

From

Theory Building and
Data Analysis in the
Social Sciences, 1984, by

Asher, Herbert
Weisberg, Herbert
Kessel, John
Shively, Phil

HERBERT F. WEISBERG

Scaling Objectives and Procedures

A wide variety of scaling techniques are available for the analysis of the structure of a set of variables. These techniques generally analyze the similarities among a set of variables from a geometric perspective. Scaling techniques introduced in this chapter include Guttman scaling, unfolding analysis, cluster analysis, factor analysis, component analysis, and multidimensional scaling.

We shall use the term "dimensional analysis" to encompass a wide variety of data reduction techniques, the best known of which are Guttman scaling, cluster analysis, factor analysis, and multidimensional scaling. These techniques focus on the similarities (or the differences) among a set of variables, rather than tracing through causal processes as is the concern of the techniques described in the previous section of this book. Most statistical analysis deals with estimating the effects of independent variables on dependent variables, but that distinction between independent and dependent variables is not important in the dimensional realm. Instead, the scaling question is how similar (or different) is a set of variables. Scaling techniques analyze the structure of a set of variables when causation is not being considered,¹ and they do so on the basis of a geometric representation of the data.

The causal problem is so much more familiar than the scaling problem that it is worth pausing to give some examples of situations leading to dimensional analysis. Two basic situations can be distinguished. First, the researcher sometimes simply wishes to describe the dimensionality of a set of variables—a "dimensioning" concern. For example, the legislative researcher might want to determine whether a single liberal-conservative dimension underlies all roll call votes in a particular Congress, whether there are separate domestic and international dimensions, or whether there are actually several separate dimensions within the domestic and/or foreign realms. Similarly, the student of comparative politics might want to discover the dimensions underlying a party system in some nation.

Alternatively, the researcher sometimes wishes to derive unidimensional indices on which individuals can be scored—a use of scaling out of a concern

¹ Procedures developed by Jöreskog which combine causal and dimensional concerns will be described briefly at the end of this chapter.

with "positioning" individuals on scales with dimensional integrity.² For example, a survey researcher might want to combine several questions designed to measure political cynicism into a single scale so as to study the effect of that attitude on political participation, but a necessary first step would be to check whether all of the questions asked tap the same dimension of cynicism. Similarly, a researcher interested in the effects of social-economic status on political behavior might first want to see whether different indicators of the person's social-economic status are interchangeable or whether there are important differences among them. In these examples, the dimensionality is examined as a prelude to positioning the individuals studied on an underlying dimension, rather than out of a pure interest in the dimensionality *per se*.

These examples illustrate the variety of interests which motivate scaling, from data reduction and object classification to measurement improvement and assessing the commonality among a large number of measures. Dimensional analysis is often performed in an exploratory mode, seeing which dimensions are required to account statistically for a body of data. As such, it can be atheoretical, with the danger of accepting any result as meaningful even if it has no substantive significance. A case of this type of misuse of scaling would be collecting a few hundred indicators of the properties of different nations, seeking the dimensions underlying those indicators, and then giving interpretations of the dimensions that are obtained without external validation of the interpretations. Dimensional analysis is better when it is performed in a confirmatory mode, testing whether the data at hand conform to prior structural hypotheses. Unfortunately, the number of dimensional hypotheses we have at present is very limited, so exploratory analysis is more prevalent than confirmatory analysis.

Scaling makes direct use of a geometric metaphor, such as the concern in the above paragraphs as to whether a set of data can be explained in terms of a single dimension or require multiple explanatory dimensions. The meaning of this geometric metaphor is explored directly by Weisberg in the next chapter. In that chapter, Weisberg also presents some alternative interpretations of dimensionality and of unidimensionality that are inherent in different scaling techniques. Also, he provides further speculation on the philosophical underpinnings and implications of dimensional analysis.

Before proceeding to that chapter, it is useful to be familiar with some of the more basic scaling procedures, and these will be described in the remainder of this chapter. A wide variety of scaling techniques will be introduced in this

²The distinction between "dimensioning" and "positioning" was suggested by Donald Stokes (personal communication).

chapter, so the reader can have some familiarity with the concepts employed in the next two chapters and compare the techniques described in those chapters with other common dimensional procedures. Some of the purposes of scaling will also be explained in this chapter, both directly and by means of example, so the reader can better understand why these techniques are useful and what types of questions they are designed to answer.

Guttman Scaling: Unidimensionality as Cumulation

One of the most commonly used scaling techniques is Guttman scaling. To show the reasons that it is useful, we shall first describe a simple "non-dimensional" procedure for data reduction.

Index Construction

The simplest way of combining several indicators into a single measure is to score each one and then add up the scores. By way of example, say that we are studying political participation, by which we mean not only voting but talking about politics, taking part in political campaigns, and joining political groups.³ No single indicator would suffice here, since political participation involves a large number of acts, those just listed and many more. But if we want to study political participation as a whole rather than its separate facets, we must somehow combine those separate indicators. The "index construction" solution would be simply to count up how many activities a particular person performs. If only the four participatory acts listed above were included, then a total score of 4 would mean that the person participated in all four acts, 3 would indicate that the person performed three of those activities, and so on down to zero for the person who participated in none. Once respondents are thus scored on this index, the index could be correlated with other variables, so that one might study the relative effects of political cynicism and political interest on political participation.

A potential problem with this procedure is that some of the items in the index might measure different things from the other items. Perhaps one of the activities is really very different from the others and does not belong in the same index. Or perhaps political participation actually consists of distinctly different types of activities, with different types of people participating in the active "gladiatorial" activities (taking part in campaigns and group membership) than

³The indicators employed here are those used by Matthews and Prothro (1966) in their study of political participation in the South.

in the more passive "spectator" activities (voting and political discussion).⁴ If two of the indicators tapped one component and the other two tapped a different component, adding the four together in a single index would be like adding apples and oranges. Minimally, the index construction process has to be modified to check whether all the individual indicators indeed measure the same thing.

Beyond this, the index construction process equates the importance of the different constituent items, treating them as if they were of equal difficulty. People would receive a score of 1 on the above political participation index if they participated in only one of the four activities which have been listed, regardless of which activity that is—as if just taking part in a campaign is the same level of political participation as merely talking to somebody about that campaign. The simple index construction process is inadequate because it does not check whether the items are of unequal difficulty. But items can be of unequal difficulty while still measuring the same dimension, as in this example. This is the precise logic which leads to Guttman scaling.

