ELEMENTARY LINKAGE ANALYSIS FOR ISOLATING ORTHOGONAL AND OBLIQUE TYPES AND TYPAL RELEVANCIES[1]

LOUIS L. McQUITTY
Michigan State University

SEVERAL of the terms of the above title are analogous to key concepts in factor analysis; types correspond to factors, and typal relevancies correspond to factor loadings. These terms have similar meanings to the corresponding ones in Thurstonian factor analysis except for one fundamental difference; the terms of elementary linkage analysis are here applied to a theory of psychological structure which is basically different from simple structure as developed by Thurstone (16, p. 181). However, data are sometimes so interrelated that the difference in theory does not express itself in empirical results; linkage analysis sometimes yields results very similar to, or even identical with, those of rotated factor-analytic solutions.

Advantages of elementary linkage analysis are its provision for investigating a particular theoretical position, its speed, and its objectivity. A fifteen variable matrix can be analyzed into objectively determined types in five to ten minutes, all operations with only pencil and paper. The original solution of a linkage analysis gives the desired structure; no rotation is required.

Need for the Method

Elementary linkage analysis is a method of clustering. It can be used to cluster either people or items, or any objects, for that matter, which have distinctive cluster-characteristics. In order to make our discussion somewhat concrete, rather than abstract, we shall talk of cluster methods as procedures

[1] This article was completed while the author was on the Faculty of the University of Illinois.

for classifying people. What we will say applies equally to items or other objects which might be classified by the methods.

The need for cluster methods with particular characteristics has been specified by other authors. Thorndike (15), for example, has emphasized that a method of clustering is needed which will yield from the data both the number of clusters required and the membership most appropriate for each person. Earlier, Cattell (2) reviewed cluster methods and classified them under four headings: (a) The Ramifying Linkage method, (b) The Matrix Diagonal method, (c) Correlation Profile Correlation (Tryon's method), and (d) The Approximate Delimitation (or Convergence) method. Cattell points out that "All methods require, directly or indirectly, that the experimenter shall begin by fixing some lower limit to the magnitude of correlation coefficient which will be accepted as qualification for entry to a cluster. . . . When some minimum of this kind has been fixed . . . , cluster search becomes first a matter of looking for linkages, a linkage being defined as a 'significant' correlation, i.e., something above the agreed minimum."

A defect of the above methods is that they depend on an arbitrary decision for a determination of the lower limit for admission to a cluster. Elementary linkage analysis remedies this defect by defining the linkage as the largest index of association which a variable has with any or all of the other variables. A more sophisticated version defines it as the largest index of association which a variable has with a composite of all the characteristics of the members of a cluster. As will be shown later, every variable is assigned to a cluster in terms of its highest index of association. Consequently, the lower limit of associations which is used in building the clusters is determined by the data exclusively.

Cattell (2) summarizes other limitations of methods of clustering. All of the methods are laborious, and some of them are not exhaustive and consequently miss some of the minor clusters. Despite these limitations, Cattell recognizes that cluster methods are valuable in "the reduction of an almost endless variety of tests and ratings to a comparative small number of representative variables." Even though he then prefers to perform the more laborious task of factor analysis on the reduced set of variables, he points out that some investigators prefer a reduction into clusters only or into both clusters and factors. And even Cattell states that "it may be desirable to publish analyses of correlation matrices both in cluster analysis and factor analysis form; for the former will preserve results in a shape suitable for immediate collation with those of other researches."

The above survey emphasizes values for clustering methods and states certain characteristics which a good method should have. All of these characteristics are possessed by elementary linkage analysis, as will be shown later, but before developing the method it is helpful to develop still another purpose for which clustering methods are required. They are needed for the investigation of typal theories, as will be emphasized by outlining a typal theory of human behavior.

A Typal Theory

This typal theory can be introduced by contrasting it with the theory of simple structure developed by Thurstone (16, p. 181). Instead of seeking a simple structure of human behavior in the Thurstonian sense, linkage analysis seeks a typal structure. A typal structure is defined as one in which every member of a type is more like some other member of that type (with respect to the data analyzed) than he is like any member of any other type.

Once the members of a type have been isolated, it is then possible to define a prototype. A prototype is some composite of the characteristics possessed by the members of the type. Operationally, it may be defined as the centroid of the characteristics possessed by the members, or it may be defined as the characteristics possessed in common by all members of the type. If the centroid is used as the operational definition of a prototype, it is then possible to compute the relevancy that each member of the type has to the prototype. Relevancies to prototypes are computed by means of the same operations as the first factor loadings of a factor analysis. The only difference is that these operations in linkage analysis are applied to submatrices rather than the entire matrix. The first step of elementary linkage analysis selects the members of the first type. The coefficients of relationships between these members constitute the entries of the first sub-matrix. First factor loadings (computed on the variables of this matrix exclusively) constitute the relevancies for the first prototype. The members of the second type are then selected. Coefficients of relationships between these members constitute the second sub-matrix. The factor loadings on the centroid of this sub-matrix constitute the relevancies for the second prototype. Analogously, relevancies for subsequent prototypes are computed until every variable has been classified into a type.

