster draws its members from almost all of the I-level subclasses. The high match does not suggest that the Se youth exist as an independent class. They are, in fact, submerged in a very large, continuous, and relatively homogeneous cluster of normal "mature" youth.

Exactly the same fate occurs for the I-level class called "cultural identifier" (Ci). This small I-level class is swallowed up within natural.

cluster 5 (mature youth). The same arguments apply regarding the "independent" existence of Ci as a class within this data space.

In summary, the following represents the relations between the natural classes and the I-level classes:

Natural Class	I-level Classes Subsumed within the Natural Class
1	I_4Nx
2	I_4Nx (weak relation)
3	I_4Na
4	I_2(Aa + Ap), I_3(Cfm + Cfc + Mp)
5	I_4(Se + Ci)
6	I_2(Aa + Ap), I_3Cfc
7	---
8	I_4Na

3. The Natural Clusters and the Simplified I-Level Classes

Table 4 indicates the association between the natural clusters emerging from the UPGMA analysis and the simplified set of the I-level classes. Again, although there is a significant statistical relationship between these two partitions the actual strength of the association, as measured by the symmetric Lambda scores, is very low. It may again be noted that with UPGMA cosine analysis (Table 4(a)) most of the highest I-level subjects fall into one single empirical cluster (i.e. cluster 5) while the lowest I-level cases all fall into natural cluster 6. As might be expected, cluster 5 contains very few delinquents, while cluster 6 contains the majority of high delinquents as shown in Table 4. Table 5 indicates the relationship between I-level and delinquency (i.e. lower I-level categories contain more delinquency and higher I-levels contain less delinquency). However, these observations should not overshadow the fact that the UPGMA-cosine clusters bear a stronger relationship to delinquency than do the I-level classifications (see below).

Table 4

UPGMA Partitions of JES1 vs. I-level Categories

(a)

UPGMA-cos level 190	I-level Categories 1	2	3	4	Total	
1	-	1	11	-	12	
2	-	2	11	1	14	L_1 = .139
3	-	3	31	-	34	L_2 = .161
4	-	11	2	-	13	
5	-	22	18	15	55	L_s = .147
6	7	19	29	-	55	
7	-	1	5	-	6	
8	-	3	4	-	7	
Outliers	-	2	1	-	3	
Total	7	64	112	16		

(b)

UPGMA-cos level 193	I-level Categories 1	2	3	4	Total	
1	-	4	42	-	46	
2	-	2	11	1	14	L_1 = .198
3	7	30	31	-	68	L_2 = .046
4	-	22	18	15	55	
5	-	5	5	-	10	L_s = .138
6	-	1	5	-	6	
Total	7	64	112	16		

(c)

UPGMA-d level 190	I-level Categories 1	2	3	4	Total	
1	-	21	60	2	83	
2	3	26	20	-	49	L_1 = .164
3	-	7	11	13	31	L_2 = .149
4	-	1	17	-	18	
5	1	5	-	-	6	L_s = .158
Outliers	3	4	4	1	12	
Total	7	64	112	16		

Table 5

I-level Categories vs. Delinquency Levels in JES1

I-level Category	Delinquency Levels 1	2	3	4	5	Total
1	2	-	-	2	3	7
2	31	16	7	7	3	64
3	51	30	18	7	6	112
4	13	3	-	-	-	16
Total	97	49	25	16	12	

$L_1 = 0$, $L_2 = .010$, $L_s = .005$

Chi-square = 33.63, D.F. = 12, Significance = .0008

4. MIKCA Partitions and I-Level Classifications

Table 6 indicates the relationship between the MIKCA analyses and the I-level classes (simplified to 4 levels). The relationships are fairly strong at the extreme ends of the I-level scales; however, only in Table 6(d) do all of the lowest I-level subjects fall into one single natural cluster. In this same table 12 out of 16 highest I-level subjects fall into one cluster. In all of the cluster vs. I-level comparisons, however, the middle range I-levels are broadly disseminated among the natural clusters (see Tables 3,4,and 6). These results indicate that the I-level classes are heterogeneous in their membership, have no relation to the natural clusters, and have no particularly clear boundaries.

How Relevant are the Jesness Dimensions in Explaining Delinquency: Global Relations

In this section we briefly examine the general relevance of the present variables (i.e. the Jesness dimensions, and the related I-level class assignments in explaining the delinquent behavior of these junior high school youth (see Table 7).

Table 6

MIKCA Partitions of JES1 vs. I-level Categories

(a) MIKCA #9 (Random Start)

	I-level Categories 1	2	3	4	Total	
1	-	8	22	-	30	
2	-	1	38	3	42	$L_1 = .223$
3	4	22	8	-	34	$L_2 = .391$
4	-	7	8	13	28	
5	3	4	8	-	15	
6	-	-	21	-	21	$L_s = .283$
7	-	22	7	-	29	
Total	7	64	112	16		

(b) MIKCA #11 (UPGMA-cos Start)

	I-level Categories 1	2	3	4	Total	
1	-	1	28	1	30	
2	-	6	20	1	27	$L_1 = .184$
3	-	20	13	-	33	$L_2 = .253$
4	-	10	12	14	36	$L_s = .208$
5	4	9	23	-	36	
6	3	17	4	-	24	
7	-	1	12	-	13	
Total	7	64	112	16		

(c) MIKCA #5 (Random Start)

	I-level Categories 1	2	3	4	Total	
1	3	20	28	-	51	$L_1 = .169$
2	-	7	38	3	48	$L_2 = .034$
3	-	21	19	-	40	
4	-	10	12	13	35	$L_s = .119$
5	4	6	15	-	25	
Total	7	64	112	16		

(d) MIKCA #7 (UPGMA-cos Start)

	I-level Categories 1	2	3	4	Total	
1	-	3	42	1	46	$L_1 = .221$
2	-	4	19	3	26	$L_2 = .172$
3	-	25	13	-	38	
4	-	13	10	12	35	$L_s = .203$
5	7	19	28	-	54	
Total	7	64	112	16		

Table 7

Pearson Correlations Between Jesness Inventory Scales, Delinquency and I-Level in the First Jesness Subset (JES1)

(N=199)

	Delinquency	I-Level	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11
Delinquency													
I-Level	-.296												
1. Social maladjustment	.517	-.355											
2. Value orientation	.507	-.407	.794										
3. Immaturity	.109	-.329	.281	.100									
4. Autism	.507	-.341	.742	.667	.416								
5. Alientation	.422	-.535	.647	.811	.184	.522							
6. Aggression	.480	-.271	.738	.849	.028	.635	.580						
7. Withdrawel	-.092	.014	.379	.350	.067	.243	.140	.349					
8. Social anxiety	-.166	.244	.230	.179	-.013	.080	-.102	.277	.565				
9. Repression	-.058	-.279	.056	-.077	.622	.105	.035	-.202	-.090	-.219			
10. Denial	-.209	.109	-.517	-.739	.076	-.418	-.574	-.649	-.563	-.443	.246		
11. Asocial index	.272	-.114	.691	.237	.113	.288	.139	.260	.199	.050	.054	-.029	

Regression of delinquency with the 11 inventory items:

multiple r .686

r^2 .470

Regression of I-level with the 11 inventory items:

multiple r .645

r^2 .416

Variance explained in delinquency by Jesness dimensions - .47
Variance explained in delinquency by I-level classification = .09

Correlation and Regression of the Jesness Dimensions Against Delinquency

The global correlation between the Jesness delinquency score and the I-level assignments is R= -.296. Thus, as expected, the higher I-level designation implies a lower delinquency score. This gives an R^2 of .09 (i.e. less than 9% of the variance in the delinquency score is explained by I-level subclasses.

Multiple regression analysis was performed between the 11 Jesness inventory dimensions and the Jesness delinquency measure. This gave a multiple R of R=.686, with R^2 = .470. Thus all of the inventory items taken together can explain 47.0% of the variance in delinquency scores. They clearly offer more predictive power than the I-level alone. This is highly encouraging, and suggests that this set of dimensions are highly relevant for the study of delinquent behavior in this age group.

Table 7 indicates the zero-order correlation between each of the Jesness dimensions and the self-reported delinquency score. It can be seen that a number of dimensions correlate at a level higher than R = .50 with the delinquency scores. These are: Social maladjustment, Value orientation, and Autism. Aggression correlates .48 with the delinquency scale.

Discriminant Analysis: Predicting the Delinquency Classes Using the Jesness Dimensions

We utilized the five classes of delinquent behavior that were constructed from the Jesness self-reported delinquency scale in a discriminant analysis using the 11 inventory dimensions. Wilk's Lambda was used in stepwise selection of the predictor variables. Almost 54% of cases were correctly classified into their appropriate delinquency levels using only 6 inventory dimensions (Table 8). These were: Social maladjustment, Autism, Aggression,

Table 8

Discriminant Analyses on Delinquency Level Groups in JES1 Based on the 11 Jesness Inventory Variables

Discriminant Analysis Predicted Level	Actual Delinquency Level					
	1	2	3	4	5	Total
1	69	13	3	1	0	86
2	18	17	5	1	0	41
3	7	13	9	4	2	35
4	1	4	5	5	3	18
5	2	2	3	5	7	19
Total	97	49	25	16	12	

(53.8% of cases correctly classified)

Withdrawal, Social Anxiety and Repression. Treating classes 4 and 5 as the "high delinquency" group, 20 of the 28 (or 71%) are correctly identified. This result again confirms the high salience that this inventory has for the prediction and explanation of delinquent behavior.

The Relation Between I-level Classes and Delinquency Levels

Table 5 illustrates the relationship between the four I-level categories and the various delinquency levels of our sample. Again, these partitions have a very low predictive association with each other (Lambda = .005). The relationship, however, reaches statistical significance and it can be seen that many low I-level youth do, in fact, show the highest delinquency scores. Furthermore, none of the highest I-level class (i.e. class 4) show any

substantial delinquency. The "neurotic" youth in class 3 are spread right across all of the delinquency levels. Many of these youth indulge in fairly high levels of delinquent behavior. Additionally, some of the most delinquent youth fall into the higher maturity levels.

The Natural Clusters and Delinquency Levels

A stronger relationship is found between the natural classes and delinquency than there is between I-level classes and delinquency. The relationship between the natural UPGMA cosine partition and delinquency is characterized by a symmetric Lambda score of .138. This is substantially higher than that shown between I-Level and Delinquency (See Table 9). In this table two levels of the UPGMA cosine analysis (i.e. the 8 cluster and the 6 cluster levels) were associated with the delinquency levels. As might be expected there is a stronger relationship with a higher number of clusters.

Most of the highly delinquent youth fall into one single cluster in each cluster solution. This highly delinquent cluster, although containing many delinquents, also contains youth who are clearly not delinquent. Thus homogeneity is low regarding levels of delinquent behavior.

The UPGMA-distance partition (Table 9(c)) assigns a much smaller proportion of higher delinquent students to a single cluster. The high delinquency youth are therefore more scattered across all of the natural clusters. Cosine as a measure of similarity therefore appears to produce the better predictive results than d. It might be noted that this holds not only for predicting delinquency (Table 9) but also for predicting I-level (Table 3).

Table 10 shows the relationship between the MIKCA natural clusterings and delinquency levels. The four different MIKCA analyses in this table vary according to starting solution and number of clusters. Firstly, none of these

Table 9

UPGMA Partitions of JES1 vs. Self-reported Delinquency Categories

(a) UPGMA-cos level 190 × Delinquency Levels

	1	2	3	4	5	Total	
1	7	5	-	-	-	12	
2	11	2	1	-	-	14	L_1 = .208
3	13	13	7	-	1	34	
4	5	5	-	3	-	13	L_2 = .039
5	41	9	2	2	1	55	
6	15	9	12	9	10	55	
7	1	1	3	1	-	6	
8	2	4	-	1	-	7	L_s = .138
Outliers	2	1	-	-	-	3	
Total	97	49	25	16	12		

Chi-square = 85.5 d.f. = 32 Probability = .000

(b) UPGMA-cos level 193 × Delinquency Levels

	1	2	3	4	5	Total	
1	20	18	7	-	1	46	
2	11	2	1	-	-	14	L_1 = .191
3	20	14	12	12	10	68	L_2 = .029
4	41	9	2	2	1	55	
5	4	5	-	1	-	10	L_s = .120
6	1	1	3	1	-	6	
Total	97	49	25	16	12		

(c) UPGMA-d level 190 × Delinquency Levels

	1	2	3	4	5	Total	
1	44	22	12	3	2	83	
2	15	13	7	9	5	49	L_1 = .078
3	25	5	1	-	-	31	L_2 = .029
4	8	7	2	1	-	18	
5	-	1	1	1	3	6	L_s = .055
Outliers	5	1	2	2	2	12	
Total	97	49	25	16	12		

Table 10

MIKCA Partitions of JES1 vs. Self-reported Delinquency Categories

(a) MIKCA #9

(Random Start)	Delinquency Levels 1	2	3	4	5	Total	
1	5	8	11	4	2	30	
2	26	12	4	-	-	42	L_1 = .102
3	14	9	3	3	5	34	L_2 = .078
4	24	2	1	1	-	28	
5	2	2	3	4	4	15	L_s = .093
6	10	9	1	1	-	21	
7	16	7	2	3	1	29	
Total	97	49	25	16	12		

(b) MIKCA #11

(UPGMA-cos Start)	Delinquency Levels 1	2	3	4	5	Total	
1	13	10	6	-	1	30	
2	20	5	2	-	-	27	L_1 = .166
3	12	9	6	5	1	33	L_2 = .059
4	29	5	1	1	-	36	
5	3	9	9	7	8	36	L_s = .125
6	13	5	1	3	2	24	
7	7	6	-	-	-	13	
Total	97	49	25	16	12		

(c) MIKCA #5

(Random Start)	Delinquency Levels 1	2	3	4	5	Total	
1	15	18	8	5	5	51	
2	30	13	5	-	-	48	L_1 = .115
3	13	11	7	6	3	40	L_2 = .029
4	31	3	1	-	-	35	
5	8	4	4	5	4	25	L_s = .080
Total	97	49	25	16	12		

(d) MIKCA #7

(UPGMA-cos Start)	Delinquency Levels 1	2	3	4	5	Total	
1	19	17	8	1	1	46	
2	21	4	1	-	-	26	L_1 = .131
3	19	9	4	5	1	38	L_2 = 0
4	26	7	1	1	-	35	
5	12	12	11	9	10	54	L_s = .077
Total	97	49	25	16	12		

partitions relate to delinquency as well as the UPGMA cosine partition. (See Tables 9 and 10). This probably stems from the fact that the MIKCA analyses are confined to the use of Euclidean distance. These solutions also tend to scatter the highly delinquent youth more broadly across many natural clusters. In the cosine solutions, in contrast, the delinquents tended all to be in one of the natural cluster groups. Examining the subtables it can be seen that when the MIKCA solution utilized the UPGMA cosine starting points, a stronger relationship to delinquency is found. When using these centroids virtually all the delinquents are again located into one of the natural clusters. In MIKCA analysis No. 7, 10 out of 12 highly delinquent youth are classified into a single natural cluster. These results indicate the importance of the starting solutions in the iterative relocation analysis. The cosine starting solutions thus converge upon solutions which have a stronger relationship to the initial starting points than did the two random start solutions. Unfortunately, these MIKCA partitions which used cosine starting points were (clearly) local optima (analyses 11 and 7, in Table 10). The random start partitions although converging on better optimization criteria, do not relate as well to delinquency (See Table 10).

The Relationship Between the I-level Designation and the Jesness Inventory Items

Regressing the 11 inventory dimensions against the I-level gives a multiple R of .645 (with an R^2 of .416). Thus 41% of the variance in the I-level designation is explained by a linear relationship with the inventory dimensions.

DESCRIPTION OF THE "CORE" CLUSTERS

In this section we provide descriptions of the psychological profiles of the "core" clusters which we believe represent relatively dense areas of the measurement space. It is important to note that there are no clear discontinuous boundaries between these "core" clusters. The "boundaries" are rather blurred and fuzzy with no marked empty space between the clusters. The structure of this data conforms to a model of undulating density levels with some areas more dense than others. Our following descriptions, therefore, are based on the predominant psychological features of these "denser" areas of the measurement space (see Tables 11 and 12).

CLUSTER 1

Psychological Description: WITHDRAWN, DEPRESSED, NEUROTIC YOUTH (6.0% of the sample)

Most of these youth were classified as "neurotics" in the I-level identification process. They are split almost equally between the two subclasses of neurotics. Our statistical evaluation tables indicate that this natural cluster has excessively high scores for being withdrawn and depressed, exhibiting social anxiety within interpersonal relationships, and exhibiting a tendancy to solve their social problems without regard for social customs or rules. On the other hand they do not exhibit immaturity, repression, or denial. They tend not to suppress feelings or emotions and they tend more than most of the youth in the sample to admit the presence of negative feelings. They are, on the other hand, critical of themselves and of others, and are quite open to acknowledging unpleasant events. These youth have a much lower delinquency score than is average for this sample.

Table 11

CENTRAL TENDENCY SCORES FOR EACH NATURAL CLUSTER
USING JESNESS DIMENSIONS

	CLUSTER NUMBERS							
	1	2	3	4	5	6	7	8
NUMBER OF CASES	12	14	34	13	55	55	6	7
AVERAGE SQUARE DEVIATIONS	3.98	5.2	4.83	5.05	8.30	10.46	2.76	3.12
CLUSTER HOMOGENEITY	.67	.57	.62	.39	.35	.18	.75	.72
VARIABLE NAME								
1. Social maladjustment	.66	-.53	-.50	.62	-.82	1.06	.02	-.46
2. Value orientation	.30	-.33	-.23	.24	-1.02	1.06	.17	.49
3. Immaturity	-.72	.41	-1.13	.51	.08	.66	-.56	-.40
4. Autism	-.22	-.33	-.69	.03	-.65	1.09	.97	.05
5. Alienation	-.06	-.55	-.33	.60	-.73	.86	-.27	.79
6. Manifest aggression	.25	-.05	-.03	.05	-1.05	1.00	.55	.12
7. Withdrawal-Depression	1.38	.24	.11	.14	-.79	.41	-.76	-.27
8. Social anxiety	1.42	1.16	.18	-.42	-.64	.15	-.10	-.40
9. Repression	-.66	.42	-1.09	1.20	.48	.09	-.87	-.35
10. Denial	-.73	-.32	-.16	.05	1.10	-.71	.36	-.72
11. Asocial index	.99	-.98	-.35	.91	-.18	.44	-.21	-1.42
12. I-level	.64	.41	.28	-.73	.27	-.48	.10	-.14
13. Self-reported delinquency	-.43	-.53	-.05	.03	-.47	.69	.53	-.04

Table 12

HOMOGENEITY COEFFICIENTS FOR JESNESS VARIABLES FOR EACH NATURAL CLUSTER

VARIABLE NAME	CLUSTER NUMBERS								VARIABLE HOMOGENEITIES
	1	2	3	4	5	6	7	8	
1. Social maladjustment	.72	.77	.86	.81	.58	.33	.94	.78	.61
2. Value orientation	.83	.85	.64	.72	.40	.68	.89	.78	.63
3. Immaturity	.77	.63	.67	.31	.35	.05	.56	.95	.39
4. Autism	.66	.25	.64	.78	.56	.43	.80	.91	.55
5. Alienation	.72	.64	.22	.47	.05	.69	.78	.89	.42
6. Manifest aggression	.75	.89	.64	.83	.58	.41	.72	.87	.61
7. Withdrawal-depression	-.04	.54	.50	.31	.51	.03	.82	.61	.34
8. Social Anxiety	.80	.71	.39	.66	.30	.08	.82	.27	.36
9. Repression	.57	.63	.75	.52	.32	.079	.91	.82	.41
10. Denial	.63	.70	.70	.60	.25	.54	.78	.76	.52
11. Asocial index	.67	-.22	.54	.55	.46	-.05	.54	.58	.31
12. I-level	.81	.30	.85	.61	-.39	-.14	.72	.62	.16
13. Self-reported delinquency	.78	.73	.62	.37	.57	-.76	.43	.52	.21

Structural Description

This is clearly one of the "core" clusters in our analysis. Both the UPGMA-cosine, the MIKCA with cosine start, and the MIKCA with random start identify this cluster. The clearest overlap, of course, is seen in the comparison of the UPGMA-cosine and the MIKCA-cosine start partitions.

The cluster is compact with homogeneity of H=.67. Furthermore, the group is highly homogeneous on virtually all of the variables except for the Jesness dimensions 7 and 9. Examining the cosine matrix and the intercluster distance matrix, it can be noted that this cluster is relatively isolated. It is closest to the mature youth of cluster 3, both according to the cosine value (.32) and distance value (2.8). The data therefore suggest that this is a homogeneous small cluster occupying a relatively distinct region of the data space and showing homogeneity, not only in a general sense but across the majority of the profile variables.

CLUSTER 2

Psychological Description: SHY, NERVOUS CONVENTIONAL YOUTH (7.0% of sample)

This natural cluster has an extremely high score for social anxiety. This implies that these youth feel nervous, shy, sensitive to criticism and experience emotional discomfort in their interpersonal relationships. The other outstanding aspect of their profile is their very negative score on dimension 11 (i.e. the asocial index). A negative score here implies that the youth pay close attention to social customs and roles and would not contravene these. This is confirmed by the virtual absence of delinquent behavior among these youth (the majority are in level 1).

Structural Description

This cluster was identified by UPGMA-cosine and by MIKCA with a cosine start, but it combined with a segment of cluster 3 in the MIKCA random start partition. The global homogeneity score of .57 indicates a relatively compact small cluster. On the Asocial Index, however, this cluster exhibits a large amount of heterogeneity. The intercluster centroid distance and cosine matrixes are not informative regarding this cluster. The cluster seems to be an average distance from all other cluster centroids with none of the cosine values reaching particularly high or low levels.

CLUSTER 3

Psychological Description: INSIGHTFUL, MATURE, YOUTH (17.1% of the sample)

The present cluster according to the Jesness I-level classification consists primarily of neurotic "acting out" youth. These youth are also relatively nondelinquent. The natural cluster shows extremely low scores on Jesness dimensions 3 and 9. The first of these scores implies that they are quite mature. They do not exhibit the lack of insight and naivity that would be implied by a high score on the immaturity dimension. The very low score on dimension 9 (i.e. repression) indicates that they recognize and are fairly accurate in labeling their emotions. They can acknowledge negative feelings and are capable of being critical of themselves. The low score on the autism scale indicates that these youth do not distort reality in accordance with their personal desires or needs.

Structural Description

This cluster is recurrently identified in the UPGMA (cosine), the MIKCA (cosine.start) analysis, and the MIKCA (random start) analysis. It is reasonably homogeneous (H=0.62).

CLUSTER 4

Psychological Description: REPRESSED HOSTILE YOUTH (6.5% of the sample)

This cluster is atypical primarily in regard to high scores on Jesness dimensions 9 and 11. This implies a strong tendency toward the repression of feelings and emotions. The asocial scale index is high (the standardized cluster score is .91) implying that these youth tend to disregard social customs or social roles as they deal with their personal and social interrelationships. This is somewhat consistent with the scores for delinquent behaviors since this cluster contains a higher proportion of delinquents than the three earlier clusters (particularly clusters 1 and 2). These youth exhibit relatively high scores for social maladjustment and alienation. They tend to show negative self-concept, a tendency to blame others for problems, to feel unhappy, worried or misunderstood, and to be bothered by feelings of hostility. These youth also exhibit an alienated, estranged and relatively distrustful attitudes toward others, particularly those in authority. Their scores indicate relatively poor interpersonal relationships, criticism and intolerance of others, and a tendency to project hostile feelings. This natural cluster has little homogeneity according to I-level designation. It is completely fragmented across four of the Jesness classes (primarily conformists and manipulators).

Structural Description

This small cluster is located in a position in a space which is rather close to the larger cluster 6. It is identified correctly in both the UPGMA-cos analysis and the MIKCA analysis using the UPGMA-cos seed points. The cosine matrix also indicates that cluster 4 has a rather large cosine value in relation to the centroid of cluster 6 (cosine=.49). Turning to the

homogeneity scores, however, it can be seen that cluster is reasonably homogeneous (H=.58), particularly on the social maladjustment score (H=.81).

CLUSTER 5

Psychological Description: MATURE,CONFIDENT NON-CRITICAL YOUTH (27.6% of the sample)

This cluster also shows virtually no relationship to the Jesness I-level classifications. It gathers members from all the I-level classes and has no clear tendency towards any particular I-level class. A point worth noticing, however, is that it contains virtually all of the highest sub-classes of the I-level in this sample (i.e. these are the situational, emotional and the cultural identifier). Thus, although it contains many other youth classified in the lower stages of the I-level, it also is biased toward the most mature members in this sample. It also has no members at the lowest category of the I-level. The relationship to delinquency supports these findings. Forty-one out of the 55 members of this group fall into the nondelinquent category. On the other hand it is a large cluster and some members are scattered right across all levels of the delinquency spectrum. For instance, it contains one member who is at the highest stage of delinquency. Turning to the psychological description it can be seen that this cluster is defined by atypical scores on a number of dimensions. It exhibits atypically low scores on Jesness dimensions 1, 2, 5, 6, and 7. This profile implies that these youth have good self-concept, do not feel unhappy, worried or misunderstood, and are not hostile. They do not share the attitudes and opinions of lower social class persons. They do not go around looking for trouble, thrills, etc. They are not alienated. Their interpersonal relationships tend to be good. They tend not to be critical nor intolerant of others. They have only

minimal tendency towards manifest aggression. They are not aware of particularly unpleasant feelings, anger or frustration. They do not illustrate lack of satisfaction with self or others and do not have a tendency toward isolation. However, a very high score on denial suggests a reluctance to acknowledge unpleasant events or feelings. This score indicates that they may see their parents as without fault and to admit no conflict with them. They may deny personal inadequacies or habits and are often unwilling to criticise others. As mentioned above, these youth are predominantly nondelinquent.

Structural Description

This natural cluster is picked up by the UPGMA-cosine and both MIKCA partitions, but was split into two parts in the MIKCA random start partition. An examination of these two parts revealed that they have generally similar profiles (cos = 0.43). However, one part is characterized by extreme scores on good behavior (no delinquency), social adjustment, absence of alienation, and interestingly, much higher scores for denial and depressive-withdrawal. The total cluster, however, appears to define one of the more populated and denser regions of the measurement space. The boundaries between this and some of the other clusters are not particularly dramatic. It has a cluster diameter of 7.8 standard deviations which is one of the largest diameters of any of the clusters in this analysis. The intercentroid cosine matrix, however, indicates that the cluster does not lie in the same direction as any of the other clusters in the analysis. All of the cosine values are close to zero or are negative. A further point to be noticed from the cluster cosine matrix is that this cluster lies in a virtually opposite direction in a measurement space from clusters 1 and more particularly, cluster 6. This can

also be seen clearly from the cluster graph where cluster 5 occupies the right region of the measurement space and clusters one and 5 occupy the left region of the space. This is a relatively nonhomogeneous cluster (H=.35). It therefore contains fairly large diversity of youth as was seen from the examination of its two subtypes. This diversity can also be seen by examining Table 3 which illustrates that the members of this cluster are spread across 7 different I-level classes. Table 3 also indicates that most of the more mature youth are members of this cluster. The 'variable-by-cluster' homogeneity scores also confirms the relative heterogeneity (virtually none of the H-values for the specific variables exceeds a level of H=0.6). This region of the measurement space therefore can be construed as a large relatively uniform region of the measurement space characterized by a continuity from fairly high levels of maturity to middling or low levels of maturity but not the lowest.

CLUSTER 6

Psychological Description: SOCIALLY MALADJUSTED YOUTH (27.5% of the sample)

This cluster also contains a range of I-level designations. Fully 7 different I-level subclasses are represented in its membership (Table 3). Over half of the members of this group however are classed as 'neurotic acting-out'. A second point to note is that very few of the three highest levels of the Jesness type designation are involved in this cluster whereas most of the lowest levels are classified into this cluster. This low maturity is confirmed since this cluster also contains the most serious delinquents in the sample. Ten out of 12 youth who are classified into delinquent level 5 are members of the present cluster. This cluster is the most delinquent of all the present natural clusters (Table 4). The fact that this naturally

occuring cluster finds its members broadly spread across the Jesness I-level assignments are artificially imposed upon a continuous distribution. The natural cluster can be seen at the extreme left side of the cluster graph.

These youth have extremely high scores on social maladjustment (i.e. inadequate and disturbed socialization, an inability to meet environmental demands, negative self-concept, unhappiness, distrust of authority, extrapunitive, hostile, and accepting of behavior regarded as antisocial). The high score on value orientation indicates that they do in fact share the opinions of persons in the lower socioeconomic classes (i.e. gang orientations, toughness, and a tendancy to look for trouble and thrills). A high score on autism indicates a tendancy to distort reality in order to fit with personal desires and needs. They see themselves as smart, good-looking, tough, and so forth. There is also an extremely low score on denial. This implies that they are not reluctant to acknowledge unpleasant events or aspects of reality. They are quite ready to see fault in their parents and admit conflict with them. They are not unwilling to criticise others. This profile fits well with the low I-level scores for these individuals, and their high levels of delinquent behavior.

Structural Description

The cluster is large and relatively heterogeneous. It emerges, however, as a recurringly identified core cluster. It is identified by three separate methods: UPGMA (cosine), MIKCA (cosine start), and the MIKCA (random start). It is the most heterogeneous cluster of all the core clusters (H=0.18). The MIKCA (random start), in fact, also breaks this cluster into separate parts. However, inspection of the centroids of these two parts led to the conclusion that they are clearly parts of the same cluster. They have high similarity to

each other, as assessed both by the formal similarity measurement (cos = 0.74) and by the subjective interpretation of the "meaning" of the profiles. On certain Jesness dimensions, however, the large cluster illustrates good homogeneity. These are the high score for lower class value orientation; alienation and estrangement; and the absence of denial. Virtually all of the members of this cluster illustrate these three tendencies and in each case the homogeneity level for the particular dimension are above H=0.5.

Discriminating between the "Natural" Clusters

A useful approach to understanding the multivariate differences between the natural clusters is to conduct multivariate discriminant analysis between them. Therefore, using the same Jesness dimensions, a stepwise discriminant analysis was conducted between the various groups. Clusters with size less than ten were eliminated from this analysis as being relatively unstable. Thus, the present analysis will provide descriptive information on the discriminating features between the remaining six major "core" types.

Table 13(a) indicates the results of the various step numbers as the separate Jesness dimensions were included into the discriminating procedure. The first two variables--value orientation and immaturity--both are major definers of the first discriminant function. This discriminant function separates the two large highly delinquent and nondelinquent clusters (see the cluster graph below). This dimension would seem to be defined as an orientation toward middle class, conventional values as opposed to a rejection of middle class values. However, since the immaturity scale loads equally heavily upon this discriminant function, it cannot be assumed that the function is purely a measure of acceptance or rejection of middle class values.