Cumulative Scaling

Guttman scaling tests for a difficulty dimension underlying a set of data. If there is a difficulty dimension, then only certain data patterns should occur—ones which are "cumulative," such as those in the top five rows of Table 1. Only five patterns of political participation fit this scale: some people not participating at all, some performing only the easiest activity, some only the two easiest, some all but the hardest, and some all four. But cumulation implies that no person should "pass" a harder item while "failing" an easier item, so performing a "harder" activity (like taking part in a campaign) without performing the "easier" activities (like voting) constitutes "error" in the scale.

If all the items fit a cumulative pattern, we would have assurance that they constitute a single dimension. Actually, a perfect scale would be unlikely, as there is always some error (whether random error, measurement error, or violation of the unidimensional model). For example, a person whose only political activity is participating in a campaign (as in the last row of Table 1) would contribute error to the scale. A low amount of error would not invalidate a scale, but if the amount of error became high we would have to conclude that the items were not all unidimensional. If most of the error were found to be due to a single item, that item could be deleted from the scale. If, instead, the errors result because the different items measure, for example, two different dimensions,

⁴The distinction between "gladitorial" and "spectator" activities is due to Milbrath (1965).

TABLE 1
Hypothetical Guttman Scale of Political Participation

Person	Activity				Score	Errors
	Voting	Political Discussion	Group Membership	Campaigning		
1	Yes	Yes	Yes	Yes	4	0
2	Yes	Yes	Yes	No	3	0
3	Yes	Yes	No	No	2	0
4	Yes	No	No	No	1	0
5	No	No	No	No	0	0
6	No	No	No	Yes	0	1

Number of items = 4

Amount of missing data = 0

Number of valid responses = (number of items) (number of people) - (missing data)
= 4 × 6 - 0 = 24

Total number of errors = 1

Guttman's coefficient of reproducibility = 1 - (number of errors)/(valid data)

= 1 - 1/24 = .96

then separate scales can be prepared and people scored separately on those two scales. In any case, it is important to have some statistic summarizing the quality of a scale.

Louis Guttman developed a "coefficient of reproducibility," which measures the proportion of responses which fit the scale, as illustrated in the calculations below Table 1. He recommended accepting as unidimensional only those scales with reproducibilities of at least .90 under his method of counting scale error. However, his measure has an upward bias; its minimum value is not .00 but depends on the proportion of people passing each item, so that in some cases the minimum reproducibility is actually quite high (.60 or .70 or even greater). To handle this problem, a "coefficient of scalability" has been developed. It normalizes the reproducibility coefficient by its minimum, given the item marginals. A scalability of at least .60 is usually required for a good scale. Actually, many other measures of scale quality have been proposed, and there is no single agreed-upon measure. But at least there is agreement on the need to measure fit to the cumulative model. Computer programs are available to assess the degree to which a set of variables fit a Guttman scale, in particular the GUTTMAN SCALE procedure in SPSS.

There are not good direct procedures for separating scales in a set of multidimensional data. A correlational approach is often employed for this purpose. The key to this approach is that if two items are perfectly cumulative,

their cross-tabulation will contain a cell without any cases—since no one will pass the harder item while failing the easier one. Table 2 labels the cells in the fourfold table obtained by crosstabulating two dichotomous variables. If there is a perfect cumulative Guttman scale relationship between the two items, the c cell should equal zero. In terms of the measures of association described in earlier chapters, Yule's Q and some other "one-way measures of association" have the value of 1.0 in this case, which can be interpreted as perfect association. As a result, one way to locate Guttman scales among a large set of variables is to look for subsets of variables which all have very high Q values with one another. Automated procedures for such a search will be described later in this chapter.

To summarize, Guttman scaling is a technique for checking whether dichotomous data are unidimensional. It can be extended to multicategory data, mainly by first dichotomizing the variables. If the variables do not form an acceptable scale, an item with excessive error can be deleted. Or we can check whether the variables contain more than one subset of scalable variables, such as two separate dimensions. Guttman scaling employs a single theory of unidimensionality—the form of cumulation shown at the top of Table 1—but there are other forms of unidimensionality even for dichotomous data. The next chapter discusses some of those alternative models of unidimensionality, especially the proximity scale.

Unfolding Analysis: Unidimensional Preference Orders

The second procedure to be described in this chapter can be dealt with more quickly. It is Coombs's unfolding analysis, which is applicable to preference order data. If we have people order a set of alternatives from their first choice to

TABLE 2
Cross-Tabulation of Dichotomous Variables*

Easier Item	Harder Item		Total
	Yes	No	
Yes	a	b	$a + b$
No	c	d	$c + d$
Total	$a + c$	$b + d$	1.00

*Since the row item is the easier item, more people pass it than the column item. Thus, $a + b > a + c$, or $b > c$. If the items are perfectly cumulative, no one passes the harder item while failing the easier item, so $c = 0$. Yule's $Q (= \gamma) = (ad - bc)/(ad + bc)$, which, if $c = 0$, reduces to $(ad - 0)/(ad + 0) = ad/ad = 1.0$. Thus $Q = 1.0$ for cumulative variables.

their last, then the individual preference orders can be studied to obtain information about the underlying preference space.

Part of the idea of unfolding can be understood from the case of three alternatives. Say that there are three parties in a country and that public attitudes toward them are unidimensional. People at the left end of the polity would prefer the left party L the most, the center party C second, and the right party R least. People at the right end of the polity would like the right party R the most, center party C second, and the left party L least. Some people in the middle might be closest to C, then to L, then to R. Other people in the middle might be closest to C but then next closest to R and then L. This leads to four preference orderings occurring: LCR, CLR, CRL, and RCL (See Figure 1).

But, if the data are indeed unidimensional, no one would be closest to the left party L, second closest to the right party R, and furthest from the middle party C. And no one would be closest to the right party R, second closest to the left party L, and furthest from the middle party C. That is, two preference orders should not occur: the two with the middle party in third place, LRC and RLC. Thus, to scale a set of data, we would check how many parties are the last place choices of people. If everyone likes either L or R least, the data are unidimensional. (Or if they all like L or C least, then the data are unidimensional but the party ordering should be LRC instead of LCR; if they all like C or R least, then the data are unidimensional with the party ordering CLR.) But if L, C, and R all get substantial numbers of last place votes, then the data are not unidimensional.