The members which represent the first type are selected in terms of the original index of association. Those representative of subsequent types can be selected either in terms of the original indices, or in terms of residuals, where the latter term is defined to have an analogous meaning to the one which it possesses in factor analysis, viz., indices expressing the degree of association between variables after that variance which they have in common with extracted types has been withdrawn.

Whether an investigator should use original indices of association or residuals in selecting subsequent types depends on the psychological theory being investigated, as can be shown by a brief consideration of one aspect of most typological theories. Most typological theories attribute a characteristic "form" or "style" to the members of a type. This form is said to have a unifying and structuring effect; it is said to play a role in giving meaning to the members of a type. Consequently, the meaning of a member can be completely understood only if seen in the light of the characteristic style of the type. From this point of view, and in conformance with clinical theory, a given response has different measurement and diagnostic indicants depending on the type of subject who gives it. For example, both a mental patient and a healthy minded individual might say *yes* to the question: *Do you usually feel at ease?* The patient might answer in the affirmative, because he is unconsciously an ill-at-ease person who rationalizes; his *yes* response might indicate that he is a rationalizing type of person. The healthy individual, on the other hand, might reply in the affirmative because he is a well-adjusted type of individual who is in all respects at ease most of the time. His *yes* would reflect mental health. Under the circumstances of these two cases, the same answer would have been given for two quite different reasons. Since the answer would have been given for different reasons, it would seem reasonable to argue that it might be possible to find that the answer has different meanings in terms of the assessment of the two individuals. This differential meaning in the two cases might be indicated by the particular patterns of responses given by the two subjects to a number of items related to mental health status. The possibility of finding these differential patterns would depend on the ability of an investigator first, to select items which would yield differential patterns, and, secondly, to analyze the responses to them in such a manner as to reveal that the patterns are differentially related to mental health status. In such an approach, it would be necessary to realize that there might be many different patterns which reflect mental health and mental illness. As a consequence, large numbers of subjects would have to be studied in order to reveal the statistical significance of such patterns.

From the point of view just developed, patterns of responses would be associated with characteristic styles of behaving. Two different styles might yield identical responses to some few items, but these responses could be shown to be component parts of different patterns; they would therefore have different meanings.

If an investigator desired to retain the full implication of the differential meanings of patterns of responses in relation to the peculiarity of the styles of behaving which produced them, he would not subtract from one pattern the variance that it has in common with another (as represented by identical answers to a few items). Instead, he would interpret each pattern separately in relation to other characteristics of the subjects who yielded it.

Despite what has just been said about the presumed role of "form" in giving structure and meaning to the members of types, it is also reasonable to assume, by way of contrast, that unique characteristics of types are determined exclusively by the particular elements of which they are composed, and that each element has the same meaning and measurement indicant irrespective of the combination in which it occurs. When this position is accepted in conjunction with a theory of parsimony, an investigator would be expected to account for as much variance as possible by each type; he would use the original indices of association for selecting the first type only, and residuals for all subsequent types. This parsimonious version of elementary linkage analysis corresponds to the isolation of orthogonal factors, and the form version corresponds to the isolation of either orthogonal or oblique factors as indicated by the data.

The fundamental difference between factor analysis and elementary linkage analysis is in terms of the assumed structure which is being investigated; Thurstonian factor analysis is designed to isolate simple structure, but linkage analysis is designed to isolate typal structure (as already defined). However, the results from an elementary linkage analysis (based on both original indices of association and residuals) can be rotated to a simple structure, and the method represents a less laborious approach than is usually required in most methods of factor analysis. The reduction in labor is attributable to the fact that elementary linkage analysis can be used, as just outlined, to combine an objective clustering with factoring and therefore is in this usage a multiple group method.

The above discussion applies to indices between either people or tests, and it should, therefore, be recognized that there are differences of opinion first, as to whether simple structure should be sought in interrelationships between people, and, secondly, as to the conditions under which it is appropriate even to compute indices of relationship between people over tests or test items. These issues are well discussed by Cattell (4) and Stephenson (13), and it is not a purpose of this paper to review them. An investigator who is considering interrelationships between people over tests or test items would be advised to consider the positions set forth by Cattell and Stephenson.

Areas of Application

Elementary linkage analysis will first be developed as it pertains to the isolation of types; it will then be illustrated and extended to the determination of typal relevancies. Before developing elementary linkage analysis, it is appropriate to mention some areas of application.

