5

The PARAFAC Model for Three-Way Factor Analysis and Multidimensional Scaling

Richard A. Harshman and Margaret E. Lundy

INTRODUCTION AND OVERVIEW

Motivation for PARAFAC

One of the oldest and most troublesome issues in factor analysis is the "rotation problem." Because the two-way factor model is inadequately constrained by the data, there are an infinite number of possible solutions consistent with any given data set. These alternatives correspond to different rotations (or more general transformations) of the coordinate axes in the factor space. To overcome this rotational indeterminacy, factor analysts have appealed to special criteria, such as "simple structure," to guide the selection of a preferred solution. But these rotation criteria are not part of the factor model itself, and the additional assumptions involved often seem hard to defend on empirical grounds. A further difficulty arises because of disagreement as

---

We would like to express our appreciation to the many colleagues (especially J. D. Carroll and J. B. Kruskal) whose stimulating comments over the last few years have contributed toward the development of the ideas presented in this chapter and the one that follows. We are particularly indebted to Henry G. Law, who (along with the other editors of this volume) initially stimulated the writing of this article, encouraged us as it was progressing, and showed extraordinary patience and forbearance in seeing it through to its final completion.

Portions of this research were supported by the Natural Sciences and Engineering Research Council of Canada, Grant #A7896. Additional funds were provided by intramural grants from the Faculty of Social Sciences, University of Western Ontario.

to which additional assumptions—or which rotation criteria—are most appropriate for determining the orientation of factors. Since alternative factor rotations sometimes lead to different and conflicting empirical generalizations, the problem of rotation is a serious one for those who seek to apply factor analysis to scientific problems.

PARAFAC is a three-way factor analysis/multidimensional scaling procedure that was developed in order to overcome the rotation problem, at least for certain classes of data. It uses three-way data in order to obtain richer information about the underlying factors than is possible with two-way data, thus providing an empirical basis for determining the "true" factor axes. The method generalizes Cattell's idea of parallel proportional profiles (1944), which is described below, into a particular strong three-way model of factor variation. (This model is closely related to the CANDECOMP model of Carroll and Chang 1970, which is also described below.) When the PARAFAC model is fit to data, a unique orientation of the factor axes is obtained as a direct consequence of determining the best-fitting factors for a given data set. Thus, if the model is appropriate, it eliminates the need for an additional factor rotation process based on controversial factor orientation criteria; the location of axes is an *intrinsic characteristic* of the factor solution itself.

Theoretical arguments (to be discussed below) indicate that intrinsic axis solutions provide a stronger basis for discovery of empirically meaningful factors than is typically provided by other criteria, such as principal components orientation, rotation to simple structure, and so forth. These arguments seem to be supported by results obtained when the model is applied to real data; the unrotated PARAFAC intrinsic axis factors are often quite meaningful (for instance, Gandour and Harshman 1978; Haan 1981; Harshman, Ladefoged, and Goldstein 1977; Harshman and Papcun 1976; Kettenring 1983; Meyer 1980; Snyder, Walsh, and Pamment 1983; Terbeek 1977; Dawson 1982; Harshman and Reddon 1983; Sentis, Harshman, and Stangor 1983; Trick 1983; Weinberg and Harshman 1980). Thus, it appears that PARAFAC can indeed provide a solution to the rotation problem in many situations.

In addition to providing meaningful intrinsic axes, PARAFAC has the extra advantage of directly decomposing the three-way array without requiring that it be "collapsed" into a two-way version. This is not a distinctive feature, of course, since PARAFAC shares this advantage with all the other three-way analysis procedures discussed in this volume. But the PARAFAC-CANDECOMP model does provide a particularly simple description of the structure of a three-way array, involving only one set of factors that are common to all three modes. In some cases, however, a more complex description is needed, and Tucker's three-mode model or one of the other models discussed in this volume might be preferable.

Limitations of PARAFAC and Attempts to Reduce Them

It is important to stress that the PARAFAC model for three-way arrays is not a completely general one. As we will discuss below, Tucker's three-mode model is more general and can thus provide solutions for data that cannot be adequately analyzed by PARAFAC. The reason for this is that Tucker's model is both more complex and weaker in its underlying assumptions. This creates the advantage of giving Tucker's model greater generality; however, it also creates a disadvantage, as it causes the model to be underdetermined by the data. Consequently, the solution obtained with Tucker's model does not have the intrinsic axis property; hence, after fitting the model, one must perform a series of factor rotations to obtain interpretable axes. This requires using additional rotational criteria to select a preferred solution, and one must face the same controversies concerning rotation that plague two-way factor analysis. As will be argued below, one would like to have intrinsic axis solutions whenever feasible. Therefore, one might adopt the following strategy: first, try to obtain an intrinsic axis solution by fitting PARAFAC; then, if additional structure seems present, or if the intrinsic axis solution is uninterpretable, proceed to apply the more general models, rotating axes to aid interpretation.

In the last five or six years, considerable effort has gone into broadening the applicability of PARAFAC, and several useful findings have been made:

a. By means of data preprocessing, the model has been extended to embrace a wider class of three-way data, while still retaining the important intrinsic axis property.
b. In some situations, however, meaningful solutions are not obtained, even with careful preprocessing, indicating that the extended PARAFAC model is apparently not general enough. For such cases, specially constrained methods of PARAFAC analysis have been developed (for instance, requiring orthogonality of axes in one mode) that often allow an interesting *subset* of the variance to be fit; useful intrinsic axis solutions can then be obtained, although additional systematic variance remains and should be analyzed with a more general model.
c. Diagnostic procedures have been developed that will alert the investigators to problems with the data or the analysis procedures (see appendix C in this volume).
d. And finally, still under development are even more general models that will allow the intrinsic axis property and other advantages of PARAFAC to be obtained with still wider classes of data.

Overview

Before discussing these more recent developments (see chapter 6), we will first consider the basic forms of the PARAFAC three-way factor analysis model: (a) the raw data or profile data form; (b) the form for analysis of covariance matrices; and (c) the form for multidimensional scaling (essentially equivalent to the INDSCAL model). Next, we will examine the intrinsic axis or "uniqueness" property of the model, briefly discussing its limitations and our interpretation of its significance. We will then consider the relationship between PARAFAC and some other models, including two-way factor analysis and Tucker's three-mode factor analysis. Finally, in chapter 6 we will consider the more recent PARAFAC extensions and other developments mentioned above.

THE BASIC PARAFAC MODEL

PARAFAC1 Applied to Raw Score or Profile Data

Generalizing the Two-Way Factor Model

The Traditional Factor Model. Let us begin by recalling the basic two-way factor-analytic model. The factor structure underlying a particular data point x_{ij} can be represented in scalar form as

$$x_{ij} = \sum_{r=1}^{q} (a_{ir} f_{jr}) + e_{ij} \tag{5–1}$$

and in matrix form as

$$\mathbf{X} = \mathbf{A}\mathbf{F}' + \mathbf{E} \ , \tag{5–2}$$

where $\mathbf{X}$ is a two-way data array with n rows and m columns. Although $\mathbf{X}$ might be any kind of data, let us assume, for convenience of discussion, that the n rows represent attributes (variables), and the m columns represent entities (cases, such as persons); thus, x_{ij} represents the value of the ith variable obtained for the jth case (person). For a model in terms of q factors, $\mathbf{A}$ is an n by q matrix of factor weights on variables, often called factor "loadings," and $\mathbf{F}$ is an m by q matrix of factor weights on cases or persons, often called factor "scores." The term a_{ir} represents the weight of the ith variable on the rth factor; similarly, f_{jr} represents the weight of the jth case on the rth factor. The weights represent the degree to which the factor is expressed in the particular variable or case. The size of the contribution of factor r to x_{ij} is thus a function of its importance both for that variable and for that person. $\mathbf{E}$ represents a matrix of random error terms (and "specific factor" contributions).