The second discriminant function does not account for an equal amount of variance as the first discriminant function. Table 13(b) indicates that it

accounts for 20% of the trace and has an Eigenvalue of 1.11. Nevertheless, it still has a very strong canonical correlation with the overall taxonomy (canonical correlation = 0.72). This function is defined primarily by repression and social maladjustment. Groups scoring high on this function; e.g., groups 1 and 3, would be expected to show an absence of repression. They will not suppress feelings or emotions and tend to be capable of acknowledging consciously to themselves the presence of negative feelings. This function, however, clearly incorporates a social maladjustment component, indicating a tendency to feel unhappy, worried, and misunderstood, to distrust authority and blame others, and so forth. Clearly the two clusters scoring high on this; i.e., clusters 1 and 3, are characterized at differential levels by these two major defining dimensions. Cluster 1 more clearly indicates the social maladjustment component and cluster 3 indicates more clearly the absence of repression. The third discriminant function is defined largely by the asocial index (discriminant function coefficient = -1.07). It is appropriate to note that clusters 1 and 4 score at the extreme levels of this function. The components of the asocial index; i.e., the projection of hostility and the tendency to solve problems without regard for social customs or rules, are shown by both of these clusters. Hostility, in particular, is shown by cluster 4. The fourth discriminant function is much more difficult to interpret and seems to be a combination of the asocial index, social maladjustment, and value orientation. This latter function serves primarily to separate cluster 2 from all of the other clusters.

The cluster graph shown below indicates the projection of the major six core clusters into the discriminant space. The ellipses have been drawn to take into account the boundaries of the clusters and the relative orientation of the clusters in discriminant space. This diagram indicates the relative

Table 13(a)

Discriminant Analysis using Jesness Dimensions Against the UPGMA-cosine Clusters

Step Number	Jesness Dimension	F to Enter	Wilk's Lambda	Change in Rao's V	Significance
1	Value Orientation	47.81	.36	334.71	.000
2	Immaturity	20.38	.20	185.61	.000
3	Social Anxiety	14.49	.13	122.20	.000
4	Antisocial Index	11.00	.09	93.88	.000
5	Repression	8.92	.07	97.75	.000
6	Autism	5.42	.06	71.09	.000
7	Withdrawal-Depression	3.96	.05	56.89	.000
8	Manifest Aggression	2.52	.04	39.95	.000
9	Denial	1.92	.04	30.09	.000
10	Alienation	1.16	.04	12.20	.094
11	Social Maladjustment	0.61	.04	6.63	.469

Table 13(b)

Discriminant Analysis Using Jesness Dimensions Against the UPGMA-cosine Clusters.

Number Removed	Eigen-value	Canonical Correlation	% Trace	Wilk's Lambda
0	3.24	.87	58.0	.039
1	1.11	.72	20.0	.168
2	.57	.60	10.2	.356
3	.39	.53	7.0	.559
4	.18	.39	3.3	.779

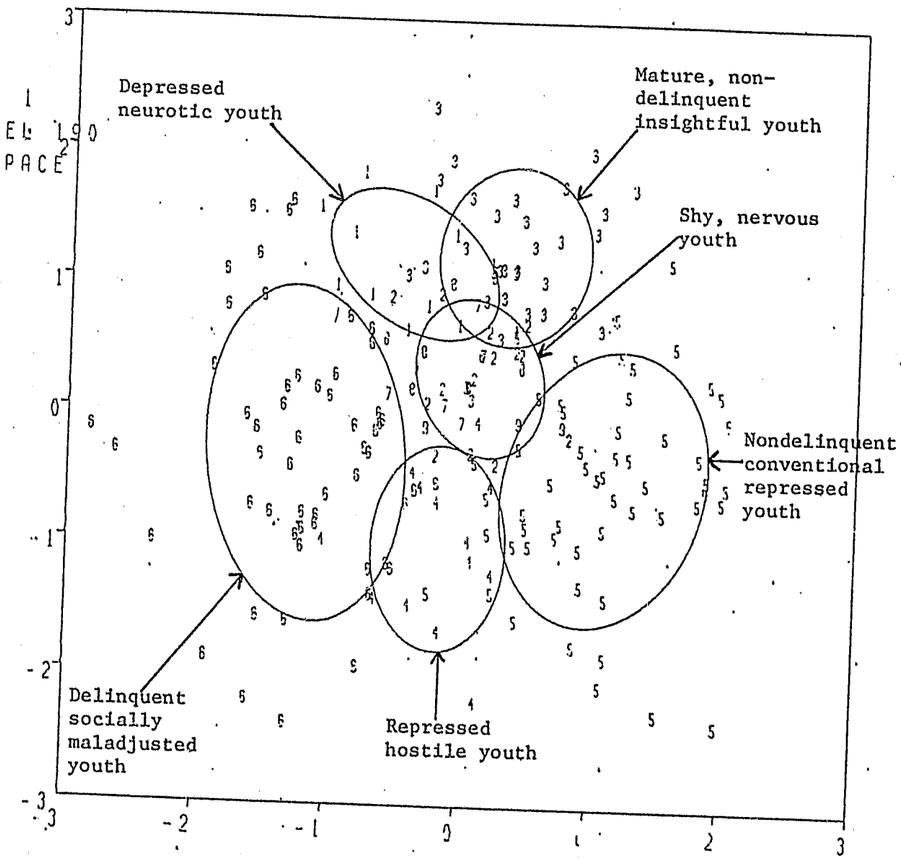
Table 13(c)

Standardized Discriminant Function Coefficients for the First Four Functions

Jesness Dimension	Discriminant Function No. 1	2	3	4
1 Social Maladjustment	-.08	.56	.36	-.81
2 Value Orientation	-.60	-.12	.19	.73
3 Immaturity	-.60	-.52	.29	-.34
4 Autism	-.32	-.37	.56	.59
5 Alienation	-.07	-.18	-.63	.26
6 Manifest Aggression	-.51	.01	.16	-.39
7 Withdrawal-Depression	-.23	.32	-.52	.27
8 Social Anxiety	-.04	.24	-.38	-.57
9 Repression	.31	-.71	-.53	-.57
10 Denial	.34	-.11	-.00	.59
11 Asocial Index	-.15	-.54	-1.07	.86

CLUSTER GRAPH

JESNESS SUBSET 1
UPGMA-COS. LEVEL 1.90
DISCRIMINANT SPACE



positioning of the clusters and tends to clarify the meaning of the major dimensions of discriminant space. The diagram uses the first and second discriminant functions. Thus, the first major separating dimension allows the high level of separation between the conventional well-behaved youth in clusters 3 and 5 and highly delinquent youth of cluster 6.

An Evaluation of the I-level Classification

In this final section we briefly examine the statistical properties of the I-level classification assessed across the various Jesness dimensions. In this section we will particularly examine, first, the homogeneity characteristics of each I-level class. We will then examine the degree of similarity and overlap between these classes, and finally, we will briefly examine the major defining features of each I-level class.

Homogeneity Characteristics of each I-level Class. Table 14(a) indicates homogeneity coefficients computed for each Jesness dimension across each of the nine subclasses. There is also an overall homogeneity coefficient computed for each class across all variables and an H-coefficient computed for each variable across all classes.

The first important observation is that all of the major central classes show extremely poor homogeneity coefficients. It is only the classes at the extreme end of this distribution; i.e., the I2 asocial aggressive at the one extreme and the situational emotional reaction and the cultural identifiers at the other extreme, that even show reasonable homogeneities. Of these three subclasses, only the cultural identifier has a homogeneity coefficient greater than 0.50. A second observation is that in comparing these homogeneity levels with those of the natural clusters, the natural clusters are clearly far more homogeneous. This suggests the conclusion that the natural clusters show much

more respect for the internal mutual similarity of class members than do the I-level classification breakdowns. Yet, a major structural feature of any classification is that the members of any class be mutually similar within classes and mutually dissimilar between classes. We will now examine this latter issue.

Table 14(a)

HOMOGENEITY COEFFICIENTS FOR EACH I-LEVEL CLASS OVER EACH JESNESS DIMENSION

VARIABLE NAME	I-LEVEL TYPES									VARIABLE HOMOGENEITIES
	Aa	Ap	Cfm	Cfc	Mp	Na	Nx	Se	Ci	
1. Social maladjustment	.35	.08	.24	.50	.28	.09	.25	.46	.73	.23
2. Value orientation	.75	.43	.19	.46	.25	.18	.38	.70	.74	.30
3. Immaturity	.66	.44	.42	.29	.34	.08	.46	.60	.08	.26
4. Autism	.63	.31	.10	.50	.42	.13	.28	.47	.67	.27
5. Alienation	.84	.58	.27	.50	.40	.18	.21	.62	.56	.30
6. Manifest aggression	.49	.12	.18	.49	.07	.12	.44	.63	.77	.24
7. Withdrawal-Depression	.11	.65	.14	.26	.12	.13	.05	.63	.81	.17
8. Social Anxiety	.62	.51	.22	.28	.11	.25	.32	.45	.28	.26
9. Repression	.07	.42	.38	.07	.25	.29	.27	.16	.67	.26
10. Denial	.10	.26	-.15	.39	.24	.07	.19	.45	.31	.16
11. Asocial Index	-.08	-.18	-.02	.26	.23	-.08	.02	-.09	.37	.03
OVERALL TYPE HOMOGENEITY	.41	.33	.18	.36	.25	.13	.26	.46	.54	

Table 14(b)

CENTRAL TENDENCY SCORES FOR EACH I-LEVEL CLASS ACROSS THE JESNESS DIMENSIONS

	I-LEVEL NAME								
	Aa	Ap	Cfm	Cfc	Mp	Na	Nx	Se	Ci
	1	2	3	4	5	6	7	8	9
NUMBER OF CASES	12	11	48	79	75	265	78	21	11
AVERAGE SQUARED DEVIATION	5.9	6.7	8.8	6.9	8.2	9.5	8.0	5.6	4.6
CLUSTER HOMOGENEITY	.41	.33	.18	.36	.25	.13	.26	.46	.54
VARIABLE NAME									
1. Social maladjustment	2.2	1.8	-.31	.15	.15	-.03	-.12	-1.2	-1.1
2. Value orientation	1.8	1.3	-.37	.35	.18	.08	-.32	-2.0	-1.3
3. Immaturity	1.5	2.0	.67	.36	.19	-.42	-.14	.40	.01
4. Autism	2.2	2.0	-.45	.11	.16	.04	-.26	-1.4	-.76
5. Alienation	1.3	1.3	-.00	.66	.49	-.11	-.59	-1.4	-1.4
6. Manifest Aggression	1.5	1.0	-.65	.04	.01	.22	-.12	-1.7	-1.3
7. Withdrawal-Depression	1.0	1.1	.06	-.56	-.15	-.08	.80	-.37	-.52
8. Social Anxiety	.29	.44	-.16	-.49	-.69	.03	1.14	-.02	-.47
9. Repression	.13	.85	1.13	.37	.21	-.46	-.16	.73	.94
10. Denial	-1.1	-.74	.61	.10	.02	-.11	-.36	1.2	1.2
11. Asocial Index	1.0	.59	-.15	-.01	.12	-.07	.00	.23	-.42

The Interrelationship between the I-level Classes. Our examination of the similarity relationships between the various I-level classes is based on the data presented in Table 14(c). This table indicates the angular similarity between the nine centroids as measured by the cosine of the angle between them and the distance or dissimilarity between them as measured by the Euclidean distance between them.

At the lowest level of the I_2 class, a major similarity is seen between the asocial agressive and the asocial passive (Aa and Ap, with a cosine of 0.96 and a distance of 1.37). These figures indicate a strikingly high level of mutual similarity between these two I-level subclasses. The data suggest that they are virtually indistinguishable. A second extremely high similarity is found between the I_3 subclasses, passive conformist and the manipulators (Cfc and Mp, with a cosine of 0.87 and a Euclidean distance of 0.59). Again, this would suggest that these two classes are virtually indistinguishable.

Next, at the I_4 level, the two neurotic groups--neurotic acting-out (Na) and neurotic anxious (Nx) show a strong similarity using the distance metric but not using the cosine metric.

The two I_4 classes, situational emotional reaction (Se) and cultural identifiers (Ci) show a striking similarity with a cosine of 0.94 and a distance of 1.34. This data would also suggest that these two profiles are virtually indistinguishable. An interesting third member of the two I_4 classes is the I_3 class Cfc (cultural conformists). This class shows a high similarity to both the situational emotional reactors and the cultural identifiers. It has a cosine of 0.70 with both of these classes and a somewhat larger Euclidean distance from the two centroids (see Table 14(d)). These three groups; i.e., Se, Ci, and Cfc, would appear to be located in very

Table 14 (c)

Similarity Coefficients between the Nine I-level Subclasses

Inter-Cluster Centroid Distances

	Nx	Na	Mp	Se	Cfc	Aa	Cfm	Ap	Ci
1 Nx	0	1.691	2.512	3.802	2.739	5.189	2.488	4.750	3.541
2 Na	1.691	0	1.358	4.100	1.608	4.831	2.363	4.581	3.507
3 Mp	2.512	1.358	0	4.164	.592	4.437	1.835	4.012	3.542
4 Se	3.802	4.100	4.164	0	4.269	8.050	2.846	7.326	1.341
5 Cfc	2.739	1.608	.592	4.269	0	4.449	1.845	3.979	3.636
6 Aa	5.189	4.831	4.437	8.050	4.449	0	5.569	1.372	7.589
7 Cfm	2.488	2.363	1.835	2.846	1.845	5.569	0	4.778	2.341
8 Ap	4.750	4.581	4.012	7.326	3.979	1.372	4.778	0	6.894
9 Ci	3.541	3.507	3.542	1.341	3.636	7.589	2.341	6.894	0

Cosine of the Angle (at the Origin) Between Cluster Centroids

	Nx	Na	Mp	Se	Cfc	Aa	Cfm	Ap	Ci
1 Nx.	1.000	.124	-.886	.200	-.876	-.139	-.096	-.112	.042
2 Na	.124	1.000	-.339	-.392	-.401	-.107	-.672	-.361	-.332
3 Mp	-.886	-.339	1.000	-.301	.872	.375	.162	.387	-.214
4 Se	.200	-.392	-.301	1.000	-.280	-.799	.701	-.651	.939
5 Cfc	-.876	-.401	.872	-.280	1.000	.336	.246	.385	-.190
6 Aa	-.139	-.107	.375	-.799	.336	1.000	-.361	.958	-.830
7 Cfm	-.096	-.872	.162	.701	.246	-.361	1.000	-.101	.705
8 Ap	-.112	-.361	.387	-.651	.385	.958	-.101	1.000	-.685
9 Ci	.042	-.332	-.214	.939	-.190	-.830	.705	-.685	1.000

Table 14(d)

Mutually Closest Connections Between the Nine I-level Subclasses

Pairs of Subclasses	Inter-centroid Similarities: Cosine	Euclidean Distance
I_4(Nx) and I_4(Na)	0.12	1.69
I_4(Na) and I_3(Mp)	-0.34	1.35
I_3(Cfc) and I_3(Mp)	0.87	0.59
I_2(Aa) and I_2(Ap)	0.96	1.37
I_4(Se) and I_4(Ci)	0.94	1.34
I_3(Cfc) and I_4(Se)	0.70	2.84
I_3(Cfc) and I4(Ci)	0.70	2.34

similar regions of the psychological space of the Jesness dimensions. Finally, the neurotic anxious class of I_4 shows a strong relationship to the manipulators of I_3 (see Table 14(d)).

These similarity connections indicate that there is very high similarity between certain adjacent members of the I-level hierarchy. These adjacent extremely high mutual similarities consist of: Aa and Ap, Cfc and Mp, Na and Nx, Se and Ci. Each of these represents a high level of mutual similarity such that it would be difficult to tell these groups apart given the numerical data. Yet, at the same time, they tend to support the I-level classification as it is presented with these groups being adjacent to each other. A further support for the I-level classification as given is that all of the mutual similarities occur with adjacent subclasses and there are only weaker kinds of connections between the basic I_2, I_3, and I_4 levels. Of these cross-level similarities, the manipulators of I_3 and the neurotic acting-out group (Na) show a high similarity (the Euclidean distance score of 1.35 with a very poor cosine level of 0.34). However, a further cross-I-level similarity is that between the cultural conformists of I_3 and the cultural identifiers (Ci) and the situational reaction (Se) of the I_4 level. The most basic finding of this group, therefore, would not contravene the high similarity between adjacent I-level subclasses, but would tend to indicate that, certainly as given, the I-level subclasses are virtually indistinguishable for certain pairs.

Profiles of Each I-level Class in Terms of the Jesness Dimensions. *

I_2: Asocial Agressive. This cluster is characterized by high scores for social maladjustment, low class value orientation, immaturity, autism, alienation, manifest aggression, withdrawal-depression, and a virtual absence of denial. It also has a very high score on the asocial index. The cluster itself is reasonably homogeneous (h = 0.41).

I_2: Asocial Passive. This cluster also shows highly atypical scores on a large number of the Jesness dimensions. The profile is very similar to that of the Asocial Aggressive. In fact, there is virtually no difference between these profiles. The suggested pairing of 1 and 2 together in the above section dealing with similarity pointed toward their indistinguishability. Examining the Jesness dimensions for both of these profiles it would be virtually impossible to separate between them (see Table 14(b)).

I_3: Passive Conformist. This I-level class shows a jointly high score for repression, denial, and immaturity. The cluster homogeneity of h = 0.18 is extremely poor, indicating that there is a very high hetereogeneity within the cluster. Examining the homogeneity coefficients for each separate Jesness dimension it is found that the cluster has reasonable homogeneity only for the immaturity score and the repression score. The fact that the cluster shows homogeneity on these two particular variables is somewhat comforting because these are the two characteristics on which the cluster clearly has a high level of atypicality. Otherwise, this is a very loose and straggly collection of heterogeneous youth.

I_3: Cultural Conformist. This I-level class is characterized by a fairly high score for alienation, coupled with a high score for repression. It has a somewhat atypically high score for lower class value orientation and immaturity. It also has low scores for social anxiety and a low score for withdrawal-depression. This indicates that these youth are relatively sociable, not particularly depressed, and not particularly anxious.

The cluster shows a fairly high level of homogeneity for social maladjustment, value orientation, autism, alienation, manifest aggression, and denial. This is one of the more homogeneous classes, with an H level of 0.36. This H level, however, is not particularly impressive.

I_3: Manipulator. This cluster shows a bland profile which is characterized only by an above average score for alienation, coupled with a very low score for social anxiety. These youth are not nervous, nor are they shy, nor are they particularly sensitive to criticism.

This Manipulator group shows relatively low homogeneity levels for each of the separate Jesness dimensions. The highest homogeneity coefficients are those for autism and alienation, both of which barely pass the 0.40 homogeneity level.

I_4: Neurotic Acting Out. These youth are characterized by a low score for immaturity and a relatively low score for repression. The profile is relatively bland, with no highly atypical features. This is partially a result of the fact that this cluster is by far the largest of the I-level classes. The cluster shows a serious absence of homogeneity, not only a the general cluster level (h = 0.13), but also for virtually all of the separate Jesness dimensions.

I_4: Neurotic Anxious. This cluster shows a very high score for social anxiety. Again, this implies that they feel nervous, shy, and experience tension. They also have a very high score for withdrawal-depression. Finally, there is a fairly low score for alienation. This indicates a relative absence of being critical or intolerant of others and a tendency not to behave in delinquent ways.

I_4: Situational Emotional Reaction and I_4: Cultural Identifier. The profiles of these two clusters are virtually identical. They have atypically low scores for social maladjustment, value orientation, autism, alienation, manifest aggression, and withdrawal-depression, while for repression they have an atypically high score and also high scores for denial.

The two clusters are both highly homogeneous and are virtually indistinguishable in terms of psychological statements that could be made on the basis of the Jesness dimensions.

CONCLUSIONS

The following conclusions can be drawn from the above analyses regarding the basic issues that were set forth in the introduction.

1. Does the Jesness Data Contain a "Natural" Typological Structure?

The evidence from our study of the general taxonomic structure of this sample of young adolescents suggests that six "core" types are recoverable from the personality dimensions stemming from the Jesness instrument. These are briefly identified as follows:

A. Mature, insightful, aware (17%)

B. Mature, well-adjusted, conventional, high denial (28%)

C. Shy, nervous, socially anxious, conventional (7%)

D. Withdrawn, depressed, self-critical (6%)

E. Repressed, hostile, antisocial, alienated (6%)

F. Socially maladjusted, low maturity, hostile, self-critical (28%)

2. How Strong is This "Natural" Structure?

The boundaries between these classes are weak, diffuse, and are characterized by many "intermediate" cases. The different classification techniques show a certain amount of disagreement on exactly where the class boundaries are drawn. This leads to some inconsistent classification of the cases which fall into these boundary regions. On the other hand, there is substantial overlap in the recovery of central members of the "core" clusters. This indicates that whereas the boundary conditions are weak and

relatively diffuse structures, the central tendencies and homogeneity structure of each of the natural clusters is quite strong. The direct measurement of homogeneity coefficients also confirms this conclusion. Such "core" clusters are also repeatedly identified by different cluster analytic methods with quite different approaches to the problem of categorizing. This gives encouraging evidence that the clusters are not an artifact of the statistical approaches.

3. Does the "Natural Structure" Support or Confirm the I-level Designation?

The cross-classifications between these two ways of partitioning the data indicates only low agreement. There is some agreement only at the extremes of the I-level classification where 15 out of 16 of the I_4Se and I_4Ci are classified into the larger "mature" cluster B. However, this larger natural cluster contains many other members who come from a broad range of I_3 and I_4 subcategories. Similarly, at the low maturity end of the I-level system, all of the I_2 members are classified into the large antisocial cluster F. Again, however, a broad range of I_2, I_3 and I_4 youth have been allocated into this large natural cluster. The natural groupings at the two extremes are, therefore, much larger than the I-level classes and the latter are totally submerged within the larger natural clusters.

In the mid-ranges of the cross-classification there is virtually no agreement. The two partitions almost totally violate each other. The natural clusters are unevenly spread across many different I-level classes, and vice versa.

If the natural "core" clusters can, in fact, be taken as representative of relatively pronounced homogeneities and higher sample densities (albeit with somewhat fuzzy boundaries), then a case can be made that the I-level classes artifically cut across these homogeneities, taking their memberships from a

number of quite different natural clusters, and cutting across the natural boundaries.

Furthermore, if the I-level boundaries were stronger or more clear-cut than the natural boundaries, then they presumably would have been picked up by the cluster analytic methods. Yet, none of the methods used in this study succeeded in replicating the I-level boundaries. Since we know that the natural boundaries are not particularly strong, or discontinuous, then the implication is that the I-level boundaries are probably even weaker. In a sense, our work replicates that of Butler and Adams (1963), although using different methods and samples of youth. The conclusions, therefore, are that the I-level classification must be regarded as having very weak, largely continuous boundaries between classes, and at the same time weak internal homogeneity with all of its classes except for the I_2 class and the I_4 (Se + Ci). The boundary conditions and internal homogeneities of all the intermediate I-level classes are hard to justify given the nature of the present findings.

4. The Concurrent Validity of the Jesness Dimensions Regarding Delinquency

The Jesness inventory dimensions have excellent predictive validity in relation to the delinquent behavior scale as assessed on the present sample. Five of the dimensions have zero-order correlations of beyond R=0.4 in relation to the delinquency scale, and jointly scored they correlate R=.64 with the delinquency scale. This suggests that the Jesness dimensions would provide an excellent attribute space within which to construct a typology of adolescents that would have high relevance to delinquent behavior.

5. The Concurrent Validity of the I-level and "Natural" Typology Regarding Delinquent Behavior

The correlation between the I-level and the delinquency scale is R=.296 (about 9% of the variance); the Jesness dimensions on their own accounted for over 40% of the variance in the delinquency score.

A second comparison indicated that the natural classification has far greater predictive accuracy regarding the five delinquency classes than did the I-level (Lambda's for the two comparisons were 0.138 and .005 respectively). On the other hand, both sets of results were statistically significant indicating that the I-level does, in fact, have some nonchance association to levels of delinquency.

BIBLIOGRAPHY

Ball, G.H.
1970 "Classification analysis." Technical report. Stanford Research Institute: Menlo Park, California.

Brennan, T.
1972 "Numerical taxonomy: Theory and applications." Doctoral dissertation. University of Lancaster: Lancaster, England.

Butler, E.W., and S.W. Adams
1966 "Typologies of delinquent girls: Some alternative approaches." Social Forces, 44, 401-407.

Carbonell, J.
1980 "Cross-classification of the MMPI and I-level systems." Paper presented at the International Differential Treatment Association Annual Conference, Estes Park, Colorado.

Everitt, B.
1974 Cluster analysis. Halsted Press: New York.

Jesness, C.F.
1974 "Sequential I-level classification manual." American Justice Institute: California Youth Authority.

Kornfeld, J.P., T.G. Eynon, W.A. Hetzner, M.E. Pfeiffer-Kornfeld, W. Bates, and T. Baumer
1975 State-of-the-art of offender classification in the U.S.A. IIT Research Institute: Chicago.

Loevinger, J.
1976 Ego development.

Molof, M.J., and C.F. Jesness
1972 "Project sequil: The development of a sequential I-level classification system." Yearly report. American Justice Institute: California Youth Authority (Contract A 147-70).

Pauker, J., and S. Hood
1979 "Classifying and the impossible child and adolescent. Unpublished paper. Family Court Clinic: Toronto, Ontario.

Sarbin, J.R.
1952 "A preface to a psychological analysis of the self." Psychological Review, 59, 11-23.

Smith, D.E.
1974 "Relationships between the Eysenck and Jesness personality inventories." British Journal of Criminology, 14(4), 376-84.

Sullivan, C., M.Q. Grant and J.D. Grant
1957 "The development of interpersonal maturity: Applications to delinquency." Psychiatry, 20(4), 373-85.

Werner, E.
1975 "Psychological and ethnic correlates of interpersonal maturity among delinquents." British Journal of Criminology, Delinquency, and Deviant Social Behavior, 15(1), 51-68.

CHAPTER 13

A CLASSIFICATION OF PROBATION YOUTH

The present chapter again tests and demonstrates the use of a variety of taxometric methods on a set of real criminological data. In the present case we attempt to clarify the taxonomic structure of a sample of probation youth. A local probation department used two psychometric instruments to assess each probationer at intake. It should be noted that the two psychometric instruments that are used in the present study formed only a segment of a larger range of data that were collected on each youth. For our present purposes, however, they provide an extremely useful vehicle for evaluating the use of selected classification techniques on real criminological data.

The "Experience Survey" (Cartwright and Howard, 1973) is particularly interesting in that it represents a theoretically well developed elaboration of the main factors which impinge upon a youth to bring about, or to diminish, delinquent behavior. The main factors are described more fully below, and the system itself is described in Cartwright and Howard (1973). The present data also includes the general personality descriptive system for youth, the High School Personality Questionnaire developed by Cattell, and based upon the well known 16PF system (Cattell and Eber, 1970).

Thus the present exercise is descriptive in its intent as well as being concerned with theory development. The typological analysis, particularly of the E-S scale, should be suggestive of new multivariate hypotheses regarding the joint interaction of the various E-S factors in differentially leading to delinquent behavior.

The Data

The primary classificatory variables were 18 Experience Survey scale scores (Cartwright & Howard) and 14 HSPQ scale scores on 148 juvenile offenders between the ages of 12 and 17 (the sample includes one 18 year old).

Descriptions of each of these scales are found below. There were 118 males and 30 females included in the sample. The 148 subjects were the entire population of youthful offenders handled by the Boulder County Court between the years 1972 and 1976, with the exception of twenty-one cases which were eliminated because one or both sets of classificatory measures were missing. Of the 148 cases, 97 (or 65.5%) had recorded at least one previous offense in addition to the presenting offense.

In addition to the above data, there are data on the disposition of the case, conditions of probation, revocation orders, and conditions of revocation. Additional demographic information includes birth order, number of siblings, days absent from school, grade point average, number of persons in the home, age of youth, percentages of fathers and mothers in the home, occupations of father and mother, and the distributions of the sample according to presenting offense.

Cartwright and Howard's Experience Survey

The theory which provides a basic framework for the experience survey consists of a formal model of the "forces leading to commission of a delinquent act" (Cartwright and Howard, 1973). Certain conditions are seen to give rise to forces that may exist in a causal relationship to delinquent activity. The formal model takes into consideration four general aspects of these causal conditions: the recency of the condition, the source of the condition, the types of causal relations the condition had with delinquency, and the social context in which it arose. Attention to these four aspects in relation to delinquency presupposes a psycho-social orientation to criminal behavior: delinquency (or the relative lack of it) is conceived of as resulting from forces which impinge on and motivate individual behavior, having been precipitated in specific psychological and social environments.

According to Cartwright and Howard (1973), delinquent behavior may evolve in three (not necessarily exclusive) contexts: family, school, or city. Within each of these contexts, the 'force' which will eventually lead to a delinquent act may arise from either social or individual sources (that is, either other people or the individual ego may "initiate the force"). This force is further posited as bearing five kinds of causal relations to the delinquent act: permission, instigation, facilitation, diversion, or prohibition. The components of these five categories, divided into their social and personal aspects, are presented in Figure 1 (from Reuterman, 1968). This figure could be replicated for each of the three contexts: school, city and family, to provide the total set of postulated causes of delinquency.

The Experience Survey Scales. The experience survey itself basically asks the subject to report certain experiences for a specified period of time, most often the past year. Each item of the survey, as utilized by Cartwright and Howard (1973) begins by asking, "During the past year, how often ...?" The format for response is divided into six categories, which are numbered from zero to five: never, a few times, sometimes, fairly often, quite often, and very often indeed.

Each item on the scale, according to Cartwright and Howard (1973), must meet criteria of test-retest reliability and external validity, the latter of which is achieved through positive correlation with a scale of self-reported delinquent behavior. Reuterman (1968) presents coefficients for validity with each item utilized in his version of the scale. All items (in the case of Cartwright and Howard 1973) were also subjected to examination for inter-rater agreement in regard to their "proper cell membership" within the model (a

FIGURE 1

Two-fold Classification of Factors in the Experimental Survey (ES)

	Social	Individual
Instigators	(a) people who lure (b) people who tempt (c) people who force (d) movies (e) magazines	(a) wish for kicks (b) wish for status (c) wish for respect (d) desire for material goods (e) role image of tough guy or smart guy
Facilitators	(a) fences and other assistants (b) opportunity in open doors, cars, etc. (c) density of population	(a) skills (b) natural abilities (c) emotional control
Permittors	(a) people who don't care (b) people who are afraid to interfere (c) people who want to avoid involvement	(a) mental retardation (b) ignorance (c) emotional instability (d) alienated attitudes
Diverters	(a) church youth clubs (b) parks and recreation (c) work program (d) schools' interesting activities	(a) hobbies (b) habits of being busy (c) social interest
Prohibitors	(a) police (b) courts (c) teachers (d) strong parents (e) vigilant citizens	(a) conscience (b) fear of punishment (c) self-concept (d) long-term image for career, position in society, etc.

"cell" being each of the 30 categories which result from crossing the six contexts and sources with the five relations of motivating forces to delinquency; one cell would thus be Social City Instigator, another would be Individual School Prohibitor, etc.). Raters were required to place items in the proper cell of the model, and only items with an "acceptable consensus" were retained.