A search for an objective method capable of classifying either tests or persons (or both) into a number of types which is determined by the data represents a continuing interest among psychologists. Various methods of clustering have been applied to isolate (a) categories of similar items in test construction research (7), (b) categories of similar people in developing typological theories of behavior (14, pp. 153–189), (c) clusters of descriptions of self and others in the study of psychotherapy (12, pp. 222 and 223), (d) groups of tests as a step in the multiple group method of factor analysis (1), (e) key tests which are representative of groups of tests (18), (f) categories of similar jobs (15), and (g) for other similar purposes. Elementary linkage analysis is appropriate to all of these purposes and has the advantages of being simple, objective, rapid, and appropriate for matrices of all reasonable sizes, even up to 200 variables.

The Method

Definition of Type

The number of types (or families) represented by any set of interassociations (or distances) among people or tests can be expected to be influenced by the definition of types. Consequently, the definition of types is of prime importance. In fact, it is the particular definition of a type here given that makes the solution to the number of types, as well as typal membership, both simple and objective.

In order to make the definition of type realistic, it will be given in terms of persons rather than in the abstract; it could be given equally well in terms of tests, test items, or jobs. A type is here defined as a category of persons of such a nature that everyone in the category is in some way more like some other person in the category than he is like anyone not in the category. In terms of coefficients of correlations between persons, every person in a type would have a higher correlation with some other person in the type than he would with anyone not in the type.

Index of Association

The exact meaning of the definition just given depends on both (a) the index of association to which it is applied, and (b) the kind of data which is analyzed. Two kinds of data can be recognized: ordered and unordered (17). In most quantitative analyses of test results, the data are first ordered to linear continua and studies are often made of the interrelationships of standings on several continua. However, it is also possible to take an index of agreement on unordered data, i.e., merely to count the number of items on which two subjects agree in the answers they give to a test (19).

If an investigator is concerned with the isolation of types on linear continua (by classifying together those who have similar profiles) he has the option of selecting an index which is influenced by "shape" only, or one which is influenced by both "shape" and "level" (3, 5).

In deciding whether to analyze ordered or unordered data, an investigator should recognize that two individuals may obtain the same standing on a psychological continuum by two very different patterns of responses to the individual items of a test (19). As a consequence, the isolation of types from ordered data may miss much of the individual differences in "style" as discussed above.

A disadvantage of analyzing unordered data is the fact that responses to individual items are generally known to be relatively unreliable, and this is one of the reasons for favoring the isolation of types on profiles of standings. However, even though responses to individual items may be in general unreliable, they may nevertheless have "differential reliabilities" across people, i.e., a particular response though generally unreliable may be highly reliable for the members of a particular type of persons; all members of this type may give this response invariably, even though other individuals fluctuate with respect to it (9).

Elementary linkage analysis, as here developed, can be applied to anyone of the indices indicated above. The choice of an index should be made by the investigator in relation to its characteristics and the purpose of his study.

Mathematical Development

Using the definition of a type given above, it is possible to prove both that there is an objectively determinable number of types represented by every matrix of interassociations between variables, and that each variable has an objectively determined typal membership.

Let $r_{i_1j_2}$ represent the index of association between any two variables such that variable i_1 has its highest r with j_2, and j_2 has its highest r with i_1. Such a pair of variables is said to be reciprocal. By the definition of a type, the variables of every reciprocal pair belong to the same type.

Every matrix has at least one reciprocal pair of variables, because every matrix has one or more highest r's. If the highest r of a matrix is for $r_{i_3j_4}$, then both i_3 has its highest r with j_4, and j_4 has its highest r with i_3; the two variables are therefore reciprocal.

A matrix may have several reciprocal pairs of variables, because i_1 may be highest with j_2 and the reverse, even though i_3 is highest with j_4 and the reverse, etc., and this can continue so long as the pairs do not have common members.

Assume a matrix in which there are $n/2$ reciprocal pairs. Consequently, as shown above, $0 < n/2 < m/2 + 1$, where m is the number of variables in the matrix, and n equals the number of variables which are in reciprocal pairs. Everyone of these variables is assigned to the type represented by the pair of which it is a reciprocal member.

Let any variable which has been assigned to a type at any given stage of the analysis, be represented by i_t, and any which has not yet been assigned to a type be represented by i_n. Then for every matrix which has one or more i_n's, there is an i_t with which each i_n has its highest r. This fact is proven by assuming the contrary to be true and then showing that it leads to an inconsistency. Assume that there is no i_t with which an i_n has its highest r, then the highest r's for i_n's would mediate between i_n's only. Consequently, some one of these highest r's would be the largest of all of them, and this would make its members reciprocal. But all variables which are reciprocal have already been identified and included as i_t's. Consequently, once all reciprocal pairs have been identified and assigned to types, every i_n must have its highest r with an i_t.

After the reciprocal pairs are identified, all that is necessary is to assign the i_n's to the types which contain the i_t's with which they have their highest r's. Initially, when only reciprocal pairs have been assigned to types, the first i_n will necessarily be assigned to a type on the basis of a highest r with a variable which is a member of a reciprocal pair. Thereafter, an i_n is assigned to a type either (a) because it has its highest r with a member of a reciprocal pair or (b) because it has its highest r with a non-reciprocal variable which has already joined a type. The method continues in this manner until all variables are assigned to types.