This technique interprets preference orders as individuals picking up the underlying dimension at their most preferred ("ideal") point and then folding the dimension at that point to obtain their ordering of points (they prefer the alternatives nearer them regardless of direction—hence the folding idea). Since the analysis procedure seeks an underlying dimension which is consistent with the observed preference orders (the folded scales), it is termed "unfolding."

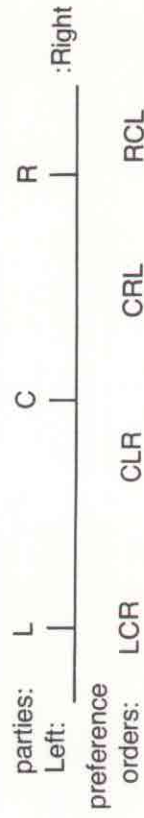


FIGURE 1

Preference Orders of a Party Dimension

For example, we could check whether Americans view partisanship as unidimensional by asking them to rate Democrats, Independents, and Republicans. If these groups were viewed as unidimensional with Independents in the middle, no partisan should like the other party more than Independents. In fact, most partisans like the opposite party more than Independents (see Table 3), so the preferences are not consistent with a Democrat-Independent-Republican ordering. They are actually more consistent with the ordering Republican-Democrat-Independent, but it is more appropriate to say that they are not unidimensional.

Unfolding analysis extends to more than three alternatives, though it rapidly becomes more complicated. With four alternatives, the occurrence of certain preference orders rather than others begins to give some information about the spacing of alternatives along a single dimension. We might find, for example, that the two left alternatives are closer to each other than are the two alternatives on the right. This is termed "ordered metric" information—information about the order of interpoint distances. Technically, this is a level of measurement between ordinal and interval, as it yields more than an ordering of the points but less than their precise locations on a dimension. The more alternatives being scaled, the more ordered metric information becomes available. With enough variables, the ordered metric information places severe limits on the possible

locations of alternatives in the space, so that nonmetric unfolding of preference order data virtually pinpoints the variables in the space.

Unfolding analysis also extends to more than one dimension, but it is too complicated to calculate by hand. Fortunately, the nonmetric multidimensional scaling procedure mentioned later in this chapter and described in the last chapter can be used to unfold preference orders in one or more dimensions. But it basically just employs the principles mentioned already. If alternatives have fixed points in a space of limited dimensionality, then certain preference orders should not occur. As a result, analyzing the pattern of preference orders can lead to information about the spatial structure. Moreover, the exact pattern of preference orders which occurs can place severe limits on the possible locations of alternatives in the space, so that nonmetric analysis of ordinal preference orders can yield spatial solutions nearly as precise as interval analysis provides.

Cluster Analysis: Separating Sets of Variables

Another important family of analysis techniques is designed to identify "clusters" of variables. A cluster of variables is a set of variables that are more similar to each other than to other variables. Sometimes a large set of variables can be decomposed into 2, 3, 4, 5, or 6 clusters, and it can be useful substantially to be able to talk about those clusters rather than the larger set of variables. A difficulty with cluster analysis is that there is no single well-defined procedure for clustering. Instead, the vague definition given here for clusters is compatible with a very large number of clustering approaches. As a result, there are literally hundreds of clustering routines that have been proposed over the years. Each has some advantages over the others, or at least its originator felt that it did. We cannot explain all clustering approaches in this short section, but the ideas behind cluster analysis can be introduced.

Simple Clustering

One clustering approach is to build up a cluster, starting with the most similar pair of variables—that with the highest correlation. Then the variable with the highest correlations with those first two variables is added to the cluster, followed by the variable with the highest correlation with those first three variables, and so on until none of the variables remaining to be clustered has high correlations with the variables already in the cluster. At that point, the first cluster is complete, and a second cluster is obtained by repeating the above process using only the unclustered variables. Further clusters can then be built

TABLE 3
Frequencies of Preference Orders Based on Thermometer Ratings of Democratic Party, Republican Party, and Political Independents, January–February, 1980*

Order	Frequency
Compatible with Dem-Indep-Rep Dimension	
DIR	15%
IDR	14
IRD	10
RID	13
Total	52%
Not Compatible with Dem-Indep-Rep Dimension	
DRI	31%
RDI	17
Total	48%

*Preference orders are listed in first choice, second choice, third choice order. The total number of respondents with untied preference orders is 507. If each possible preference order of three alternatives were equally likely, two-thirds of the preference orders would be consistent with a single dimension. This table is taken from Table 1 (page 39), of Weisberg (1980) with the data based on the Center for Political Studies 1980 American national election survey.

according to the same procedure until all variables have been clustered or no remaining unclustered pair of variables has a high correlation.

This is a very simple clustering algorithm, but as stated so far, it is actually incomplete. Decisions have to be made about how high correlations have to be to put a variable in a cluster. The computer program for this type of clustering⁵ permits the analyst to choose the limit after which variables are not added to a cluster and the limit after which new clusters are not formed. These are arbitrary decisions, but they can affect the clustering—different values would yield different sets of clusters.

Clustering Criterion

Additionally, the description in the above paragraphs did not define what it means to have "the highest correlations with variables already in a cluster." Does that mean the highest correlation with any variable in that cluster? If so, clusters are formed such that each variable has a high correlation with at least one variable in the cluster but not necessarily all of them. This possibility, called "single linkage," is useful in some clustering applications. For example, all that is necessary for the diffusion of a rumor through a communications network is for each person in a network to communicate with one other person in that network.

Alternatively, "the highest correlations with variables already in a cluster" could mean requiring high correlations with every variable in that cluster—or more precisely, that the new variable's minimum correlation with variables in the cluster is greater than the minimum correlation between the variables already in the cluster and any other unclustered variable. This condition leads to tight clusters, with a single low correlation being sufficient to prevent a variable from entering a cluster. The result, called "complete linkage," is a very common clustering criterion.

As a compromise between single and complete linkage, we could also use "average linkage," adding to a cluster the variable with the highest average correlation with the variables already in the cluster. This is a reasonable compromise, but it does operate at the interval level of measurement, while the other criteria use only ordinal properties of the correlations.

Hierarchical Clustering

Regardless of the linkage criteria chosen, the clustering approach described so far is limited in that it builds one cluster at a time. By contrast, another

⁵The computer program referred to here is the CLUSTER program in the OSIRIS computer package.

common clustering algorithm, Johnson's (1967) hierarchical clustering procedure,⁶ can build several clusters at once. The pair of variables with the highest correlation starts the first cluster. Then, depending on which correlation is higher, either a second cluster is started with the pair of unclustered variables with the highest remaining correlation or a third item (having the highest correlations with the variables already in the cluster) is added to the first cluster.⁷ At each following stage, a new cluster is formed, another item is added to one of the existing clusters, or existing clusters are merged, depending again on which correlation is higher.