Each scale consists of a number of items which are then divided into six batches, one batch for each of the context and source points (social-city, individual-city, social-school, individual-school, social-family, individual-family). Within each of these groups, there are placed an even number of items which are judged to correspond with each of the five causal relations: instigation; facilitation; permission; diversion, and prohibition. Each relevant variable is thus identical to a single "cell", as defined above. Of the items corresponding to each cell, half were phrased with an adult as the agent or referent, and half with peers, friends, or other children as the agent or referent. This procedure is followed partially in order to determine whether or not peer influence in reference to delinquency actually exists and if so where it can be found. The scales and subscales are designed to discover relevant factors existing within each cell.

The subscales used for this research are the 18 scores resulting from the six contexts (social-city, individual-city, etc.) described above for (1) adults as agents, (2) peers as agents, and (3) a composite pattern of experience (Cartwright & Howard, 1973). A high score on any of these 18 scales indicates a high proneness to delinquency contributed by that component.

The names of these 18 scales are as follows:

1. Social-City
2. Social-School
3. Social-Family

4. Individual-City
5. Individual-School
6. Individual-Family
7. Peer-Social-City
8. Peer-Individual-City
9. Peer-Social-Family
10. Peer-Individual-School
11. Peer-Social-Family
12. Peer-Individual-Family
13. Adult-Social-City
14. Adult-Individual-City
15. Adult-Social-School
16. Adult-Individual-School
17. Adult-Social-Family
18. Adult-Individual-Family

Multiple-factor System[1]

Definitions

Instigators — Provide the push or impulse to action

A. Social: Direct personal initiation of delinquent act, luring to immorality, daring to steal, forcing to carry weapons, inviting the person to commit delinquency.

Indirect personal provocation, unjustly depriving, stressing, aggravating, frustrating the person to the point where some kind of retaliation against the aggressor or a substitute is likely.

Impersonal provocation, such as poverty "an added spur to dishonesty" (Burt): overexciting movies, magazines, stimulating to violence, immorality, etc.

B. Individual: Conscious wishes for goods, excitement, reputation, etc..: fulfilment of personal ideal

Unconscious conflict generates tension for which release is sought through delinquency; delinquency is displaced aggression or symbolic search for affection; act is compensatory for some inferiority.

Facilitators — Provide the means for carrying the impulse into actual behavior

---

[1]Revised and expanded from "A new multiple-factor approach to delinquency", by Desmond S. Cartwright, Nicholas A. Reuterman, and Richard D. Vandiver. University of Colorado, 1966, Mimeo.

A. Social: Assistance and opportunities. People show the person how to commit the act, or provide necessary information or tools or other means. Opportunities arise from unguarded property, left around, keys in car, doors unlocked; dark alleys, crowded areas, places to hide, etc.

B. Individual: Physical capabilities (e.g. small enough to get through little window; skills (e.g. can make a zip gun); emotional qualities (e.g. can stay cool under pressure)

Permittors — Provide sanction or tacit encouragement for an act

A. Social: Deviant example set by parents or citizens; parents, teachers, other code-bearers lax or indifferent in discipline; people don't care or afraid to interfere in violations; friends offer protection, strengthen rationalizations, etc.: enforcement inefficient or corrupted; societal anomie.

B. Individual: Personal alienation; don't-care, so-what, nothing-to-lose-attitude; emotional instability, mental deficiency, ignorance of laws.

Diverters — Militate against commission of delinquent acts by preoccupying time, interest and energy in other ways

A. Social: Recreational programs that "keep youth off the streets" and involve them in wholesome activities; work in the usual employment sense and in the special sense of organized work programs, again designed to occupy and provide remunerative satisfaction; school and pre-employment training; special enterprises such as Big Brothers; encouragement by members of the family to engage in wholesome activities, both intra- and extra-family.

B. Individual: The personal side of involvement in work, school or recreation: cultural interests, ability to develop sustained attention and investment of energy in constructive enterprise, hobby, sport; ability to develop trusting, admiring, affectional relationships with others, especially with persons representing stable norms of society, teachers, coaches, and so on.

Prohibitors — Directly forbid, block or threaten penalties for committing delinquent acts; or otherwise directly oppose delinquency

A. Social: Parental supervision, control, discipline; proper supervision counsel and enforcement from other authorities such as teachers and police; citizen intervention; church teachings.

B. Individual: Conscience, personal scruples, disgust at violations; fear of being caught, punished, loss of social status; stake in community and sense of long-run importance to own self-image as good citizens; stake in family and complete unwillingness to risk bringing trouble.

Social Factors — Are those which arise outside the person, outside of his skin, so to speak.

The ultimate origins may be in some economic or physical conditions of the environment, even the wider environment of a nation. But in the present scheme the factors must be in some way impinging quite directly upon the individual. Thus a general state of anomie in the nation would have to result in there being particular people who display normless behaviors in the immediate local environment of the child in question before such conditions could be included as social factors for the present analysis.

Individual Factors — Are those which are within the person himself, in his physical condition, his thoughts, feelings and impulses, and also in his behavior.

A deficiency in physical stature may itself facilitate entrance through exceptionally small passageways, and thus be a factor. If a compensatory striving for importance arises, it is this striving which is the factor. If a child wants money and steals it, the wish as such is the individual instigator here; regardless of the fact that the child's family might be very poor. The poverty of the family would be another factor, a social one. And though the poverty of the family and the child's individual wish for money might seem to be closely connected in this instance, they might not be in another instance, where for example the wish for money derives mainly from symbolic needs for resolution of some inner conflict over loss of love to a sibling. Analytically, the wish for money and the state of poverty are separable and must be kept separate in the present system.

It is expected that, in the usual case, a considerable number of individual factors and also a large number of social factors will be at work in the production of a given delinquent act.

Cattell's High School Personality Questionnaire (HSPQ)

The High School Personality Questionnaire (HSPQ) is an instrument developed in order to measure the basic personality factors in high school age individuals. The factors used in this scale are drawn from the more general 16PF scale, (Cattell, Eber, and Tatsuoka, 1970). The common structure of these two sets of scales makes it possible to generate comparisons of the personality structure of people at different ages. The HSPQ, however, does not include factors L, M, N or Q_I, which are part of the 16PF; it does include factors D (undemonstrative vs. excitable) and J (zestful vs. circumspect) which are both specific to the HSPQ.

The fourteen scales of the HSPQ, then, are summarized as follows:

Scale	Brief Description
1. Factor A	Sizothymia (reserved, detached, critical) vs. Affectothymia (warmhearted, outgoing)
2. Factor B	Intelligence (Low B-, High B+)
3. Factor C	Ego strength (Low C-, High C+)
4. Factor D	Temperament (phlegmatic D-, Excitable D+)
5. Factor E	Submissive vs. Dominant
6. Factor F	Surgency (Sober C-, Happy-go-lucky F+)
7. Factor G	Superego strength (Disregards rules G-, Conscientious G+)
8. Factor H	Threctia (Shy) vs. Parmia (Adventurous)
9. Factor I	Harria (Tough-minded) vs. Premsia (Sensitive)
10. Factor J	Zeppia (Zestful) vs. Coasthenia (Reflective)
11. Factor O	Untroubled Adequacy vs. Guilt Proneness
12. Factor Q_2	Group Dependency vs. Self Sufficiency
13. Factor Q_3	Self-Sentiment (Uncontrolled Q_3-, Controlled Q_3+)
14.Factor Q_4	Ergic Tension (Low Q4-, High Q4+)

RESEARCH SEQUENCE

The data were subjected to a number of clustering routines. Table 1 below lists all of the analyses together with the characteristics of each. Dendrograms (trees) were created for all of the hierarchical clustering results. The trees and the levels of the trees to be subjected to further analysis were chosen using four criteria. These criteria were: (1) clusters formed early in the process, (2) clusters remaining intact for a reasonable time before being combined with other clusters (long branch line), (3) clusters containing more than four members, and (4) clusters where similar groupings were arrived at by several different methods. The first three criteria for this selection procedure are obviously quite subjective. We shall discuss the fourth criterion in more detail in the following paragraphs.

By this procedure, three partitions of the data were chosen for further analysis; the first clustering was selected from the results of UPGMA clustering on the 18 ES measures using Euclidean distance (a three group solution), the second and third partitions were chosen from the results of UPGMA clustering on the HSPQ measures using the correlation coefficient (a four group and a five group solution).

Results were evaluated in four steps: (1) Alternative "natural" partitioning levels were selected for further analysis by visual inspection of the dendrograms produced by the clustering routines; (2) these partitions were compared pairwise by cross-tabulation of class memberships, and by computing several agreement coefficients; (3) each "most likely" partition was then subjected to a number of statistical tests to determine adequacy of the implied typologies, using both the original classificatory variables and a set of standard demographic data describing the offenders, and finally, (4) each constructed typology was "validated" against an actual behavioral

classification based on the presenting offense of the subject. This comparison also produced a set of cross-tabulations and associated measures of agreement between the constructed personality types and the offense classification.

Table 1

Analytic Methods Used to Cluster
the Boulder County Probation Department Data

Method	Variables	Similarity	# Groups
	ES Clustering		
*UPGMA	ES (18)	Euclid. Dist.	3
HMODE (30 outliers)	ES	Euclid. Dist.	5, 3
NMODE	ES	Euclid. Dist.	one mode
NMODE	ES	Euclid. Dist.	2
	HSPQ Clustering		
*UPGMA (group ave.)	HSPQ (14)	Correlation	(11), 5, 4
HMODE (30 outliers)	HSPQ	Correlation	9, 6
UPGMA	HSPQ	Euclid. dist.	7, 3
HMODE (25 outliers)	HSPQ	Euclid. dist.	single group
NMODE (126 outliers)	HSPQ	Euclid. dist.	single group N = 22

* partitions selected for further analyses

RESULTS

Comparison of Clusterings of Experience Survey (ES) Data

Natural mode search (NMODE) found 2 clusters (78 and 10 points respectively) and 60 outliers. NMODE tends to define clusters more strictly than other methods.

The tree produced by the group average method indicated 3 clusters of 98, 22, and 18 points respectively, with 10 outliers.

The hierarchical mode search (HMODE) tree indicated a 3-cluster solution, although the 5-cluster level of the tree was also considered. The 3-cluster solution compares most favorably with those produced by the other methods. Clusters contain 97, 14, and 5 points respectively, with 32 outliers.

Partition matching

The tables below show the number of points held in common by clusters from different partitions. In Table 2, for example, cluster 1 of the group average partition contains 22 points. Fourteen of those fell in cluster 1 of the HMODE partition, 5 in cluster 3 of the HMODE partition, and 3 were classified as outliers by HMODE. Cluster 2 of the group average partition contains 98 points, 89 of which fell in HMODE's cluster 2, while the other 9 were classified as outliers by HMODE. There is substantial agreement between the two partitions on clusters 1 and 2, but cluster 3 in each case seems to represent a small piece broken from one of the two main clusters. This may indicate that there are only two clusters with a number of points loosely scattered around and between them or that a third, and possibly even a fourth, cluster exists quite close to the other two. All 10 of the points left as outliers under the group average method were also classified as outliers by HMODE. The value of symmetric lambda for the cross-tabulation is .55.

Table 2

Cluster Overlap Matrix Between
Two Partitionings of the ES Data

GROUP AVERAGE
(UPGMA)

	Cluster	1	2	3	Outliers (Unclassified)	Totals
	1	14	0	0	0	14
HMODE	2	0	89	8	0	97
	3	5	0	0	0	5
Outliers (Unclassified)		3	9	10	10	32
Totals		22	98	18	10	148

Lambda (Symmetric) = .455

Table 3 shows that HMODE agrees completely with the large cluster produced by NMODE, but splits the small cluster into two parts. All 32 of HMODE's outliers were classified as outliers by NMODE as well. Symmetric lambda is .521 — higher than in the other tables partly because of the greater agreement among outliers.

Table 3

Cluster Overlap Between Two Partitions of ES Data

NMODE

	Cluster	1	2	Outliers (Unclassified)	Totals
	1	0	6	8	14
HMODE	2	78	0	19	97
	3	0	4	1	5
Outliers (Unclassified)		0	0	32	32
Totals		78	10	60	148

Lambda (Symmetric) = .521

The clustering methods show substantial agreement on one large dense cluster. But they disagree somewhat on the one or two smaller, sparser peripheral clusters, indicating that boundaries between these sparse clusters are not clearly defined.

Suspecting that some combination and improvement could be obtained on the sample by specifying three clusters (K=3) and computing clusters by K-means (MIKCA) methods, two runs of MIKCA were obtained using different initial cluster centers. These runs turned out to give two only slightly different solutions. MIKCA - trial A seemed to place centroids much closer together than the centroids obtained by UPGMA, HMODE, or NMODE. How these MIKCA runs agreed with each other and with the other clustering results is summarized in the next section.

The three basic clusters identified within this data are shown below as a plot in discriminant space.

EVALUATION OF ES CLUSTERING RESULTS

The evaluation of the clusters chosen for further scrutiny consists of two tasks: (1) to examine the structural characteristics of each partition and (2) to summarize and interpret the substantive characteristics of each cluster in the various partitions. The following three ES partitions have been chosen for these more detailed evaluations:

1. Gp Average - Euclid. d on ES variables, 3 group solution
2. MIKCA (K-means) - on ES variables, 3 groups - trial A
3. MIKCA (K-means) - on ES variables, 3 groups - trial B

Structural Characteristics of Solutions

Comparisons of these partitions, including various measures of agreement are shown in Tables 4 and 5.

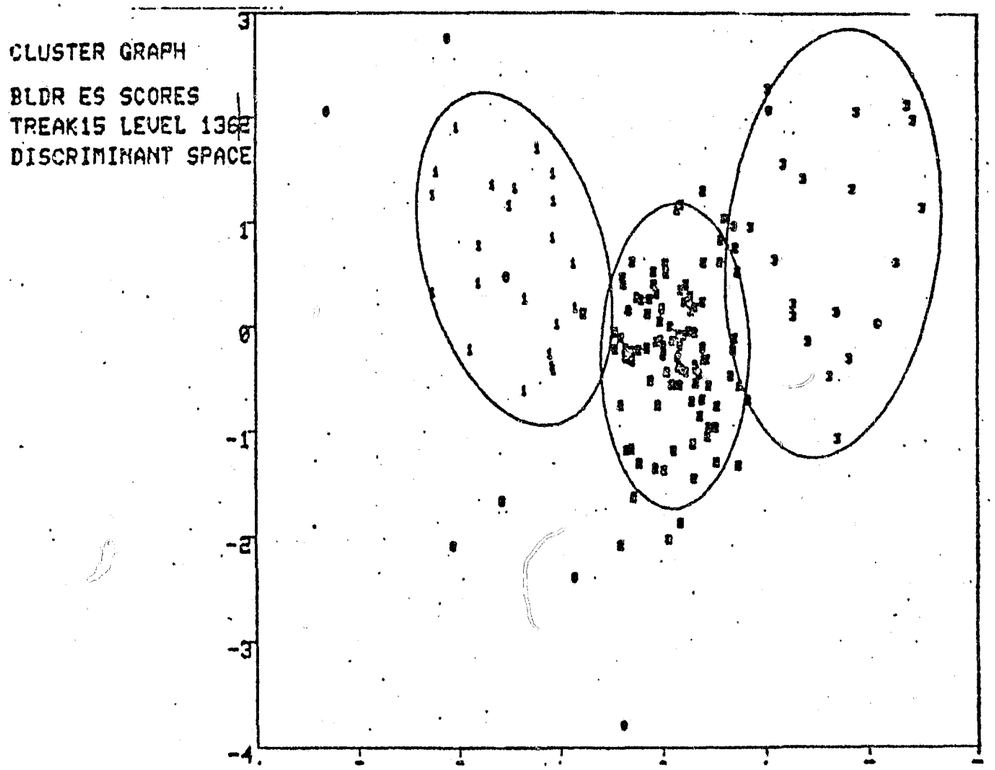


Table 4

Partition Matching Coefficients (ES Comparisons)

	Rand	Jaccard	Cramer	Lambda
NMODE,ES(2)vs.UPGMA,ES(3)	.767	.526	.621	.367
NOTE: (NMODE's 2 clusters agree well with 2 of the Gpav ES clusters. Large number outliers lowers the coefficients.)				
HMODE,ES(5)vs.UPGMA,ES(3)	.731	.489	.635	.385
HMODE,ES(3)vs.UPGMA,ES(3)	.834	.692	.648	.455
HMODE,ES(3)vs.NMODE,ES(2)	.838	.633	.727	.521
MIKCA,ES(3A)vs.UPGMA,ES(3)	.592	.381	.526	.190
MIKCA,ES(3B)vs.UPGMA,ES(3)*	.909	.834	.904	.755
MIKCA,ES(2)vs.UPGMA,ES(2)	.595	.425	.511	.231
MIKCA,ES(3A)vs.MIKCA,ES(2)	.690	.517	.677	.458
MIKCA,ES(3A)vs.MIKCA,ES(3B)	.591	.409	.571	.236
MIKCA,ES(2)vs.MIKCA,ES(3B)	.582	.437	.581	.224

*MIKCA trial B with K=3 produced clusters almost identical with UPGMA - 3 group solution.

Variable-by-cluster homogeneities, as well as the centroid vectors, distribution indices, etc., for each partition were computed for several possible clustering solutions. The results are presented in evaluation tables 6 and 7. These tables indicate that the highest cluster and variable homogeneities are found in the UPGMA - 3 group solution using ES scores. They are generally higher even than either of the MIKCA - 3 group solutions also using ES scores.

Homogeneity Characteristics of the Selected ES Partitions

Empirical measures used to assess the homogeneity of the partitions are the Gamma measures; the homogeneity measures discussed earlier. The contribution of each variable to the clustering is assessed by examining their homogeneities both within and across the clusters. All of these considerations on the partitions chosen are presented and interpreted in the next section. Homogeneity measures and distribution indices of the clusters are shown below.

Table 5

Homogeneities and Distribution Indices for
UPGMA - Euclidean Distance - 3-group Solution

	Cluster 1	2	3
N	22	98	18
Homogeneity	.507	.482	.412
Dist. Index	.536	.851	.203

The homogeneities of variables and variable-by-cluster statistics are found in the evaluation tables below. The analysis indicated that the larger cluster (N=98) is more compact than the other two based on the ES scores. Looking at the variable means associated with each cluster, it is relatively easy to see that cluster 1 (N=22) is characterized by low ES scores, cluster 2 (N=98) by ES scores near the mean and cluster 3 (N=18) by high ES scores. Since high ES scores are supposed to be associated with greater amounts of delinquent behavior, we would expect to find that group 3 represents a population of significantly higher delinquency. This expectation will be examined in later paragraphs when we look at the composition of the clusters in more detail. It is interesting that NMODE methods could detect groups 1 and 2 but could not detect group 3.

Table 6

Boulder County Probation Data - ES Scores, Standardized
No outliers removed (N=148)
Method: UPGMA Similarity: Euclidean Distance

ES Variables	Cluster 1 N=22		Cluster 2 N=98		Cluster 3 N=18		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
Soc. City	-1.064	.343	.130	.538	1.123	.293	.475
School	1.218	.632	.184	.644	1.215	.599	.631
Family	-1.168	.480	.210	.332	.341	.505	.379
Indiv.							
City	-1.065	.582	.016	.596	1.536	.554	.589
School	-1.313	.493	.121	.695	1.245	.688	.662
Family	-1.390	.677	.091	.684	1.348	.415	.648
Peer:							
Soc. City	.969	.421	-.019	.536	1.248	.264	.482
Ind. City	-954	.657	.017	.410	1.329	.311	.437
Soc. School	-1.111	.468	.098	.554	1.275	.461	.528
Ind. School	-.984	.154	.099	.462	1.035	.828	.461
Soc. Family	-.958	.195	.157	.355	.653	.493	.347
Ind. Family	-1.114	.505	.117	.469	1.046	.559	.486
Adult:							
Soc. City	-.469	.428	.105	.161	.511	.317	.224
Ind. City	-.641	.580	-.017	.464	1.285	-.049	.415
Soc. School	-.870	.494	.181	.352	.398	.113	.343
Ind. School	-.908	.686	.062	.489	1.141	.329	.500
Soc. Family	-1.124	.674	.186	.378	.500	.423	.431
Ind. Family	-1.182	.651	.059	.558	1.304	.350	.545
	Dist. Index= .536		Dist. Index= .851		Dist. Index= .203		
	MSD= 8.470	H=.507	MSD= 9.229	H=.482	MSD= 9.995	H=.412	

Table 7

Boulder County Probation Data - ES Scores, Standardized
N=148

Method: MIKCA-B Similarity: Mahalanobis Distance

ES Variables	Cluster 1 N=24		Cluster 2 N=104		Cluster 3 N=20		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
Soc. City	-1.342	.389	.094	.500	1.122	.364	.463
School	-1.447	.426	.108	.545	1.174	.585	.531
Family	-1.221	-.001	.106	.303	.395	.529	.284
Indiv.							
City	-1.156	.423	.035	.586	1.570	.494	.547
School	-1.423	.118	.086	.643	1.261	.538	.543
Family	-1.366	.401	.056	.623	1.347	.317	.546
Peer:							
Soc. City	-1.044	.380	.012	.460	1.314	.136	.403
Ind. City	.972	.611	-.033	.343	1.335	.370	.390
Soc. School	-1.403	.604	.103	.562	1.148	.050	.500
Ind. School	-1.129	-.267	.080	.380	.941	.735	.323
Soc. Family	-1.187	.196	.151	.267	.641	.545	.293
Ind. Family	-1.208	.291	.083	.370	1.018	.507	.376
Adult:							
Soc. City	-.813	-.090	.088	.147	.520	.368	.138
Ind. City	-.807	-.202	-.043	.460	1.189	-.033	.286
Soc. School	-1.188	.155	.214	.315	.312	.135	.265
Ind. School	-1.104	.288	.023	.442	1.205	.220	.387
Soc. Family	-1.344	.562	.200	.306	.572	.404	.360
Ind. Family	-1.294	.572	.048	.516	1.305	.325	.499
	Dist. Index= .715		Dist. Index= .852		Dist. Index= -.107		
	MSD= 12.597	H=.270	MSD= 10.135	H=.431	MSD= 10.841	H=.366	

Conclusions regarding ES Classification

The results of this section of the work on the probation sample indicate that within the context of the ES variables, the present sample is not strongly clustered. The agreement coefficients are relatively low, indicating that the various cluster methods could not find stable and reliable partitions. The three-cluster level that was selected for special study is shown by the evaluation tables to represent simply three different levels of delinquency proneness. Furthermore, the homogeneity coefficients of these three clusters are not high. None of them has a homogeneity much in excess of H=0.50. The cluster graph confirms our sense that within the ES attribute space we are dealing with a rather continuous distribution.

CLUSTERING RESULTS FOR THE HSPQ

Group average and hierarchical mode search (HMODE) methods were also used to cluster data on the 14 HSPQ variables. Both Q-correlation and distance were used to measure similarity between subjects. The latter attempt did not produce any reliable clusters. Apparently, the profile shapes, not the levels of the personality measures, are taxonomically significant.

Agreement among HSPQ partitions

Together with the overlap matrices, various measures of overall agreement among clusters are available. Table 8 shows only the significant results of a large number of partition comparisons. The homogeneity measures for each variable by group for selected solutions are found in Tables 9 and 10. Inspection of the group average clustering led to the choice of a 5-cluster solution, with clusters containing 56, 40, 19, and 14 points respectively and 3 points left as outliers. The best level of the HMODE tree appears to be the 6-cluster level, where clusters contain 52, 41, 10, 4, 4 and 4 points. It is

possible that the three 4-point clusters represent only minor variations in a generally sparse background of points, however, indicating the possibility of fewer clusters. HMODE left 33 outliers.

Partition matching

From the overlaps between clusters in the group average and HMODE partitions, it is apparent that there is substantial agreement on the two large clusters. HMODE's cluster #1 matches fairly well with the group average cluster #2, and HMODE's #2 matches well with group average's #3. The smaller clusters, however, show virtually no relationships across the two partitions. Such instability suggests that these are not true clusters, but chunks of fairly uniform background noise instead. (Symmetric lambda for this comparison is .516). Table 8 below summarizes the results of these comparisons:

Table 8

Partition Matching Coefficients
HSPQ Comparisons

	Rand	Jaccard	Cramer	Lambda
UPGMA,HSPQ,r(11)vs.HMODE,HSPQ,r(9)	.881	.389	.558	.514
UPGMA,HSPQ,r (5)vs.HMODE,HSPQ,r(9)	.809	.327	.546	.485
UPGMA,HSPQ,r (5)vs.HMODE,HSPQ,r(6)	.815	.419	.511	.516

Cluster Profiles based on HSPQ

The gamma and homogeneity measures as well as means, etc., are geared to a "points-in-space" model and, therefore, do not reflect the cluster characteristics of the solution from the HSPQ variables which was based on similarity of profile shape. It has already been noted that the ES clusters and the HSPQ clusters are not directly comparable. It will then be necessary to evaluate the two HSPQ clustering partitions by comparing the profiles of the cluster centroids rather than by comparing the clusters in Euclidean space.

Table 9

Boulder County Probation Data - HSPQ, Standardized
No Outliers Removed (N=148)

Method: UPGMA Similarity: Correlation Coefficient

HSPQ Variables		Cluster 1 N=19		Cluster 2 N=56		Cluster 3 N=40		Cluster 4 N=14		Cluster 5 N=16		Variable Homo-geneity
		Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
Reserved/Warm(+)	A	.920	.565	-.347	.212	.364	.350	-.085	-.767	-.605	.211	.202
Dull/Bright(+)	B	-.808	.321	.101	.349	-.109	-.090	.953	.068	.201	.047	.164
Feelings/Stable	C	.155	.832	-.626	.384	1.001	.219	-.416	.547	-.292	.647	.442
Undemonstrative/ Excitable	D	-.127	.568	.765	.596	-.762	.428	-1.094	.272	.316	.316	.483
Obedient/ Assertive	E	-.282	.413	.120	-.147	-.050	-.276	.097	.343	-.075	.195	-.024
Sober/ Enthusiastic	F	.080	.213	.422	.238	-.312	-.208	.195	.746	-.947	.263	.164
Disr. Rules/ Conscientious	G	.100	.244	-.616	.431	.696	.263	-.375	.218	.724	.296	.324
Shy/Adventurous	H	.027	.513	-.308	-.003	.696	.266	.108	.045	-.580	.479	.196
Tough/Tender	I	.125	.622	-.571	.110	.451	.260	.050	-.204	.743	.281	.207
Zestful/ Circumspect	J	-.605	.352	.102	-.120	-.259	.102	.502	.208	.491	.498	.103
Self-Assured/ Apprensive	O	.139	.754	.634	.266	-.948	.382	-.245	.156	.195	.582	.386
Group Dep./ Self-sufficient	Q_2	.035	.692	-.112	-.181	-.249	-.094	.693	.292	.295	.184	.043
Uncontrolled/ Controlled	Q_3	.014	.485	-.535	.292	.723	.127	-.369	.377	.452	.115	.261
Relaxed/Tense	Q_4	.190	.092	.629	.386	-.786	.329	-.321	.208	-.003	.431	.320
		Dist. Index= .536		Dist. Index= .783		Dist. Index= .719						
		MSD= 6.948	H=.476	MSD= 10.989	H=.201	MSD= 11.645	H=.147	MSD= 10.669	H=.179	MSD= 8.863	H=.325	

CONTINUED

7 OF 9

Table 10

Boulder County Probation Data - HSPQ, Standardized
No Outliers Removed (N=148)
Method: UPGMA Similarity: Correlation Coefficient

HSPQ Variables		Cluster 1 N=59		Cluster 2 N=56		Cluster 3 N=40		Cluster 4 N=14		Variable Homo-geneity
		Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
Reserved/Warm(+)	A	.543	.360	-.347	.212	-.085	-.767	-.600	.307	.190
Dull/Bright(+)	B	-.334	-.053	.101	.349	.953	.068	.036	-.008	-.116
Feelings/Stable	C	.728	.264	-.626	.384	-.416	.547	-.110	.520	.369
Undemonstrative/ Excitable	D	-.557	.390	.765	.596	-1.094	.272	.281	.414	.460
Obedient/ Assertive	E	-.125	-.052	.120	-.147	.097	.343	-.038	.305	-.004
Sober/ Enthusiastic	F	-.186	-.091	.422	.238	.195	.746	-.811	.249	.156
Disr. Rules/ Conscientious	G	.504	.191	-.616	.431	-.375	.218	.527	.164	.281
Shy/Adventurpus	H	.480	.256	-.308	-.003	.108	.045	-.664	.513	.171
Tough/Tender	I	.346	.361	-.571	.110	.050	-.204	.572	.235	.197
Zestful/ Circumspect	J	-.370	.169	.102	-.120	.502	.208	.479	.427	.096
Self-Assured/ Apprensive	O	-.598	.246	.634	.266	-.245	.156	.168	.615	.292
Group Dep./ Self-sufficient	Q_2	-.158	.151	-.112	-.181	.693	.292	.311	.288	.056
Uncontrolled/ Controlled	Q_3	.495	.141	-.535	.292	-.369	.377	.313	.114	.217
Relaxed/Tense	Q_4	-.472	.055	.629	.386	-.321	.208	-.153	.389	.238
		Dist. Index= .826		Dist. Index= .783		Dist Index= .543				
		MSD= 11.415	H=.171	MSD= 10.989	H=.201	MSD= 10.669	H=.179	MSD= 8.970	H=.324	

The following represent the psychological descriptions of the five clusters generated by the UPGMA method using the correlation coefficient as a similarity measure. The cluster graph indicates the relationship of these clusters to each other in discriminant space.

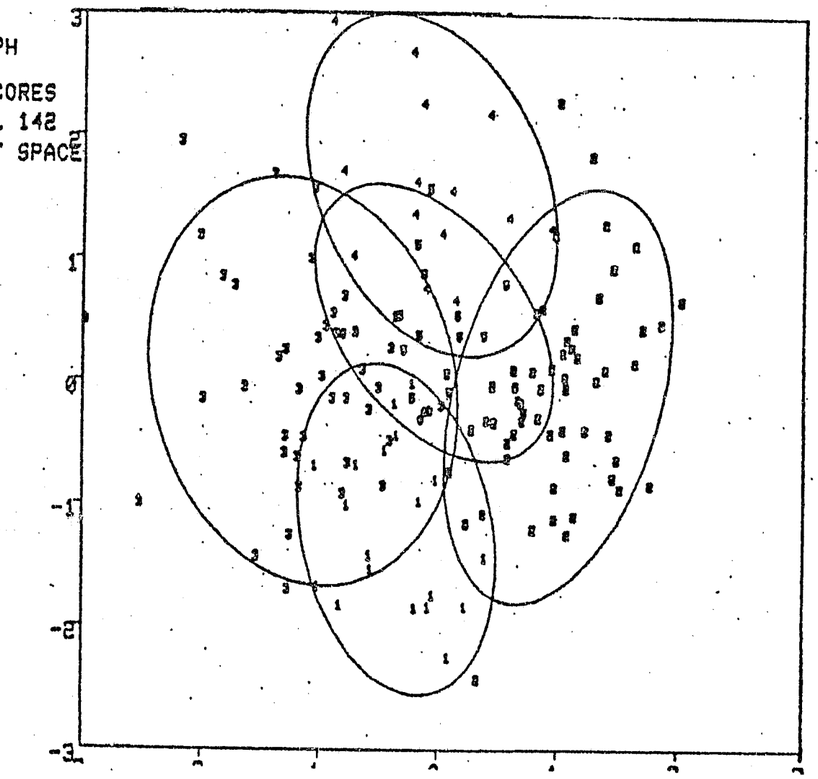
Cluster 1 (N=19) - Sociable, Outgoing but Dull. This cluster has an HSPQ profile of (A+, B-, J-). This indicates a warm, outgoing, group-joining youth who has low mental capacity.