The model (5–1) and (5–2) can be considered either a principal components model or a common factor model, depending on our assumptions about $\mathbf{E}$. Current usage among some psychometricians assigns the terms "factor" and "factor score" specific meanings, which precludes their application to dimensions obtained by fitting a principal-components-like model. However, this convention is not followed by other factor analysts, such as Horst (1965), and to simplify exposition it will not be strictly followed here. We will often use "factor" in a general sense as a latent component of the kind described by (5–1), regardless of the assumptions about the error terms. We will also use the more general terminology of Kruskal (1978, 1981, and chapter 2), who calls (5–1) a *bilinear model* because its nonrandom or structural part is a bilinear expression. (The structural part of [5–1] is bilinear because it would be linear in the a_{ir} coefficients if the f_{jr} coefficients were considered fixed or given and vice versa.) Our use of the term "factor" in the more general sense and our adoption of Kruskal's classification scheme both allow us to focus on the characteristics of the structural part of the model independent of the particular assumptions made about the stochastic (random error) part. (For more details on Kruskal's terminology, see Kruskal, chapter 2, in this volume.)

The PARAFAC Three-Way Generalization. The PARAFAC generalization is based on Cattell's (1944) idea of parallel proportional profiles, which we will discuss in more detail in a later section. We simply note here that it describes the relationship between the factor loadings on two different occasions in which the factors change their relative influence. It specifies that if on the second occasion the influence of a given factor is increased by some amount, then on that occasion all the loadings for that factor should be increased by the same proportion. Thus, if the influence of factor 2 were increased by 20%, then all the a_{i2} coefficients should be 20% larger.

One easy way to express proportional changes without rewriting all the a coefficients is to introduce a third set of loadings, or o_{kr} coefficients, corresponding to occasions. Just as the a_{ir} coefficients proportionally increase or decrease the contributions of factor r from variable to variable, the o_{kr} coefficients proportionally increase or decrease the contributions of factor r from occasion to occasion. If for occasion k the coefficient represents a 20% increase in the effect of factor r relative to some baseline, then the revised loadings might be thought of as $(o_{kr}a_{ir})$, where $o_{kr} = 1.2$. An equivalent perspective is to consider the loadings to be fixed but the factor scores increased by 1.2 on the particular occasion. Thus, we might write the factor score for occasion k as $o_{kr}f_{jr}$.

Although Cattell conceptualized his principle of parallel proportional profiles as a means of relating the factors obtained in two different factor analyses of two different data sets, we generalize his notion to apply to the simultaneous factor analysis of many different occasions (factor analysis of a three-way data array). Thus, instead of x_{ij}, we now consider the triply subscripted data entry x_{ijk}, an element from a data array organized in terms of three different ways or (in Tucker's terminology) "modes" of classification. To extend the earlier example, the three modes of the data array could correspond to variables, persons, and occasions. The three modes could also refer to stimuli, rating scales, and individuals doing the rating, or any other such classifications that would define a data cell in a three-way array.

To maintain generality, we will often refer to the three modes of a data array as "Mode A," "Mode B," and "Mode C." From the point of view of PARAFAC, there is nothing distinctive about the mathematical properties of any mode as compared to any other. Because PARAFAC will be applied to many different kinds of three-way arrays, we no longer have any particular reason for treating the expression of factors in one of the three modes differently from the expression of factors in another. Therefore, we drop the distinction between factor "loadings" in one mode and factor "scores" in another. We consider factors to have weights or loadings in all three modes, and these loadings are of the same kind in all modes, although we might choose to scale them differently in one mode or another (see below). We thus change notation, from f_{jr} for a factor score to b_{jr} for a Mode B factor loading, and from o_{kr} to c_{kr} for a Mode C factor loading. The PARAFAC generalization of the two-way factor model of (5–1) takes the following simple form:

$$x_{ijk} = \sum_{r=1}^{q} (a_{ir} b_{jr} c_{kr}) + e_{ijk} \,. \tag{5–3}$$

Here, x_{ijk} is an entry in a three-way data array; it might, for example, be the score on the ith variable obtained by the jth person on the kth testing session. The a_{ir} term represents the loading of factor r on the ith level of Mode A; in terms of our example, it would stand for the importance or size of contribution of the rth factor to the ith variable. Similarly, b_{jr} represents the loading of factor r on the jth level of Mode B and so stands for the importance or size of the contribution of the rth factor for person j. Finally, the c_{kr} coefficient stands for the loading of factor r on the kth level of Mode C and in our example data set, would represent the importance of factor r on occasion k. Kruskal (1981, 1983, chapter 2) describes (5–3) as a *trilinear model*, since there are now three different sets of loadings, and the model is linear in each set if the other two are considered fixed. (Quadrilinear and higher-order generalizations have also been envisioned [Harshman 1970, 22] but have not been implemented in the PARAFAC program. However, the CANDECOMP procedure implemented by Carroll and Chang [1970] will analyze up to seven-way tables.)

The matrix representation of the PARAFAC model is in some ways less elegant than the scalar one. The evenhanded treatment of all three modes by PARAFAC results in a three-way symmetry of the model that is easy to represent in scalar terms but harder to carry over into matrix notation.[1] To facilitate the use of conventional matrix notation, we must divide up the three-way array into a stack of two-way matrices. The direction of division is arbitrary—we could divide the array in any of several different ways. Let us adopt the convention that the three-way array is "sliced" into matrices corresponding to different levels of Mode C. Thus, the array is represented as a stack of two-way matrices, each matrix being n by m and corresponding to a Mode A by Mode B set of observations; there would be p such matrices, one for each level of Mode C. We then write an expression representing all "slices" in the array by means of describing its arbitrary kth slice, just as we wrote a scalar expression describing all the elements in the array by describing its arbitrary ijth element.

If we let $\mathbf{X}_k$ represent that n by m matrix, which is the kth slice of the n by m by p three-way array, we can then write the expression

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{D}_k\,\mathbf{B}' + \mathbf{E}_k \,, \tag{5–4}$$

where $\mathbf{A}$ is an n by q factor-loading matrix for Mode $\mathbf{A}$, $\boldsymbol{B}$ is an m by q factor-loading matrix for Mode B, and $\mathbf{D}_k$ is a q by q

diagonal matrix, with diagonal elements taken from the kth row of **C**, a p by q factor-loading matrix for Mode C. Thus, the diagonal matrix $\mathbf{D}_k$ provides weights for the kth occasion, which step up or down the sizes of the columns of **A** (or, equivalently, the rows of $\mathbf{B}'$). The r diagonal elements of $\mathbf{D}_k$ thus represent the effect of the changes in the relative importance or influence of the r factors on occasion k.

*Scaling and Interpretation of **A**, **B**, and **C** Matrices.* As we have already noted, traditional two-way factor analysis has developed different names for the Mode A versus Mode B weights. In the example described above, if we considered only one occasion, the Mode A weights would be called factor loadings and the Mode B weights would be called factor scores. In addition to distinctive names, the two sets of weights have become associated with different scaling conventions and frameworks of interpretation. These conventions enhance the meaning of the weights by allowing one set (the loadings) to be interpreted as standardized regression coefficients—beta weights—and sometimes as correlations between variables and factors, while the other set (the scores) can be interpreted as z-scores describing the amount of each factor attributed to each case.

Since PARAFAC views the three modes as being interchangeable, it seeks to treat them evenhandedly. Thus, the basic model is worked out in a more general fashion in which the weights need have no special interpretation other than describing the linear composite of factors that would predict the data. This is analogous to the interpretation assigned to regression weights when neither the predictors nor the predicted variable are standardized in any way. It also means that with appropriate scaling of the output, PARAFAC-CANDECOMP can be viewed as a three-way generalization of the singular value decomposition of a two-way matrix. (For a discussion of the singular value decomposition, see Green 1978 or Kruskal, chapter 2. For an application of PARAFAC to provide the singular value decomposition, see Reddon, Marceau, and Jackson 1982.)

However, by adopting special conventions for standardizing both the input data and the output **A**, **B**, and **C** matrices, it is possible to produce PARAFAC weights that are strictly analogous to the loadings and scores of traditional factor analysis. PARAFAC solutions obtained in this way will have one mode that can be given the same special interpretations that traditional two-way loadings are given (for instance, as factor-variable correlations in the orthogonal case), while the other two modes can be interpreted as typical z-scores analogous to estimates of factor scores. Other standardizations that provide more general interpretations are also possible. The issues involved in standardization of data and loadings to achieve different meanings require some knowledge of the effects of different kinds of preprocessing (discussed in chapter 6), and so the relevant conclusions cannot be proven here. A more detailed discussion and proofs are provided in appendix 5-1. To show what is possible, however, we present in the following few paragraphs a summary of some of the most useful results.