An Illustration of the Method of Analysis

The Analysis

A matrix of intercorrelations from an investigation by Stephenson (14, p. 169) was chosen for the purpose of the present study. It was selected for two reasons: (a) Stephenson has for many years been interested in the study of types, and it seemed appropriate that the present method be illustrated in terms of some of his data; (b) the matrix is one in which elementary linkage and factor analyses will give similar results because the rotated factors are approximately co-linear with the centroids of the two clusters of variables; the analysis will illustrate that the two methods can under certain conditions give very similar results. The illustration will also show that elementary linkage analysis is many times more rapid than Q-technique of factor analysis.

The particular matrix chosen for the illustration is one in which Stephenson reported the intercorrelations between 15 selected persons as determined from their responses to 121 Jungian statements on introversion and extraversion. The intercorrelations are reproduced in Table 1.

The steps in linkage analysis as they apply to Table 1 are outlined as follows:

1. Underline the highest entry in each column of the matrix.
2. Select the highest entry in the matrix. It has a value of

TABLE 1*

Correlations between the Responses of 15 Persons to 121 Introversion-Extroversion Items

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A		387	459	577	376	352	264	503	088	096	085	030	−005	023	030
B	387		603	493	242	431	307	392	−017	174	114	−006	049	082	159
C	459	*603*		*615*	463	508	399	522	−006	105	095	−043	080	094	099
D	*577*	493	*615*		398	*572*	*441*	*607*	087	083	138	062	100	172	063
E	376	242	463	398		557	320	500	137	190	148	107	077	140	172
F	352	431	508	572	*557*		324	467	095	089	128	013	044	015	101
G	264	307	399	441	320	324		370	−250	090	−040	−015	007	108	084
H	503	392	522	607	500	467	370		073	055	003	060	103	081	105
I	088	−017	−006	087	137	095	−250	073		357	365	251	208	197	324
J	096	174	105	083	190	089	090	055	357		392	*409*	*483*	*392*	*580*
K	085	114	095	138	148	128	−040	003	*365*	392		380	390	275	396
L	030	−006	−043	062	107	013	−015	060	251	409	380		353	333	400
M	−005	049	080	100	077	044	007	103	208	483	390	353		375	428
N	023	082	094	172	140	015	108	081	197	392	275	333	375		389
O	030	159	099	063	172	101	084	105	324	*580*	*396*	400	428	389	

* The correlation coeficients are from Stephenson (14, p. 169); the decimal points are omitted.

.615 and mediates between persons C and D. Enter these codes for these persons as shown in Type I of Figure 1. They constitute the first two persons of the first type.

3. Select all of those persons who are most like the members (C and D) of the first type. This is done by reading across the rows C and D of the matrix and selecting the underlined entries in these rows; persons of the columns thus selected are most like C and D. This step yields persons A, F, G, H, and B. They are called first cousins. Record them as shown in Type I of Figure 1.
4. Select all those persons who are most like first cousins (A, F, G, H, and B). This is done in a manner similar to that described in paragraph 3 above. For each row A, F, G, H, and B, the column which is associated with it by an underlined entry is identified. The persons of these columns are here called second cousins. Only one second cousin (E) occurred in Type I, as shown in Figure 1.
5. In an analogous fashion search for third and higher order cousins. No third order cousins occurred in Type I, and consequently no higher order cousins could occur in this type as pertains to the present matrix.
6. Excluding all the persons already classified, repeat steps two to five inclusively until all persons are classified.

Type I　　　　　Type II

FIG. 1. The types

⇄ Means a reciprocal pair of variables

→ Means that the variable at the tail of the arrow is highest with the one at the head, but the one at the head is not highest with the one at the tail.

Typological Results

The results of the analysis of Table 1 are shown in Types I and II of Figure 1. In comparing these results with those by Stephenson, we find that Stephenson obtained the same two types, containing identically the same persons as reported here. The present analysis required less than five minutes while a factor analysis of the matrix as applied by Stephenson would require several hours.

A Simply Structured Factor Analytic Solution without Rotation

Even though elementary linkage analysis has the advantage of requiring much less time, Q-technique of factor analysis results in more information; it gives factor loadings in addition to determining both the types and typal memberships. But, elementary linkage analysis can be extended to give typal relevancies which are analogous to factor loadings in ways already explained. A feature of the method is that the typal relevancies are sometimes identical with simply structured, rotated factor-analytic solutions, without the laborious necessity of rotating the typal relevancies. All that is necessary in case of data like that of the present illustration is to treat the variables of the two types in two separate matrices and to compute the first factor loadings for the variables of each submatrix. The resultant loadings are here called typal relevancies. In some cases, one or two additional steps are required in order to obtain a simply structured solution. They will be outlined later, after the results from the present analysis have been studied.