Cluster 2 (N=56) - Weak Superego and Weak Ego Strength. This cluster is characterized by the profile (C-, D+, G-, O+, Q_3- and Q_4+). Ego weakness (C-) implies that the youth worries, easily becomes emotional and upset, and gives up easily. The D+ score implies an impatient, demanding, jealous, egotistical person. G- implies an undependable, fickle person lacking in superego strength. O+ and Q_4+ imply guilt proneness, anxiety, moody, as well as tense, fretful, and driven orientation.

Cluster 3 (N=40) - Calm, Strong Youth. This cluster is characterized by the profile (C+, D-, G+, H+, O-, Q_3+, and Q_4-). These youth have higher ego strength (stable, calm, unruffled) coupled with a complacent, undemonstrative temperament. Superego strength (G+) is also demonstrated. They tend to be friendly, socially active, carefree, and self-assured. The combination of Q_3+ and Q_4- indicates they are strong in self-sentiment (controlled and compulsive) as well as relaxed, composed, and unfrustrated.

Cluster 4 (N=14) - Intelligent, Calm Youth. This small group has the profile (B+, D-, Q_2+). This indicates intelligence, phlegmatic temperament, and self-sufficient resourcefulness.

Cluster 5 (N=16) - Reserved, Introspective, Responsible Youth. This small group is characterized by the profile (A-, F-, G+, I+). Thus, they are reserved and aloof, sober and introspective, and characterized by higher than



average superego strength. The I+ score indicates a tendency toward sensitivity, dependency, and insecurity.

ES Clusters versus HSPQ Clusters: Are They Related?

The partition produced by the group average method using ES scores is compared with that using HSPQ scores in Table 11. The 98 point ES cluster is distributed almost uniformly across the five HSPQ clusters, indicating virtually no relationship between that large group and the HSPQ clusters. The dominance of this group produces a very low symmetric lambda of .099.

Table 11

Cluster Overlap Matrix
ES vs. HSPQ Clusters

		UPGMA ES CLUSTERS				
	Clusters	1	2	3	Outliers	Totals
	1	1	17	0	1	19
	2	2	39	(12)	3	56
UPGMA HSPQ CLUSTERS	3	(14)	19	2	5	40
	4	0	11	3	0	14
	5	4	10	1	1	16
	Outliers	1	2	0	0	3
	Totals	22	98	18	10	148

Lambda (Symmetric) = .099

Focusing on the two smaller ES clusters, the possibility of a relationship with the two major HSPQ clusters can be seen. ES cluster #1 falls predominantly in HSPQ cluster #3, while ES #3 falls mainly in HSPQ #2. This suggests the possibility that the two smaller ES clusters may represent two distinctly different personality types as measured by the HSPQ.

DISCUSSION

The present exercise has not yielded results that are particularly interesting, either substantively or methodologically. The classifications based on the ES seem to represent a rather continuous distribution, with the three clusters showing very strong continuity (fuzzy boundaries) with each other. Substantively, this result is easy to interpret. The large central group is bounded on two sides by two smaller extensions of the distribution. One of these has high ES scores (high delinquency proneness) and the other has low ES scores (low delinquency proneness).

The HSPQ data also seems to be a relatively continuous distribution. The clusters that were found were not highly reliable in terms of their boundaries, although moderate stability was noted in the agreement analyses. The substantive interpretation of the clusters is hampered by the fact that the homogeneities of the clusters are poor. Thus, the central profiles of these clusters are not good prototypes, or representatives, of the clusters.

A possible reason for the failure of this exercise to yield reliable results may stem from the nature of the sample. A broad age range (12 through 17) tends to add complexity and developmental diversity to the score patterns in the data. The sample size was unfortunately too small to be broken down into age-homogeneous sets, and this option was not taken. Secondly, the data was collected continually over a rather long period of time (five years). This latter characteristic might also be expected to introduce generation-effect changes into the data, which again would tend to blur any boundaries. Additionally, although the administration and testing procedures were in place continuously during this time period, it is our understanding that different "testers" had to be hired/trained/replaced during the five-year period. Thus,

an additional possibility of variation in testing procedures might also have contributed to the lack of clarity of the basic data. For these reasons, therefore, we cannot with certainty make the claim that there is no taxonomic structure within probation youth as tested with the HSPQ and the ES. The present data set gives only sketchy evidence of such taxonomic structure, and we would recommend that other probation samples be studied before this conclusion of weak structure is accepted. An obvious alternative interpretation is that this data, in fact, does not contain any clearly demarcated type-structures and that the methods are accurately reflecting this absence.

BIBLIOGRAPHY

Cartwright, D.S., and K. Howard
1973 "The psychological structure of peer-group forces in delinquency." In Multivariate analysis and psychological theory. Academic Press: New York.

Cattell, R.B., H.W. Eber, and M.M. Tatsuoka
1970 Handbook for the Sixteen Personality Factor Questionnaire (16PF). Institute for Personality and Ability Testing: Champaign, Illinois.

Reuterman, N.A.
1968 A new multiple-factor approach to delinquency and its application to types of juvenile offenders. Doctoral dissertation. University of Colorado: Boulder, Colorado.

CHAPTER 14

CLASSIFICATION BY PERSONALITY OF A PRISON SAMPLE

A large literature exists in criminology examining the issue of the personality make-up of criminals. Many classification systems have been proposed that are based upon differential personality structures (reviews are provided by Megargee & Bohn, 1979; Solomon, 1977; Ferdinand, 1966; and others). The vast majority of prior systems stem from theoretical arguments rather than empirical analysis and have serious problems in regard to verification, operationalization, and validation. For these reasons they have generally been difficult or impossible to apply in practical settings.

There are only a handful of classification systems based on personality that have been generated by multivariate taxonomic methods. The MMPI studies of offender types by Megargee and Bohn (1979) and Blackburn (1971) stand out as good examples of the empirical approach. Virtually no multivariate taxonomic work has been conducted using Cattell's 16 factor system of personality assessment on prison samples. The intent of the present chapter, therefore, is to conduct an examination of the performance of selected taxometric techniques for the analysis of personality types among prisoners. A large set of data was kindly made available to the present project by the research and administrative staff of the Colorado State Penitentiary at Canon City, Colorado. This provision of data was also made feasible by the assistance of Dr. Herb Eber.

The goals of the present chapter are multiple. First, to examine the personality differentiation which might exist within a prison sample (all males) assessed by the Cattell 16 Personality Factors (16PF) and the Clinical Assessment Questionnaire (CAQ). Second, the present analyses will be used to further our knowledge of the diverse kinds of taxometric methods as they are applied to real criminological data. The particular methods used in the present chapter are identified in the "research sequence" section below.

Personality Characteristics of Prisoners and Criminals as Identified by Cattell's 16PF Test

A large amount of prior research has examined the global personality characteristics of criminals in a variety of different settings. Most of this research is reported in Cattell, Eber, and Tatsuoka (1970). Since our present research aims to amplify the personality descriptions of prisoners according to the 16PF dimensions, we now present the major findings regarding global personality characteristics of prisoners. The following represent the major characteristics established by prior research.

Ego Weakness, Low C (Emotional Instability. It has been established that criminals and prisoners have low scores for Factor C in the 16PF. This implies that the person becomes emotional when frustrated, is easily perturbed, tends to evade responsibilities, may easily give up, tends to worry, and often gets into fights in problem situations. This absence of emotional stability indicates that the person tends to be changeable and easily upset. There is a tendency toward disorganization and uncontrollability. The person with a low score on C tends to be easily annoyed by situations, things, and people. He or she tends to be dissatisfied with family, situations, and the restrictions upon his or her life. The person may feel unable to cope with life.

Low Superego, Low Score on Factor G. A person with a low score on Factor G tends to disregard rules and generally adopt expedient strategies to achieve what he or she wants. This is in contrast with a conscientious, moralistic approach to solving problems. The low G person is fickle, frivolous, tends toward self-indulgence, indolence, and is generally undependable. A further feature, according to Cattell et al. (1970), is that this person tends to

disregard obligations to people and is not overly concerned about moral standards or rules. The 16PF handbook indicates that the G factor correlates negatively with delinquency, sociopathic tendencies, homosexuality, etc. Psychopaths, criminals, and others who are characterized by an absence of concern with conventional moral standards also have low scores.

High Autism (High Factor M). The presence of a high score on this factor indicates a tendency toward being unconventional, self-absorbed in ideas, imaginative, absent-minded, relatively impractical, easily seduced from practical judgment, and occasionally prone to hysterical swings. Cattell's 16PF handbook suggests that the high autia person has an "intense subjectivity". The handbook also mentions a "higher internal, spasmodic anxiety and conflict tensions."

Low Naivete, Low Score on Factor N. A further finding in regard to criminals of many kinds is a generally low score on Factor N. This indicates the following negative characteristics: simple tastes, lacking in self-insight, unskilled in analyzing motives, generally gregarious, has a vague and relatively injudicious mind, socially clumsy, etc. The 1970 16PF handbook indicates that convicts have a particularly high score on this factor.

On the other hand, there is some complexity in making complete generalizations regarding this personality feature. The handbook also indicates that certain "types" of delinquents have relatively high N scores. The high N score tends to indicate the practical, shrewd, suspicious, and calculating type of criminal or delinquent.

Guilt-Proneness, High Score on Factor O. A high score on Factor O indicates that the person tends toward anxiety, worry, depression, fussiness, hypochondriacal feelings, inadequate feelings, and may be lonely, brooding,

or overcome by moods. Such persons may not feel accepted in groups or free to participate. They may be considered shy, ineffective, and hinderers by others.

Self-Sufficiency, Resourcefulness, High Score on Factor Q2 . A high score on this factor indicates resourcefulness, a preference for one's own decisions, self-sufficiency, and perhaps a tendency toward introversion. The 1970 handbook indicates that this factor is found at a high level in criminals. A high score on this factor will also make a person significantly dissatisfied with group integration. Such persons tend to be rejected.

The above represents a very general review of the findings reported in the 16PF handbook regarding the personality characteristics of prisoners. The present research will attempt to go beyond this general description and examine the possibility that there may be a larger number of differentiated personality profile types among a sample of state penitentiary inmates. It should be pointed out that the above profile characteristics are averaged over many kinds of sociopathic and criminal types and that not ALL criminals may possess such characteristics.

METHODS OF ANALYSIS AND DATA

The Data

The data for this set of analyses were derived from a population of 1,031 new residents in the Colorado state correctional system since January, 1978. All measures of interest in classifying these residents were not available for all 1,031 cases. Also, after a thorough examination of the records, several incorrect items were identified (such as IQ scores greater than 200, ages less than 16 years or greater than 65 years, etc.). It was decided to eliminate these cases, along with cases with missing items, leaving a clean, complete

set of 880 cases. The cases which were eliminated were examined to determine if they were significantly different from the 880 remaining cases. Details of that inquiry are discussed in a section below.

For the purposes of cluster analysis from this relatively large data set (N=880) two approximately 20% random samples were created (PRIS1 and PRIS2). It was determined that these two samples were fair and representative of the population. The first of these samples (PRIS1, N=172) was used to estimate the classifications, and the second (PRIS2, N=180) was used for validation checks in the results of clustering.

The original population was comprised of 993 males and 38 females, aged 16 to 60, with a mean age of 27. Measures on these subjects included Nurture IQ, Reading-level, and Cattell's scales: Culture Free IQ, Sixteen Personality Factor Test (16PF), and Clinical Analysis Questionnaire (CAQ). These last three measures were collected by the Colorado Department of Corrections to provide predictions, for example, of suicide, violence, and escape. The population under study were assigned to one of the eight state facilities depending on the results of these prediction calculations and other relevant assessments. Thus, our population contains prisoners not only from the Canon City facility but also from the facilities at Buena Vista and six others. Scores from the 16PF and CAQ will provide the basis for our classification study.

Additional data on the sample of 880 cases were provided for our substantive validation study. These data will be discussed in the appropriate sections below.

Sixteen Personality Factors (16PF)

The 16PF test (Cattell and Eber, 1964) contains adult-oriented measures similar to the HSPQ measures utilized in our study of Boulder County Court data. To the 16 original factors of Cattell's test was added a scale on "open/defensive" responses. These measures are presented in STEN scores (standard-tens or deciles) normed on the general population, both males and females. In the general population, scores of 5.5 are average. The profile in Table 1 shows how the Colorado correctional population compares with the general population.

As can be seen from Table 1, our correctional population (N=1031) falls almost entirely below the general population means on all but Defensiveness and below the "normal" range on four of the factors. The prison population is apparently abnormally reserved, tough-minded, placid, and relaxed; also somewhat on the dull, submissive, sober, trusting, practical, forthright, and conservative side. Some of these differences may be due to the nearly all maleness of our sample.

Table 1

Aggregate 16PF profile of Colorado Correctional Population (N=1031)

General population means are 5.5 on each scale
"Normal" ranges lie between 4 and 7

No.	Factor	STEN	LOW	HIGH
1	A	3.9	Reserved	Outgoing
2	B	4.0	Dull	Bright
3	C	5.3	Easily upset	Calm
4	E	4.4	Submissive	Dominant
5	F	4.4	Sober, serious	Happy-go-lucky
6	G	5.7	Expedient	Conscientious
7	H	5.1	Shy, timid	Venturesome
8	I	3.8	Tough-minded	Apprehensive
9	L	4.1	Trusting	Suspicious
10	M	4.4	Practical	Imaginative
11	N	4.3	Forthright	Shrewd
12	O	3.3	Placid, serene	Apprehensive
13	Q_1	4.9	Conservative	Experimenting
14	Q_2	5.2	Group oriented	Self-directed
15	Q_3	5.1	Undisciplined	Disciplined
16	Q_4	3.5	Relaxed	Tense, driven
17*	MD	6.0	Open	Defensive

* This is a special scale developed for the particular study of prisoners.

Clinical Analysis Questionnaire (CAQ)

The CAQ (Cattell and Eber, 1964; Cattell, et al., 1970) comprises 12 scales designed to identify specific clinical problems which should be taken into consideration in placing the prisoner. Again the scores are in STENS, normed on males and females in the general population. In Table 2, next page, the differences between our population and the general population can be seen on these CAQ scores.

Table 2

Aggregate CAQ profile of Colorado Correctional Population (N=1031)

General means are 5.5 on each scale
The "normal" ranges lie below 7

SCALE	STEN	NAMES FOR CAQ	ABBREV. SCALE NAME	HIGH
1	4.0	Hypochondria	D_1	Over-concerned with health
2	4.2	Suicidal Depression	D_2	Thinks of self-destruction
3	4.3	Agitation	D_3	Restless; excited
4	3.7	Anxious Depression	D_4	Easily upset; disturbed
5	3.9	Low Energy Depression	D_5	Feels weary
6	4.8	Guilt and Resentment	D_6	Blames self; guilt
7	4.2	Boredom and Withdrawal	D_7	Bored with people; withdraws
8	5.2	Paranoia	Pa	Feels persecuted; spied on
9	4.5	Psychopathic Deviation	Pp	Condones antisocial acts
10	4.4	Schizophrenic	Sc	Hallucinates; rebels
11	4.8	Psychasphenia	As	Repetetitive thoughts/impulses
12	4.3	Psychological Inadequacy	Ps	Feels worthless

More information on the CAQ can be found in Cattell, Eber, and Tatsuoka (1970).

Data Preparation

It was noted before that a number of cases were removed from the data set because of extensive missing items in 16PF and CAQ scores. Also a number of cases were removed because of erroneous items. First, the 33 cases with erroneous data were removed. then the 118 cases with missing data. In Table 3, below, the characteristics of the 118 cases with missing data are compared with the resulting sample of 880 cases.

Table 4

A Comparison of Cases with Missing Data (N=118) with the Resulting (Clean) Data Set (N=880)

Variable	Means of Removed Cases (N=118)	Means of Retained Cases (N=880)	t	Sign. Diff.
Age	28.339(18-53)	26.952(16-60)	-1.81	N
Sex	all males	M=844,F=36	--	--
Mixture IQ	88.678	89.085	-.21	N
Culture Free IQ	96.903	100.310	-1.10	N
Reading Level	97.000	100.483	-.67	N

As can readily be observed, the differences between the cases omitted (because of missing 16PF and CAQ scores) and the cases remaining after the file was culled are not statistically significant. It can therefore be concluded that the data set was not changed in any important way by removing the faulty records. Since all cases with missing data were males, the few females (36) in the corrected data set were all retained.

The Samples

Even with faulty records removed, 880 cases is far too large a sample for the cluster-analytical software. Therefore, it was decided to work with two 20% random samples of the 880 cases. These samples were created by using the Statistical Package for the Social Sciences (SPSS) sampling facility. The samples were taken from the 880 cases independently, with replacement. The two samples resulting contain 172 (PRIS1) and 180 (PRIS2) cases, respectively.

Before using these samples, it is necessary to show how fair and representative these samples are. Table 4, below, lists the means and the t-ratios between the variables of each subsample and the original set of 880 cases. Since only one variable showed a significant difference, age, between the PRIS1 (N = 172) sample and the original data, the samples were judged fair

and reasonably representative of the population. It was decided to use PRIS1 (N = 172) as the sample on which the bulk of analysis is performed. PRIS2 (N = 180) will be used as a check on the reliability of the analysis on PRIS1. This aspect of our method is described in the following section.

Table 4

Means, Standard Deviations and t-scores; Comparing Each Subsample with the 880 Case Set

	Full Sample		PRIS1 Subsample		PRIS2 Subsample		PRIS1	PRIS2
	Mean	SD	Mean	SD	Mean	SD	t	t
Sex	(F=36, M=844)		(F=8, M=164)		(F=7, M=173)		-	-
Age	26.952	7.929	25.634	6.981	26.528	7.644	2.03*	.66
Nurture IQ	89.085	18.840	86.956	19.252	88.065	16.718	1.30	.65
Culture Free IQ	100.310	17.461	101.200	17.425	98.971	17.382	-.59	.92
Reading Level	100.483	32.251	98.827	34.270	100.667	32.593	.58	-.07
PF 1	3.881	2.201	3.837	2.263	3.956	2.098	.24	-.42
PF 2	3.958	1.944	4.180	2.068	4.050	1.912	-1.36	-.58
PF 3	5.305	2.078	5.064	2.133	5.217	2.136	1.39	.52
PF 4	4.358	2.016	4.134	2.017	4.072	2.071	1.33	1.73
PF 5	4.403	1.836	4.459	1.902	4.511	1.985	- .36	- .71
PF 6	5.155	1.708	5.145	1.743	5.106	1.881	.07	.34
PF 7	5.118	2.119	5.151	2.158	5.206	2.232	- .19	- .50
PF 8	3.770	1.641	3.866	1.626	3.872	1.708	- .70	- .75
PF 9	4.041	2.026	4.000	1.991	3.889	2.002	.24	.92
PF 10	4.425	1.908	4.698	1.810	4.456	1.886	-1.73	- .20
PF 11	4.236	2.077	4.134	2.026	4.206	2.054	.59	.18
PF 12	3.319	2.119	3.326	2.149	3.511	2.012	- .04	-1.12
PF 13	4.850	1.903	4.930	1.972	4.811	1.787	- .50	.25
PF 14	5.174	2.009	5.128	1.921	5.228	2.092	.27	- .33
PF 15	5.084	1.951	4.942	2.107	4.922	1.953	.87	1.01
PF 16	3.486	1.828	3.512	1.999	3.561	1.779	- .17	- .50
PF 17	6.003	2.003	6.041	2.081	5.911	1.998	- .23	.56
CAQ 1	4.050	1.543	4.128	1.562	4.128	1.589	- .61	- .61
CAQ 2	4.241	1.520	4.343	1.554	4.300	1.637	- .64	- .47
CAQ 3	4.283	1.827	4.134	1.842	4.400	1.777	+ .97	- .79
CAQ 4	3.732	1.740	3.924	1.761	3.772	1.812	-1.32	- .28
CAQ 5	3.942	1.548	4.000	1.672	3.906	1.512	- .44	.29
CAQ 6	4.893	1.732	5.047	1.809	5.122	1.707	-1.06	-1.55
CAQ 7	4.202	1.686	4.238	1.776	4.228	1.733	- .25	- .19
CAQ 8	5.215	2.089	5.360	2.240	4.994	2.051	- .82	1.30
CAQ 9	4.515	2.024	4.233	2.064	4.489	1.930	1.60	.16
CAQ 10	4.453	1.972	4.587	1.931	4.378	1.921	- .82	.47
CAQ 11	4.781	1.824	4.965	1.771	4.889	1.724	-1.21	- .73
CAQ 12	4.349	1.630	4.506	1.562	4.331	1.594	-1.16	.14

* significant at .05 alpha level

Analytical Sequence

In elucidating the classification structure for the 16PF and CAQ, the two inventories were analyzed separately. The analytical sequence in both cases was identical. Two clustering methods (UPGMA and HMODE) and three different similarity coefficients (correlation, cosine, and Euclidean distance) were applied to the analysis sample. Dendrograms were produced for ALL of these analyses. Certain levels of these hierarchical trees were chosen for detailed examination based on the four criteria (early formation, long branch lines, large numbers of members, and consistency of membership across different methods of analysis). The partitions chosen and their evaluations are discussed separately for the 16PF and the CAQ below.

For these classifications an experiment was conducted to find a reduced number of 16PF variables identified by discriminant analysis. The twelve variables produced extremely poor clustering results and the effort to reduce the dimensionality of the clustering problem was abandoned. This indicated to us the necessity of using the complete set of 16 factors for all analyses in this chapter.

Further partitions were created on the second (validation) sample of cases in an attempt to reproduce the clustering results of the experimental sample. These attempts at validation are discussed below in detail.

A final set of analyses was included as an illustration of how these methods may be used for prediction. External characteristics of a 16 PF pattern and a CAQ pattern are examined to determine whether the scale-scores are correlated in any way with (1) type of crime (violence, etc.), (2) need for supervision or security, or (3) rule infraction behavior while in prison

Also reported are the results of analysis of variance comparing the clusters on age, IQ, reading level, and other external variables in order to show any other distinctive characteristics of the clusters.

RESULTS

Sixteen Personality Factors Classifications

Six clustering procedures were applied to the 16PF scores of subjects in the experimental sample (N=172). Complete dendrograms for these six procedures are found in Figures 1 to 6. The procedures were:

Clustering Approach	Partition Selected
1) UPGMA - correlation	8 groups solution
2) UPGMA - cosine	8 groups solution
3) UPGMA - euclidean distance	6, 7, & 8 groups solutions
4) HMODE - correlation	7 groups solution
5) HMODE - cosine	5 groups solution
6) HMODE - euclidean distance	7 & 8 groups solutions

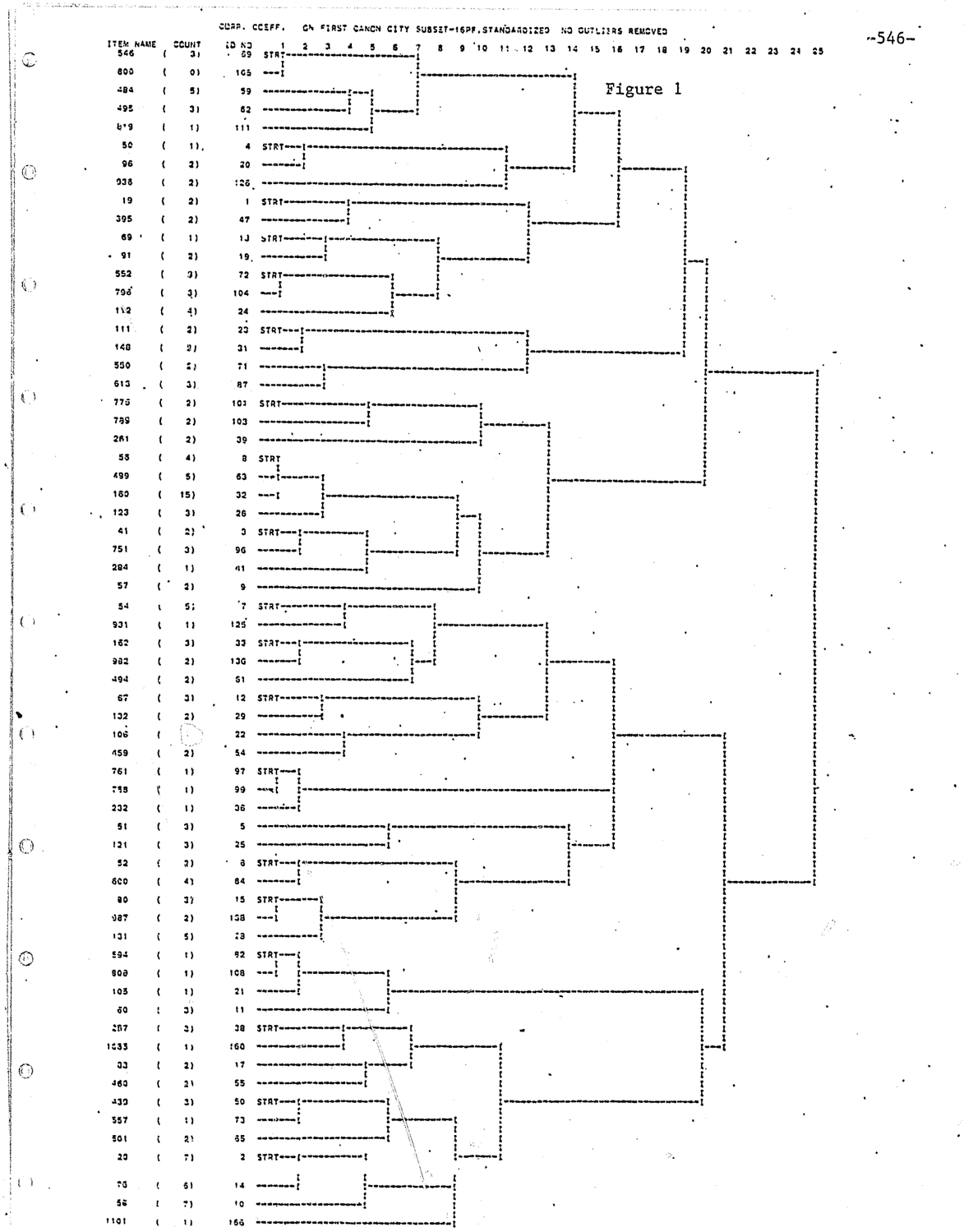
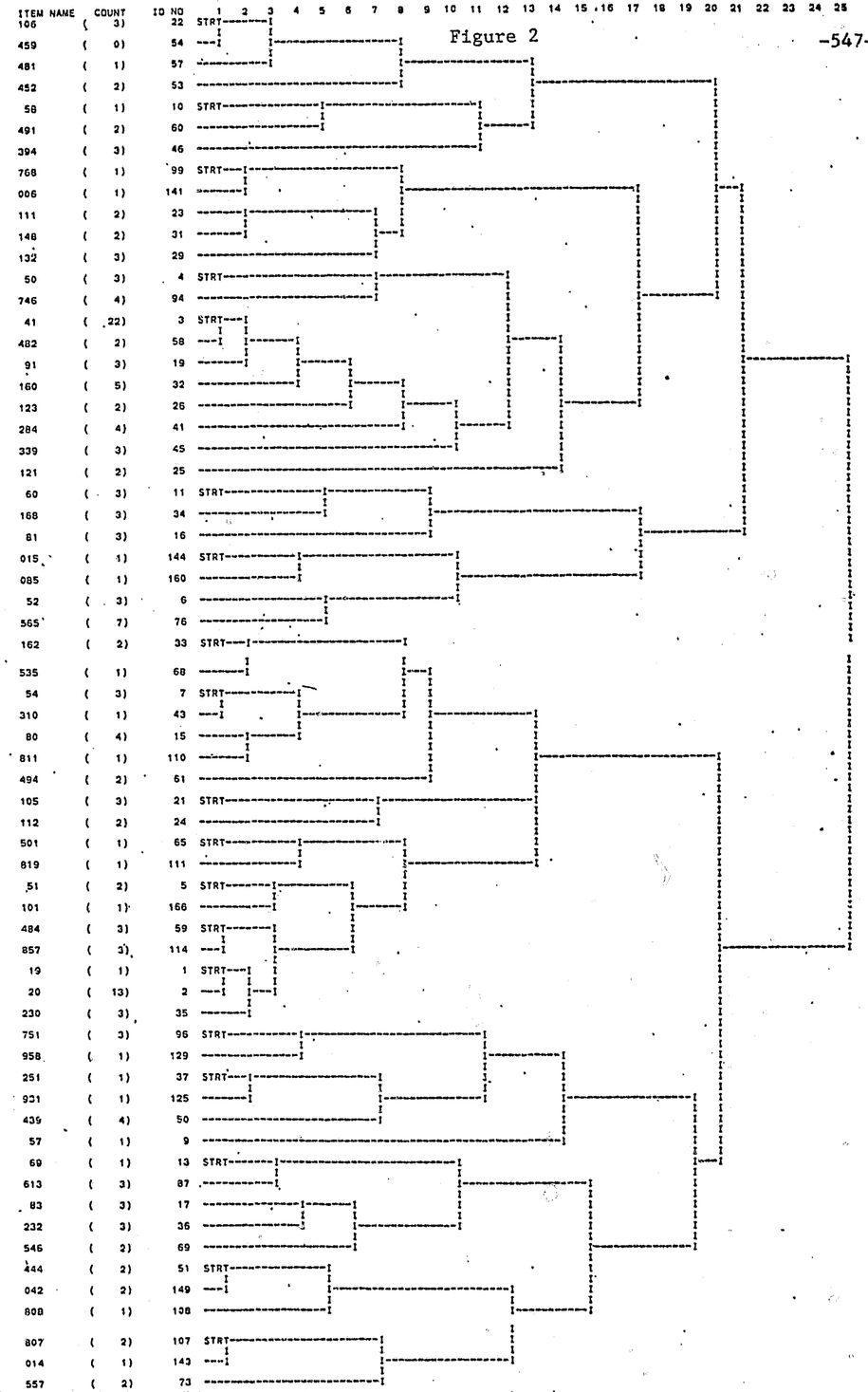