Suppose, for example, that we are planning a PARAFAC analysis of a three-way set of profile data consisting of a group of variables measured on a set of cases under several different conditions. Furthermore, suppose that we wish Mode A, the variable weights, to be interpretable as loadings in the traditional sense, with Modes B and C weights interpretable as factor or component scores for the cases and conditions. To permit this we must standardize the data in a certain way before the analysis, and then we must standardize the loadings in a certain way after the analysis. Before analysis, the data must be mean-standardized or "centered" (that is, have means removed according to the procedures described in chapter 6) across levels of Modes B and/or C. In addition, the data must be size-standardized within levels of Mode A so that each level of Mode A has a variance of 1.0. After the analysis, the output loadings must also be standardized. The mean-squared loading for Modes B and C must be set to 1.0 so that Mode A loadings will reflect the scale of the data.

If these things are done, the Mode B and/or C loading matrices will be composed of z-scores, with each column having zero-mean and unit variance. This leads to two possible interpretations. In one, the Mode B and/or C weights are themselves factor scores of the traditional kind.[2] In the other, they represent *average* factor scores, while the actual factor scores are taken to be estimated by the $b_{jr}c_{kr}$ products. For simplicity of discussion, we only consider the second interpretation in this chapter. From this perspective, each PARAFAC weight (such as b_{jr}) corresponds to an average (that is, root-mean-square) factor score. For example, b_{jr} will give the root-mean-square score for factor r across all observations at level j of Mode B. This root-mean-square average will be equal to the standard deviation of the factor scores at that level, if the other mode is centered (for instance, b_{jr} will be the standard deviation of the factor scores at level j of Mode B if Mode C is centered, and vice versa). For each factor, the outer product of its Mode B and C PARAFAC weights will give a table of traditional z-score factor scores, with rows corresponding to levels of Mode B, columns to levels of Mode C, and entries ($b_{jr}c_{kr}$) corresponding to the individual factor score of case j in condition k for factor r.

If the entries in the factor score table for factor r are uncorrelated with the corresponding entries in the table for factor r^*, then factors r and r^* are "orthogonal" in the traditional sense (assuming zero factor score means as before). In terms of our example, this will happen whenever the two factors have orthogonal loadings in Mode B or C (or both). Thus, to have an "orthogonal solution," in which every table of factor scores is orthogonal to every other, it is sufficient that either the Mode B or C PARAFAC loading matrix have mutually orthogonal columns. (In fact, this latter condition is sufficient to produce orthogonal factor scores even when the factor score means are not zero.)

In a suitably standardized orthogonal PARAFAC solution, Mode A loadings can be interpreted as correlations between variables and underlying factors, just as in an orthogonal two-way solution. Conversely, if the tables of factor scores have correlated entries, the PARAFAC Mode A loadings will be interpretable as standardized regression coefficients—beta weights—in the same way as the traditional factor loadings that appear in the factor pattern matrix of an oblique two-way solution.

By applying the same preprocessing and standardization procedures to other combinations of modes, it is possible to interpret either Mode *B* or *C* PARAFAC loadings as factor loadings of the traditional kind. And even if the data are not centered in the necessary way, it is still possible to generate interpretations of a more general kind (in terms of mean-squares contributed by factors, and so forth). (These various combinations of input data processing and output loading standardization may seem complicated to select and cumbersome to employ in any actual analysis, but experience shows that this is not the case. They are available as defaults and/or easily specified options in the recent versions of the PARAFAC program; hence, they can be applied and the effects of different standardizations can be compared with little or no effort.) For further discussion of the different types of preprocessing and output standardization, including mathematical proofs of some of the assertions made above, see chapter 6 and appendix 5–1.

The Conceptual Model Underlying PARAFAC

Behind any *mathematical* model created for data analysis there must lie a *conceptual* model. The conceptual model provides a framework for logical or semantic interpretation of the terms in the mathematical model and thus determines the conditions under which its application to data can be deemed reasonable. In order to fully understand the PARAFAC generalization of two-way factor analysis, we need to consider the way in which it generalizes the conceptual as well as the mathematical two-way model. This task is not as straightforward as it might first seem. As has been pointed out previously (Harshman 1970, 19–25), the two-way *mathematical* model of factor analysis is an ambiguous representation of several fundamentally different *conceptual* models. And although these conceptual models have equivalent algebraic representations in the two-way case, their three-way generalizations are quite different. For example, one gives rise to PARAFAC, another to Tucker's three-mode model, and a third to a representation that cannot be fit by any current model of three-way data arrays and so must be fit indirectly.

System versus Object Variation. We can illustrate the implications of different conceptual models by comparing two models (discussed previously in Harshman 1970) that involve a differing "locus of action" for the factors. In the first model, the *system variation model,* the factors reside in the system under study and through the system affect the particular objects; the factor influences exhibited by particular objects would thus vary in a synchronous manner across a third mode such as occasions. In the second model, the *object variation model,* separate instances of the factors can be found in each of the objects, and these within-object factors would not be expected to show synchronous variation across levels of a third mode, such as occasions.

Consider, for example, an economist doing factor analysis of several production measurements on ten industries across a number of different years. He might believe that any factors which he uncovers—such as "raw material costs" or "energy costs"—reflect aspects of the economic *system* under study. Thus, he might find it natural to assume that when a factor increases its impact on some occasion, it does so proportionally for all measures and industries through the system. This would be a case of *system variation*.

In contrast, consider a psychologist studying personality change across time. He might think that the factors he uncovers —such as introversion-extroversion—refer to a *type* of influence rather than a single *source* of variation, with the actual effect of introversion-extroversion occurring at innumerable independent loci. Indeed, each individual under study would presumably be a source of his own introversion-extroversion variation. In this case, the factors would be considered resident in the objects under study (the persons) rather than in some single system. Thus, the variations across time (in introversion-extroversion) would show a different pattern for each individual, corresponding to his idiosyncratic experiences and life history. This would be a case of *object variation*.

Mathematical Expression of the Two Models. These two different conceptions of "factor" imply different three-way factor models. To compare them, first consider this very general expression for factor variation in a three-way array, which we call the *unconstrained object variation model*:

$$x_{ijk} = \sum_r (a_{ir} f_{jkr}) + e_{ijk} \ . \qquad (5\text{–}5)$$

Because the factor score component in (5–5) is triply subscripted, there is no constraint placed on the pattern of variation of factor scores across Modes B and C of the data. (Some may question whether this is really a three-way model at all, since it is easy to fit [5–5] by conventional two-way methods: Simply place all the two-way slices of the three-way array end to end, thus "stringing-out" the data into a large two-way matrix that is i by jk, and then perform a two-way analysis on this matrix. Nonetheless, [5–5] is a reasonable model for a three-way array. Such ambiguities of classification are not uncommon when one considers extreme limiting cases of particular ideas.)

The system variation concept of factor variation would imply that

$$f_{jkr} = b_{jr} c_{kr} \ , \qquad (5\text{–}6)$$

which says that if we consider object j on occasion k, we find that the influence or expression of factor r is stepped up or down by an amount c_{kr}. This shrinkage or expansion in the importance of factor r occurs in parallel for all objects, but the effect for any particular object is proportional to the basic sensitivity of object j to effects of factor r (as designated by b_{jr}). Thus, the change in the cost of energy would affect the different measures and different industries in proportion to their energy dependency.

Another example of an appropriate application of the system variation model is semantic differential or similar data, in which each of several stimuli are rated on each of several scales by a number of different judges. It is reasonable to assume that the set of rating scales taps a smaller set of underlying attributes (factors), and that these attributes are found across all the different stimuli and are used to greater or lesser degree by most of the judges. Moreover, suppose we assume that: (a) underlying attributes change in their relative value from one stimulus to the next (consistently across scales and judges); (b) that each judge is more or less sensitive to each attribute factor (consistently across stimuli and scales); and (c) that each scale can be described as measuring some combination of these basic attributes (more or less consistently across the stimuli and judges). Then we can justify the assumption that the influence of a given factor goes up or down proportionally as one goes from one stimulus to the next, one judge to the next, and one attribute or rating scale to the next. The rating scale example of system variation is particularly useful in that it demonstrates the actual symmetry of the PARAFAC system variation model; the requirement is that factors change influence proportionally across the levels of *each* mode. We need not focus on a particular mode as representing factor loadings and the others as representing factor scores, except to facilitate discussion. It also shows how the "system" involved in the system variation model might be quite abstract. In the case of the ratings, the system is presumably the cognitive/semantic system in which the dimensions of meaning are defined; this allows a given dimension to be expressed to a greater or lesser degree by a given rater, a given stimulus, and/or a given scale.