TABLE 2

Factor Loadings and Corresponding Typal Relevancies

	Factor Loadings* Centroid		Rotated		Typal Relevancies		Difference	
	I	II	I′	II′	I″**	II″**	I″–I′	II″ – II′
A	.46	.38	.60	.00	.64		.04	
B	.48	.33	.58	.06	.64		.06	
C	.57	.46	.73	.02	.77		.04	
D	.63	.46	.78	.06	.80		.02	
E	.56	.25	.59	.17	.63		.04	
F	.52	.41	.66	.02	.70		.04	
G	.37	.36	.52	−.04	.53		.01	
H	.54	.43	.70	.02	.73		.03	
I	.30	−.36	.00	.47		.48		.01
J	.49	−.44	.08	.65		.74		.09
K	.42	−.40	.07	.58		.60		.02
L	.32	−.44	−.04	.55		.59		.04
M	.37	−.44	.00	.58		.63		.05
N	.36	.33	.06	.49		.55		.06
O	.47	.46	.06	.66		.72		.06
		Mean of absolute differences = .04						

* The factor loadings are from Stephenson (14, p. 169).

** Portions of these columns are vacant because linkage analysis does not normally compute typal relevancies for variables with respect to a cluster to which they are not assigned; it can, however, be extended to do so.

The typal relevancies from the elementary linkage analysis are shown in Table 2 along with the corresponding loadings from the Q-technique factor analysis. The first two columns of the table report the centroid factor loading for the unrotated factors, the next two columns the simply structured loadings obtained by Stephenson through a rotation; the fifth and sixth columns report the typal relevancies obtained directly from the linkage analysis without rotation, and the final column reports differences between the typal relevancies and the corresponding rotated loadings, with the loadings subtracted from the relevancies. It will be noticed that the greatest difference is only .09, and that the relevancies are universally higher than the corresponding rotated loadings; the mean of the absolute differences is .04.

Interpretation of Results

The present solution by the linkage method was computed with pencil and paper in only 65 minutes after the matrix of correlations was available; the solution did not require either the computation of residuals or the rotation of types.

The slight differences between the rotated and linkage solutions is attributable to one, two, or all of three conditions. In the linkage solution, the exact location of each typal centroid is determined exclusively and objectively by the centroid of the variables found to have membership in the type. However, in the rotated solution, the location of the corresponding factor is determined on the basis of subjective opinion. One would therefore expect a lack of perfect agreement in the corresponding indices.

Another consideration which may explain some of the difference between the two solutions pertains to the purpose of rotation to simple structure. The primary purpose is not to send factorial vectors through clusters; it is, instead, to send them through that space which places the greatest number of variables in the hyperplanes and consequently results in the greatest possible number of zero factor loadings, requiring, of course, that each variable have at least one zero loading. This purpose sometimes sends factorial vectors (as in the present case) through clusters, and to the extent that these vectors are co-linear with the centroids of the typal clusters, then typal relevancies and rotated factor loadings are to that extent similar.

A third reason for the difference between the linkage and factor-analytic solutions can be suggested. The rotation of the present factor analysis was restricted to an orthogonal solution, but the centroids of the types were not required to be orthogonal.

An Extension of the Method

Linkage analysis can be revised in such a manner that the centroids of the types will be required to be orthogonal. In this revised version, the first type and the first typal relevancies are computed in the fashion already outlined. However, subsequent types and typal relevancies are computed in terms of residual indices of association between variables rather than in terms of the original index of association. After the first type is isolated, the relevancies of all variables (not just those of the first type) are computed with respect to this type.[2] The variance accounted for in terms of this type is subtracted from the original matrix (in exactly the same manner as it is in factor analysis). The entire procedure is then repeated on the residuals until all variables are classified into types and all typal relevancies are computed.

If the above extension of the method is to be used, that particular type which accounts for the most variance should always be the only one for which typal relevancies are computed. This principle should be applied not only to the original matrix of associations between variables; it should also be applied to every residual matrix. One way in which to apply the principle is as follows: for every matrix (original and residual) isolate all types. Every type yields a sub-matrix of correlations. Estimate the variance accounted for by every sub-matrix. There are three ways to do this. One is to take the sum of the squares of the coefficients of each sub-matrix as a measure of the variance accounted for; another is to take merely the sum of the coefficients without squaring them; and a third is merely to count the number of variables represented in each sub-matrix. The first of these techniques is the most dependable, and the third is usually the least dependable. The sub-matrix which yields the largest index in terms of any one of the techniques is estimated to account for the most variance.

Everyone of the three techniques just outlined depends on the assumption that the amounts of variance accounted for by the types in terms of their sub-matrices is perfectly correlated with the amounts they would account for in the larger matrix of which they are the parts.