Figure 1

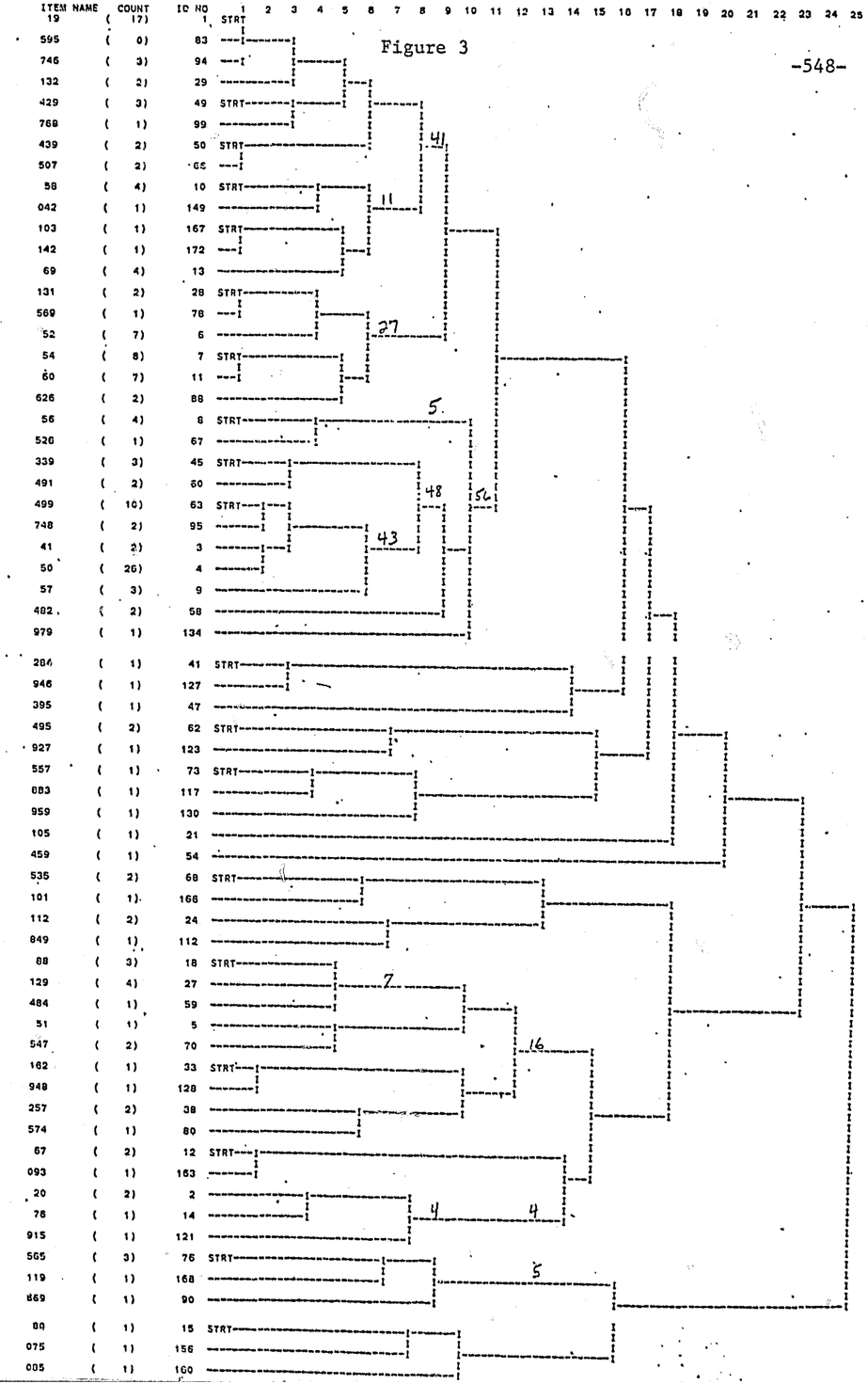
COSINE ON FIRST CANON CITY SUBSET-16PF,STANDARDIZED NO OUTLIERS REMOVED

Figure 2



EUCLID. DIST. ON FIRST CANON CITY SUBSET-16PF,STANDARDIZED NO OUTLIERS REMOVED

Figure 3



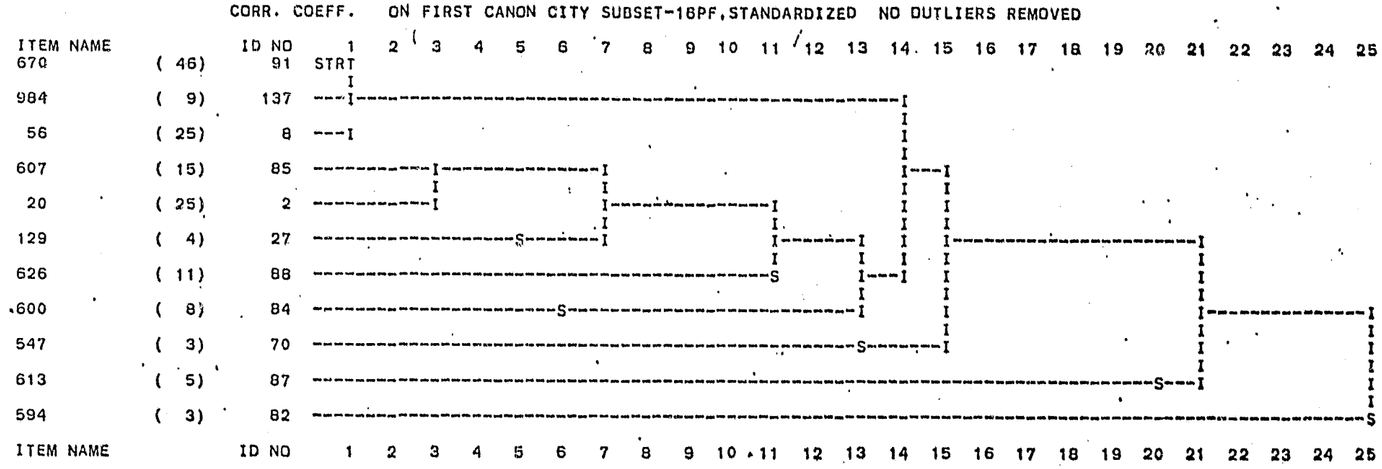

CORR. COEFF. ON FIRST CANON CITY SUBSET-16PF,STANDARDIZED NO OUTLIERS REMOVED
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
670 (46) 91 STRT
984 (9) 137
56 (25) 8
607 (15) 85
20 (25) 2
129 (4) 27
626 (11) 88
600 (8) 84
547 (3) 70
613 (5) 87
594 (3) 82
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 4

COSINE ON FIRST CANON CITY SUBSET-16PF,STANDARDIZED NO OUTLIERS REMOVED

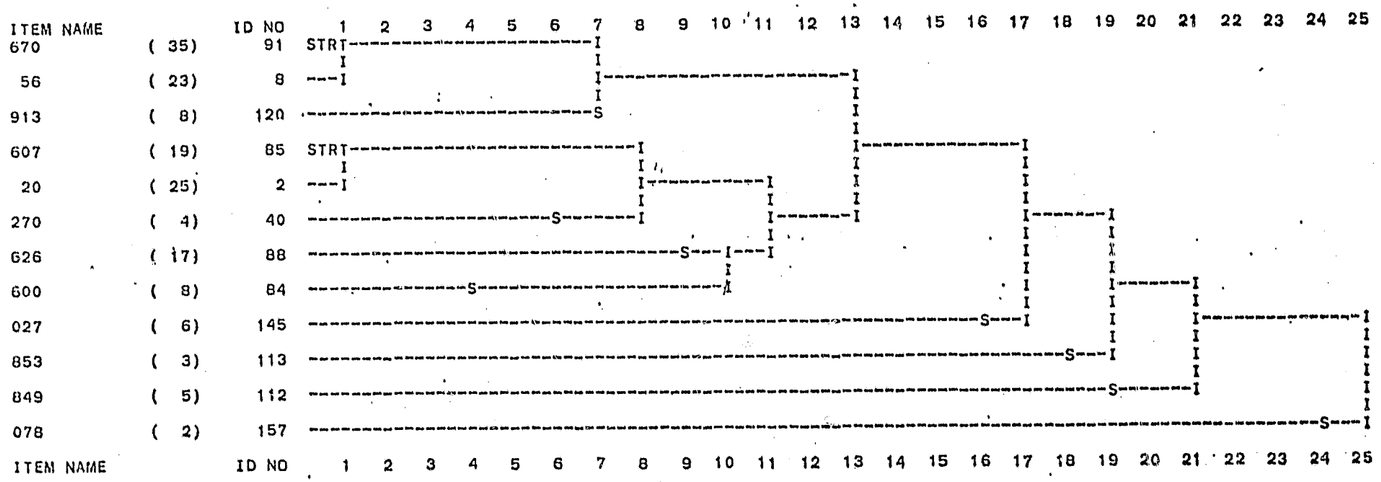


Figure 5

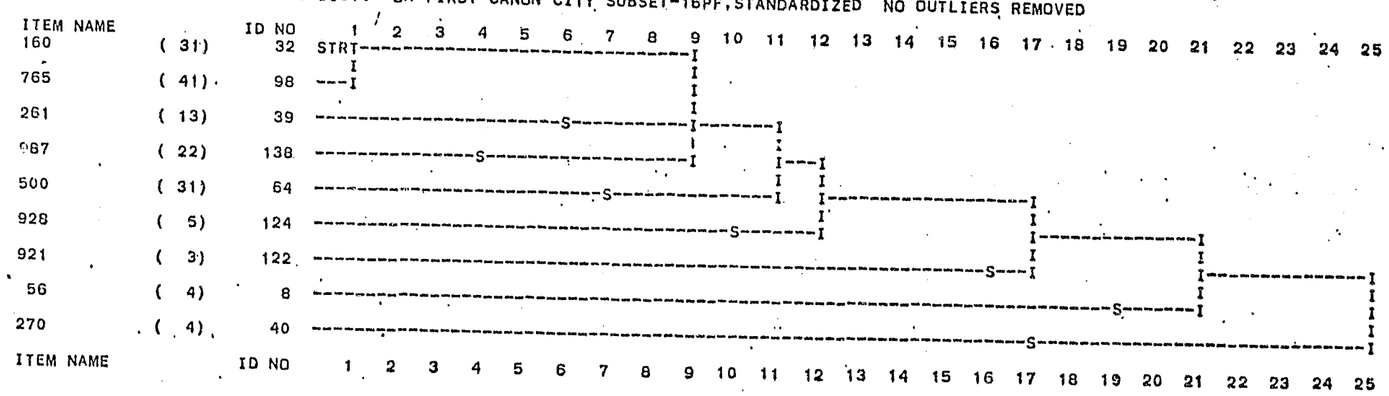

EUCLID. DIST. ON FIRST CANON CITY SUBSET-16PF,STANDARDIZED NO OUTLIERS REMOVED
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
160 (31) 32 STRT
765 (41) 98
261 (13) 39
987 (22) 138
500 (31) 64
928 (5) 124
921 (3) 122
56 (4) 8
270 (4) 40
ITEM NAME
ID NO
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 6

An inspection of the trees, using our four criteria, resulted in the choice of levels as listed on the right above. The seven-group solution from the UPGMA-Euclidean dendrogram was constructed from various levels.

Table 5 indicates the cross-partition agreement scores between these analyses. The coefficients of agreement indicate that the different approaches do, in fact, have some positive overlap with each other, but that the overlap is not generally high. The different methods have formed clusters that do not share the same boundaries. The distance-based partitions are clearly different from the angular-based similarity partitions.

It is noteworthy that the "best" agreement in this set of analyses occurs when the same clustering method (UPGMA) is compared at the same hierarchical level (8 classes) and both analyses use an angle-based similarity measure (correlation and cosine, respectively). This match is quite impressive with Rand's Coefficient equalling 0.90, and Goodman and Kinskel's Lambda equalling 0.70. This result is encouraging. At the same time the general overriding result is that this data is extremely sensitive to changes in method. The implication is that the clustering structure is weak.

Refining the Partitions: Two Attempts to Refine the Classifications

Since generally low agreement coefficients resulted from comparing the clustering solutions on 16PF measures, it was decided to attempt a refinement of the partition by reducing the size of the problem. One reason why clusters may be fuzzy in this kind of problem is that some variables are poor discriminators. Therefore, one refinement method is to eliminate the nondiscriminating variables. Using linear discriminant analysis, the set of personality criteria was reduced to 12 variables based on the failure of 5 of the variables to discriminate (e.g., in values Λ 1.0) on a linear discriminant analysis, and on these variables' performance in other studies.

The experimental subset was then reclustered on the 12 remaining variables by UPGMA-correlation and HMODE-correlation. Table 6a shows the results of comparisons between these partitions and the original UPGMA partition. Table 6b contains information about attempts to reproduce partitions by discriminant analysis. Clusters produced by the reduced variable set appeared to be no more adequate than those produced by all the factors.

Another method for refining the partitions is to submit the centroids produced by our UPGMA method to a number of K-means procedures. Six such runs were made using the K-means software (MIKCA). Runs M-1 and M-2 were made with centroids from UPGMA-correlation and UPGMA-cosine 8-group solutions; then M-3 to M-6 were made allowing starting centroids to be assigned randomly at K-levels of 6, 7, and 8. Selected pairs of these partitions were compared and the resulting agreement coefficients are reported in Table 7. This approach cannot be considered a success because the agreements (measures of partition stability) are lower than those in the original set of comparisons.

One possible reason why both of the above traditional refinement procedures failed is that they require a "points-in-space" model, while partitions based on the correlation coefficient assume a "profile" model. One need for future research is to develop refinement procedures based on correlation coefficients. The other obvious explanation is that the data is relatively unstructured and that the methods are discovering a variety of partitions based on minor idiosyncratic structures of the data or secondly, based primarily on method differences.

Table 5

PRISON - AGREEMENT COMPARISONS = Selected pairs of Partitions

	16 PF Clusters	Rand	Jacc.	CR V	∧s	Gps.
1	HMODE-r(2)vs.UPGMA-r(164)	.780	.333	.505	.344	7, 8
2	HMODE-r(2)vs.HMODE-cos(5)	.823	.472	.595	.505	7, 5
3	HMODE-r(2)vs.HMODE-d(1)	.718	.256	.417	.240	7, 7
4	UPGMA-r(164)vs.UPGMA-cos(164)	.898	.577	.708	.699*	8, 8
5	UPGMA-r(164)vs.UPGMA-d(143)	.815	.254	.526	.439	8, 8
6	UPGMA-r(164)vs.UPGMA-d(SEL)	.827	.379	.542	.464	8, 6
7	HMODE-cos(5)vs.UPGMA-cos(164)	.807	.385	.540	.461	5, 8
8	HMODE-cos(5)vs.HMODE-d(1)	.722	.260	.442	.361	5, 7
9	UPGMA-cos(164)vs.UPGMA-d(143)	.802	.223	.519	.410	8, 8
10	UPGMA-cos(164)vs.UPGMA-d(SEL)	.796	.306	.540	.408	8, 6
11	HMODE-d(1)vs.UPGMA-d(143)	.779	.225	.527	.345	7, 8
12	HMODE-d(1)vs.UPGMA-d(SEL)	.762	.279	.469	.319	7, 6
12A	HMODE-d(0)vs.UPGMA-d(148)	.811	.230	.554	.396	7, 8
12B	HMODE-d(0)vs.UPGMA-d(143)	.829	.213	.521	.396	7, 8
12C	HMODE-d(0)vs.UPGMA-d(140)	.829	.213	.550	.404	7, 8

* Indicates relatively stable solution

Table 6a

a. Comparisons between Clusters on Reduced Set of Variables and 17 Original Variables

16PF Partitions	Rand	Jacc.	∧s	# Gps.	Outliers
UPGMA-r(12var) vs HMODE-r(12vars)	.806	.311	.450	8, 8	3, 24
UPGMA-2(12var) vs UPGMA-r(17 orig)	.809	.280	.395	8, 8	3, 0

Table 6b

b. Discriminant Analysis on 12 Variables
Results at 90% Confidence Level

	% correct	% missed	% unclass.
UPGMA-r(8gr; 17 vars)	54	0	41
HMODE-r(8gr; 12 vars)	44	5	51
UPGMA-r(8gr; 12 vars)	54	1	45

Conclusion: 12 variable subset produces worse (less stable and reproducible) clusters

Table 7

K-Means Comparison Results on the Experimental Subsample

M = MIKCA
M-1 = MIKCA run #1
M-2 = MIKCA run #2, etc.

	Comparisons		16PF	Rand	Jacc.	Λ_s	# Gps.	Outliers
1	M-1	vs M-2*	(16PF)	.833	.280	.524	8, 8	0, 0
2	M-1	vs M-5	"	.790	.183	.351	8, 8	0, 0
3	M-2	vs M-5	"	.793	.166	.287	8, 8	0, 0
4	M-3	vs M-4	"	.757	.245	.363	6, 7	0, 0
5	M-3	vsM-5	"	.780	.249	.439	6, 8	0, 0
6	M-4	vsM-5	"	.776	.196	.333	7, 8	0, 0
7	M-4	vs M-6	"	.770	.204	.380	7, 7	0, 3
8	M-5	vs M-6	"	.775	.162	.281	8, 7	0, 3

* detW is a minimum for M-2, centroids supplied by UPGMA-cosine 8-group solution.

AGREE MIKCA's with 10% Outliers 16PF

M-1 vs. M-2 16PF	.838	.284	.531	8, 8	17, 17
M-1 vs. M-5 16PF	.742	.210	.321	8, 8	17, 18
M-2 vs. M-5 16PF	.728	.158	.256	8, 8	17, 18

Evaluation of the "Stable" UPGMA 8-Level Partition

Two partitions were chosen for evaluation: (1) the 8-group UPGMA-correlation solution and (2) the 8-group UPGMA-cosine solution. The correlation and cosine based solutions were chosen because of their relative stability as shown by partition comparisons above.

UPGMA (Corr.) and UPGMA (Cos.) 8-Type Solutions

These two analyses converged upon a highly similar partitioning of the 16PF profiles at the 8-cluster level. The symmetric Lambda coefficient of 0.70 indicated substantial similarity of the two classifications. We conducted a cross-classification of the two 8-level classifications and found that certain classes were independently recovered by the two analyses. This is indicated below.

Overlap between UPGMA (r) vs. UPGMA (cos.)

Cluster No.	UPGMA (r)	No. of Points in Common	UPGMA (Cos.)	Cluster No.
1	19	17	22	7
2	39	36	47	1
3	55	48	50	2
4	18	11	12	3
5	15	9	11	4
6	4	--	--	--
7	15	6	9	6
8	7	5	9	8

This data suggests that all of these clusters, with the exception of the very small cluster 6 (N=4), are partially recovered from the similarity matrix by the UPGMA method. Given the cross-method stability of these cluster-types we have conducted evaluation studies of their psychological meanings. These are discussed below.

Tables 8 and 9 outline the geometrical relation of the clusters produced by the UPGMA (r) and (cosine) analyses. It can be seen that virtually ALL of

the correlations and cosines between the clusters are either close to zero or negative. This is highly encouraging, indicating that all of the profiles are relatively dissimilar to each other.

Table 8

Inter-Profile Correlations - UPGMA-Correlation (8-group Solution)

Clusters

	1	2	3	4	5	6	7
2	.3269						
3	-.3790	-.9172					
4	-.2270	-.0301	-.2528				
5	.0729	.2987	-.1904	-.2591			
6	.0536	.2750	-.2102	-.2191	.0256		
7	-.3134	.0636	-.2129	.2451	-.5709	.0091	
8	-.2289	-.2481	.0697	.1624	-.2393	-.2185	.2893

Table 9

Cosine of the Angle (At the Origin) Between Cluster Centroids
UPGMA-Cosine (8 Group Solution)

Clusters

	1	2	3	4	5	6	7
1							
2	-.836						
3	-.352	.270					
4	-.017	-.307	-.135				
5	.203	-.221	-.382	-.030			
6	.054	-.054	-.093	-.377	.026		
7	.070	-.192	.021	.261	-.549	-.215	
8	.226	-.358	-.151	-.299	.061	.203	-.331

Personality Types Found to be Replicated in the Prison Population

The following personality profile descriptions represent the clusters which are well-replicated between the UPGMA (correlation) clustering method and the UPGMA (cosine) clustering method. The following descriptions provide an indication of the particular internal similarity that the members of each cluster have toward one another, as well as the major features on which they are atypical in regard to the overall prison sample (see table 10).

Cluster 1, Shy, with Feelings of Inferiority (Low H). This cluster is characterized most strongly by a very low score on Factor H. Virtually all of these prisoners can be characterized as shy, timid, withdrawn, embittered, rulebound, and perhaps tormented by an unreasonable sense of inferiority. The 16PF handbook indicates that they may be slow and impeded in expressing themselves. They may not be able to keep up with all that is going on in their social situation. There may be a tendency toward schizoid malajustment. The cluster is also characterized by low scores on Factors A, F, and E, and rather high scores for Q_1 and Q_2. The internal variance on these additional factors is much higher and, hence, the cluster, although atypical on these dimensions, does not exhibit strong homogeneity among all members.

Cluster 2, Emotionally Unstable, Ego Weakness (Low Factor C). The profile of this group of prisoners clearly demonstrates ego weakness and emotional instability. The STEN score for these prisoners is 2.7, indicating a low score on ego strength. They may become highly emotional and frustrated. They are changeable, easily perturbed, evasive of responsibilities, worrying, and are prone to get into fights in problem situations.

Cluster 3, Self-Assured, Strong Ego Strength, Fearless. This large group represents a complex profile. They have a strong ego strength, indicating

calmness and unruffledness, with a tendency to avoid difficulties. Their high score on Factor Q_3 indicates that they are both strong-willed and controlled. The low score on Q_4 indicates that they tend to be composed in many situations. The low score on Factor O also indicates the tendency toward being self-assured, placid, and almost complacent. This factor also indicates a relative absence of guilt, an absence of fear, and a tendency not to care. Finally, the low score on Factor L further reinforces the easygoing, relaxed, and relatively permissive approach of these persons.

Cluster 4, Aggressive, Socially Impulsive, and Naive. This type is defined primarily by three factors: E-positive, H-positive, and N-negative. The E-positive score indicates that they tend to be independent, assertive, and perhaps aggressive and headstrong. The high score on Factor H indicates that they tend to be thick-skinned, adventurous, and socially bold. They may not see danger signals. Finally, the negative score on Factor N indicates that they may be socially clumsy, injudicious, gregarious, and perhaps lacking in self-insight and unskilled in analyzing their own motives.

Cluster 5, Lacking Intelligence. A primary feature of the present cluster is the very low score for Factor B. This indicates low mental capacity, inability to handle abstract problems, disorganized, with poor judgment, and a tendency toward low morale and quitting. A second feature of the cluster is the relatively high score for this prison sample on Factor O; i.e., a tendency toward guilt-proneness. Their STEN score of 5.6 places them substantially above most of the other prison clusters. It suggests that they may be somewhat more depressed, anxious, moody, hypochondriacal, and lonely than most of the other prisoners in this sample.

Cluster 7, Self-Confident, Cheerful, Warm, and Outgoing. This cluster is defined primarily by three variables: a low score on Factor O, and high scores on Factors A and H. The cluster has very high homogeneity on the low Factor O. This indicates self-confidence, cheerfulness, an absence of fears, and a tendency to a simple, vigorous, and perhaps insensitive approach to social interaction. The high scores on Factor A indicate a tendency to be easygoing, cooperative, attentive to people, trusting, and warm-hearted. Similarly, the high score on Factor H also indicates an adventurous, socially bold, gregarious, friendly, and impulsive approach. The negative aspect of this is that the person tends to be carefree and does not see danger signals. A final feature of this cluster is the fairly high score for Factor B. This indicates that they are fairly bright, intellectually adaptable, and perhaps of higher morale and of better judgment.

Cluster 8, Tough-Minded, Cynical, and Down to Earth. The members of this small cluster are defined by low score on Factor I, indicating a tendency to be tough-minded and cynical. They also have a very low score on Factor M, indicating a tendency to be down to earth and practical. The low score that they exhibit for Factor O indicates an active, self-assured, confident approach. Finally, their relatively low score on Q_3 indicates a compulsive, controlled person with a strong self-sentiment.

Table 10

16PF Profiles of "Matched" Cluster-Types at 8-Level of UPGMA(r) versus UPGMA (cosine) Analyses

Cattell 16PF		Cluster Number 1*	2**	3**	4*	5**	6	7*	8**	Prison Sample Mean
		N=19	39	55	18	15	4	15	7	
1	A	2.9	3.0	4.5	4.8	3.5	(un-	3.9	4.0	3.84
2	B	3.8	4.1	4.2	4.2	2.5	stable)	7.1	3.0	4.18
3	C	5.0	2.7	6.9	5.9	4.2		4.7	5.4	5.06
4	E	3.0	4.4	3.1	6.7	3.7		5.3	4.1	4.13
5	F	2.8	3.8	4.5	6.1	4.1		5.4	6.0	4.46
6	G	4.5	4.4	6.1	3.8	6.0		5.1	5.0	5.14
7	H	3.0	4.1	6.0	6.3	5.0		6.7	4.1	5.15
8	I	4.1	3.1	4.5	4.5	4.3		2.3	2.7	3.86
9	L	5.0	4.6	2.3	5.6	4.5		4.9	4.7	4.00
10	M	4.7	4.8	5.0	4.9	5.7		3.4	1.8	4.69
11	N	5.6	5.4	2.9	2.8	4.7		4.5	3.1	4.13
12	O	3.0	4.9	1.9	2.9	5.6		2.7	2.3	3.33
13	Q_1	6.1	4.3	4.5	6.6	5.1		4.9	4.7	4.93
14	Q_2	6.3	5.8	4.6	5.1	3.7		5.1	4.4	5.13
15	Q_3	4.6	3.4	6.6	3.2	5.3		4.5	6.2	4.94
16	Q_4	3.4	5.5	1.9	4.1	4.1		3.1	3.4	3.51
Cluster H-Co-efficient		.306	.163	.301	.206	.352	.523	.299	.521	

* Replicated clusters in the cross-method comparison

** Replicated clusters in the cross-sample comparison

Validation

There are two forms of validity for the results outlined in earlier portions of this chapter: internal and external. For internal validation, we propose to attempt a replication of the results on a second subset, randomly drawn (N=180) from our original 880 observations. In this way, we can evaluate how stable the group structure observed in the original sample.

For external validation, we must try to determine if the characteristics of the clusters in a given partition reflect separate and theoretically meaningful classes. We propose to accomplish this validation by calculating means and variances of selected external descriptive variables. Whether or not these descriptive means are significantly different from each other can be determined by overall F ratios and Eta-squared measures computed by analyses of variance procedures (ANOVA). These ANOVA summary tables are included in this chapter, together with some discussion of the significant differences among clusters.

Internal Validation

In order to validate internally the results on the 16PF clustering, an attempt was made to replicate the results outlined above using the second (randomly drawn, N=180) subset. Before clustering, 10% of the extreme cases, based on a "points-in-space" model were removed.

Clustering was accomplished, again, by UPGMA-correlation and by HMODE-correlation clustering routines. Comparisons between two partitions from these trees were made with the following results:

Methods	Rand	Jacc.	Cramer V	Λ_s	Groups
UPGMA-r vs. HMODE-r	.764	.363	.457	.507	8, 7

Agreement coefficients are not extraordinarily high, but high enough to indicate some stability.

The most important evidence for internal validation is the comparisons among clusters identified on the experimental subset (PRIS1) and clusters on the second subset (PRIS2). Table 10 shows the intercentroid correlations which provide a measure of how closely the characteristics of clusters of one data set match those of another data set. This time we are looking for high positive correlations in order to match similar clusters. An examination of the matrix of Table 11 might lead to a four-group classification as follows:

PRIS1 Clusters	N	PRIS2 Clusters	N	Q Correlation
1 & 2	(58)	1	(56)	.539, .864
3	(55)	3	(59)	.864
5	(15)	4 & 5	(20)	.471, .539
8	(7)	8	(10)	.587

Table 11

Comparisons of 8-Group Solutions (UPGMA-corr.)-corr.) from both the Experimental and (PRIS1) and Validation Subsets (PRIS2)

INTER-CENTROID Q-CORRELATIONS

PRIS1 Clusters	PRIS2 Clusters 1	2	3	4	5	6	7	8
1	.503*	.116	-.564	.207	.173	-.016	.342	-354
2	.894	.391	-.905	.214	-.069	-.225	-.072	-.182
3	-.928	-.466	.858	.052	.061	.190	.212	.116
4	.159	.212	.069	-.504	-.241	.043	-.466	-.018
5	.174	.311	-.465	.478	.577	.004	.199	.147
6	.333	-.238	-.257	-.082	-.115	.040	-.141	-.151
7	.025	.022	.233	-.603	-.419	-.014	-.390	.153
8	-.192	.059	.266	-.445	.102	-.246	-.243	.578

* Highest Q-correlations are underlined

Note that not only are fairly high Q-correlations accounted for by this matching, but cluster-sizes match very well. The matching of cluster 1 in PRIS1 can, in fact, be improved by allowing a comparison between PRIS1 at the 8-cluster level with PRIS2 at the 9-cluster level. This comparison indicates

that at the higher level of the PRIS 2 tree, the cluster is perfectly well matched across the two samples.

Table 12

Comparison of 8-Group Solution from Experimental Subgroup (PRIS1) and the 9-Group Solution from the Validation Subset (PRIS2)

INTER-CENTROID Q-CORRELATIONS

PRIS1 Clusters	PRIS2 Clusters 1	2	3	4	5	6	7	8	9
1	.094	.116	.717	-.564	.207	.173	-.016	.342	-.354
2	.700	.391	.797	-.905	.214	-.069	-.225	-.072	-.182
3	-.847	-.466	-.721	.858	.052	.061	.190	.212	.116
4	.624	.212	-.305	.069	-.504	-.241	.043	-.466	-.018
5	-.077	.311	.346	-.465	.478	.577	.004	.199	.147
6	.166	-.238	.382	-.257	.082	-.115	.040	-.141	-.151
7	.302	.022	-.230	.233	-.603	-.419	-.014	-.390	.153
8	.004	.059	-.309	.266	-.445	.102	-.246	-.243	.578

* Highest Q-correlations are underlined.

A slightly different matching, yielding 5 clusters, is suggested by Table 11, as follows:

PRIS1 Clusters	N	PRIS2 Clusters	N	Q correlation
1	(19)	3	(28)	.717
2 & 4	(58)	1	(28)	.700, .624
3	(55)	4	(59)	.858
5	(33)	5 & 6	(20)	.478, .577
8	(7)	9	(10)	.578

Q-correlations may be somewhat better, but cluster sizes diverge a bit more than the first (4-cluster) matching. It will be seen later on, however, that external relationships (particularly with violence) may cause us to prefer this particular partition.

External Validation

Tables 13a and 13b show the results of analysis on external variables. The first five columns are the cluster means on the selected external variables and the three right-hand columns are the ANOVA results. Among the criteria are included the 12 CAQ variables, since in a sense they also are external to the 16PF measures on which the clusters were identified. Cluster 2 + 4 is labeled 0 in this table. The column headed "Sig." is a list of alpha levels at which the compared F-ratios are significant, and the last column estimates the proportion of variance in the extended criterion accounted for by the classification.