In contrast to the system variation model, we might consider the object variation model to be represented as follows:

$$x_{ijk} = a_{ir}(b_{jr} + v_{jkr}) + e_{ijk} \; . \tag{5–7}$$

Here, the basic factor score for object j on factor r is modified on occasion k by a variation v_{jkr}; this variation is not necessarily similar to the variation of any other object.

Other Types of Data. Data that requires the generality of the unconstrained object variation model (5–5) or the object variation model (5–7) cannot be adequately described by more restricted models such as PARAFAC or Tucker's three-mode model. However, there are intermediate forms of variation in which these more restricted models might be useful. For example, many longitudinal data sets may contain a combination of object and system variation. To analyze such data, one might fit the system variation part and treat the object variation as "error." One might instead choose to fit the data initially with a model that does not distinguish the two parts (for instance, [5–12], below; also see the discussion in Harshman and Berenbaum 1981 and application in Haan 1981).

There are also forms of systematic three-mode variation that are less general than (5–7) but more general than (5–3). Some of these will be captured by the extended PARAFAC model, to be discussed below. Others require three-way models with a more complex structure, such as Tucker's three-mode model, but may be captured in part by specially constrained PARAFAC analyses (see chapter 6). Tucker's three-mode model is based on a conceptualization of "factor" that is different from either the system variation or object variation models. For Tucker, a factor does not represent a distinct additive source of variation in the data; rather, it represents an idealized aspect or pattern of variation in a given mode, which generates the data by interaction with different idealized aspects of other modes. Tucker's model thus conceptualizes a separate set of factors for each mode of the data (indeed, the number of factors in one mode need not match the number of factors in another). This model will be considered further when PARAFAC is compared with other three-way factor analysis procedures.

Because the application of PARAFAC to raw or profile data presupposes the system variation model, it is important to consider the type of variation expected in one's data before performing an analysis. If one believes that object variation is likely to be the main kind of factor variation found, then direct application of PARAFAC is not warranted. However, an indirect application of PARAFAC analysis is still possible, provided the data is preprocessed by conversion to covariances before analysis. We will show in the next section that this data transformation allows one to fit a version of the PARAFAC model appropriate for object variation data.

PARAFAC1 Applied to Covariance Data

Importance of Covariance Analysis

Because factor analysis provides a structural model for one's original observations ([5–1] and [5–2]), it would seem natural to obtain factor loadings and factor scores by directly fitting this model to a set of data. Some investigators (including Horst 1965; Kruskal 1978) have favored this approach, pointing out that it is the most mathematically and logically straightforward method of estimation. Historically, however, factor analysis has focused on analysis of correlations rather than the raw data, and this perspective has maintained its dominance to the present day, perhaps because of the reduction in computational effort that it permits. Consequently, factor analysis is usually performed by what Kruskal (1978) calls the *indirect fitting* approach: The data are first transformed into a set of correlations among variables (or cases), and then a factor model (derived from [5–1]) is fit to the correlations. It turns out that in the two-way case, these approaches are equivalent. That is, the indirect fitting of the data based on a least-squares fit of the derived factor model to the correlation matrix yields the same loadings as the direct least-squares fit of (5–1) to the z-score matrix. Thus, the distinction between fitting methods has not been given much attention (aside from Kruskal 1978 and McDonald 1979).

With three-way data, the equivalence between direct and indirect fitting no longer holds. The results of direct and indirect PARAFAC analysis differ both at the *statistical* level, in terms of the precise values of the loadings and residual errors that are obtained when fitting the system variation model to a given data set, and at a deeper *structural* level, in terms of the patterns of data variation that can be implicitly represented by the solution. The different structural implications of the two approaches were pointed out earlier (Harshman 1972b), but the differences in statistical characteristics are only now becoming known. Because we are still investigating these statistical differences, we will only briefly summarize some current findings and will focus the bulk of our discussion on the different structural models that can be fit by the two methods.

Statistical Nonequivalence. In the three-way case, direct and indirect fitting will provide identical loadings only in very special circumstances that would not be realized with real data (for instance, when factors are orthogonal in Mode *B* and the data are fit perfectly, or when all the variance is not fit by the extracted dimensions but the unextracted dimensions—systematic *and error*—are perfectly orthogonal to the extracted dimensions in both Modes *A* and *B*). With realistic cases involving fallible data and less than orthogonal dimensions, the loadings obtained by indirect fitting may show varying degrees of resemblance to those obtained by direct fitting. Generally, the resemblance is fairly close, but in a few of these cases, the differences can be large enough to substantially affect interpretation.

Currently, our interpretation is that the general equivalence of two-mode solutions obtained by least-squares direct and indirect fits depends in part on the arbitrary axis orientation that is possible in the two-way case but not in the PARAFAC three-way case. In the two-way case, it is always possible to obtain successive best fitting dimensions that are orthogonal to all previous ones in both modes. As a result, successive solutions are "nested," in the sense that the best fitting two-dimensional solution forms the first two dimensions of the best fitting three-dimensional solution, and so on. In the three-way case, the best fitting dimensions are not generally orthogonal to one another in both modes, and they cannot be made orthogonal because of the intrinsic axis property. Thus, the nesting of solutions does not generally occur. And since in the indirect solution Mode *A* is represented twice, Mode *C* loadings are squared, and Mode *B* is not represented at all (see below), the degree and pattern of effective variance overlap of dimensions will be different from the direct fitting solutions. Hence, reduced-rank approximations (for example, the two-dimensional approximation of a three-dimensional data set) obtained by direct versus indirect fitting methods will involve different compromises and so will not generally be the same. The only circumstance in which the compromises would be the same is when the nonarbitrary PARAFAC orientation of axes happens to provide orthogonal dimensions in both Modes *A* and *B*.

Nonequivalent Structural Implications. As we shall see below, both direct and indirect methods can be used to fit the system variation model to three-way data. However, indirect fitting does not make use of the same information in the three-way array, and it involves different assumptions about the patterns of three-way data variation. As a result, it can also be used to fit non–system-variation versions of the PARAFAC structural models. For example, the indirect fitting approach permits PARAFAC (and Tucker's three-model model) to handle object variation data and thus provides a means of fitting three-way models that have much greater generality. At the same time, the simplest PARAFAC implementation of the indirect fitting approach requires assumptions about factor orthogonality that are not required by the direct fitting approach. These structural differences will be considered in detail in the following sections.

The differences in statistical and structural characteristics of direct and indirect fitting in the three-way case make it important to consider in some detail the implications of analysis of covariances as opposed to raw or preprocessed profile data. Unless the investigator understands the different implications of the two approaches, he may not be able to properly determine which is most appropriate for a given problem.

Approach and Terminology

In the discussion that follows, we focus on analysis of summed cross-products and covariances rather than analysis of correlations. This gives our discussion greater generality, since correlations are a special case of covariances in which the data for each variable are scaled to have unit variance, and covariances are in turn a special case of summed cross-products, in which the data for each variable have zero-mean. There is also a more serious reason for avoiding correlations, however. As we will show below, computing correlation matrices for each of several occasions would generally impose separate scalings on each variable for each occasion, which would complicate any expression for the size of contributions of factors to a given variable across occasions. Thus, we will see that analysis of correlations is not in general appropriate for either PARAFAC, Tucker's three-mode, or other current three-way factor analysis models.

The analysis of covariances by PARAFAC can be thought of in either of two ways: (a) as fitting to covariances the same model used for raw data ([5–3] and [5–4]); or (b) as fitting to the covariances a model *derived* from (5–4), one which represents the structure that would underlie covariance matrices if they were computed from raw data with structure described by (5–4). From the first perspective, the model is unchanged but the solution is said to take on a "special form" (for example, Mode A loadings equal Mode B loadings). From the second perspective, the special form of the solution corresponds to the special form of the derived model. In either case, the same computational algorithm can be used to perform the analysis, since the special form emerges as a result of the data rather than as a result of any constraints imposed by the analysis procedure. Thus, we consider the two perspectives to be interchangeable and so will sometimes talk of the "PARAFAC1 model for covariances" and at other times talk of the "PARAFAC1 model" in a more general way that is meant to include its application to raw data, covariances, and even multidimensional scaling applications.