After the type which is estimated to account for the greatest amount of variance in the original matrix has been determined, the relevancies of all variables on its typal centroid are computed. Other types indicated by this matrix are thereafter ignored. Using the relevancies of all the variables on the first typal centroid, the first residual matrix is computed (just as

[2] This can be done using an approximate method developed by Dwyer (6). Application of the Dwyer method to elementary linkage analysis is unusually simple, because the submatrices are so chosen that only one type is used to account for the group variance within each submatrix.

in factor analysis) and the procedure is repeated on the first residual matrix. The analysis continues until residuals are estimated to be due to chance alone (just as in factor analysis).

Critique of the Method

Every statistical method designed to reduce a set of variables to more fundamental constructs (whether they be types, factors, or whatnots) invites the question as to the stability of the constructs. In the final analysis this crucial question usually has to be put to empirical tests, for the answer generally depends in part on how the variables are in fact interrelated by nature. Are they held together by types or factors, for example? An index of association between variables is but one manifestation of their fundamental interrelationships. The index does not reveal conclusively which of two or more constructs had best be assumed in order to comprehend the manifest interrelationships. One test of the relative value of two constructs is their stability.

A more fundamental test of constructs is their value for prediction. The truth of this statement is realized by recalling that if two tests are of equal validity but unequal reliability (stability) the less reliable one offers greater hope for increased validity; this is especially true if the more reliable test is near its saturation point with respect to stability, and the less reliable one is considerably below a corresponding degree of stability.

Previous typological analyses by the author (11) have produced findings related to the reliability and validity of types. These studies applied a hierarchical system of classification whereby individuals were classified into "species," then "species" into "genera," "genera" into "families," etc., until every subject of a sample was classified into but one of two top classes. Since the lower levels of classes represented many types, each with few subjects but relatively many characteristics in common, they were called molar types; the higher classes, which represented few types each with many people but only few characteristics in common, were called global types. In the studies thus far completed the molar types have given more valid predictions than the global types.

Though the above studies did not investigate the reliabilities of molar versus global types, one would suspect that the latter types are probably (though not necessarily) the more reliable ones, even though they were found to be the less valid ones. Anyway, the above studies indicate a need for a method whereby molar types can be isolated in a large matrix; this would put more subjects in the molar types and thus probably give them both increased reliability and validity. Linkage analysis was developed to fulfill this need. On the basis of the results from the above mentioned studies, it is expected that elementary linkage analysis, when applied to matrices of less than 100 subjects, will yield types with relatively low reliabilities but with promising validities, and when the method is applied to larger matrices it will yield both increased reliabilities and validities.

Despite the potentiality of elementary linkage analysis as just indicated, a severe criticism of the method would seem to be, at least at first glance, the fact that the number of types depends exclusively on the number of reciprocal pairs. A pair of variables i and j are reciprocal, as stated earlier, if both (a) i has its highest index of association with j, and (b) j, on the other hand, has its highest index with i. It would seem that the number of reciprocal pairs might change from sample to sample even when they are for the same universe.

Even though the method thus far has been described as if the number of types depends exclusively on the number of reciprocal pairs of variables, it can be revised in two different ways so that the number of types will depend on a broader base. The first way is rapid but subjective. The second is objective but more tedious.

Imagine an analysis which produced a type such as the one reported in Figure 2. The figure contains but one reciprocal pair, but the two almost separate clusters argue for two types; they could be divided on the basis of subjective opinion of the objective evidence. Variable 7 could go either with variable 8 or 2, depending on whether r_{78} or r_{72} were larger, or it could be omitted in computing the two typal centroids and a portion of its variance assigned to each type in accordance with its relevancies, thus recognizing a mixed type.

$$\begin{tikzcd}
& 3 \arrow[d] & 6 \arrow[d] & & 10 \arrow[d] & 12 \arrow[d] & \\
4 \arrow[r] & 1 \arrow[r, shift left] & 2 \arrow[l, shift left] & 7 \arrow[l] & 8 \arrow[l] & 9 \arrow[l] & 13 \arrow[l] \\
& 5 \arrow[u] & 8 \arrow[u] & & 11 \arrow[u] & 14 \arrow[u] &
\end{tikzcd}$$

FIG. 2. A hypothetical type or types

Versions of Linkage Analysis

Several versions of linkage analysis have been developed by the author in order to meet special needs. Three of these will now be previewed because of their relationships to the version developed herein; they will be neither illustrated nor described in detail because of their complexities. They will, instead, be reserved for subsequent papers. In order to distinguish the version already outlined from the later ones, it is called elementary linkage analysis (because it is the simplest of all the versions).