This process can build several clusters at once, each from its nucleus out. The process continues until all items fit into a single cluster. For n variables, the process gives a set of n clusters, a set of $n - 1$ clusters, a set of $n - 2$ clusters, . . . , a set of 3 clusters, a set of 2 clusters, and finally the 1-cluster solution. Since the clusters at any later stage are actually made up of combinations of those at earlier stages, this is termed a "hierarchical" procedure. Which stage gives the correct solution? The researcher is likely to look at auxiliary statistics to decide that, particularly at the proportion of high correlations that are within the clusters already formed at a stage rather than between variables in different clusters. The higher that statistic, the sharper the separation between the clusters. Generally, that statistic is high for a very large number of clusters, eventually falls as the number of clusters goes down, and then goes back up for a very small number of clusters. The best solutions seem to be those just before the statistic falls or just when it comes back up.

The types of clustering described so far are highly arbitrary. Although there is no formal mathematical justification for a particular clustering algorithm, clustering can give researchers a quick, if only partially adequate, view of their data's internal structure. As we have described it so far, clustering is actually nondimensional, but we have included cluster analysis in this chapter because it can be used for scaling purposes (see below) and because the next procedure to be described in this chapter can be considered a dimensional version of cluster analysis.

Coefficients for Clustering

The reader might have noticed that we have discussed clustering correlations in this section without indicating which correlation coefficients are to be used. The reason is that it depends—different correlations can be used for different

⁶This technique is available in the HICLUSTER program in the OSIRIS computer package.

⁷Any of the criteria mentioned above—single linkage, complete linkage, or average linkage—can be employed in judging the correlation of unclustered variables with those in the cluster.

purposes. With interval data, Pearson's r can certainly be used in clustering. The result would be to give sets of variables which vary together to a considerable degree. With noninterval data, two-way measures of association, like Kendall's τ (τ_{au}) would give similar results.

One interesting application of clustering is to cluster "agreement scores," measures of how often responses to two variables are identical. This application is most common when the focus is on clustering individuals, rather than variables, such as clustering legislators to see which legislators tend to vote together as blocs. The agreement score could simply count the number of times a pair of legislators vote together, and clustering could be applied to the matrix of agreement scores.

Another useful application is to cluster Yule's Q (or other measures of one-way association) as an approximation to Guttman scaling. It was mentioned in an earlier section that variables which are cumulative would have a high Q coefficient with one another. A set of variables with high Q 's with one another would form a good quality Guttman scale. Cluster analysis thus can be used on Yule's Q to separate sets of variables which form Guttman scales from a larger set of multidimensional data. For example, Clausen and Cheney (1970) cluster a large number of legislative roll call votes to find what unidimensional Guttman scales exist in them.

Tree Analysis

Mathematical psychologists have recently developed an alternative to hierarchical clustering that represents similarity data as "additive trees" (Sattath and Tversky, 1977). An additive tree is a chart in which dissimilarity between objects is represented by the length of the path joining them. For example, Figure 2 shows an additive tree representation for the data of Table 4.

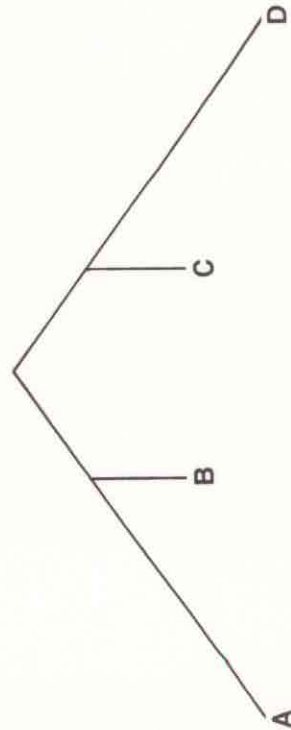


FIGURE 2
Additive Tree Representation of Table 4*

TABLE 4
Dissimilarities among Four Variables

	A	B	C	D
A	—	5.0	5.6	7.2
B		—	4.6	5.7
C			—	4.9
D				—

* Adapted from Table 1, Sattath and Tversky, p. 320.

There are two clusters: A and B on the left path in one cluster and C and D on the right path in the second. Additionally, the most similar (least dissimilar) variables are those closest together in the center—B and C—while the least similar (most dissimilar) are those at the opposite ends—A and D.

Additive tree analysis is related to multidimensional scaling as presented later in this chapter, except that the relations that count are those along the paths. The technique is most useful for classification purposes, when different paths along the tree can be different categories—subdividing the total set of variables into a few subsets, each of which is further subdivided. This procedure represents the data relations more completely than does conventional clustering.

Blockmodel Analysis

Another new clustering procedure is a "blockmodel" approach to analyzing relationships among individuals or other units. It decomposes a matrix of relations into blocks of individuals with similar relations to other blocks. Blocks need not be internally cohesive in this model; all that matters is that members of a block have similar relations with other blocks. The blockmodel procedure has been presented in a Workshop article by Panning (1982).

In a political example, Panning (1982) presents a blockmodel analysis of transfers among Senate committees to determine sets of committees with similar membership transfer relationships. He finds four sets of committees: six top committees (Foreign Relations, Appropriations, Agriculture, Finance, Judiciary, and Commerce), four at a second rung (Armed Services, Banking and Currency, Interior, and Labor and Public Welfare), three at a third rung (Public Works, Governmental Operations, and Rules and Administration), and two at the bottom (Post Office and District of Columbia). Membership transfers follow a general upward flow, except for equal flow back and forth from the second and fourth rung committees. The analysis does not differentiate within

the four sets of committees, for all that matters is that there are similar relationships among the four sets. The analysis is hierarchical, moving individuals (or committees in this example) from set to set to try to maximize the fit between the actual data and the blockmodel. This technique seems to be particularly appropriate for "dominance" situations, as when seeking a hierarchy of importance—here of committees. And the blockmodel procedure can handle asymmetric relations, as when the number of moves from committee A to committee B does not equal the number from committee B to committee A. It is too early to gauge the importance of the new blockmodel and additive tree approaches. However, it is clear that new and more applicable clustering procedures are being developed. The researcher interested in partitioning data into clusters can seek high-powered techniques that display many aspects of the data relations at once. The development of these techniques indicates a continued interest in mathematical psychology in finding appropriate geometric, cluster, and dimensional representations of complex data relations.