We refer to "PARAFAC1" rather than simply "PARAFAC" in order to distinguish the model of (5–3) and (5–4) and its derivatives (such as [5–10], below) from generalizations such as (5–9), which is called "PARAFAC2" (Harshman 1972b), and (T1–7) (see Table 5–1), which is called "PARAFAC3." These latter generalizations require different computer algorithms and can no longer be considered trilinear models (for instance, PARAFAC2 is quintilinear, at least from one perspective).

Deriving the PARAFAC1 Covariance Model from the Raw-Score Model

A General Expression for Cross-Products. Suppose we begin with data that have PARAFAC latent structure, as described in (5–4), and compute cross-products among variables for each occasion, as follows:

$$\mathbf{C}_k = (\mathbf{X}_k \mathbf{X}_k') \ . \qquad (5\text{–}8)$$

By substituting the PARAFAC representation of $\mathbf{X}_k$ from (5–4) into (5–8), we get

$$\mathbf{C}_k = (\mathbf{A}\,\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k)\ (\mathbf{A}\,\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k)' \ ,$$

and if we assume that the error is orthogonal to the systematic part, the cross-products of error and systematic terms drop out, leaving

$$\mathbf{C}_k = (\mathbf{A}\,\mathbf{D}_k\mathbf{B}')\ (\mathbf{A}\,\mathbf{D}_k\mathbf{B}')' + (\mathbf{E}_k\mathbf{E}_k') \ .$$

By taking transposes and regrouping terms, we obtain

$$\mathbf{C}_k = \mathbf{A}\,\mathbf{D}_k\,(\mathbf{B}'\mathbf{B})\ \mathbf{D}_k\,\mathbf{A}' + \mathbf{E}_k\,\mathbf{E}_k' \ ,$$

and if we let $\mathbf{W} = (\mathbf{B}'\mathbf{B})$, the matrix of cross-products among Mode *B* factor loadings, we get the general PARAFAC model for cross-products:

$$\mathbf{C}_k = \mathbf{A}\,\mathbf{D}_k\mathbf{W}\,\mathbf{D}_k\mathbf{A}' + \mathbf{E}_k\mathbf{E}_k' \ . \qquad (5\text{–}9)$$

For maximum generality, we have developed (5–9) in terms of summed cross-products. Several special cases should be noted. If the $\mathbf{X}_k$ matrices are individually row-centered (so that the mean for each variable in each matrix is zero), then $\mathbf{C}_k$ is the deviation sums of squares and cross-products matrix for occasion k. By taking the further step of dividing each entry in $\mathbf{C}_k$ by m, the number of entries in each row of $\mathbf{X}_k$, then $\mathbf{C}_k$ represents the covariances among variables on occasion k and (5–9) gives the general PARAFAC representation of the structure underlying a three-way array of covariance matrices.

PARAFAC2. The model (5–9) corresponds to the PARAFAC2 generalization of PARAFAC (Harshman 1972b). PARAFAC2 is a nonorthogonal factor model for summed cross-product and covariance matrices and will be discussed briefly later in this chapter. The standard algorithm used to fit (5–4) to data cannot be used to fit this model to data.

The $\mathbf{W}$ matrix, the only new component to the PARAFAC2 model, can be interpreted as describing obliqueness or nonindependence among the factors in the mode over which covariances were computed (Mode *B*). When the $\mathbf{X}_k$ are row-centered and multiplied by $1/m$ so that the $\mathbf{C}_k$ become covariance matrices, the

$\mathbf{B}'$ matrix in the underlying PARAFAC representation of $\mathbf{X}_k$ is also row-centered. If we absorb the $1/m$ multiplier of $\mathbf{X}_k$ into $\mathbf{B}$, then the $\mathbf{B}'\mathbf{B}$ matrix becomes a matrix of covariances among the Mode B factor loadings.

If the columns of $\mathbf{B}$ (rows of $\mathbf{B}'$) are further scaled so as to have mean-square of 1.0, with compensatory rescaling of the $\mathbf{D}_k$ (which is always possible; see appendix 5–1), then $\mathbf{W}$ becomes a matrix of factor intercorrelations, and (5–9) corresponds to a three-way generalization of the two-way oblique factor analysis model. The $\mathbf{A}$ matrix is the same n by q matrix of factor loadings that appears in (5–4), $\mathbf{D}_k$ is the same q by q matrix that gives the weights of the factors on occasion k, and $\mathbf{W}$ is a q by q matrix that in this case gives the cosines of the angles among the factors. For this special case, $\mathbf{W}$ could be designated Φ to correspond with traditional notation for the factor intercorrelation matrix. Finally, $\mathbf{E}_k\mathbf{E}'_k$ gives the residual covariances not fit by the model.

PARAFAC1. If we make the additional assumption that $\mathbf{W} = \mathbf{I}$ (that is, that the factors are orthogonal in Mode B), then the $\mathbf{W}$ matrix in (5–9) disappears, and the model can be written as follows:

$$\mathbf{C}_k = \mathbf{A}\mathbf{D}_k^2\mathbf{A}' + \mathbf{E}_k\mathbf{E}'_k \; . \tag{5–10}$$

This is called the "PARAFAC1 model for covariances" (or other cross-product data). The numeral "1" designates that it corresponds to the simplest PARAFAC model and can be fit by the same algorithm as is used to fit (5–4) to a raw data matrix. Thus, as noted earlier, it can be considered an alternative application of the original PARAFAC procedure, one in which the solution takes on a special form. When this model is fit to covariance data by the same general three-way PARAFAC algorithm used to fit (5–4) to profile data, the first and second tables of factor loadings (which in [5–4] would correspond to Mode A and Mode B loadings, respectively) start out containing different values but end up at convergence to be identical to one another; this is a natural consequence of the symmetry of Modes A and B that occurs when the input data are covariances. Furthermore, the Mode C matrix obtained by analyzing the $\mathbf{C}_k$ will contain the *squares* of the factor weights for occasion k, namely, the squares of entries that would be obtained if the raw $\mathbf{X}_k$ matrices were analyzed. This occurs because each covariance is a mean of cross-products, and in each cross-product, the Mode C weight occurs twice and hence is squared.

Consequences of Indirect versus Direct Fitting

If the raw data can be assumed to have a structure similar to (5–4) and the factors are orthogonal in the mode over which covariances are computed (in the example above, if $\mathbf{B}'\mathbf{B}$ is diagonal), then the PARAFAC1 model for covariances (5–10) is appropriate and the factor loading matrix $\mathbf{A}$ obtained by indirect fitting will usually be quite similar to the matrix obtained by direct factoring of the raw data. However, indirect fitting does not optimize the same fit criterion as direct fitting: One is least-squares to the covariances, the other least-squares to the profile data. As noted earlier, these criteria are not strictly equivalent in the three-way case and can sometimes lead to differences in the resulting solution that are large enough to affect interpretation. Thus, some thought should be given to selection of the approach that would seem most appropriate for any given problem.

The solution obtained by indirect fitting will entail some initial loss of information, because loadings for Mode *B* of the raw data do not occur. (However, they can be estimated after the analysis by regression methods.) On the other hand, there may be some economy of computation in the indirect fitting procedure if there are many more levels to Mode *B* than to Mode *A*. These are not usually the main considerations, however. When the direct fitting model is deemed appropriate and gives interpretable results, it would often seem more straightforward and thus preferable.

There is one respect in which the indirect fitting model might sometimes seem too restrictive. If the factors are *not* orthogonal in the mode over which covariances are computed, then the PARAFAC1 model for covariances is not strictly appropriate and the solution will be distorted to some extent. If the divergence from orthogonality is not great, then the distortion will be quite small and will not affect the interpretation of the factors. If, however, the underlying factors are actually quite oblique in the mode over which covariances were computed, then the solution will be considerably distorted and a PARAFAC1 analysis may be misleading or uninterpretable. Thus, the necessity of assuming that the factors are orthogonal in the mode over which covariances are computed can occasionally be a serious disadvantage of the indirect fitting approach. Note that direct fitting of (5–4) to the raw data does not make any assumptions concerning the
lumns) orthogonality of factor-loading matrix **B**. Thus, it would usually
the seem preferable to use direct fitting when other considerations are roughly equal.