Centroid-Linkage Analysis

Centroid-linkage analysis is an objective method for solving such problems as that illustrated in Figure 2, where an investigator wonders whether more than one type is involved. Centroid-linkage analysis has all of the main characteristics of elementary linkage analysis and holds the same general relationships to factor analysis as does elementary linkage analysis. Instead of combining variables into clusters on the basis of links between pairs, it clusters variables about typal centroids on the basis of links with them. The centroids are built up gradually in the course of the analysis. Every pair of variables i and j defines a trial centroid. The typal relevancies of i and j are computed on the trial centroid C_{ij} for all i's and j's ($i \neq j$). The highest relevancy is selected. Suppose it is for i_1 on $C_{i_1 j_2}$, then it is equalled by j_2 on $C_{i_1 j_2}$. Both i_1 and j_2 are then classified with $C_{i_1 j_2}$. In fact, $C_{i_1 j_2}$ replaces them in the matrix and the analysis proceeds in a similar fashion as just described. In computing the relevancy for the $C_{i_1 j_2}$ column, the relevancy of $C_{i_1 j_2}$ on $C_{i_1 j_2 i_3}$ is the same as that of i_3 on $C_{i_1 j_2}$, viz., the factor loading of variable i_3 on the centroid of variable $i_1 j_2 i_3$. Using analogous methods throughout, the analysis proceeds until all variables are classified with centroids of types. Any inconsistencies of classification are corrected in the same manner as in agreement analysis (10).

Hierarchical Linkage Analysis

When every variable has been classified into a type, the resultant types can be called species. However, the analysis does not need to discontinue here. The same procedure can continue to classify the species into genera, and then the genera into families, etc., until every variable is classified into one of two major types. The resultant types at each level of classification can be studied in relation to indices of both reliability and validity to determine which levels are most promising for specified purposes.

Differential Linkage Analysis

Either centroid or elementary linkage analysis can be revised to select patterns of responses to differentiate between two or more categories of people. Assume for illustrative purposes that we have a sample of N normals numbered 1, 2, 3, $\cdots N$ and a sample of P patients lettered $a, b, c \cdots P$. The subjects of these two categories have responded to a common test of r items. The problem is to select sets of items which will best differentiate all types of patients from all types of normals. The approach assumes both (a) that items have predictive values when treated in combinations which they do not reveal when analyzed separately in relation to a criterion, as emphasized in the Meehl paradox (8), and (b) that items have different configural values for various types of subjects. As a result of the latter assumption, patterns of responses on several different sets of items would be helpful in differentiating all normals from all patients. One type of patient would be best differentiated from normals in terms of one set of items, but a quite different set of items would be required to select the members of another type of patient, etc. Analogously, patterns of responses on different sets of items would be required to differentiate different types of normals from patients.

Differential linkage analysis, as now to be outlined, can be used in selecting sets of items to differentiate types of persons in one category from those in another category, such as the normals and patients just mentioned. Using the responses of the subjects (both patients and normals) to the r test items, an index of association (such as a coefficient of correlation) is computed between every pair of subjects. These indices of association are assembled into a matrix, where either the normals are listed first followed by the patients, or vice versa. Then for every normal, both the normal and the patient most like him are underlined in the columns of the matrix, and likewise for patients. Two complete linkage analyses are then performed, one for normals and one for patients. The first pertains only to that portion of the matrix which gives indices between normals, and the second only to that portion which gives indices between patients, except that whenever a normal is classified in the first analysis, the patient most like him is classified with the normal, and in the second analysis whenever a patient is classified, the normal most like him is classified with the patient. A type of normal subjects (with associated patients) which would result might be such as shown in Figure 3. In addition to other normal types with associated patients, patient types with associated normals would result from such an analysis.

Each type would be analyzed to produce a set of items with a scoring key. Those which derived from normal types could be called normal sets and keys, and those which derived from patients could be called patient sets and keys. The two sets for every type can be combined into a single category of items with a scoring key designed to differentiate some normals and patients. The method of developing a scoring key will be described with respect to the normal type of Figure 3. Find the

$$\begin{tikzcd}
& (h)3 \arrow[d] & 5(e) \arrow[d] & \\
(a)12 \arrow[r] & (b)4 \arrow[r, shift left] & 9(d) \arrow[l, shift left] & 7(g) \arrow[l] \\
& (m)8 \arrow[u] & 15(c) \arrow[u] &
\end{tikzcd}$$

Numbers represent normals.

Letters represent patients; the patients are most like the normals with whom they are associated; they were selected as the most similar patients from a large sample of patients.

FIG. 3. A typological classification of normals with associated patients from hypothetical data

items on which all of the normal subjects of the type agreed in their answers, or at least reached a specified limit of agreement. Do likewise for the associated patients. Find which items are common to the two selections. Of the common items find which ones have the same predominant answer (meeting the criterion of agreement in selecting the items) for both normals and patients. Eliminate these items from the set which originated with the normals and from the set which originated with associated patients. The items which remain in the two sets constitute the desired category of items. In scoring them, the ones which derived from the normals are given a plus one when the typical, normal type of answer is given. Those items which derived from the associated patients are given a minus one when the patient type of answer is given. All other answers are given a score of zero. In a like manner an analogous category of items and scoring key would be prepared for both (a) every other normal type with associated patients, and (b) every patient type with associated normals. Criterion scores for admitting subjects to types would be computed and the keys would be tried out on cross-validational subjects.