Factor Analysis: Dimensional Clustering

Factor analysis can be thought of as a dimensional clustering technique. Whereas the clustering procedures described so far view a variable as either in a cluster or not in it, factor analysis allows for the possibility that a variable is only partially in a cluster. The "factor loading" of a variable indicates the extent to which it partakes of the factor cluster—its correlation with the factor dimension.

Common Factor Model

The factor model interprets the standardized Z_{ij} score of a person i for variable j as a linear function of the person's position (F_m) for each factor p and the extent to which the variable measures the factor (a_{jp}). In the one-factor case, for example, the person's score is

$$Z_{ij} = a_{j1}F_{i1} + e_{ij}$$

where the e is an error term. Different people have different values for the variable because of their different positions for the factor, the F_i . The reader might note that this is very similar to a regression equation, except that the "independent" variable (the F_i) is not observed and is just hypothesized to exist. In the more general multi-factor case, the model is that the person's score is

$$Z_{ij} = a_{j1}F_{i1} + a_{j2}F_{i2} + \dots + a_{jm}F_{im} + e_{ij}$$

where m is the number of factors. This is like a multiple regression equation, except that again the F terms are not observed.

The solution of the factor model requires some auxiliary assumptions. In particular, given that the factors are just hypothetical anyhow, we might as well assume that they are standardized and uncorrelated. Those further assumptions are sufficient to yield a decomposition of the correlation between two variables j and k which is known as the "fundamental factor theorem":

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \dots + a_{jm}a_{km}$$

With one factor, this means that the correlation between two variables is just the product of their loadings on the underlying factor. With more factors, it means that the correlation is a linear sum of the product of their loadings on the respective factors. Factor analysis attempts to locate the variables in a dimensional space so as to satisfy this fundamental factor theorem.

In a geometric interpretation, factor analysis attempts to locate the variables on dimensions such that the correlation between a pair of variables (r_{jk}) equals the product of their distances from the origin ($d_j d_k$) times the cosine of the angle between them ($\cos \Theta_{jk}$). If the variables are perfectly correlated ($r = +1.0$), factor analysis would place them at the same point one unit from the origin, and they would be unidimensional. If the variables are independent ($r = .0$), factor analysis would locate them on orthogonal (right angle) dimensions.

For example, let us say that the correlations among five variables (labeled A-E) are as in Table 5. It is clear from the correlations that variables A and B have much in common, as do variables D and E, but those two sets are fairly distinct from each other. Furthermore, variable C is related to both sets. A geometric solution for these correlations is in Figure 3. The location of the variables on the dimensions are their "factor loadings." The loadings in this figure are given in Table 6.

The correlations can actually be "reproduced" from their loadings through use of the fundamental factor theorem. Variables A and B, for example, both

TABLE 5
Hypothetical Correlation Matrix Variable

	A	B	C	D	E
A	—				
B	.60	—			
C	.30	.50	—		
D	.00	.32	.50	—	
E	-.32	.00	.30	.60	—

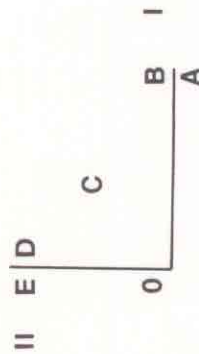


FIGURE 3
Common Factor Analysis of Table 5

have loadings of .8 for factor I and thus a product of .64. For factor II, A has a loading of $-.2$ and B has a loading of $+.2$, with a product of $-.04$. The sum of these products of loadings is then $.64 + (-.04) = .60$, the correlation given between variables A and B. Similarly, the other correlations can be reproduced exactly from the factor loadings.

In this example, it is possible to satisfy the correlation by locating the variables in a space of just two dimensions. Were that impossible, more than two dimensions would be used. As it is, two factors are required. Factor I is interpreted by examining what variables A and B have in common and factor II by examining what variables D and E have in common. Variable C is here related to both factors.

In a factor analysis, the use of Pearson's r correlations indicates that the variables are essentially standardized—means of zero and variances of one. If there were 5 variables, there would then be 5 units of variance being analyzed. Some of this variance is unique to each variable, but some is common variance. The type of factor analysis described in this section is termed the common

TABLE 6
Factor Loadings from Figure 3

Variable	Factor		Communality
	I	II	
A	.8	$-.2$.68
B	.8	.2	.68
C	.5	.5	.50
D	.2	.8	.68
E	$-.2$.8	.68
Sum of Squares	1.61	1.61	3.22

factor model.⁸ The amount of the variance of a variable which is accounted for by a factor is determined by squaring that variable's loading on the factor. Thus, variable A has a loading of .8 on factor I, so 64% of its variance is accounted for by that factor. Since its loading on factor II is $-.2$, 4% of its variance is accounted for by that factor. The sum of the squared values is known as the communality of the variable. Here the third column of the Table 6 shows that 68% of the variance of variable A is being accounted for by the factor analysis. This is less than 100% because variable A has some unique variance. In this case the variable with the least common variance is variable C.

How effective is the overall factor analysis? The sum of the communalities in the table is 3.22 units of variance. This is 100% of the common variance of the 5 variables, but it is not all of their total variance. Five standardized variables have 5 total units of variance, so that analysis accounts for $3.22/5.00 = 64.4\%$ of the total variance. When the common factor model is employed, the common variance should be a reasonable proportion of the total variance. An analysis that accounted for 95% of the common variance would not be very meaningful if the common variance were only 10% of the total variance.

Another way of looking at the factor loading matrix is to examine the sum of squared loadings in each column. The first factor accounts for .64 of the variance in variables A and B, .25 for variable C, and .04 for variables D and E. When summed, this factor accounts for 1.61 units of variance. Similarly, factor II accounts for 1.61 units of variance. The column sum of squares (or "eigenvalue") shows the importance of the factor, and here the two factors are of equal importance.

An important question with factor analysis is how many factors to employ. The usual means of handling this decision is to perform a preliminary "component analysis" (to be described in the next section)—a factor analysis which analyzes the total variance of each variable. As many factors are extracted as there are variables being analyzed. If each variable contributes one unit of standardized variance, then a factor improves on the original variables only if it accounts for more than one unit of variance. This logic leads to "Kaiser's criterion"—the use of as many factors as have a sum of squared loadings (eigenvalues) greater than one. Once the number of factors is determined, then the common factor solution is obtained using that number of factors.⁹

⁸This corresponds to the PA2 type of factor analysis in SPSS.