Table 13a. Cluster Means of External Variables
METHOD: UPGMA - Correlation - 16PF Clusters
on PRIS1 Subsample

	Cluster					ANOVA Results		
	0	1	3	5	8			
Criterion	N=58	N=19	N=55	N=33	N=7	F	Sig.	ETA^2
Violent	.1552	.5263	.2545	.1818	.2857	3.0292	.0192	.0676
Murder	.2441	1.1053	.0909	.5152	0	2.3439	.0568	.0532
Age	1.8793	2.0000	2.1455	1.9394	2.1429	.8492	.4959	.0199
Nurture	82.7547	81.8421	95.3725	84.8276	80.1429	3.9919	.0041	.0939
Culture	97.7500	98.3333	108.1887	96.1333	103.0000	3.6428	.0072	.0859
Reading	95.2037	85.2778	110.7308	93.3548	97.4286	2.7365	.0308	.0652
CAQ 1	5.4828	3.3684	3.2545	3.7879	3.4286	27.7617	.0000	.3994
CAQ 2	5.4828	3.8947	3.3636	4.3333	3.8571	19.9743	.0000	.3236
CAQ 3	4.2586	3.0000	4.0182	4.5758	5.0000	2.9079	.0233	.0651
CAQ 4	5.1207	3.4737	2.7818	4.1212	3.2857	18.5092	.0000	.3072
CAQ 5	5.3276	3.6842	2.6545	4.0606	4.1429	31.0697	.0000	.4267
CAQ 6	6.2069	4.7895	3.8545	5.2121	4.7143	16.5948	.0000	.2844
CAQ 7	5.3103	4.5789	3.1455	3.8788	4.7143	14.6975	.0000	.2604
CAQ 8	6.7241	5.6316	3.7091	5.6364	5.0000	18.4308	.0000	.3063
CAQ 9	3.7586	4.2105	4.8545	3.9394	4.7143	2.3465	.0566	.0532
CAQ 10	5.7586	4.3158	3.3818	4.6970	4.5714	14.1360	.0000	.2529
CAQ 11	6.1379	5.1579	3.6182	5.1818	4.2857	21.7785	.0000	.3428
CAQ 12	5.6379	4.4211	3.5091	4.3030	4.1429	19.2076	.0000	.3151
R SCORE	1.3793	1.1053	1.0364	1.6667	.4286	.2378	.9166	.0057
INOFF	.4310	.3158	.3091	.4848	.2857	.2368	.9173	.0056
S SCORE	.0847	.0496	.0525	.0768	.0226	.2086	.9334	.0050
SECURE	.6034	1.0000	.5091	.7576	.5714	.6240	.6460	.0147
STATUS	.3276	.4737	.2182	.5152	.0000	1.4247	.2279	.0330
CONTROL	.1379	.2632	.1455	.1212	.2857	.7483	.5604	.0176
W SCORE	.0744	-.1501	-.0757	.1289	-.1619	.4699	.7578	.0111

Table 13b. Cluster Means of External Variables
METHOD: UPGMA - Distance - CAQ Clusters

	Cluster							ANOVA Results		
	0	1	2	3	4	5	6			
Criterion	N=4	N=40	N=10	N=80	N=22	N=9	N=7	F	Sig.	ETA^2
Violent	.2500	.2000	0	.3125	.0909	,2222	.4286	1.6834	.1280	.0577
Murder	0	.6000	0	.2875	.1364	0	.8571	.7310	.6253	.0259
Age	2.2500	1.9750	1.8000	2.0500	1.7727	2.1111	2.2857	.6656	.6776	.0236
Nurture	73.7500	83.9167	74.4000	91.7703	86.4000	85.0000	81.0000	2.1631	.0496	.0787
Culture	101.6667	91.7368	98.3333	107.8974	100.0000	80.3750	111.4000	7.5191	.0000	.2277
Reading	102.0000	85.0000	91.3333	109.2208	88.4000	81.3750	120.1667	3.6614	.0020	.1241
PF 1	3.0000	3.3750	2.4000	4.3375	4.0000	3.4444	3.2857	1.8839	.0864	.0741
PF 2	5.5000	3.2250	6.0000	4.4375	3.8182	3.4444	5.4286	4.3826	.0004	.1375
PF 3	2.0000	4.2750	3.1000	6.2750	5.1364	2.3333	3.5714	17.3581	.0000	.3870
PF 4	4.2500	4.4500	5.3000	3.6875	3.7273	4.4444	6.5614	3.5561	.0024	.1145
PF 5	3.0000	4.4000	4.5000	4.7250	3.8636	4.1111	4.8571	1.1256	.3496	.0393
PF 6	4.7500	4.6750	3.9000	5.4875	6.2273	4.4444	3.4286	5.3930	.0000	.1640
PF 7	4.2500	4.5500	4.5000	5.8625	4.7273	4.3333	4.2857	2.9719	.0088	.0975
PF 8	2.5000	3.7250	2.5000	4.1250	4.1818	4.2222	3.0000	2.7343	.0147	.0904
PF 9	4.2500	4.8750	5.4000	3.2125	3.6818	5.0000	5.5714	6.4371	.0000	.1897
PF 10	4.0000	4.7500	4.3000	4.7250	4.7727	4.5556	5.0000	.2298	.9665	.0083
PF 11	4.7500	4.6250	5.2000	3.3375	5.0455	5.3333	4.1429	4.8102	.0002	.1489
PF 12	6.7500	4.1750	5.7000	2.1750	3.9091	4.2222	3.2857	12.9460	.0000	.3201
PF 13	3.2500	5.0750	3.9000	4.9625	5.0455	4.7778	6.0000	1.3598	.2338	.0471
PF 14	6.2500	5.1500	5.6000	4.8750	5.0000	5.8889	6.0000	1.0539	.3926	.0369
PF 15	3.5000	3.9000	2.2000	5.7750	5.3182	4.7778	5.1429	8.9255	.0000	.2450
PF 16	6.7500	4.5000	5.7000	2.5375	3.2273	4.5556	3.5714	12.7580	.0000	.3169
PF 17	2.5000	4.9250	2.2000	7.1375	7.0000	5.5556	5.0000	27.9904	.0000	.5044
R SCORE	1.0000	.8500	3.5000	1.2500	.9545	.6667	2.2857	.7802	.5866	.0276
INOFF	.5000	.3250	1.2000	.3750	.1818	.1111	.5714	1.5290	.1716	.0527
S SCORE	.0268	.0411	.3066	.0685	.0210	.0175	.0736	1.8820	.0867	.0641
SECURE	2.5000	.3750	.7000	.6875	.5455	.2222	1.4286	2.5648	.0211	.0853
STATUS	1.0000	.2750	.5000	.2875	.2727	.2222	.8571	1.4359	.2038	.0496
CONTROL	.7500	.0500	.1000	.2000	.1364	.0000	.2857	3.2295	.0050	.1051
W SCORE	-.1119	-.0968	.7932	-.0053	-.0772	-.1510	.0408	1.2596	.2788	.0438

In comparing the two cluster matchings above, and two other internal validation procedures, we conclude that there exist within the data four or five fairly stable and reproducible clusters. These clusters may represent natural classes of penitentiary inmates.

An examination of this data reveals that all but one CAQ score take on significantly variant mean values across the classification. Violence is, apparently, significantly higher in cluster number 1 than in the other clusters. Nurture-IQ, Culture-free IQ and Reading scores are significantly different (at alpha of .05). Particularly high scores are found in cluster no. 3. Cluster number 1, associated with violence, shows low scores in Nurture IQ, and Reading, and an average culture-free IQ level. None of the scores on security, status, and control differ significantly among clusters. Not significant are any of the scores on rule infractions: INOFF, internal rule infraction (offender =1, non-offending prisoner = 0); RSCORE = number of rule infractions; SSCORE scaled rule infraction score; WSCORE, weighted rule infraction score.

PRISONER PERSONALITY TYPES BASED ON THE CAQ

Six clustering procedures were applied to the 12 CAQ scores in the experimental sample. Selected dendrograms are shown below. The six procedures are the same as those used for clustering on the 16PF variables. An inspection of the trees led to the selection of various levels for further analysis. These levels and the agreement coefficients between them are found in Table 14a.

From these comparisons, two partitions were chosen as "best": the UPGMA-cosine (5-group solution) and the UPGMA-Euclidean (6-group solution), asterisked in Table 14a. The high symmetric lambda scores of these particular solutions prompted us to select them for further examination.

Refining the Partitions

The method used to refine the partitions was the method of reclustering using the K-means procedure. Again, two sets of starting centroids for K-means partitions were suggested by the solutions involving UPGMA. Table 14b shows the results of comparing the K-means (MIKCA) solutions with the solutions from UPGMA, and comparing them with each other. As with the 16PF analysis, the MIKCA solutions with starting centroids suggested by UPGMA produced much better agreements with the UPGMA solutions. The solutions based on random starting centroids apparently found local clustering which agreed poorly with all of the other solutions. However, the generally positive agreement coefficients prompted an examination of the actual overlap matrices between all of these partitions. This led to a discovery of certain clusters which continually recurred across these analyses. We will later describe these "replicated" clusters and suggest the overall data structure that is implied--collectively--by this cumulative approach to analysis.

Table 14a

PRISON - AGREEMENT COMPARISONS = Selected pairs of Partitions

CAQ Clusters	Rand	Jacc.	Cr V	⌐s	Gos.
HMODE-r(3)vs.UPGMA-r(SEL)	.795	.420	.567	.526	6, 5
HMODE-r(3)vs.HMODE-cos(2)	.870	.610	.639	.586	5, 5
HMODE-r(3)vs.HMODE-d(1)	.735	.154	.515	.308	10, 5
HMODE-r(3)vs.HMODE-d(3)	.784	.334	.510	.351	8, 5
UPGMA-r(SEL)vs.UPGMA-cos(167)	.781	.414	.585	.551	5, 6
UPGMA-r(SEL)vs.UPGMA-d(134)	.788	.177	.578	.425	13, 6
UPGMA-r(SEL)vs.UPGMA-d(163)	.733	.318	.465	.444	6, 6
HMODE-cos(2)vs.UPGMA-cos(167)	.798	.472	.613	.587	5, 5
HMODE-cos(2)vs.HMODE-d(3)	.819	.395	.591	.460	8, 5
UPGMA-cos(167)vs.UPGMA-d(163)	.817	.529	.655	.621	6, 5
HMODE-d(1)vs.UPGMA-d(134)	.895	.219	.561	.440	13, 10
HMODE-d(3)vs.UPGMA-d(163)	.790	.363	.5235	.391	6, 8
UPGMA-d(SEL)vs.HMODE-d(5)	.804	.474	.540	.394	10, 6
UPGMA-d(SEL)vs.UPGMA-cos(154)	.796	.441	.709	.525	10, 5
UPGMA-d(SEL)vs.UPGMA-cos(SEL)*	.827	.483	.749	.555	10, 5
UPGMA-d(SEL)*vs.UPGMA-cos(SEL)*	.845	.563	.723	.621	6, 5
UPGMA-d(SEL)*vs.HMODE-d(3)	.796	.366	.540	.420	6, 8
UPGMA-d(SEL)*vs.HMODE-d(5)	.802	.508	.467	.435	6, 6

* partitions chosen for evaluation and further analysis

Table 14b

K-Means Comparison Results on the Experimental Subsample

M = MIKCA
M-1 = MIKCA run #1
M-2 = MIKCA run #2, etc.

Comparisons	Rand	Jacc.	Cr V	⌐s	# Gos.
M-1(CAQ)vs.UPGMA-d(s)CAQ	.863	.573	.792	.695	6, 6
M-6(CAQ)vs.UPGMA-d(s)CAQ	.767	.286	.597	.467	8, 6
M-2(CAQ)vs.UPGMA-cos(s)CAQ	.905	.682	.806	.774	6, 5
M-6(CAQ)vs.UPGMA-cos(s)CAQ	.771	.283	.595	.475	8, 5
M-1 vs. M-2 CAQ*	.883	.590	.752	.696	6, 6
M-1 vs. M-4 CAQ	.819	.384	.657	.556	6, 6
M-2 vs. M-4 CAQ	.796	.339	.597	.479	6, 6
M-3 vs. M-4 CAQ	.818	.360	.718	.565	5, 6
M-3 vs. M-5 CAQ	.799	.307	.681	.478	5, 7
M-4 vs. M-5 CAQ	.823	.340	.696	.518	6, 7
M-4 vs. M-6 CAQ	.833	.315	.707	.526	6, 8
M-5 vs. M-6 CAQ	.864	.395	.748	.608	7, 8

* MIKCA solutions with 10% Outliers Removed

Evaluation and Interpretation of the CAQ Prisoner Personality Profiles

The generally positive agreement coefficients between the trial cluster analyses on this data (see Tables 14) indicated that although the various partitions did not overlap exactly, there was some substantial discovery of similar kinds of structures. The agreement tables and evaluation tables for all of these partitions allowed us to identify particular clusters that consistently reappeared across all of these multiple solutions. Two partitions in particular were chosen for more intense scrutiny. These were (1) the UPGMA-cosine solution at the 5-cluster level and (2) the UPGMA-Euclidean distance solution at the 6-cluster level. These particular partitions had shown a reasonably good stability as exhibited by the Goodman and Kruskal lambda coefficients. The evaluation tables below include these and a number of other analyses in which the recurrent profile types are marked with the nomenclature A, B, C, D, E, F, G. Types A through G represent the particular recurrent profiles that were discovered by scrutiny of all of the analyses.

The Hierarchical Structure of the CAQ Personality Profiles

The schematic diagrams (Figures 7a and b) indicate the general personality structure that emerged from the set of analyses. These are particularly indicated by the evaluation tables, which are also included in this section.

A first basic division in the CAQ personality profiles is into depressed, nonpsychotic prisoners (accounting for about 30% of the overall sample). In contrast, a second very general class of nondepressed but psychotic prisoners is found. The nondepressed, psychotic group is substantially larger, accounting for about 55% of the overall sample. The two natural dendograms that we include also in this section indicate this basic split into depressed

and nondepressed prisoners. Both of these dendograms were generated by the UPGMA analyses. The first utilized the cosine measure of similarity while the second utilized the correlation coefficient as a measure of similarity. In both cases, long branch lines separate a number of depressed subclasses from nondepressed subclasses. A second point to note is that both dendograms indicate the smaller size of the depressed group. The depressed group, is, in both cases, the bottom half of each dendogram.

The schematic tree diagram indicates more clearly the substantive meetings of each of the various replicated types. Following from the basic division into depressed and nondepressed prisoners, a number of different subdivisions are formed. Initially the depressed group splits into two closely related subtypes. The first subtype, F, has relatively positive scores for all of the first seven variables, indicating an overall depression. The score on profile F, however, are not as extreme as those illustrated in profile G. In this latter profile, extremely severe scores are found for CAQ numbers 1 and 2, 4 through 8, and 10 and 12. It is noteworthy that neither of these two depressed subtypes has a severe score for CAQ dimension 9, which indicates psychopathic deviation.

At the top end of the diagram the nondepressed, psychotic group basically splits into three different subtypes. All three of these subtypes (A, B, and C) are characterized by positive scores for CAQ variable 9, psychopathic tendencies. Variations between the three clusters basically stem from the different combinations of other CAQ variables which are found in the three subtypes. Subtype A basically has a somewhat positive score on CAQ dimension 3, indicating agitation and restlessness. This profile is, in fact, the largest of the various clusters. About one-third of all prisoners seem to

fall into this large profile. Clusters B and C add different kinds of CAQ dimensions to the basic psychopathic tendencies. It is noteworthy that in both B and C elements of depression creep into the profile. These will be fully discussed in the sections below.

A third subtype emerges in a number of analyses as a result of two clusters having high scores on the joint combination of CAQ dimensions 6 and 11; i.e., guilt and resentment (CAQ 6) and psychasphenia (CAQ 11). In many of our analyses these two clusters emerged as separate subtypes with cluster D having particularly high scores on CAQ dimension 11, while cluster E was separated primarily by its high score on CAQ dimension 6. In other analyses, however, the two clusters were merged together, forming one larger cluster accounting for about 17% of the overall sample and characterized by jointly high scores on CAQ dimensions 6 and 11. These two subtypes have relatively average scores on the depression variables and we would generally classify these two subtypes into the depressed side of the measurement space.

A further manner of illustrating the structure of this data set can be seen in the cluster graph of the clusters as plotted into discriminant space. The cluster graph (Fig.8) below indicates that when the 5-cluster level of the UPGMA analysis with cosine is plotted the clusters are basically spread out according to the first dimension separating depressed prisoners (cluster 1 in the graph) from the large nondepressed group (cluster 2 in the graph). Between these two large clusters we have clusters 5, 3, and 4. It is interesting to note that, in our schematic diagram, clusters D and E represent clusters 3 and 5. The cluster graph indicates the closeness of 3 and 5, their relative sparsity, and their location halfway between the depressed and non-depressed groups. The graph also indicates the relative overlap between groups D and E; i.e., 3 and 5 on the graph.

Figure 7a

Schematic Taxonomic Structure of Personality Patterns of Prisoner Sample Based on the Clinical Assessment Questionnaire (CAQ)

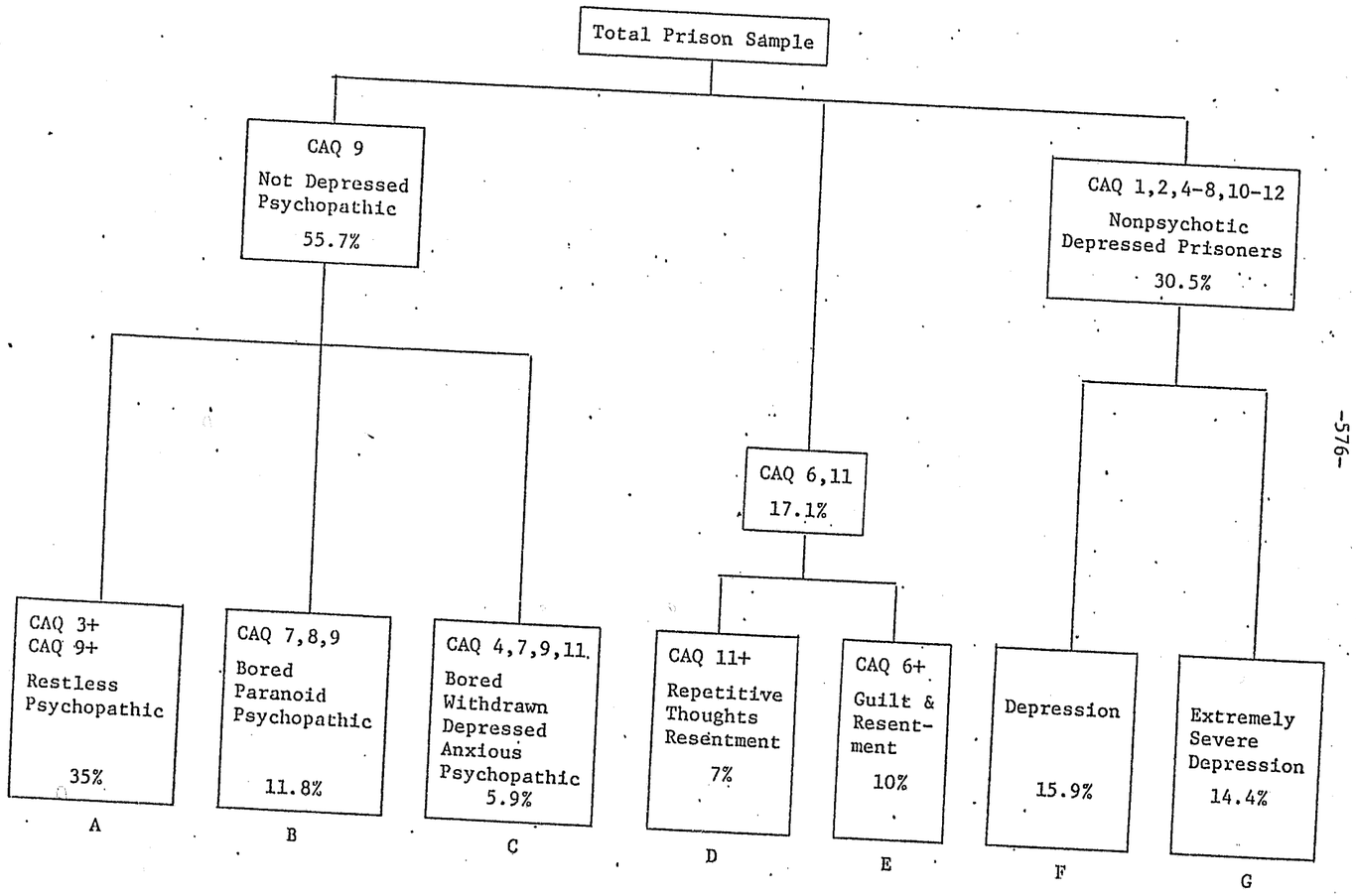
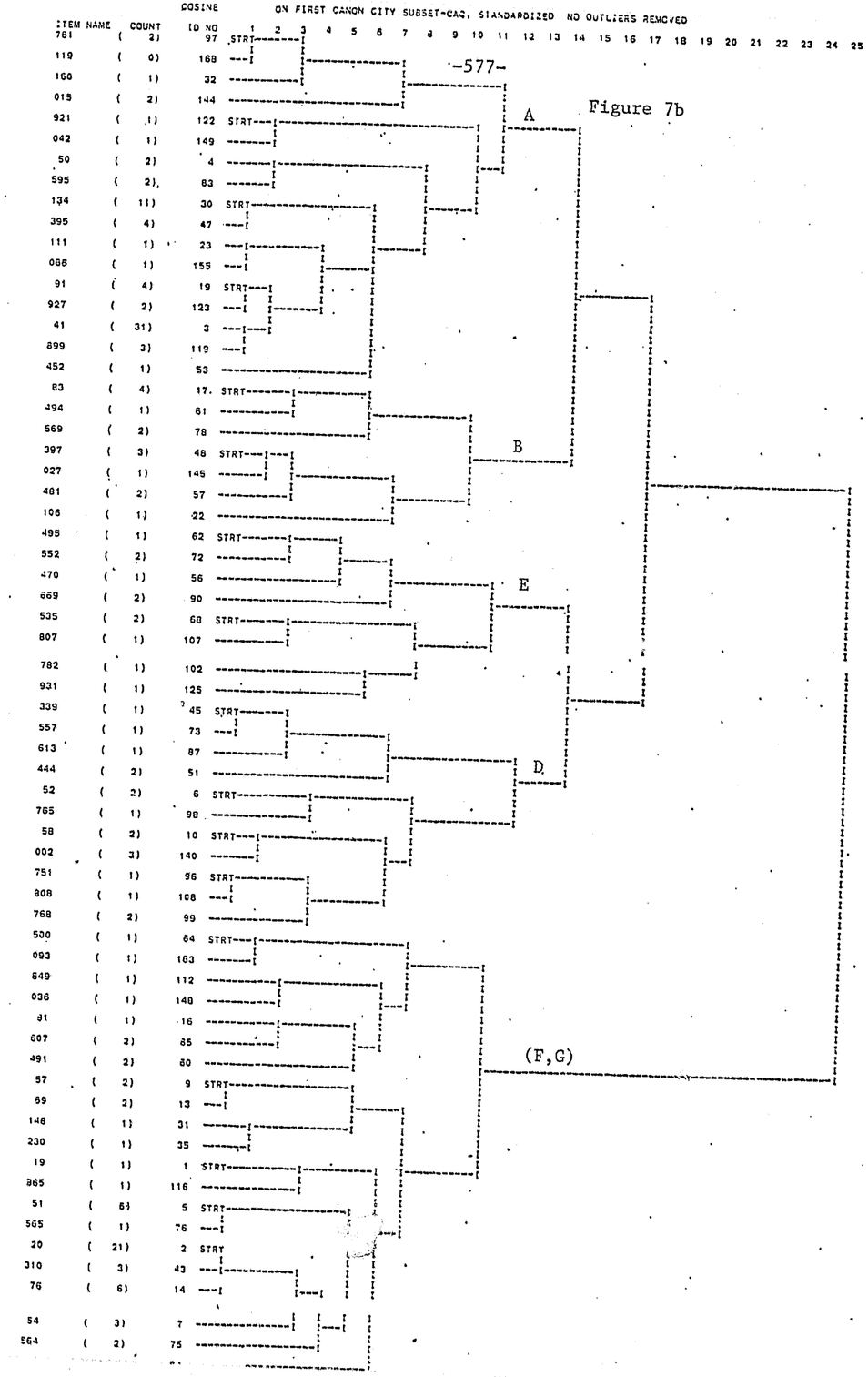


Figure 7b


COSINE
ON FIRST CANON CITY SUBSET-CAD, STANDARDIZED NO OUTLIERS REMOVED
ITEM NAME COUNT ID NO 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
761 (2) 97 STRT
119 (0) 168
160 (1) 32
015 (2) 144
921 (1) 122 STRT
042 (1) 149
50 (2) 4
595 (2) 83
134 (11) 30 STRT
395 (4) 47
111 (1) 23
088 (1) 155
91 (4) 19 STRT
927 (2) 123
41 (31) 3
899 (3) 119
452 (1) 53
83 (4) 17. STRT
494 (1) 61
569 (2) 78
397 (3) 48 STRT
027 (1) 145
481 (2) 57
106 (1) 22
495 (1) 62 STRT
552 (2) 72
470 (1) 56
669 (2) 90
535 (2) 60 STRT
807 (1) 107
782 (1) 102
931 (1) 125
339 (1) 45 STRT
557 (1) 73
613 (1) 87
444 (2) 51
52 (2) 6 STRT
765 (1) 98
58 (2) 10 STRT
002 (3) 140
751 (1) 96 STRT
808 (1) 108
768 (2) 99
500 (1) 64 STRT
093 (1) 163
649 (1) 112
036 (1) 148
81 (1) 16
607 (2) 85
491 (2) 80
57 (2) 9 STRT
69 (2) 13
148 (1) 31
230 (1) 35
19 (1) 1 STRT
865 (1) 116
51 (64) 5 STRT
565 (1) 76
20 (21) 2 STRT
310 (3) 43
76 (6) 14
54 (3) 7
564 (2) 75
A
B
E
D
(F,G)

CLUSTER GRAPH

PRISON SAMPLE 1
CAN SCORES - CUS
DISCRIMINANT SPACE

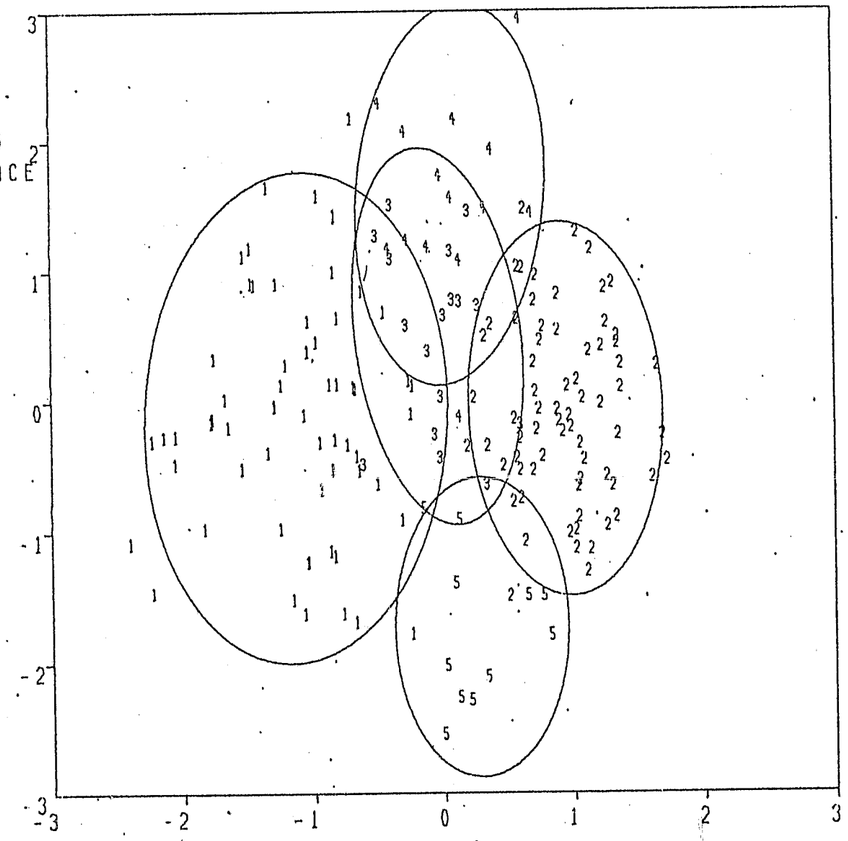


Figure 8

Descriptions of the Various CAQ Personality Types

Cluster A. Cluster A is not depressed. It is characterized by low scores on virtually all of the depression dimensions (CAQ 1 through 7). As represented in the various analyses, cluster A has average scores for CAQ dimension 3. This is a depression dimension and indicates a tendency toward restlessness and agitation. Another recurrent feature of cluster A is the above-average score for CAQ dimension 9, indicating a tendency toward psychopathic deviation. This cluster is fairly homogeneous, with most of the H-coefficients in the various analyses being close to or above 0.60.

Cluster B. This cluster also does not indicate serious levels of depression. However, it is even more extreme than cluster A in regard to the above-average score for CAQ dimension 3; i.e., restlessness and agitation. The main defining features of the cluster, however, are high scores on dimensions 7, 8, and 9. These three dimensions indicate that the persons in this cluster tend to be withdrawn, disinterested in people, paranaoid with feelings of persecution, and finally, as in all of these nondepressed clusters, they have a very strong tendency toward psychopathic deviation. This cluster is a highly reliable cluster in that it is picked up by virtually all of the analyses.

Cluster C. This cluster is identified clearly in only two of the analyses. It is characterized again by a tendency toward psychopathic deviation. It is generally not classified as a depressed cluster, although the two versions identified by these analyses indicate a high score for CAQ dimension 4 (anxious depression). These persons are disturbed and easily upset and emotionally volatile. The other definers of the cluster are the high scores on dimensions

7, 9, and 11. The CAQ 11 dimension again indicates psychopathic depression. The dimensions 7 and 9 indicate that these persons are disinterested in other people, withdrawn and somewhat bored, and also have high scores on dimension 11; i.e., psychasPhenia.

Cluster D. This small group is also difficult to distinguish from its near neighbor, cluster E. It is characterized by marginally high scores on dimension 6 and extremely high scores on dimension 11. The latter indicates that the persons in this group suffer from recurrent, repetitive thoughts and emotional impulses surrounding these thoughts. The dimension "As" is called psychasPhenia. The cluster generally has low or average scores on the depression variables, indicating that although depression may be present, it is not held at severe levels. The score for psychopathic deviation is below average.