Sometimes, however, the restriction on the solution imposed by the implicit orthogonality assumption in the covariance model (5–10) is a help rather than a hindrance. Certain difficult data sets—in which the patterns of factor variation do not closely approximate those of any PARAFAC model (see chapter 6)—tend to produce degenerate solutions when the raw data is analyzed directly using the PARAFAC1 raw data model (5–4); very highly correlated factors occur and interpretation is not feasible. These degenerate solutions appear to be less likely when indirect fitting using (5–10) is employed, in part, perhaps, because of the stability resulting from the implicit orthogonality assumption, which tends to block the appearance of highly correlated factors.

The most important advantage of indirect fitting has not yet been discussed, because is it not apparent from our derivations thus far: By analyzing cross-products or correlations rather than raw data, one can estimate the parameters for models more general than (5–4), while still retaining the desirable intrinsic axis property. In particular, one can obtain meaningful intrinsic axis PARAFAC solutions when the raw data follows the object variation model, and it is even possible to fit data sets in which a different sample of cases is measured at each level of Mode *C*, so that there is no direct correspondence whatsoever between the factor score for a given case on one occasion and the score on the next.

Deriving the PARAFAC1 Model for Covariances from More General Assumptions

Sampling from Several Populations. Let $\mathbf{X}_k$ be an n by m_k data matrix, consisting of measurements of n variables on m_k cases. Assume that there are p such matrices obtained by sampling from p different populations, and that while the number of cases in the samples may differ, all samples are measured on the same n variables. If we postulate that these variables tap the same set of factors in all the samples, but expect that the relative importance of the factors may change across populations and therefore samples, then we might represent the underlying structure of such data as follows:

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{D}_k\,\mathbf{B}'_k + \mathbf{E}_k \;. \tag{5-11}$$

Here, $\mathbf{A}$ is a common n by q factor-loading matrix for all samples; $\mathbf{D}_k$ is a diagonal q by q weight matrix that gives the relative importance of the q factors on the kth occasion; and $\mathbf{B}_k$ is an m_k by q matrix of person weights or factor scores for sample k. $\mathbf{E}_k$ is an n by m_k matrix of error terms for occasion k.

In its raw data form, (5–11) cannot be fit by the PARAFAC model. One could perform a two-way analysis, by "stringing out" or concatenating all the matrices, to obtain a single matrix with n rows and as many columns as cases in all samples combined. Of course, the two-way analysis of this array will not have the intrinsic axis property. However, the data can be *indirectly* fit by PARAFAC if they are first converted to cross-products (covariances if the $\mathbf{X}_k$ were all row-centered). This analysis *will* have the intrinsic axis property.

Let $\mathbf{C}_k$ be the kth such cross-product or covariance matrix, defined as in (5–8). By substituting the expression $\mathbf{X}_k$ from (5–11) into (5–8) and then applying the same steps that follow (5–8), we obtain:

$$\mathbf{C}_k = \mathbf{A}\,\mathbf{D}_k\,(\mathbf{B}'_k\,\mathbf{B}_k)\,\mathbf{D}_k\,\mathbf{A}' + (\mathbf{E}_k\,\mathbf{E}'_k) \;.$$

If we let $\mathbf{W}_k = \mathbf{B}'_k\,\mathbf{B}_k$, then

$$\mathbf{C}_k = \mathbf{A}\,\mathbf{D}_k\,\mathbf{W}_k\,\mathbf{D}_k\,\mathbf{A}' + (\mathbf{E}_k\,\mathbf{E}'_k) \;. \tag{5-12}$$

This is a general expression for the factor structure of the kth matrix of summed cross-products. If the data arrays had been row-centered, then this would be an expression for deviation cross-products, and if $\mathbf{C}_k$ is divided by m_k, then $\mathbf{C}_k$ is a covariance matrix, and $\mathbf{W}_k$ can be standardized to represent correlations or cosines among the factors in Mode B for occasion k. This then becomes a three-way generalization of the oblique factor model. In this generalization, the angles among the factors are not expected to remain constant across different levels of Mode C. Hence, this is more general than the PARAFAC2 model derived earlier by assuming system variation.

The model (5–12) is similar to Carroll and Chang's IDIOSCAL model (1972). It is more general than Tucker's three-mode model, although it is a special case of a variant of Tucker's model that Kroonenberg and de Leeuw (1980) call the "Tucker2" model. The representation of (5–12) was also discussed by Meredith (1964). The relations among such models will be considered in a later section.

In many cases, it will be reasonable to assume that the factors maintain roughly the same intercorrelations across the different levels of Mode *C*; hence, $\mathbf{W}_k = \mathbf{W}$, a common matrix of angles among factors for all covariance matrices. In this case, (5–12) reduces to (5–9), the PARAFAC2 model.

If we assume that in all samples the factors are orthogonal in Mode *B*, then $\mathbf{W}_k = \mathbf{I}$ and (5–12) reduces to (5–10), the PARAFAC1 model for covariance matrices. Thus, the PARAFAC1 model for covariances can be derived from assumptions far less restrictive than the system variation model from which it was originally obtained. Consequently, PARAFAC1 can be used to estimate the factors underlying raw data of the very general form (5–11), provided that indirect fitting is used.

Object Variation in One Population. In the case where all $\mathbf{B}_k$ matrices are measured on the same subjects, and so are all the same size, then (5–11) can be taken as a representation of a three-way data array with object variation. Thus, PARAFAC1 can also be used to analyze object variation data, by means of indirect fitting.

However, in all these applications of PARAFAC1, there is the need to adopt an orthogonality assumption that was not part of the system variation model: We must assume that the factors are orthogonal in the mode over which the covariances were computed (in these examples, Mode *B*).

Issues of Scaling and Interpretation

Scaling of Loadings and Data. As noted earlier, nonstandardized loadings can be interpreted as describing relative importance of different variables for a given factor by simply comparing their sizes within a column of the loadings matrix, provided the data sum of squares is similar across variables. However, if one wants to make additional interpretations of loadings in terms of correlations between factors and data and so forth, and if one wants PARAFAC loadings to be directly comparable to the loadings that are obtained in two-way factor analysis of correlation matrices, then special standardization of the data and of the output loadings is required. In essence, the covariance matrices must be scaled so that the average covariance matrix would have unit diagonals and therefore be interpreted as a correlation matrix. The loadings must be scaled so that the identical Mode *A* and Mode *B* tables jointly reflect the scale of the data, with the Mode *C* table having column means of 1.0. (This arrangement is discussed in appendix 5–1.) With this rescaling, factor loadings of the traditional kind are obtained. It is important to note, however, that one *cannot* simply proceed by converting the covariance matrices individually to correlation matrices and then analyzing the correlation matrices.

Inappropriateness of Correlations. If, in the above derivations, we had proceeded to rescale the covariances to obtain correlation matrices, no simple PARAFAC representation would have been possible. Such rescaling destroys the comparability of the $\mathbf{A}$ matrix across levels of Mode C. To see this, let us write the expression for a correlation matrix $\dot{\mathbf{C}}_k$. Let $\dot{\mathbf{D}}_k$ be an n by n diagonal matrix with diagonal entries equal to the reciprocal of the square roots of the diagonals of $\mathbf{C}_k$. We can write an expression for the kth correlation matrix as:

$$\dot{\mathbf{C}}_k = \dot{\mathbf{D}}_k \ \mathbf{A} \ \mathbf{D}_k \ \mathbf{W} \ \mathbf{D}_k \ \mathbf{A}' \ \dot{\mathbf{D}}_k + \dot{\mathbf{D}}_k \ \mathbf{E}_k \ \mathbf{E}_k' \ \dot{\mathbf{D}}_k \quad . \tag{5–13}$$

Thus, $\dot{\mathbf{D}}_k$ rescales the entries in $\mathbf{C}_k$ from covariances to correlations. However, this causes the factor-loading matrix $\mathbf{A}$ for each occasion to be row-rescaled. In order to analyze correlations, we would need a model with a separate $\mathbf{A}_k$ matrix for each occasion, permitting $\mathbf{A}_k = \dot{\mathbf{D}}_k \mathbf{A}$, or with special row-rescaling $\dot{\mathbf{D}}_k$ parameters in the model to modify the factor loading matrix for each occasion. None of the existing models for three-way data—including PARAFAC and Tucker's model—incorporate such parameters. They do not really need to, however, when use of covariances avoids the problem completely. It should be clear from the above, then, that correlations are not only inappropriate as input for PARAFAC—they are inappropriate for *any* currently existing three-way factor analysis procedures. Similarly, Jöreskog (1971) points out that correlations are not appropriate input for his methods of factor analysis in multiple populations.