⁹Common factor analysis is actually iterative. An estimate of the common variance of each variable is made (possibly by using the proportion of its variance that can be accounted for statistically by the other variables, as shown by the R^2 value with that variable as dependent and the rest as independent). The factor analysis is performed using that estimate as the correlation of the variable with itself. The variable's communality in that analysis—the proportion of its variance

We should note that the mathematics underlying factor analysis assumes use of Pearson's r as the measure of correlation. Since factor analysis makes interval level assumptions, there would be no gain in moving down to an ordinal level measure of association.

Factor Rotation

A complication is that the placement of the axes in the factor space is arbitrary. It is easiest to interpret the factors if they are rotated so that the variables tend to have high loadings on some factors and low on the rest. This amounts to seeking to place the axes through clusters of variables: "simple structure rotation." The solution given above is of this type and would correspond to the rotated solution in a factor analysis.

Actually, the factors need not be at right angles to each other. The solution shown above uses orthogonal axes (ones at right angles to each other) and corresponds to the commonly used "varimax solution." But the factors could be correlated, since there is no reason to expect that separate clusters of variables are completely independent of one another. Correlated factors require an examination of the "oblique axes." The best way to interpret correlated factors is to examine the oblique factor "pattern" which projects a point on the dimensions by drawing parallels to the other dimensions. This is illustrated in Figure 4, where variable X would be located on factor II but not factor I, since its projection is high only on factor II.¹⁰

Factor Analysis Problems

Factor analysis is often portrayed as a mindless data-crunching exercise. That is true in the sense that any statistical analysis routine is mindless—running a hundred cross-tabulations or regressions can be equally mindless. Factor analysis is probably most useful for testing an hypothesis about the

accounted for by the factors—is used as a new common variance estimate. The factor analysis is repeated until the communality estimates stabilize.

¹⁰If we had used the other system of projection, perpendicular projection of X onto the two axes, we would find that X has high values on both factors, which does not help us interpret the results. The perpendicular projection system is known as the factor "structure." It corresponds to the correlation between the variable and the factor, and it is employed in factor scoring applications.

In orthogonal space, the factor pattern and factor structure are identical, which is a good reason for using the simpler orthogonal solution unless there is a substantial correlation among the oblique factors. Often this means performing an oblique analysis, checking whether the factor correlations are high, and (if not) performing an orthogonal analysis.

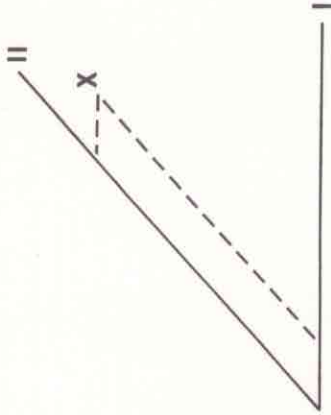


FIGURE 4

Oblique Factor Space

structure of a set of variables. This is termed the "confirmatory mode," in contrast to the "exploratory mode." A common application of the confirmatory mode is testing whether the common variance of a set of variables can be accounted for with a single underlying factor. However, we do not have many good structural dimensional hypotheses, so this theoretical mode of factor analysis is of limited usefulness.

Factor analysis is also often portrayed as plagued by indeterminacies, such as the choice of rotation. There are indeed indeterminacies, but the basic factor model is really the same basic general linear model which underlies regression analysis except that the "independent variables" are not observed but just hypothesized. Fortunately there are conventional ways of handling the indeterminacies that exist. These solutions may be arbitrary, but cluster analysis also involves a number of arbitrary decisions. The advantage of factor analysis over cluster analysis is that it allows variables to belong to the dimensional clusters to a partial degree. If we are willing to operate at the interval level of measurement, this method yields a more realistic portrayal of the data.

The final problem is one of interpretation of the solution. Factor analyses are often misinterpreted, as the users bend the interpretations to fit their desires. But that is not a problem with the technique so much as one with the researchers. No statistical technique can provide its own substantive interpretations. There are good ways to interpret the factors (generally by scoring the people on the factors and relating those scores to other variables as a means of testing different interpretations), but often we do not have enough data available to provide separate tests of the interpretations.

The discussion of factor analysis could go on endlessly. Instead, we shall stop here and turn more briefly to some related techniques.

Jöreskog Procedures

A special form of factor analysis should be mentioned that departs from the basic restrictions of this chapter. Jöreskog (1975) has developed a set of procedures for "exploratory factor analysis," "confirmatory factor analysis," "analysis of covariance structures," and "estimation of linear structural equation systems" that break down the conventional distinction between causal and scaling procedures. Basically, since the same general linear model is involved in regression and factor analysis, it can be generalized to handle the mixture of the two.

We shall give a single example to illustrate one Jöreskog procedure, but the reader should realize that the Jöreskog procedures are more general than this example. The Michigan elections studies include a large number of issue questions. It is reasonable to expect that these different issues tap a smaller number of underlying common dimensions, and factor analysis can be used to check for this possibility. In turn, these different dimensions can have different relationships to demographic variables, as can be determined by regression analysis using factor scores as dependent variables.

Knoke (1979) applied the Jöreskog procedure for the "estimation of linear structural equation systems with multiple indicators of unmeasured variables" (LISREL) to the 1972 Center for Political Studies national election survey. Specifically, he hypothesized a particular structure to the issue questions—making this a confirmatory rather than exploratory factor analysis—with some issues tapping only a social dimension (particularly attitudes toward marijuana and sex roles), some tapping only a racial dimension (specifically attitudes on minority rights), some both (attitudes toward rights of the accused), some only an economic dimension (attitudes on government action on inflation and on progressive taxation), and some both the racial and economic dimensions (business, government guarantees of jobs, and government health insurance). The analysis confirmed that structure, finding a sizable (.62) correlation between social and racial factors, a moderate (.37) correlation between racial and economic, and none (.08) between social and economic.

Furthermore, the analysis found different causal patterns for the different factors. The social factor was most affected by age (younger people more liberal on social questions), education and occupation (people of higher socioeconomic status more liberal). The racial factor was most affected by race (not surprising) and education (the more educated more liberal again) with some age effect (younger people more liberal). The economic factor was most affected by income (people of lower income more liberal), race (blacks more liberal), education (the less educated more liberal this time), and age (older people now the more liberal).