Cluster E. This cluster is also characterized by an absence of psychopathic deviation and an absence in general of depression. It is, however, characterized by extremely high scores on one component of depression--CAQ dimension 6--indicating guilt, resentment, and a tendency to blame self. Like cluster D, depression is generally low or absent and psychopathic deviation is low or absent. In certain analyses this cluster merges with cluster D, forming one general cluster characterized by high scores on CAQ dimensions 6 and 11.

Clusters F and G. These two clusters might be classified together since they have very high scores on virtually all of the CAQ dimensions, from dimension 1 through 12, with the exception of dimensions 3 and 9. The low score on dimension 9 indicates that these persons are not psychopathic in their general attitudes. The low score on dimension 3 indicates that they are not particularly restless or agitated. On virtually all of the other CAQ dimensions,

however, these persons are at the extremely high levels. The differentiation between clusters F and G basically stems from the fact that many of the cluster analytic methods separated the two groups according to degree of severity, with F having above average scores for all of the depression dimensions and G having extremely above average scores. These differentiations can be seen in the evaluation tables included in this section.

Table 15

First Canon City Subset - CAQ, Standardized

No Outliers Removed (N=172)
Method: UPGMA Similarity: Cosine

CAQ Variables	Cluster 1 (F,G) N=61		Cluster 2 (A) N=69		Cluster 3 (D) N=17		Cluster 4 (B) N=14		Cluster 5 (E) N=11		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	1.031	-.010	-.601	.924	-.383	.840	-.676	.971	-.489	.896	.587
2	.824	.003	-.631	.818	-.183	.561	-.313	.627	.072	-.604	.397
3	.061	.276	.124	-.101	-.967	.191	.742	.558	-.566	.442	.150
4	.900	.262	-.690	.686	-.224	.632	-.322	.672	.095	-.062	.481
5	1.000	.371	-.763	.738	-.035	.666	-.385	.857	-.217	.551	.599
6	.799	.403	-.779	.677	.072	.571	-.657	.725	1.180	.950	.591
7	.743	.022	-.599	.674	-.499	.804	.710	.280	-.493	.475	.411
8	.813	.425	-.802	.739	.548	.675	.573	.613	-1.054	.801	.615
9	-.295	.076	.302	.239	-.654	-.082	.925	.608	-.421	.142	.173
10	.986	.431	-.687	.656	-.304	.464	-.082	.682	-.587	.659	.559
11	.739	.492	-.861	.576	.748	.653	.020	.366	.122	.631	.540
12	.831	.073	-.667	.643	-.173	.819	-.278	.781	.200	-.051	.425
	Dist. Index= .7937		Dist. Index= .7947		Dist. Index= .2839						
	MSD= 9.025	H=.235	MSD= 4.663	H=.606	MSD= 4.899	H=.566	MSD= 3.956	H=.645	MSD= 6.518	H=.403	

Table 16

First Canon City Subset - CAQ, Standardized (N=172)

Method: MIKCA Similarity: Mahalanobis Distance
(Second Analysis)

CAQ Variables	Cluster 1 (G) N=40		Cluster 2 (A) N=63		Cluster 3 (D) N=26		Cluster 4 (B) N=21		Cluster 5 (D,E) N=13		Cluster 6 (E) N=9		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	1.407	.046	-.590	.919	-.304	.772	-.509	.822	.312	.485	-.509	.898	.648
2	1.324	.154	-.660	.842	-.270	.671	-.129	.532	-.023	.628	-.149	.229	.570
3	-.032	.238	.108	-.037	-.866	.216	.703	.541	.554	.516	-.555	.378	.199
4	1.293	.500	-.723	.686	-.023	.604	-.309	.694	-.001	.491	.106	-.004	.581
5	1.241	.314	-.816	.754	.138	.648	-.057	.610	.092	.592	-.199	.642	.600
6	.942	.472	-.815	.717	.102	.626	-.578	.542	.952	.636	1.203	.941	.630
7	1.020	.625	-.635	.672	-.243	.619	.590	.456	-.134	.101	-.572	.700	.445
8	1.033	.577	-.856	.751	.440	.650	.668	.576	-.264	.530	-1.054	.751	.657
9	-.561	.321	.333	.210	-.690	.324	.833	.566	.595	.155	-.651	.387	.302
10	1.068	.360	-.715	.644	-.065	.373	.090	.493	.692	.443	-.764	.769	.510
11	.779	.487	-.945	.630	.626	.622	-.034	.528	.886	.545	.145	.622	.579
12	1.149	.070	-.730	.692	-.127	.811	-.232	.783	.563	.211	.103	.385	.524
	Dist. Index= .7722		Dist. Index= .7782		Dist. Index= .5775		Dist. Index= .4341						
	MSD= 8.224	H=.297	MSD= 4.441	H=.624	MSD= 4.869	H=.578	MSD= 4.625	H=.595	MSD= 6.154	H=.444	MSD= 4.714	H=.558	

Table 17

Second Canon City Subset - CAQ, Standardized

Outliers Removed (N=162)
Method: UPGMA Similarity: Cosine

CAQ Variables	Cluster 1 (F) N=52		Cluster 2 (D,E) N=12		Cluster 3 (C) N=6		Cluster 4 (A) N=71		Cluster 5 (?) N=8		Cluster 6 (B?) N=9		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	.902	-.378	-.109	.679	-.297	.605	-.556	.909	.173	-.200	-.506	.890	.399
2	.819	-.366	.247	.426	-.394	.822	-.530	.767	-.587	.873	-.437	.688	.371
3	.129	.295	-1.137	-.047	.095	.755	.249	-.014	-.423	.607	-1.055	.378	.168
4	.795	.041	.325	.554	1.198	.886	-.716	.713	.120	.810	-.205	.546	.482
5	1.005	.511	.537	.127	-.530	.773	-.700	.673	-.248	.686	-.279	.291	.561
6	.613	.167	.630	.304	-.138	.364	-.512	.305	-.001	.698	-.723	.591	.298
7	.820	.237	.302	.330	.711	.605	-.684	.598	-.694	.859	.477	.383	.460
8	.708	.218	-.389	.792	-.178	.103	-.552	.287	-.115	.712	.919	.552	.332
9	-.355	-.026	-1.067	.208	.836	.276	.411	.321	-.779	.804	.157	.546	.234
10	.908	.265	-.704	.855	.197	.903	-.697	.655	.695	.636	.228	.424	.537
11	.642	.442	.406	.406	.719	.778	-.552	.154	.354	.580	-.638	.216	.316
12	.933	.234	.283	.774	-.187	.719	-.655	.603	-.686	.513	.244	.366	.481
H-Coeff.		.137		.451		.633		.482		.631		.489	

Table 18

Second Canon City Subset - CAQ, Standardized (N=162)

Method: MIKCA Similarity: Mahalanobis Distance
(Validation on MIKCA Partition using UPGMA-R Centroid)
Second Analysis

CAQ Variables	Cluster 1 (D) N=27		Cluster 2 (F,G) N=33		Cluster 3 (E) N=18		Cluster 4 (C) N=11		Cluster 5 (A) N=44		Cluster 6 (B) N=29		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	-.172	.436	1.399	-.165	-.046	.386	-.468	.764	-.553	.870	-.387	.822	.517
2	-.123	.376	1.086	-.740	.161	.474	-.499	.849	-.569	.797	-.169	.561	.334
3	-.047	.311	.050	.103	-.759	-.172	.391	.580	.229	-.161	-.038	.078	.064
4	-.205	.231	.946	-.243	.103	.429	.890	.696	-.636	.691	-.321	.472	.356
5	-.356	.339	1.182	.423	.391	.229	-.325	.649	-.838	.701	-.140	.489	.490
6	.154	.422	.760	.039	.557	.194	-.218	.448	-.562	.224	-.418	.361	.256
7	-.382	.569	1.010	.135	.087	.367	.456	.578	-.806	.680	.203	.157	.415
8	.075	.586	.796	.135	-.600	.729	-.431	.231	-.891	.752	.913	.611	.535
9	-.334	.404	-.459	-.203	-.875	.354	.940	.534	.409	.160	.399	.652	.262
10	.587	.625	1.171	.296	-.783	.750	-.088	.676	-.929	.857	.050	.592	.633
11	.591	.574	.918	.581	.301	.435	.595	.744	-.941	.576	-.580	.478	.555
12	-.357	.608	1.275	.244	.361	.642	-.390	.468	-.838	.750	.076	.561	.558
V-Coeff.		.457		.050		.401		.601		.575		.486	

Table 19

First Canon City Subset - CAQ, Standardized

No Outliers Removed (N=172)

Method: UPGMA Similarity: Cosine

CAQ Variables	Cluster 1 (F_1) N=33		Cluster 2 (G) N=21		Cluster 3 (A) N=80		Cluster 4 (E) N=21		Cluster 5 (B) N=7		Cluster 6 (F_2) N=7		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	.461	.433	1.991	.307	-.626	.947	-.326	.817	-.539	.902	.558	.727	.740
2	.325	.531	1.741	.193	-.639	.863	-.006	.530	-.037	.211	-.221	.724	.641
3	.108	.472	-.279	-.046	.192	-.030	-1.003	.229	.781	.916	.393	.663	.166
4	.817	.318	1.314	.358	-.639	.686	-.173	.436	-.363	.601	.286	.478	.530
5	.707	.766	1.709	.382	-.748	.760	-.028	.768	-.085	.830	.342	.779	.719
6	.410	.546	1.211	.422	-.710	.619	.632	.584	-.894	.709	.922	.825	.588
7	.429	.405	1.448	.156	-.550	.665	-.456	.665	1.394	.607	-.054	.215	.530
8	.894	.628	1.179	.781	-.635	.529	-.203	.284	.923	.810	-.544	.772	.571
9	-.274	.301	-.851	.610	.402	.333	-.782	.027	1.133	.229	.787	.184	.312
10	.841	.568	1.200	.252	-.621	.630	-.477	.750	.140	.068	.880	.132	.542
11	.531	.617	1.038	.569	-.649	.310	.261	.475	-.382	.713	1.065	.744	.457
12	.472	.410	1.414	-.072	-.636	.678	-.049	.607	-.049	.609	.499	.356	.508
H-Coeff.		.500		.326		.582		.514		.600		.550	

Table 20

First Canon City Subset - CAQ, Standardized

No Outliers Removed (N=172)
Method: UPGMA Similarity: Euclidean Distance

CAQ Variables	Cluster 1 (F) N=40		Cluster 2 (G) N=10		Cluster 3 (A) N=80		Cluster 4 (E) N=22		Cluster 5 (G) N=9		Cluster 6 (B) N=7		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	.478	.492	1.711	.472	-.626	.947	-.344	.818	2.550	.647	-.539	.902	.776
2	.230	.528	2.547	.811	-.639	.863	.072	.419	.995	.850	-.037	.211	.694
3	.158	.503	.090	.014	.192	-.030	-.961	.226	-.555	-.139	.781	.916	.167
4	.725	.318	1.122	.107	-.639	.686	-.189	.457	1.305	.695	-.363	.601	.531
5	.643	.754	1.914	.618	-.748	.760	-.082	.717	1.794	.463	-.085	.830	.732
6	.499	.562	1.080	.660	-.710	.619	.653	.595	1.141	.202	-.894	.709	.586
7	.344	.357	1.949	.644	-.550	.665	-.441	.676	.929	-.066	1.394	.607	.550
8	.643	.354	1.313	.909	-.635	.529	-.262	.242	.980	.646	.923	.810	.490
9	-.088	.134	-1.082	.583	.402	.333	-.795	.069	-.597	.648	1.133	.229	.278
10	.848	.512	1.715	.556	-.621	.630	-.516	.728	.617	.345	.140	.068	.572
11	.625	.604	1.145	.507	-.649	.310	.275	.495	.770	.762	-.382	.713	.457
12	.477	.416	1.917	.157	-.636	.678	-.062	.623	.815	.100	-.049	.609	.544
	Dist. Index= .8044				Dist. Index= .6462		Dist. Index= .3900						
	MSD= 6.304	H=.461	MSD= 5.364	H=.503	MSD= 4.948	H=.582	MSD= 5.665	H=.505	MSD= 6.085	H=.430	MSD= 4.110	H=.600	

Table 21

First Canon City Subset - CAQ, Standardized
(N = 172)

Method: MIKCA Similarity: Mahalanobis Distance

CAQ Variables	Cluster 1 (F) N=25		Cluster 2 (G) N=24		Cluster 3 (A) N=64		Cluster 4 (D,E) N=34		Cluster 5 (B) N=14		Cluster 6 N=11		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	.533	.642	1.919	.276	-.622	.932	-.383	.870	-.631	.946	.209	.478	.758
2	.423	.655	1.764	.210	-.704	.868	- 088	.604	-.267	.461	-.104	.601	.643
3	.231	.580	-.231	.013	.131	-.016	-.743	.141	.742	.467	.569	.480	.177
4	1.020	.395	1.274	.420	-.711	.682	-.174	.589	-.322	.672	-.009	.454	.570
5	.670	.752	1.620	.394	-.869	.830	.106	.708	-.256	.796	0	.571	.714
6	.505	.453	1.149	.513	-.829	.681	.397	.554	-.776	.783	.929	.628	.604
7	.429	.419	1.367	.154	-.627	.693	-.333	.579	.791	.336	-.288	.233	.497
8	1.071	.696	1.011	.535	-.824	.722	-.135	.360	.764	.710	-.405	.666	.616
9	-.210	.140	-.759	.599	.387	.270	-.668	.330	.960	.560	.724	.104	.322
10	.752	.676	1.228	.196	-.701	.611	-.274	.481	-.045	.433	.590	.244	.499
11	.402	.611	.958	.513	-.860	.599	.318	.266	-.101	.553	1.145	.620	.522
12	.496	.435	1.464	.038	-.744	.724	-.041	.672	-.232	.757	.433	.195	.545
	Dist. Index= .7150		Dist. Index= .5363		Dist. Index= .7918		Dist. Index= .6467				Dist. Index= .0488		
	MSD= 5.324	H=.538	MSD= 7.798	H=.322	MSD= 4.335	H=.633	MSD= 5.673	H=.513	MSD= 4.204	H=.623	MSD= 6.114	H=.440	

Reliability Analysis using Solit-Samples

The reliability of the prior results was examined by an attempt to replicate the results on a second random sample of prisoners. Specifically, the methods used were UPGMA (cos), HMODE (cos), UPGMA (Euclidean), and HMODE (Euclidean). The comparison between these partitions, at a global level, again revealed only moderate agreement coefficients; e.g.:

	Rand	Jacc.	Symm. Lambda
UPGMA (Cos) (CAQ 1) vs. HMODE-Cos (CAQ 2)	.77	.41	.54

The overlap in comparing specific profiles was, however, much more encouraging. The EVAL tables included in the present section indicate that the larger clusters A, F, G, and the combination of (D,E) tend to recur as recognizably the same profiles. This is further indicated by the following matrix:

Table 22

Comparison of 5 and 6-group Solutions (UPGMA-cosine) from Experimental (PRIS1) and Validation (PRIS2) Subsets

Inter-Centroid Cosines (At the Origin) . UPGMA-cos (CAQ 1)

		1 (F,G)	2 (A)	3 (D)	4 (B)	5 (E)
UPGMA-cos (CAQ 2)	1	.994	-.967	-.121	-.298	-.219
	2	.247	-.340	.532	-.661	.633
	3	.096	-.121	-.199	.415	-.116
	4	-.975	.977	-.030	.359	.132
	5	-.102	.071	.495	-.420	.140
	6	-.180	.140	.265	.341	-.369

The depressed types (F, G) are obviously totally replicated. Similarly, the large nondepressed psychopathic group is well-replicated. Clusters D and E, which are so closely related in the original analysis, again find two matches in 2 and 5 of the CAQ 2 solution. Cluster B (bored paranoid), which lies within the center of the cluster distributions (see earlier cluster graph) is again close to more than one cluster in the cross-sample comparison.

The conclusion of this comparison is that although the profiles are not ALL matched exactly. We have evidence that the major differentiations of personality type are, in fact, replicated across the samples.

Table 23

Second Canon City Subset - CAQ, Standardized (N=162)

Method: MIKCA Similarity: Mahalanobis Distance
(Validation on MIKCA Partition Using UPGMA-R Centroid)
(Third Analysis)

CAQ Variables	Cluster 1 (D,E) N=61		Cluster 2 (F) N=17		Cluster 3 (G) N=18		Cluster 4 (A) N=66		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	-.044	.451	.300	-.042	1.874	-.138	-.547	.886	.511
2	-.009	.526	.805	.373	1.615	-.733	-.639	.874	.512
3	-.055	.211	-.166	.018	-.201	-.234	.149	-.140	-.002
4	.112	.394	.455	.394	1.609	.274	-.660	.647	.484
5	.273	.530	.711	.579	1.437	.190	-.827	.740	.583
6	.329	.449	-.074	.538	1.471	.649	-.687	.419	.468
7	.285	.287	.133	.368	1.258	.096	-.641	.509	.365
8	-.041	.459	1.266	.682	1.031	.280	-.569	.334	.411
9	-.184	.298	-.346	.208	-.787	-.555	.474	.235	.168
10	.089	.476	.883	.645	1.461	.093	-.709	.644	.520
11	.341	.433	.676	.582	.997	.548	-.761	.358	.431
12	.150	.561	.469	.347	1.574	.124	-.689	.565	.491
	MSD= 6.811	H=.423	MSD= 6.878	H=.391	MSD= 10.773	H=.049	MSD= 5.841	H=.506	

Table 24

Second Canon City Subset - CAQ, Standardized

Outliers Removed (N=162)

Method: UPGMA . Similarity: Euclidean Distance

CAQ Variables	Cluster 1 (D,E) N=62		Cluster 2 (F) N=10		Cluster 3 (A) N=67		Cluster (G) N=11		Variable Homogeneity
	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	Centroid	Homo-geneity	
1	-.054	.454	.455	.467	-.549	.904	1.855	-.160	.611
2	.090	.505	.991	.336	-.607	.842	1.180	.483	.643
3	-.029	.353	-.142	-.512	.126	-.084	-.039	-.056	.070
4	.194	.359	.397	.511	-.655	.674	1.562	.358	.510
5	.296	.522	.901	.565	-.766	.704	1.319	.732	.622
6	.357	.438	-.050	.634	-.705	.498	1.158	.803	.505
7	.201	.252	.079	.621	-.610	.513	1.286	.129	.384
8	-.039	.409	1.391	.761	-.476	.187	1.179	.599	.347
9	-.272	.304	-.208	.135	.454	.280	-1.087	.247	.278
10	.105	.417	1.050	.712	-.682	.633	1.516	.429	.534
11	.359	.500	.823	.652	-.793	.440	.880	.494	.483
12	.089	.586	.705	.178	-.595	.468	1.531	.594	.507
	MSD= 6.790	H=.425	MSD= 6.246	H=.422	MSD= 5.852	H=.505	MSD= 6.679	H=.388	

Concurrent Validation

Tables 25 and 26 show the results of analyses of variance on the external variables. This time the 16PF variables can be listed among the "external" criteria.

a. UPGMA-cosine - 5-group solution.

All but Factors B, F, I, M, and Q_1 differ significantly among the groups in this classification, as do the scores on Nurture IQ, Culture free, and Reading tests. Cluster 3 seems to contain a significant number of low scores on the three intellectual tests.

b. UPGMA-Euclidean - 6-group solution

16PF scales significantly differing across this classification are F, M, Q_1, and Q_2. Again, the intellectual measures all differ significantly with groups 1, 2, and 5 measuring particularly low on all three. The "security" and "control" measures differed significantly among the groups of the partition.

We stress that these ANOVA's should be tested with extreme caution, and should be used descriptively. The eta-squared values indicate a more realistic measure of strength of association.

Table 25. Cluster Means of External Variables

METHOD: UPGMA - Cosine - CAQ Clusters

	Cluster					ANOVA - Results		
Criterion	1 N=61	2 N=69	3 N=17	4 N=14	5 N=11	F	Sig.	ETA^2
Violent	.1803	.3043	.1176	.3571	.1818	1.3617	.2494	.0316
Murder	.3934	.1304	0	.9286	.9091	1.8349	.1244	.9421
Age	1.9344	2.1304	1.8824	2.0714	1.6364	1.2048	.3107	.0280
Nurture	81.8750	92.1905	81.1250	84.5385	94.1818	3.1018	.0173	.0746
Culture	92.2037	107.3284	89.6667	109.7692	113.6364	12.1699	.0000	.2390
Reading	87.1579	107.9077	84.6250	109.7692	113.3636	4.7024	.0013	.1070
PF 1 - A	3.3115	4.4783	4.4706	2.9286	4.4545	3.2514	.0134	.0723
PF 2 - B	4.1148	4.3478	3.4706	4.9286	3.6364	1.2859	.2776	.0299
PF 3 - C	3.5410	6.3043	5.6471	4.9286	5.0000	20.1942	.0000	.3260
PF 4 - E	4.6393	3.4783	3.9412	5.5000	4.0000	4.8328	.0010	.1037
PF 5 - F	4.4262	4.6957	4.0000	4.5000	3.8182	.8294	.5082	.0195
PF 6 - G	4.4426	5.6087	6.0000	4.5000	5.6364	6.0607	.0001	.1268
PF 7 - H	4.5738	5.9130	4.6471	4.7143	4.9091	3.9009	.0047	.0855
PF 8 - I	3.4098	4.1739	4.0588	3.7143	4.3636	2.2299	.0679	.0507
PF 9 - L	5.0164	2.9565	4.4706	4.3571	3.7273	11.3096	.0000	.2131
PF 10 - M	4.6885	4.6232	5.2353	4.7857	4.2727	.5585	.6931	.0132
PF 11 - N	4.8852	3.1884	5.4706	4.1429	3.8182	9.2750	.0000	.1818
PF 12 - O	4.6230	2.0725	3.2941	2.9286	4.5455	17.1141	.0000	.2907
PF 13 - Q_1	4.8197	4.8841	5.0588	5.6429	4.7273	.5557	.6952	.0131
PF 14 - Q_2	5.3607	4.6812	5.4118	6.2857	4.7273	2.7479	.0300	.0618
PF 15 - Q_3	3.6721	6.0435	4.7059	5.1429	5.1818	13.3720	.0000	.2426
PF 16 - Q_4	4.8852	2.3768	3.3529	3.2857	3.5455	17.9200	.0000	.3003
PF 17 - MD	4.2459	7.2609	7.1765	6.2857	6.2727	32.2107	.0000	.4355
R SCORE	1.5410	1.4203	1.0588	.2143	.2727	.5572	.6940	.0132
INOFF	.4918	.3768	.2941	.1429	.2727	.4614	.7640	.0109
S SCORE	.0897	.0732	.0422	.0095	.0057	.5295	.7142	.0125
SECURE	.5574	.5797	.5882	.9286	1.2727	.9962	.4113	.0233
STATUS	.3607	.2319	.2353	.5000	.7273	1.4450	.2214	.0335
CONTROL	.0984	.1739	.1765	.2143	.2727	.8024	.5253	.0189
W SCORE	.0753	.0431	-.1343	-.1932	-.1969	.4522	.7707	.0107

Table 26 Cluster Means of External Variables
METHOD: UPGMA - Distance - CAQ Clusters

Criterion	Cluster 0 N=4	1(F) N=40	2(G) N=10	3(A) N=80	4(E) N=22	5(G) N=9	6(B) N=7	ANOVA Results F	Sig.	ETA^2
Violent	.2500	.2000	0	.3125	.0909	.2222	.4286	1.6834	.1280	.0577
Murder	0	.6000	0	.2875	.1364	0	.8571	.7310	.6253	.0259
Age	2.2500	1.9750	1.8000	2.0500	1.7727	2.1111	2.2857	.6656	.6776	.0236
Nurture	73.7500	83.9167	74.4000	91.7703	86.4000	85.0000	81.0000	2.1631	.0496	.0787
Culture	101.6667	91.7368	98.3333	107.8974	100.0000	80.3750	111.4000	7.5191	.0000	.2277
Reading	102.0000	85.0000	91.3333	109.2208	88.4000	81.3750	120.1667	3.6614	.0020	.1241
PF 1 - A	3.0000	3.3750	2.4000	4.3375	4.0000	3.4444	3.2857	1.8839	.0864	.0741
PF 2 - B	5.5000	3.2250	6.0000	4.4375	3.8182	3.4444	5.4286	4.3826	.0004	.1375
PF 3 - C	2.0000	4.2750	3.1000	6.2750	5.1364	2.3333	3.5714	17.3581	.0000	.3870
PF 4 - E	4.2500	4.4500	5.3000	3.6875	3.7273	4.4444	6.5614	3.5561	.0024	.1145
PF 5 - F	3.0000	4.4000	4.5000	4.7250	3.8636	4.1111	4.8571	1.1256	.3496	.0393
PF 6 - G	4.7500	4.6750	3.9000	5.4875	6.2273	4.4444	3.4286	5.3930	.0000	.1640
PF 7 - H	4.2500	4.5500	4.5000	5.8625	4.7273	4.3333	4.2857	2.9719	.0088	.0975
PF 8 - I	2.5000	3.7250	2.5000	4.1250	4.1818	4.2222	3.0000	2.7343	.0147	.0904
PF 9 - L	4.2500	4.8750	5.4000	3.2125	3.6818	5.0000	5.5714	6.4371	.0000	.1897
PF 10 - M	4.0000	4.7500	4.3000	4.7250	4.7727	4.5556	5.0000	.2298	.9665	.0083
PF 11 - N	4.7500	4.6250	5.2000	3.3375	5.0455	5.3333	4.1429	4.8102	.0002	.1489
PF 12 - O	6.7500	4.1750	5.7000	2.1750	3.9091	4.2222	3.2857	12.9460	.0000	.3201
PF 13 - Q_1	3.2500	5.0750	3.9000	4.9625	5.0455	4.7778	6.0000	1.3598	.2338	.0471
PF 14 - Q_2	6.2500	5.1500	5.6000	4.8750	5.0000	5.8889	6.0000	1.0539	.3926	.0369
PF 15 - Q_3	3.5000	3.9000	2.2000	5.7750	5.3182	4.7778	5.1429	8.9255	.0000	.2450
PF 16 - Q_4	6.7500	4.5000	5.7000	2.5375	3.2273	4.5556	3.5714	12.7580	.0000	.3169
PF 17 - MD	2.5000	4.9250	2.2000	7.1375	7.0000	5.5556	5.0000	27.9904	.0000	.5044
R SCORE	1.0000	.8500	3.5000	1.2500	.9545	.6667	2.2857	.7802	.5866	.0276
INOFF	.5000	.3250	1.2000	.3750	.1818	.1111	.5714	1.5290	.1716	.0527
S SCORE	.0268	.0411	.3066	.0685	.0210	.0175	.0736	1.8820	.0867	.0641
SECURE	2.5000	.3750	.7000	.6875	.5455	.2222	1.4286	2.5648	.0211	.0853
STATUS	1.0000	.2750	.5000	.2875	.2727	.2222	.8571	1.4359	.2038	.0496
CONTROL	.7500	.0500	.1000	.2000	.1364	.0000	.2857	3.2295	.0050	.1051
W SCORE	-.1119	-.0968	.7932	-.0053	-.0772	-.1510	.0408	1.2596	.2788	.0438

Comparing the 16PF Classification with the CAQ Classifications

Cross-classifications were conducted between selected 16PF classifications and CAQ classifications. The overlap in group membership was only moderate. Rand Coefficients varied around 0.75 while the Symmetric Lambda scores were in the range of .15 to .36. Nevertheless, all scores tended to reach statistically significant levels of associations, indicating that, although class memberships were quite different, there was some positive but weak relation between the two instruments when used for classification.

CONCLUSIONS

Cluster analytic methods have had virtually no use in creating personality classifications systems using Cattell's 16PF and the CAQ. The present chapter has demonstrated that differential structuring of personality does, in fact, exist within prison samples. Different analytical clustering methods were applied to an extensive set of prisoner data. Personality types were discovered to exist and, to a certain extent, replicated both across different samples, using different clustering algorithms and using different similarity coefficients.

It is crucial to acknowledge, however, that although certain clusters were recurrent, the matches were not exact; certain persons were classified into different classes by different methods, and in some instances the methods showed differential sensitivity to different kinds of clusters.

The major structures in the data set--the large depressed cluster and the large psychopathic cluster--were identified by ALL methods in ALL samples. The finer subtypes, however, were not always identified. Sometimes they would be merged together. This occurred with the two subdivisions (F and G) of the

depressed prisoners. However, this division was picked up a sufficient number of times to lead us to the hypothesis that two subtypes of depressed prisoners exist and may be regarded as a reliable finding. In general, the two cross-sample validation exercises and the cross-method studies led us to reduce the number of clusters. It was consistently the case that we would drop certain clusters if cross-method and cross-sample evidence for their existence was not available.

The cross-method and cross-sample tests--the latter in particular--seem to be much more stringent tests than the concurrent validity test. This suggests that reliability work should certainly precede external validation work with real data sets. This is particularly important since the present analyses indicate that establishing evidence for the concurrent validity of taxometrically generated classifications is simply too easy to obtain. The frequent practice of conducting statistical tests of differences between the emergent clusters invariably leads to the establishing of "significant differences" and the noncompliance with the assumptions of the parametric significance test is all too easily ignored. Yet undoubtedly this practice will continue until the applied statistical researcher discovers a more appropriate methodology. The computation of Eta^2 values is a useful, perhaps partial, antidote to misinterpretations of significance testing. These bring the relation between the classification and the external variables to a more "descriptive" level, indicating the percent of variance in the external variable that is "explained" by the classification.

The sizes of the "violence" effect on the 16PF clusters can be judged approximately by the size of the Eta-squared statistic: Violence Eta^2 = .068 and Murder Eta^2 = .053, or about 6.8% and 5.3% of the variance

CONTINUED

8 OF 9

respectively. The effect size in control produced by the CAQ Euclidean distance solution is reflected by an Eta^2 of .105 or 10.5% of the variance.

At the substantive level the present results are exciting. They suggest that a reliable start can be made in working toward a taxonomy of prisoners based on personality. The link between personality and particular kinds of crimes may receive some help from the present work. For example, the cross-validation findings of the 16PF classification showed that one of those clusters contained a higher than expected number of prisoners convicted of violent crimes, including murderers. This same cluster averaged significantly lower on the intellectual measures, Nurture-IQ, Culture-free IQ, and Reading.

BIBLIOGRAPHY

Blackburn, R.
1971 "Personality types among abnormal homicides." British Journal of Criminology, 11(1), 14-31.

Cattell, R.B., H.W. Eber, and M.M. Tatsuoka
1970 Handbook for the Sixteen Personality Factor Questionnaire (16 PF). Institute for Personality and Ability Testing: Champaign, Illinois.

Ferdinand, T.N.
1966 Typologies of delinquency - a critical analysis. Random House: New York.

Megargee, E.I., and M.J. Bohn, Jr.
1979 Classifying criminal offenders: A new system based on the MMPI. Sage Publications: Beverly Hills, California.

Solomon, H.M.
1977 Crime and delinquency: Typologies. University Press of America: Washington, D.C.