One of the motivations for use of correlation coefficients is to remove arbitrary differences in the scale of variables by setting all variables to a constant unit mean-square. This objective can be accomplished within the context of covariance analysis by standardizing the total variance of each variable across all levels of Mode C, rather than the separate variances within each level. Thus, we would use an unsubscripted $\dot{\mathbf{D}}$ matrix in (5–13), for which the ith diagonal entry would be:

$$\dot{d}_{ii} = \left(\frac{1}{p} \sum_{k=1}^{p} c_{iik} \right)^{-1/2} .$$

In fact, this is a built-in PARAFAC option and has become a standard procedure to use for PARAFAC analysis of covariance matrices; it is called "equal average diagonal standardization."

Principal Components versus Common Factor Models

A further difference between direct and indirect fitting is that indirect fitting allows a three-way generalization of either the common factor or principal components models to be fit to the data, whereas direct fitting only allows the three-way principal components model to be fit. In this chapter, we have not stressed the distinction between the *principal components* and the *common factor* models for factor analysis. In general, we agree with Harris (1975) that from a data-analytic standpoint, the distinction is not critical, since the two different factor estimation methods usually give very similar sets of factor loadings.

From a theoretical standpoint, however, it seems clear that the principal components model ignores an obvious source of bias that is corrected by the common factor model. Consider any of our expressions for the structure underlying a covariance matrix, such as equation (5–10). The off-diagonal elements of the error covariance matrix $\mathbf{E}_k\mathbf{E}'_k$ will tend to be small and randomly distributed about zero, due to the small covariances that arise by chance among finite samples of random errors. They will not bias the solution. The diagonals, however, will be systematically larger and all positive, since each diagonal represents the covariance of an error component with itself, or, in other words, the error variance for a variable. There is a good argument, then, for ignoring the diagonal elements of a covariance data matrix, since they are likely to contain larger error components and be systematically biased upward from their "true" values (defined as the contribution of the factors before the error component is considered).

In some factor analysis procedures, it is not easy to fit off-diagonal cells and ignore the diagonal cells. In such cases, an alternative approach is sometimes used: The diagonal values are modified before the analysis to more closely represent an unbiased estimate of what they would equal if it were not for the inflating effect of the error. These communality estimates may be based on the squared multiple correlation between the variable and all the other variables in the data set or on some other estimation procedure. One of the more common approaches is to iterate on the diagonals. This method uses the results of a factor analysis to estimate the size that the diagonal cells would have if they only contained common factor variance; it replaces the diagonal with these estimates and then refactors the modified matrix. This procedure is supposed to be iterated until it converges, although tests have demonstrated that actual convergence usually takes a larger number of iterations than have commonly been applied in the past. In PARAFAC, the method used for ignoring the diagonals—and any other cells identified as containing missing data—can be viewed as a modified version of this iterative estimation procedure (see below).

In practice, the bias introduced by including the inflated diagonal values in the data set often has very little effect on the factor loadings (as Harris [1975] demonstrates). In some Monte Carlo studies (including Velicer, Peacock, and Jackson 1982), the principal components solutions have actually recovered underlying structure as well or better than common factor solutions, perhaps because of greater stability arising from the component model's fewer parameters.

Whether there will be an appreciable difference between factors determined using the principal component versus common factor models will depend on the characteristics of the data being analyzed. When one is analyzing large matrices, the diagonal cells constitute such a small proportion of the total data that the systematic bias introduced by including them in the analysis has a minimal effect on the final solution. Furthermore, unless the communality estimates are widely divergent, the pattern of loadings is not noticeably changed, even with smaller matrices (Harris 1975). Nonetheless, there are certain circumstances in which a difference between the two methods may be noticeable. If the covariance matrix is small—for instance, less than 15 or 20 variables—and if the estimated communalities are quite different from one another, so that relations among diagonal elements would be altered by communality estimation, then the common factor model might be expected to give an appreciably more accurate solution.

A method for fitting the common factor model by direct decomposition of the raw data array has been suggested for two-way data by McDonald (1979) but has not yet been generalized to the three-way case. However, when doing indirect fitting by means of covariances, the common factor model is easily fit in the three- as well as the two-way case, by simply ignoring the diagonal and/or placing communality estimates in the diagonals of the covariance matrices.

The PARAFAC computer program allows either model to be fit to covariances. If the user decides to fit the common factor model, the program takes the same approach as Harman and Jones' (1966) MINRES procedure; namely, parameters of the PARAFAC model are estimated by fitting only the off-diagonal elements of the covariance matrices. One of the program's analysis options allows the user to ignore the diagonal cells of each covariance matrix using the same technique that PARAFAC uses to ignore cells with missing data (by continuous iterative reestimation within the alternating least-squares algorithm; see Harshman 1972b). The method appears to work quite well. In fact, our experience suggests that this method may work better than more traditional iteration methods commonly used with two-way arrays, converging more rapidly to accurate estimates of the appropriate values (as tested by Monte Carlo experiments). This may in part be due to the fact that reestimation occurs repeatedly during the factor estimation phase, not just at the end. It may also be in part attributable to the stronger information provided by the three-way data.

Initial applications of the common factor model to real data problems have confirmed our theoretical expectations. With a large problem (Haan 1981), including or ignoring the diagonals made very little difference in the final **A** matrix—at most, a small change in the second decimal place of a factor loading (based on comparisons conducted in our laboratory but not reported in Haan's article). This was to be expected, since the data (which was from a longitudinal study of personality) consisted of 12 matrices, each of which was 86 by 86; thus, less than 2% of the data cells were ignored when the common factor model was estimated. Furthermore, examination of the communality estimates showed that most were of comparable size. Conversely, the common factor model provided a noticeably more interpretable solution in a three-way analysis of the WAIS and WAIS-R intelligence test standardization data (Harshman and Reddon 1983). This latter data set had those characteristics that might lead one to expect a difference; the 9 covariance matrices were small—only 11 by 11—and the communality estimates for the 11 variables turned out to differ considerably in size both within and across occasions.

The PARAFAC1 Model Applied to Multidimensional Scaling of Proximity Data

The Relationship between Factor Analysis and Metric MDS

Multidimensional Scaling (MDS) is sometimes used in a broad sense to mean any procedure that represents a large matrix of observations or relationships in terms of a small set of underlying dimensions. PARAFAC is classed as a multidimensional scaling procedure in this broad sense by Carroll and Arabie (1980). We shall use "multidimensional scaling" in a more restricted sense, however, meaning a procedure that derives a dimensional representation of relationships among entities from data on their proximity (such as pairwise similarity or dissimilarity). From this perspective, MDS is distinguished by the fact that it takes as input a type of data that is not directly appropriate for factor or component analysis: interpoint "distances," dissimilarities, or other distancelike quantities. Distances are not suitable for factor analysis because a distance cannot be directly decomposed into the additive contributions of dimensions. As the Pythagorean theorem demonstrates, it is the *squared* distances along dimensions that add together to produce the *squared* distance between two points in a space.

Nonetheless, factor analysis and multidimensional scaling are very closely related. In fact, metric multidimensional scaling can be accomplished by performing a factor (or principal component) analysis of distancelike data that has been appropriately preprocessed. The preprocessing turns dissimilarities into squared interpoint distances and then converts these into scalar products, which are essentially equivalent to covariances. Factor analysis of the scalar products then proceeds in the same way as analysis of covariance matrices. Thus, the application of factor or component analysis to MDS problems involves a further example of indirect fitting: an MDS distance model is indirectly fit to the original distancelike data by directly fitting a factor or component model to the covariancelike matrices obtained after the data is preprocessed.

The conversion of dissimilarities to scalar products has been described elsewhere (Kruskal and Wish 1978; Spence 1977; Torgerson 1958) and will therefore only be briefly recounted here. Three steps are involved for each subject: (a) an additive constant is estimated which converts that subject's dissimilarities into distances (that is, from interval-scale to ratio-scale measures of subjective distance between stimuli); (b) each entry in the distance matrix is then squared; and (c) the matrix of squared distances is then double-centered (row means are removed and then column means are removed from the residuals, so that the resulting matrix has both rows and columns summing to zero). This last step removes undesired constants that entered at the squaring stage and also adjusts the resulting scalar products so that their origin is at the centroid of the configuration of stimulus points in the space. The resulting values are all multiplied by -.5 for geometric reasons, as explained in Torgerson (1958).