The success of this analysis rests on the ability of the analyst to make many assumptions, including which variables tap which dimensions as well as reliability estimates for some of the variables. Instead of allowing the analyst to explore a set of data without any prior expectations (as when a computer grinds out all possible crosstabulations of a set of variables), the researcher must be familiar with the data. The researcher must have an idea of what to look for in the data and where it is likely to be found. This is a good development, making it more difficult to capitalize on random features of the data and more necessary to develop good theories. The Knoke example nicely illustrates the interplay of theory and data analysis which is to be desired in any data analysis enterprise.

Component Analysis: Factoring Total Variance

There are actually several factor analysis "models." In addition to the "common factor model" described in the previous section, the "principal component model" is frequently employed.¹¹ It is particularly useful when trying to ascertain the extent to which a set of variables covary with one another. It differs from the common factor model in analyzing the total variance of the variable, not just its common variance.¹²

The principal component procedure seeks to locate a first component dimension in the space that accounts for a maximum proportion of the variance of the variables. If more than one component is required, the procedure locates the second to account for a maximum proportion of the variance not accounted for by the first component, subject to the constraint of orthogonal dimensions. This explanation holds for the *unrotated* component solution.

The first unrotated principal component thus serves as a general component, indicating the maximum proportion of the total variance which can be accounted for by a single component. If we wanted to construct a general index from the variables, we could use their loadings on the first component as their appropriate "weights." This would weight variables which tap the main underlying syndrome more heavily than those which do not.

If we wished to perform a component analysis on the example of Table 5 above, we would find the loadings in Figure 5 and Table 7. The principal component would correspond to the bisector of the two factors in the preceding figure, and there would be a second component. The first is the main compo-

¹¹This corresponds to the PA1 type of factor analysis in SPSS.

¹²Operationally, this is achieved by analyzing the correlation matrix with values of 1.0 along the main diagonal (corresponding to correlations of unity between the variable and itself), whereas the common factor model puts the commonality estimates into the main diagonal in an iterative analysis.

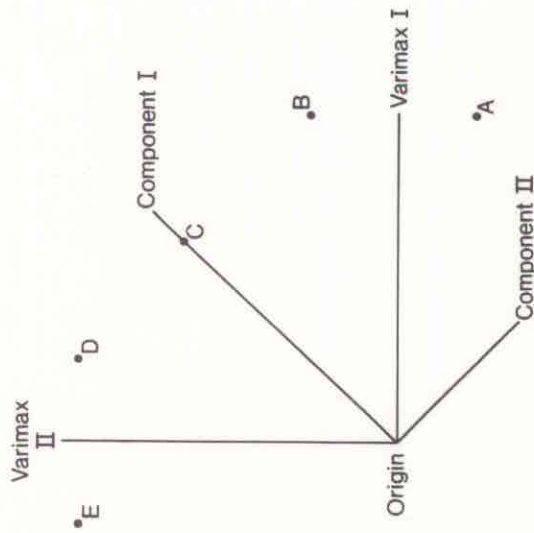


FIGURE 5
Component Analysis of Table 5

ment, since its sum of squares is larger (1.86 compared to 1.36), but the second is still to be retained, as its sum of squares (eigenvalue) is at least greater than 1.00.

This solution can be interpreted to mean that 1.86/5.00 or 37% of the total variance can be accounted for by a single component. All variables load on that component, though *B*, *C*, and *D* have the highest loadings. However, a second component exists, with *A* and *E* having higher loadings on it than the first component. That second component stresses the opposite character of *A* and *E* (and to a lesser extent *B* and *D*), with *C* not being related at all to the second component.

Model Choice

When is the component model used rather than the factor model? If the interest is in the extent to which a single component underlies a set of variables, the statistics underlying the component model make it highly appropriate. In more exploratory analyses of the structure of a set of variables, the common factor model is more likely to be used.

Incidentally, the mathematical difference in terms of the models is merely

TABLE 7
Factor Loading Matrix for Component Analysis (Figure 5)

Variable	Factor		Communality
	I	II	
A	.42	-.71	.68
B	.71	-.42	.68
C	.71	.00	.50
D	.71	.42	.68
E	.42	.71	.68
Sum of Squares (eigenvalue)	1.86	1.36	

that there is no error term included for the component model, and there are as many components as variables. So, for n variables, the decomposition of Z_{ij} is

$$Z_{ij} = a_{1j}F_{1i} + \dots + a_{nj}F_{ni}$$

Once again, the mathematics of the procedure is based on the use of Pearson's r correlations. Other coefficients are not usually employed.

Dichotomous Data

It should be noted also that the logic behind factor analysis fails for dichotomous data when Pearson's r is used (or its equivalent statistics with dichotomous data, Φ (phi) or Kendall's τ (tau) on a Guttman scale structure. With dichotomous data, Pearson's r would not equal 1.00 if the variables cumulate perfectly in a Guttman scale structure. As a result, factor analysis of Pearson's r 's on dichotomous data can yield multiple factors when the data fit a single Guttman scale (Clausen, 1964) and indeed will obtain more factors the more variables there are on that scale (Weisberg, 1968).

An interesting application, however, emerges from the utility of one-way measures of association like Yule's Q for analyzing Guttman scale structures. Since high Q values would be obtained between cumulative variables and zero Q values between independent variables, the component analysis procedure is actually a fairly effective means of separating multiple Guttman scales in a set of data (MacRae, 1970; Weisberg, 1968). This usage is unconventional but seems to work fairly well.

Multidimensional Scaling: Nonmetric Similarity Analysis

A fairly new procedure for analyzing the dimensional structure of a set of data is nonmetric multidimensional scaling. It seeks a spatial solution of

minimum dimensionality such that the more similar a pair of variables, the closer together they are in the space. While there are also some metric versions of multidimensional scaling, the nonmetric version is able to locate variables in a space quite well with merely ordinal interpretations of the similarity data. Since Rabinowitz explains the procedure in detail in the last chapter of this book, we shall not describe it further here.