Summary

This paper develops and illustrates a rapid and objective method for clustering variables into types; it should prove especially helpful in clinically oriented studies. An elaboration of the method results in obtaining typal relevancies for the members of every type. The centroids of the types may be required to be orthogonal with respect to one another or they may be allowed to be either oblique or orthogonal as determined by the data.

The method is analogous to factor analysis except for the fact that it results in a typal structure instead of attempting to achieve a simple structure. A type is defined as a category of persons (or other variables) of such a nature that everyone in the category is more like someone else in the category than he is like anyone not in the category. This is the definition used in the simplest version of the method, called elementary linkage analysis.

A more sophisticated version, called centroid-linkage analy-

sis, defines a type as a category of persons (or other variables) of such a nature that everyone in the category is more like the centroid of the type than he is like the centroid of any other type. Both of these two versions provide that a response to a test item can be found to have different validities and reliabilities for different types of subjects. In this sense linkage analysis is more consistent with clinical theory than is factor analysis. A typal structure is the isolated types as defined above together with the relevancies of the members of every type to the centroids of their types.

In the case illustrating the method, the types were isolated from a fifteen variable matrix in only five minutes with but pencil and paper. Only one additional hour was required to compute the relevancies of the variables on the typal centroids of the two types into which they were classified.

An elaboration of the methods of this paper yields hierarchical linkage analysis. The latter method classifies variables first into species, then the species into genera, the genera into families, etc., until every variable has been classified into one of but two major categories at the top level of classification. A different kind of elaboration of the method results in differential linkage analysis; it is appropriate to the selection of sets of items designed to differentiate between the types of two or more major classes of subjects (such as normals and mental patients) in terms of their responses to selected sets of items.

REFERENCES

1. Adcock, C. J. "A Note on Cluster-Directed Analysis." *Psychometrika*, XVII (1952), 249–253.
2. Cattell, R. B. "A Note on Correlation Clusters and Cluster Search Methods." *Psychometrika*, IX (1944), 169–184.
3. Cattell, R. B. "r_p and Other Coefficients of Pattern Similarity." *Psychometrika*, XIV (1949), 279–298.
4. Cattell, R. B. "The Three Basic Factor-Analytic Research Designs—Their Interrelations and Derivations." *Psychological Bulletin*, XLIX (1952), 499–520.
5. Cronbach, L. J. and Gleser, Goldine C. "Assessing Similarities between Profiles." *Psychological Bulletin*, L (1953), 456–473.
6. Dwyer, P. S. "The Determination of the Factor Loadings of a Given Test from the Known Factor Loadings of Other Tests." *Psychometrika*, II (1937), 173–178.

7. Loevinger, Jane, Gleser, Goldine C., and DuBois, P. H. "Maximizing the Discriminating Power of a Multiple-Score Test." *Psychometrika*, XVIII (1953), 309–317.
8. Meehl, P. E. "Configural Scoring." *Journal of Consulting Psychology*, XIV (1950), 165–171.
9. McQuitty, L. L. "A Method of Pattern Analysis for Isolating Typological and Dimensional Constructs." San Antonio, Texas: Headquarters Air Force Personnel and Training Research Center, *Research Bulletin* TN-55-62, (1955).
10. McQuitty, L. L. "Agreement Analysis: Classifying Persons by Predominant Patterns of Responses." *British Journal of Statistical Psychology*, IX (1956) 5–16.
11. McQuitty, L. L. "Isolating Predictor Patterns Associated with Major Criterion Patterns." EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, XVII (1957), 3–42.
12. Seeman, J. and Raskin, N. J. "Research Perspective in Client-Centered Therapy." In O. H. Mowrer, *Psychotherapy Theory and Research.* New York: Ronald Press, 1953, pp. 205–234.
13. Stephenson, W. "Some Observations on Q Technique." *Psychological Bulletin* XLIX (1952), 483–498.
14. Stephenson, W. *The Study of Behavior.* Chicago: The University of Chicago Press, 1953.
15. Thorndike, R. L. "Who Belongs in the Family." *Psychometrika*, XVIII (1953), 267–276.
16. Thurstone, L. L. *Multiple-Factor Analysis.* Chicago: The University of Chicago Press, 1947.
17. Tiedeman, D. V. "On the Study of Types." In S. B. Sells, *Symposium on Pattern Analysis.* Randolph Field, Texas: Air University, USAF School of Aviation Medicine, Pp. 1–14.
18. Wrigley, C., Marsh, J. E., and Twery, R. "A Factor Analysis of the Air Force Factor Reference Battery." San Antonio, Texas: Headquarters Air Force Personnel and Training Research Center, *Research Bulletin*, AFPTCR-56-137 (1956).
19. Zubin, J. "The Determination of Response Patterns in Personality Adjustment Inventories." *Journal of Educational Psychology*, XXVIII (1937), 401–413.