CHAPTER 15

MAKING PREDICTIONS FROM EMPIRICAL CLASSIFICATIONS USING CLASS MEMBERSHIP AS A DUMMY VARIABLE

In the previous chapter, a series of prisoner classifications were developed by cluster analysing two subsamples of inmates of Colorado State Penitentiary. Cattel's 16PF scores and Clinical Assessment Questionnaire were used as bases for the classifications. The purpose of this chapter is to evaluate the utility of these classifications in predicting the incidence of several prisoner characteristics: (1) likelihood of violent behavior, (2) probability of requiring maximum security (control), and (3) the tendency to commit rule infractions. The ability of classifications to predict these external criteria are compared with the results of traditional linear least-squares regression analyses.

If the conditions of linear regression can be met, one would expect least-squares prediction to be more accurate than predictions based on categories derived from clustering. This is because categorization procedures usually are not aimed at optimizing the relation to a single external criterion variable. If, as is often the case in real applications, these conditions are not met, classifications based on theoretically relevant personality or clinical measures may perform as well, or better, than the traditional regression approach. For example, murderers and even the violent offenders are a very small proportion of the cases in the Canon City data. The standard regression approach often performs poorly when the criterion incidence (variance) is so small. Criterion characteristics are quite varied; therefore comparisons between the methods may help decide which approach produces the best prediction results for a given practical application.

METHOD

Two statistical measures can be used to compare results of prediction methods: (1) R-squared: the proportion of variance in the characteristic to be predicted which can be explained by the predictors (personality measures or categories), and (2) a measure of significance: which is used to assess the probability that a reported R-squared is actually larger than zero. These two measures are standard. They can both be computed as a part of a regression procedure or from the F-ratio in ANOVA (analysis of variance), and they are strictly comparable. Thus an R-squared of .05 from a regression procedure (interpreted as 5% variance explained) is always larger, for example, than a .04 implied R-squared value computed from the ANOVA based on a classification.

Actually, an even more direct procedure can be used. "Dummy" variables with values of zero, if the observation is not in a specified group, and of one, if it is, can be created. Then OLS linear regressions can be used to compute R-squared and significance measures on the continuous predictors (e.g., 16PF scores) and then on the "dummy" variables created by the classification being used. In this way, standard errors of R-squared and an error interval around the R-squared (e.g., $R^2 = .05 \pm .013$) can be computed. Regardless of whether the R-squared is significant, say at the .05 level, these measures can be computed.

In comparing the two prediction procedures, R-squared values and 95% error intervals can be computed from a regression using continuous predictors and from a regression using classification "dummies". The better prediction will have a larger R-squared (proportion of variance explained) and a smaller error interval. Even small values of R-squared can be compared in this manner.

CRITERIA

In order to compare prediction methods, the criterion variables to be predicted must be identified. Eleven of the subjects in the experimental subset were convicted of murder, thirty-one more were convicted of violent offense (not murder), and the remaining 130 were imprisoned for some other kind of offense. Additional information is available on the murderers: a coded rating (from 1 to 8) of the seriousness of the offense. Secondly, it is noted on the data file which prisoners were assigned to maximum security (N=27). Finally, there are extensive records of rule infractions for a substantial number of the subjects in the sample. From this information, four criterion variables can be defined: (1) level of violence, (2) non-violent/violent/murderer, (3) control (maximum security = 1, otherwise = 2), and (4) rule infractions.

1. Level of Violence

Since the seriousness ratings of murder charges have been coded in the Canon City data (from 0 to 8) it is possible to combine this information with the violence information (1 = violent crime, 0 = not a violent crime), to create a composite variable approximating the "level of violence". This new measure varies between 0 (for all non-violent offenses) and 9 (for 1st degree murder). All subjects convicted of violent crimes which were not associated with murder received a score of 1. This measure has the advantage of trying to account for the degree of violence and intent to violence as assessed by the prisoner evaluation staff. It also introduces more variance in the criterion variable, a statistically desireable property. The disadvantages include the extremely skewed distribution of scores (130 zeros, three nines, for instance) and the non-interval nature of the scale.

2. Ranked Measure of Violence

Since it may be hard to establish truly interval "levels of violence", a ranked measure can be constructed as follows:

0 = non-violent, non-murder offense

1 = violent offense, not murder

2 = all types of murder

This measure has the disadvantage of lumping murders all the way from third degree (perhaps only incidental to the crime) with convicted first degree murderers, assigning them a score of 2. This reduces the variance but helps alleviate the skew in the distribution.

3. Control

Twenty-seven prisoners have been assigned maximum security. The criterion "control" is a simple binary variable on which maximum security prisoners are assigned a score of one; all others a zero. It could be very useful to be able to predict a prisoner's tendency to require maximum security.

4. Rule Infractions

A useful criterion may be the tendency of prisoners to break rules. A simple additive variable is created by summing all the rule infractions for every subject with disciplinary records. An attempt to scale this variable, or to normalize it by dividing by the length of time (in months) over which the records spanned, failed to improve the correlations with the predictors.

PREDICTING VIOLENCE

Table 1 shows the results of predicting levels of violence and ranked measures of violence using (1) the classification "dummy" variables and (2) the traditional regression approach. Note that the prediction of the ranked measure is much better than the prediction of the level of violence using the

classification as a predictor. The regression predictions produce somewhat larger R^2 when the predictors are the 16PF scale scores. These R^2 values, however, were not significant at the .05 level, due to the size of the standard error of R. In addition to the usual R^2, we report the R^2 adjusted downward for the number of degrees of freedom. Since traditional regressions based on 16PF scores, in this case, contain many more predictors than the classification variables, the downward adjustment of the associated R^2 is much more severe.

The final results of these comparisons can be summarized as follows:

- Predicting from a classification based on 16PF scores is slightly better for this kind of data than is OLS regression on the original scores. Classification information results in a smaller standard error of R^2 than does the entire set of continuous interval variables.
- The only information needed to make the prediction is whether or not subjects are in cluster no. 1 of the UPGMA-correlation 8-group solution. In other words, membership in the 19 person subgroup created by the clustering analysis procedure is more predictive than the entire range of 16PF interval scores.
- When classification is used as a predictor, a ranked measure of violent activity is predicted more accurately (above 10.65% of variance) than a measure of "level of violence (about 7.69% of variance).
- In this case, the regression approach was no more successful than the classification approach in predicting a level of violence. In fact, the 95% error interval was largest (and included R2 = 0) for the regression.

Table 1

Predicting to Violence and Murder using Classification Based on 16PF Scores vs. Direct Linear Regression on 16PF Scores, Nurture Scores, and Age

a. Predicting the Level of Violence

	F	Implied R^2	R^2 adjusted	95% error	Signif.
1. From classification*	6.08	.0769	.0524	± .039	.003
2. From regression	1.71	.1890	.0782	± .091	.078

b. Predicting Ranked Measure (non-violent, violent non-murderer, murderer)

	F	Implied R^2	R^2 adjusted	95% error	Signif.
1. From classification	6.88	.1246	.1065	± .035	.000
2. From regression	1.62	.1809	.0690	± .073	.060

* Using UPGMA-correlation (8-group solution) Predictor=membership in cluster no. 1 (N=19)

Further confirmation of the relationship between violence and classification based on 16PF scores can be provided as follows. Consider that the ranked violence variable is actually a categorical variable and examine the distribution of nonviolent, violent offenses, and murders across the clusters of the UPGMA-correlation partition. We can condense all the clusters into the relevant groups, clusters 1 and 2, and a third category, "all others".

Table 2 shows the results of cross-tabulating ranked violence scores with these cluster categories both for the experimental subset and the validation subset. These cross-tabulations reveal a marked tendency for violent nonmurders to be found in group two and the murderers in group one. In the experimental subgroup both violence and murder were underrepresented in the other clusters. In the validation subset, violent nonmurders also occurred at about the expected rate in all clusters except clusters 1 and 2.

There is no doubt, then, that clustering using UPGMA with the correlation coefficients as measures of similarity produce a group containing more than the expected number of murders and a group with more than the expected number of prisoners serving for violent (nonmurder) offenses. One possible interpretation of these tables is that, (1) given membership in cluster 1, probability of being a murderer ranges from .12 (validation set) to .37 (experimental set) and (2) given membership in cluster 2, probability of being violent ranges from .21 (validation set) to .25 (experimental set). It can be verified that both probabilities are much lower in other clusters. Probability of being either violent or a murderer, given membership in cluster 1, may be as high as .89.

Since these groups were created by correlation, it is instructive to examine the profiles of these two groups.

Type 1: (N=19) - High Violence/Homicide Tendency. It may be recalled that this type is characterized by 16PF profile (A-, F-, H-, O-, Q_1+, and Q_2+). Outstanding features are their withdrawn, shy, embittered and strong sense of inferiority. They may be slow and repressed in expressing themselves. Schizoid maladjustment is also noted in this cluster.

Type 3: (N=55) - High Propensity to Violence. This cluster has the 16PF profile (C+, L-, N-, O-, Q_3+, Q_4-). Major features include tendencies toward calmness, strength of will, self-assurance, an absence of guilt, fearlessness, and a tendency not to care too much, or to be fussy. The O- indicates that these persons are prone to "acting out" their maladjustment rather than suffering it as an internal conflict (see also Cattell et al., 1970).

Table 2

Cross-tabulations: Cluster Memberships with Nonviolent Offenses, Violence, and Murders for Partitions of Experimental and Nonexperimental Subsets by UPGMA-8 Group Solutions

1. EXPERIMENTAL SUBSET — 172 cases

	Cluster (8-group solution) 1	2	All others	
Non-violent offenders	2 (1.5%)	38 (29.2%)	90 (69.3%)	130
Violent offenders	10 (32.3%)	(14) (45.3%)	7 (22.6%)	31
Murderers	(7) (63.6%)	3 (27.3%)	1 (9.1%)	11
	19 (11.0%)	55 (32.0%)	98 (57.0%)	

Chi square = 71.03 (4 d.f.), signif. = .000

2. VALIDATION SUBSET — 180 cases

	Cluster (8-group solution) 1	2	All others	
Non-violent offenders	47 (32.4%)	43 (29.7%)	(55) (37.9%)	145
Violent offenders	3 (13.9%)	(12) (52.2%)	8 (34.8%)	23
Murderers	(7) (58.3%)	3 (25.0%)	2 (16.7%)	12
	57 (31.7%)	58 (32.2%)	65 (36.1%)	

Chi square = 9.972 (4 def.), signif. = .040

Predicting Control

Table 3 shows the results of attempts to predict prisoner control (tendency to require maximum security). Again, the classification approach is more successful. The regression identifies only about 3.6% of the variance, but with an error of 5% on either side. The classification information is based on whether the case is identified as an outlier (cluster 0) and explains about 6.2% of the variance. Also this classification is derived from CAQ scores using Euclidean distance as a measure of similarity.

Table 3

Predicting to Prisoner Control (Maximum Security) using Classifications from CAQ Measures vs. Direct Linear Regressions on CAQ Variables

	F	Implied R^2	R^2 Adjusted	95% Error	Signif.
From classification *	2.230	.1130	.0624	± .054	.029
From regression	1.529	.1035	.0358	± .050	.119

* Using UPGMA-euclid. dist. (6 group solution). The prediction is made on whether or not the observation is classed as an "outlier", i.e., all "dummy" regression coefficients are negative.

Predicting Rule Infraction Behavior

The number of rule infractions appeared to be somewhat correlated with membership in cluster 2 of the UPGMA-Euclidean partition of CAQ scores ($r = .1456$, $p = .057$). As shown in Table 4, this correlation corresponds to a prediction, based on classification, of about 1.5% of the variance of rule infractions. This is a very low predictive ability, and marginally significant. But again, the regression on CAQ variables produces about the same R^2 (only slightly higher) but with an error range an order of magnitude larger.

It is doubtful whether prediction of this criterion from 16PF or CAQ scores could be in any way useful. What the comparison shows, however, is that the classification prediction is more likely than the regression approach to identify a very small effect.

Table 4

Predicting to Rule Infraction Behavior from Classifications Based on CAQ Measures vs. Direct Linear Regressions on the CAQ Measures

	F	Implied R^2	R^2 Adjusted	95% Error	Signif.
From Classification *	3.672	.0212	.0154	± .016	.057
From Regression	1.412	.0963	.0281	± .162	.165

* Using UPGMA-Euclidean dist. (6 group solution). The prediction is based on whether or not the observation is in cluster no. 2.

VALIDATING THE PREDICTIONS

All of the results reported above were obtained from the experimental subset (N = 172). The reliability of these results were tested by attempting to replicate the procedures on a second random subset of prisoners (N = 180). Table 5 summarizes these results. In general, only one prediction attempt was successful. Using the 9-group solution, cluster no. 1 of the validation subset (N = 29) was most similar to cluster no. 1 of the experimental subset (N = 19). It was expected that prediction efforts developed on the experimental set would be somewhat less successful on the validation set--especially with relationships as weak as they are in this study. The essential purpose however, was to compare the performance of the two methods. The adjusted R^2 of .027, while somewhat disappointing, does demonstrate that the clustering approach is again more successful than OLS regression. The 95%

Table 5

Validating the Prediction Results using Random Data Set No. 2 (N=180)

Prediction	F	Implied R^2	R^2 Adjusted	95% Error	Signif.
1. Violence from Classification[1]	5.578	.032	.027	± .021	.019
Violence from Regression in 16PF	1.058	.119	.007	± .046	.400
2. Violence (Ranked Measure) Clasif.[1]	2.048	.012	.006	± .010	.154
Violence (Ranked Measure), Regression	.643	.076	.000	± .405	.868
3. Control from Classification [2]	.184	.003	.000	± .223	.907
Control from Regression on CAQ	.907	.065	.000	+ .018	.541
4. Rule Infraction from Classification[3]	1.035	.018	.001	± .005	.378
Rule Infraction from Regression	.631	.046	.000	± .044	.813

NOTES:

1 - Using UPGMA-correlation (9-group solution). Predictor = membership in cluster no. 1 (N=29).

2 - Using UPGMA-Euclid. distribution (6-group solution). Again, the prediction is made on whether or not the observation is classed as an "outlier".

3 - Using UPGMA-Euclidean distribution (6-group solution).

4 - The only significant prediction at alpha .05.

error intervals tend again to be about half the size of the error generated by the regression approach.

The major methodological difficulty in validating a classification prediction is to insure that the partition on the validation subset matches as closely as possible the partition on the experimental subset. Comparisons need to be made at several levels of both trees. In the present case, the 9-group solution on the validation subset compared most favorably with the 8-group solution on the experimental subset. No other comparisons were made. If the comparison between partitions is good, the researchers can be sure that prediction results based on classification will be well replicated.

Although the other predictive exercises are inconclusive there is a trend for R^2 from classification to be higher than from regression. The attempt to predict "control" is an obvious exception.

CONCLUSIONS

The proportion of variance in the experimental data which can be explained using the 16PF scores, or the CAQ scores, on our several criteria is quite small. When predicting level of violent behavior from 16PF scores, only about 7.69% of the variance can be explained using traditional regression techniques. Using classification based on 16PF scores, about 10.65% can be explained. Predicting from classification, therefore, fared somewhat better than traditional OLS regression. Predictions of control status of prisoners (tendency to require maximum security) and tendency to break rules were much less predictable. The results which could be considered significant were produced by the UPGMA-euclidean distance classification method applied to CAQ scores. Only about 2-3% of the variance could be explained. Again, the classification information provides a better set of predictors.

When the criterion variable is a rankable or categorical variable a cross-tabulation with the classification (cluster memberships) can provide a great deal of information. Clusters may have to be artificially merged to provide meaningful categories for cross-tabulation.

The most important substantive results of this prediction study may be the identification of two clusters in the data which could provide meaningful characteristic profiles of murderers and violent offenders.

Suggestions for future studies of this type:

- Secure samples with larger numbers of violent offenders and murderers.
- Secure more precise <u>behavioral</u> descriptions of the subjects, offenses; i.e., better criterion measures.
- When establishing the predictive classifications, utilize comparisons of several (not just one or two) K-levels or levels of the tree.

Validation using a second set of randomly selected observations on which to test the clustering and predictions derived from the first set of observations is a rigorous and necessary process. The results of deriving predictions and testing them on the same data set unnecessarily inflates the experimenter's confidence in the results. Much could doubtless be gained, rather than simply reclustering on the second data set, from trying a number of cluster <u>assignment</u> procedures. The preliminary results of our current inquiry would certainly suggest that further investigations of predicting from empirical classifications would be a worthwhile enterprise.

The present exercise of using the class-membership as a dummy variable in regression approaches to prediction is simply one approach to typological prediction. It is acknowledged that there are many different ways of using type-membership for predictive purposes. We would also, therefore, propose greater examination of the use of typological methods for predictive purposes,

particularly given the relative success of the typological approach in the present study. The work of Tryon and Bailey (1970), Armstrong and Andress (1970), Sonquist and Morgan (1963), Forgey (1965), and Gower (1974) would all represent useful avenues for the exploration of the use of classifications for prediction.

BIBLIOGRAPHY

Armstrong, J.S., and J.G. Andress
1970 "Exploratory analysis of marketing data: Trees vs. regression." Journal of Marketing Research, 7, 487-92.

Cattell, R.B., H.W. Eber, and M.M. Tatsuoka
1970 Handbook for the Sixteen Personality Factor Questionnaire (16PF). Institute for Personality and Ability Testing: Champaign, Illinois.

Forgy, E.W.
1965 "Analysis of multivariate data." Biometric Society Meetings, Riverside, California.

Gower, J.C.
1974 "Maximal predictive classification." Biometrics, 30(4), 643-54.

Sonquist, J.A., and J.N. Morgan
1964 The detection of interaction effects. Survey Research Center Monograph No. 35. Institute for Social Research, University of Michigan: Ann Arbor, Michigan.

Tryon, R.C., and D.E. Bailey
1970 Cluster analysis. McGraw-Hill: New York.

CHAPTER 16

CONCLUDING REMARKS AND RECOMMENDATIONS

The present research project has covered a rather broad range of topics. These will not all be summarized here, and the reader is referred to the concluding statements in each of the earlier individual chapters. The objective of the present chapter is to draw attention to some of the more general conclusions and recommendations regarding classification methods for criminology and the criminal justice system.

THE DIFFUSION AND UTILIZATION OF NUMERICAL CLASSIFICATION INTO THE CRIMINAL JUSTICE SYSTEM

This research project has paid attention to both innovative taxometric methodology as well as the on-going role of classification in the justice system. We have, therefore, an interest in the diffusion and acceptance of this new methodology into criminological research and the justice system. The following represent factors that seem critical in the diffusion of numerically based classifications. We note that diffusion implies both utilization as well as acceptance of such classifications.

Scientific Quality versus User-Acceptance

The criteria associated with the scientific quality of classification systems are markedly different from the criteria that govern user-acceptance of a classification. Scientific quality is governed primarily by the criteria of theoretical coherence, descriptive accuracy, reliability of measurement, generalizability across samples and range of applicability. Basic research almost totally emphasizes scientific quality. User-acceptance of a classification, on the other hand, is based primarily on factors such as: simplicity and parsimony, ease of learning and remembering, face validity

(ease of recognition clinically), an effective nomenclature, clinical utility (clear and predictively valid differential interventions), and ease of use (easy identification of unknown cases).

These two general goals are not always consistent with each other, e.g., scientific quality may demand large numbers of sub-types, or mixed dimensional categorical structures. These may become quite complex. This increase in complexity may severely disrupt the clinical acceptance of such classifications (see Skinner and Blashfield, 1980).

The classifications stemming from taxometric methods are not likely to immediately have the features which would aid acceptance. We emphasize that although the scientific quality of such numerically based classifications is liable to be higher than that of the currently existing criminological classifications, they may not be accepted because of inattention to the basics of user-acceptance. The advocates of this new technology must not only ensure the scientific quality of their classifications, but must attend to the criteria that govern acceptance. Brewster Smith (1980) notes that:

> . . . any new taxonomy should have clinical utility as well as scientific value . . . anything less might not be worth the time and money.

Fragmentation and Proliferation of Numerical Studies

The development of criminological classifications of high scientific quality (broad range, power, and so on), will represent an enormous undertaking. A well integrated research effort will be required to effectively deal with this task. A systematic research effort is one in which different classification studies can be related to each other, and through which

there can be a steady cumulative building of knowledge. A serious danger to the diffusion and acceptance of numerically based methods is the possibility of a fragmented, non-systematic research effort in which large numbers of ad hoc isolated numerical classifications are generated. Such a non-integrated proliferation of numerical systems (of perhaps dubious quality) would promote confusion, skepticism and would not enhance user-acceptance. Since the methodology of numerical classification is still under rapid development, with some major challenges still unsolved, we must also reserve judgment on the scientific quality of its products.

Timing and the Perceived Need for Innovation

Innovative approaches to classification will not be accepted into institutional practice unless there is a climate of acceptance and a perceived need to change (Skinner and Blashfield, 1980). In criminology we have already discovered (see chapters 2, 3 and 4 of the present volume) a pervasive dissatisfaction with current classification systems and related processes. This has sometimes generalized into over-reaction and to a specific focus on the negative effects of classification (labelling, stereotyping, etc.). Yet CJS practitioners cannot avoid classification, since it is embedded in almost all phases of management, treatment and decision-making.

The present climate in criminal justice systems agencies is, therefore, mixed. There is a perceived need on the one hand for better classifications, but on the other, skepticism, vested interests, inertia, etc. The benefits that would stem from higher quality classification systems would seem to inevitably propel some systems towards innovation and experimentation.

Towards Improved Classification Systems in Criminology

Guidelines for improvement in both the construction of classification systems and their applied uses require an articulation of the relation between the end-use of a classification system and that system's structural characteristics. As mentioned earlier, this articulation is at the earliest stages and much more attention should be given to this task. The following recommendations are made tentatively, acknowledging the complexity of these interrelationships.

User-purposes should be explicated and refined. We have stressed that end-uses in criminal classification are often implicit and confused, and as a result may work at cross-purposes. The user often has multiple purposes. These should be unravelled and clarified. User purposes must guide the structuring of new classifications, and it is invariably disastrous if these are not made explicit. This also underlines the necessity of the basic research worker having a clear idea of the goals of the criminal justice system practitioners. Classification research must be well integrated with everyday practice.

Definition of attribute spaces should be upgraded. The sets of variables entering the attribute space in criminological classification have often been of very poor quality. Inadequate input data inevitably undermine the quality of the resulting classification, irrespective of the methodology used for clustering. Therefore, perhaps a foremost challenge in ALL areas of criminological classification is the development of improved, refined, reliable, and coherent attribute spaces. Irrelevant variables should be

identified and eliminated. Precision of measurement should be generally upgraded. There should be an ongoing attempt to provide a clearer articulation of the relation between criminological theories and the specification of the attribute space for classifications. Taxometric studies, in turn, may be helpful in guiding the search for new relevant variables (see Enc, 1976; Sneath and Sokal, 1973). There is, therefore, a necessity to work for consistency between user-purpose and the selection of appropriate attribute spaces. Theory development, descriptive research, and prediction may all imply different kinds of attribute spaces.

There should be consistency between methodology and the user's purpose. We have stressed in earlier chapters that the "blind" use of classification methods - in which many of the assumptions and operational processes of the methodology remain implicit - will hinder improvement in developing good quality classifications. The explication of the nature of both grouping techniques, similarity assessment and data type is required to allow the user to approach consistency between the methodological options, the data, and the user's purpose. The requirement of such consistency cannot be overstressed (see also Mojena, 1977; Ball 1970). Our chapters dealing with comparative studies of similarity assessment and grouping techniques fit into this general task of elucidating the operating styles of these clustering methods.

Several clustering methods should be used in creating classifications. The current ambiguities emerging from both the theoretical comparisons of clustering methods, and the Monte Carlo comparisons, obviate clear guidance to the applied researcher. The absence of clear-cut mathematical criteria,

or a single statistic, that would serve as an undisputed indication of cluster validity (see Dubes and Jain, 1979; Blashfield et al. 1978) also compounds the problems of applications research. Blashfield et al. (1978) write:

> . . . it is impossible to recommend any single validation approach as the best for use in applied research. Simply put, not enough is known about these various approaches to support their uncritical use.

We have come to accept the position that at the present state of methodological development the applications researcher is advised to utilize several different clustering approaches in creating a classification. He/she would also be advised to use several validation approaches in evaluating the resulting classifications and in searching for common clusters. Dubes and Jain (1979), Anderberg, (1973), and Blashfield et al. (1978) similarly imply that this may be the most defensible tactic. Indeed, the concept of repeated analyses, perhaps experimenting with different similarity coefficients, or refining the attribute space by elimination of different variables, etc., fits well with the idea that clustering and classification are tools for discovery within the general context of exploratory data analysis (see the early papers by Ball, 1970; Anderberg, 1973). Successive analyses, guided by reformulations and prior results and utilizing different methods, should help the applied researcher, given the ambiguity regarding strict statistical criteria for validation of clusters.

Different clustering methodologies used should complement each other. The various clustering techniques differ in certain important aspects. To the extent that these differences may complement each other (see Anderberg, 1973; Ball, 1970), a prudent approach is to use several methods

to deal with specific features of the overall data analysis plan. Inexpensive methods can be used first or on larger data bases, different clustering techniques can be used in the search for different kinds of clusters, and different evaluation approaches can be used to reveal different structural aspects of the classification system. We advocate the use of different procedures in these specialized tasks.

Comparing Classification Methods: Choosing a Method

Our comparisons of various clustering methods have underlined the fact that there are multiple criteria, and multiple characteristics, on which clustering methods can be compared. Increasingly, our objective has become not the choice of a "best" method, but rather to attempt to understand and delineate the peculiarities and "classification styles" of the different techniques - particularly with reference to user purposes. Dubes and Jain (1976) have noted:

> The subjective nature of the clustering problem precludes a realistic mathematical comparison of all clustering techniques.

and later:

> . . . The objective is not to choose a "best" clustering technique or program. Such a task would be fruitless and contrary to the very nature of clustering.

Yet the applied use of such techniques is still confronted with the practical necessity of choosing among the available methods. In the following remarks we relate our own results, reported in detail in the previous chapters, with the general findings of other research workers on this topic.

Ward's method. This technique has been found to perform best in a number of comparative studies (Mojena, 1977; Blashfield, 1976; and others).

Our results cast doubt on the conclusion that Ward's method gives the most accurate recovery of the known structure. In many of the comparative studies conducted in our present research Ward's method was outperformed by other methods, particularly when the data structures were highly complex. Unequal cluster sizes combined with non-spherical shapes seemed to seriously disrupt the ability of this method to correctly recover the classification structures. These findings are consistent with those of Milligan and Isaac (1980) and Sokal and Milligan (1980).

Group Average method. This method generally provided the most accurate cluster recovery in our comparisons. It generally outperformed Ward's method, HMODE, and the K-means approach. Four comparative studies (Edelbrock, 1979; Edelbrock and McLaughlin, 1979; Milligan and Isaac, 1980; and Cunningham and Ogilvie, 1972) also suggest that the Group Average method gives good recovery of clusters.

K-means methods. When the correct K level was known, and when a "fortunate" starting configuration was used, we found that these partitioning methods gave excellent results. However, we also found that they are extremely sensitive to the initial starting configuration and frequently converged upon partitions which had very little relationship to the correct data structure. In the analysis of real data, where there is little knowledge of the correct K level, these represent critical problems. The joint use of K-means with other methods (e.g., HMODE) to provide evidence of an appropriate range of K levels and approximate starting configurations is recommended.

Comparative performance of methods depends upon types of data structures. It may be noted that the Group Average method did not always outperform

the various other methods used. The results suggest that there is a highly significant interaction between type of clustering method and the form of data structures (e.g., cluster shape, relative sizes and concentrations, distances apart, noise, etc.). Currently, there is no clear understanding of the exact nature of this method by data-structure interaction (see Milligan and Isaac, 1980; Dubes and Jain, 1980). A basic reason for this absence is that only a handful of Monte Carlo studies have been conducted and many of these have utilized fairly restricted kinds of artificial data. Much more work with carefully structured and complex artificial data is required to clarify the manner in which different methods deal with different data structures.

Limitations of the Present Comparative Studies

Although the comparative studies reported in the present research project are fairly broad - particularly with reference to the data types used - we acknowledge a number of limitations. The most serious of these is the limitation on methods compared. We utilized representatives of most of the major families of clustering methods, i.e. hierarchical, iterative partitioning, and mode search methods. However, we did not examine dimensional/ordination techniques, divisive monothetic methods, inverse factor analysis, or the methods for separating mixtures. Additionally, a rigorous examination of ALL interactions between grouping technique, similarity coefficient, data type and user purpose would be a tremendous undertaking and was beyond the scope of this project. Similarly, we examined only eight of the 50-plus similarity coefficients which can be found in the clustering literature.

BIBLIOGRAPHY

Anderberg, M.R.
1973 Cluster analysis for applications. Academic Press: New York.

Ball, G.H.
1970 Classification analysis. Technical Report. Stanford Research Institute: Menlo Park, California.

Blashfield, R.K.
1976 "Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods." Psychological Bulletin, 83, 377-88.

Blashfield, R.K., M.S. Aldenderfer, and L. Morey
1978 "Cluster analysis literature on validation." Paper presented at the Classification Society Meeting, May, 1978.

Cunningham, K.M., and J.C. Ogilvie
1972 "Evaluation of hierarchical grouping techniques - a preliminary study." Computer Journal, 15(3), 209.

Dubes, R., and A.K. Jain
1976 "Clustering techniques: The user's dilemma." Pattern Recognition, 8, 247-60.

Dubes, R., and A.K. Jain
1979 "Validity studies in clustering methodologies." Pattern Recognition, 11, 235-54.

Dubes, R., and A.K. Jain
1980 "Clustering methodologies in exploratory data analysis." Advances in Computers, 19, 113-228.

Edelbrock, C.
1979 "The accuracy of hierarchical clustering algorithms: The problem of classifying everybody." Multivariate Behavioral Research, 14, 367-84.

Edelbrock, C., and B. McLaughlin
1979 "Intraclass correlations as metrics for hierarchical cluster analysis: Parametric comparisons using the mixture model. Paper presented at the meeting of the Classification Society, Gainesville, Florida.

Enc, B.
1976 "Spiral dependence between theories and taxonomy." Inquiry, 19, 41-71.

Milligan, G.W. and P.D. Isaac
1980 A review of Monte Carlo tests of cluster analysis. College of Administrative Science, The Ohio State University: Columbus, Ohio.

Mojena, R.
1977 "Hierarchical grouping methods and stopping rules: An evaluation." The Computer Journal, 20, 359-63.

Skinner, H.A., and R.K. Blashfield
1980 "Psychiatric classification and cluster analysis: A tradeoff between scientific quality and clinical acceptance." Unpublished manuscript. Alcoholism and Drug Addiction Research Foundation.

Smith, B.
1980 APA Monitor, 2(1), 33.

Sneath, P.H.A., and R.R. Sokal
1973 Principles of numerical taxonomy. W.H. Freeman and Co.: San Francisco.

Sokol, L., and G.W. Milligan
1980 "The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure." Unpublished manuscript.

END