How is multidimensional scaling used? It has been used on correlation coefficients (even Pearson's r , as in Weisberg and Rusk, 1970, and Rusk and Weisberg, 1972), but there are problems involved when there is a very small number of variables (Jones, 1974), and this procedure cannot be used to locate individuals in the same space as the variables (Rabinowitz, 1973). It can also be used on measures of similarity among the variables, or inter-individual agreement scores. For example, the number of times members of a legislature vote together can be analyzed to obtain a spatial mapping of their relative locations (Hoadley, 1980). Rabinowitz (1976) has developed a particularly ingenious measure to be used in multidimensional scaling of thermometer data, where survey respondents rate political candidates from 100° (very warm feelings) to 0° (very cool feelings).

An interesting question is how multidimensional scaling and factor analysis differ. Both give multidimensional geometric interpretations of relationships among a set of variables, but factor analysis provides a metric translation of correlation coefficients into scalar products between variables in the space, whereas multidimensional scaling provides a nonmetric representation of similarity data as spatial distances. In the two chapters which follow, Weisberg and Rabinowitz give their evaluations of the differences between these important techniques. Minimally, it is clear that both techniques will remain frequently used analysis procedures for the foreseeable future.

Multidimensional Unfolding

A special application of multidimensional scaling programs is for the purpose of unfolding analysis. In this application, the procedure seeks to locate individuals and variables together in a space such that the higher a person ranks an alternative in his preference order, the closer the variable is placed to him in the space. This provides a practical means of performing the type of unfolding analysis described earlier in this chapter, even with multidimensional data.

Individual Differences Scaling

Finally, there is an important variant of multidimensional scaling logic which allows for the possibility of different individuals weighting the dimen-

sions of a common space differently. For example, Converse's (1966) analysis of citizen preferences for French parties derives a two-dimensional structure—a left-right dimension and a cleric-anticleric dimension—but Converse goes on to show that different parties weight these dimensions differently: some parties stressing the left-right dimensions while giving little weight to the religious dimensions, others emphasizing the religious dimension while giving little weight to the left-right dimension, and still others giving fairly equal weights to the two dimensions. This same notion of individuals giving varying weights to common dimensions is embodied in Carroll and Chang's (1970) procedure for "individual differences scaling." Takane, Young, and de Leeuw (1977) have developed a particularly efficient procedure for such analysis, a procedure which can also be used for the basic multidimensional scaling analysis without the potential for local minimum problems.

Conclusions

No doubt the large number of procedures for dimensional analysis renders the area confusing. Often the nature of the available data serves to narrow down the choices. Yet it is important to consider a wide range of dimensional analysis procedures, rather than using a technique only because it has become conventional in a field without noticing that it has particular substantive implications that might not be fully warranted. There are a wide variety of alternative scaling models (see Coombs, 1964; Wolters, 1978). The chapter that follows picks up directly on this theme, illustrating some unconventional models which can be very appropriate substantively.

REFERENCES

- Carroll, J. Douglas, and Chang, J. J. 1970. "Analysis of Individual Differences Multidimensional Scaling Via an N-way Generalization of 'Eckart-Young' Decomposition," *Psychometrika*, September 1970, pp. 283-319.
- Clausen, Aage. 1964. "Policy Dimensions in Congressional Roll Calls: A Longitudinal Analysis." Ph.D. dissertation, University of Michigan.
- Clausen, Aage R., and Cheney, Richard B. 1970. "A Comparative Analysis of Senate-House Voting on Economic and Welfare Policy: 1953-1964," *American Political Science Review*, March 1970, pp. 138-152.
- Converse, Philip E. 1966. "The Problem of Party Distances in Models of Voting Change," in M. Kent Jennings and Harmon Zeigler, eds., *The*

- Electoral Process*. Englewood Cliffs, New Jersey: Prentice-Hall, pp. 175-207.
- Coombs, Clyde H. 1964. *A Theory of Data*. New York: Wiley.
- Hoadley, John. 1980. "The Emergence of Political Parties in Congress, 1789-1803," *American Political Science Review*, September 1980, pp. 757-779.
- Johnson, S. C. 1967. "Hierarchical Clustering Schemes," *Psychometrika*, September 1967, pp. 241-254.
- Jones, Bryan D. 1974. "Some Considerations in the Use of Nonmetric Multidimensional Scaling," *Political Methodology*, Fall 1974, pp. 1-30.
- Jöreskog, Karl G. 1973. "A General Method for Estimating a Linear Structural Equation System," in Arthur S. Goldenberger and Otis D. Duncan, eds., *Structural Equation Models in the Social Sciences*. New York: Seminar Press, pp. 85-112.
- Knoke, David. 1979. "Stratification and the Dimensions of American Political Orientations," *American Journal of Political Science*, November 1979, pp. 772-791.
- MacRae, Duncan, Jr. 1970. *Issues and Parties in Legislative Voting*. New York: Harper and Row.
- Matthews, Donald R., and Prothro, James W. 1966. *Negroes and the New Southern Politics*. New York: Harcourt.
- Milbrath, Lester W. 1965. *Political Participation*. Chicago: Rand-McNally.
- Panning, William H. 1982. "Blockmodels: From Relations to Configurations," *American Journal of Political Science*, August 1982, pp. 585-608.
- Rabinowitz, George B. 1973. *Spatial Models of Electoral Choice: An Empirical Analysis*. Chapel Hill, North Carolina: Institute for Research in Social Science.
- Rabinowitz, George B. 1976. "A Procedure for Ordering Pairs Consistent with the Multidimensional Unfolding Model," *Psychometrika*, September 1976, pp. 349-373.
- Rusk, Jerrold G., and Weisberg, Herbert F. 1972. "Perceptions of Presidential Candidates," *Midwest Journal of Political Science*, August 1972, pp. 388-410.
- Sattath, Shmuel, and Tversky, Amos. 1977. "Additive Similarity Trees," *Psychometrika*, September 1977, pp. 319-345.
- Takane, Yoshio; Young, Forrest W.; and de Leeuw, Jan. 1977. "Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features," *Psychometrika*, March 1977, pp. 7-67.

- Weisberg, Herbert F. 1968. "Dimensional Analysis of Legislative Roll Calls," Ph.D. dissertation, University of Michigan.
- Weisberg, Herbert F., and Rusk, Jerrold G. 1970. "Dimensions of Candidate Evaluation," *American Political Science Review*, December 1970, pp. 1167-85.
- Weisberg, Herbert F. 1980. "A Multidimensional Conceptualization of Party Identification," *Political Behavior*, no. 1, 1980, pp. 33-60.
- Wolters, Menno. 1978. "Models of Roll-Call Behavior," *Political Methodology*, no. 1, 1978, pp. 7-54.