When PARAFAC is used to perform three-way MDS, the mathematical model that is being indirectly fit to the data differs from those that we have discussed previously. It is called the "weighted Euclidean model." If, for example, we considered x_{ijk} to represent the distance between stimulus i and j as judged by the kth individual, the weighted Euclidean model would represent that distance in terms of a generalization of the Euclidean distance formula:

$$x_{ijk} = \left(\sum_r (w_{kr}(a_{ir} - a_{jr}))^2 \right)^{1/2} + e_{ijk} \ . \qquad (5\text{–}14)$$

In this formula, a_{ir} and a_{jr} represent the stimulus projection of the ith and jth stimuli onto dimension r, and w_{kr} represents the weight that subject k places on dimension r, or, in other words, the salience of that dimension for that subject.

Use of the PARAFAC program to indirectly fit the weighted Euclidean model is essentially equivalent to use of Carroll and Chang's (1970) INDSCAL procedure. Both PARAFAC and INDSCAL follow the same steps: First, transform the input data from similarities or dissimilarities to distances; then, convert the distances to scalar products; and finally, fit to the scalar products a trilinear model (5–3), which, because of the covariancelike structure of the data, takes the special form of (5–10). In the resulting representation, the **A** matrix gives the projections of stimuli onto dimensions, and the $\mathbf{D}_k^2$ matrix gives the squares of the dimension weights or saliences for the kth subject.

It is interesting that although INDSCAL and PARAFAC were developed independently and from somewhat different perspectives, both approaches led to the same trilinear model, based on related reasoning about systematic variations of dimensions underlying levels of a three-way array. We call (5–3) the "PARAFAC" model, alluding to Cattell's parallel proportional profiles idea, from which it was derived. But Carroll and Chang developed the model from somewhat different perspectives related to Horan's (1969) findings concerning individual differences in MDS; they call (5–3) the "CANDECOMP" (for CANonical DECOMPosition) model. Because of this parallelism, (5–3) is sometimes referred to as the "PARAFAC-CANDECOMP" (Kruskal 1981) or "CANDECOMP-PARAFAC" model (Carroll and Arabie 1980, 635).

Properties of Three-Way MDS Solutions

Advantages. When the trilinear model (5–3) is applied to scalar products in order to accomplish multidimensional scaling, the solutions have the same important intrinsic axis property as when it is applied to raw profile or covariance data to accomplish three-way factor analysis. This intrinsic axis property is perhaps the most important reason INDSCAL has been such a popular MDS procedure, with more than 75 successful applications to date. The success of INDSCAL in multidimensional scaling strengthens arguments for the potential value of the intrinsic axis property in factor analysis, as well.

However, MDS applications are in some respects easier than factor-analytic ones. For one thing, subtle preprocessing issues that arise in three-way factor analysis (which will be discussed in chapter 6) are avoided in MDS. Also, there are implied orthogonality constraints resulting from the indirect fitting of the MDS application that can increase the stability of the solution and help to prevent oblique axis problems that sometimes trouble factor-analytic applications. Finally, it appears from our accumulating experience that the patterns of dimensional variation in three-way profile data are sometimes more complex than provided for by the system variation model; when this occurs, the recovery of sensible solutions requires special constraints on the form of the loadings matrices. For all of these reasons, the general application of the trilinear PARAFAC-CANDECOMP model to profile data has been a more challenging problem than the application of this model to MDS data. Thus, it has taken longer to develop all the necessary theory, as well as preprocessing and other techniques, to make three-way factor analysis relatively trouble-free.

Orthogonality Assumptions. As we have pointed out previously (Harshman 1972b; Terbeek and Harshman 1972; see also Carroll and Wish 1974; McCallum 1976), the MDS application of the PARAFAC-CANDECOMP trilinear model (5–3) involves implicit orthogonality assumptions that are not required in two-way multidimensional scaling. Since fitting PARAFAC to scalar products is equivalent to analyzing covariances, it requires the same kind of assumption of orthogonality (discussed above). However, the interpretation of this assumption is slightly different in the case of MDS. When we analyze scalar products derived from dissimilarities, there is no "Mode *B*" over which we must assume uncorrelated variations in factor influence. Instead, the diagonal form of the matrix **W** reflects the special orthogonal form of the Euclidean distance formula. In the more general case, in which the axes in a space are not orthogonal, (5–10) will not describe the scalar products. Instead, a more general Euclidean formula must be used, thus corresponding to (5–9) (see Harshman 1972b).

There is some evidence that oblique perceptual dimensions may indeed arise when two or more properties by which people perceive a set of stimuli are strongly associated in the minds of those making the judgments. In an earlier unpublished study (Harshman 1973), subjects judging the similarity of objects differing only in size and weight appeared to treat the dimensions of size and weight as if they were oblique in their perceptual space. Tucker (1972) also obtained MDS results that suggested oblique dimensions.

Several three-way MDS models have been formulated to deal with the case of nonorthogonal perceptual dimensions. If we apply the PARAFAC2 model (5–9) to scalar product matrices, we can obtain a three-way MDS solution with oblique perceptual dimensions but involving the assumption of a common pattern of obliqueness across all subjects. This model may share the PARAFAC1 intrinsic axis property, although the question of uniqueness is more complex. (This will be pursued later in our discussion of PARAFAC2.) An even more general model would allow different patterns of obliquely related dimensions for each subject, as in the very general covariance model (5–12). This model corresponds to what Carroll and Chang (1972) call the "IDIOSCAL" model and Kroonenberg and de Leeuw (1980) call the "Tucker2" or "T2" model. A version of this model that puts restrictions on the different W_k is the Tucker application of three-mode factor analysis to multidimensional scaling (Tucker 1972). Neither IDIOSCAL, T2, nor Tucker's three-mode scaling provide intrinsic axis solutions.

Diagonal Cells in MDS. It often occurs that the distances or dissimilarities used for MDS input are obtained in such a way that the diagonal cells may not be directly comparable to the other cells or indeed may be set to zero by assumption and not even measured. For example, when subjects are asked to judge the pairwise dissimilarity of elements from some stimulus set, they are often not presented with pairs consisting of a stimulus matched with itself. Such stimuli are usually going to be judged to have infinite similarity or zero dissimilarity by an alert subject; hence, these trivial judgments are omitted to save time, even though they might serve a valuable "anchor" function in the experiment. When similarities are obtained by other means, such as stimulus confusions, it still may often be the case that the diagonal cells of the data matrices are subject to special considerations that make them not directly comparable to the off-diagonal cells.

The question thus arises: Would it be advisable to ignore the diagonal cells of MDS, much as is done when fitting the common factor model in three-way factor analysis? As in factor analysis, if the stimulus matrix is large and/or the subjective diagonal values that might be estimated by the dimensions (that is, the analogs of communality estimates) would be fairly uniform across stimuli, then the solution would probably not be noticeably affected by ignoring the diagonals. But if the stimulus set were small—as is common in MDS applications in which all pairwise comparisons are needed—and the subjective diagonals somehow divergent, then ignoring the diagonal might be useful. Exploration of the effects of such an approach would easily be accomplished by means of the PARAFAC option to ignore diagonal cells, but no systematic study of the effects of such a procedure has yet been attempted.

UNIQUENESS PROPERTIES OF PARAFAC-CANDECOMP

Why is the Intrinsic Axis Property Important?

Viewed geometrically, a factor analysis or multidimensional scaling solution provides two basic types of information: (a) a configuration of points in a low-dimensional space; and (b) a set of axes spanning that space. The configuration of points provides a compact description of the observed relationships among the things represented by the points. Such a description may clarify our insight by simplifying the patterns we are trying to understand, but it does not present us with anything essentially new. The axes, on the other hand, can potentially take us beyond observed relationships into new inferred ones. In the appropriate conditions, the projections of points onto correctly oriented axes could be taken to indicate relationships between observed surface variables and unobserved, but empirically real, latent variables. If a factor analysis can indicate which set of axes is most likely to approximate empirically real processes, it can provide us with genuinely novel information. But if the choice of a particular axis orientation is somewhat arbitrary, or if it is based on a criterion that is hard to defend empirically, the orientation of factor axes will be less informative and inferences from factor axes to theoretical constructs will seem dubious.

As noted in the introduction, PARAFAC was developed primarily to provide a stronger theoretical/empirical basis for determining orientation of factor axes. This advantage is important to scientists who would like to use factor analysis inferentially, either inductively, as an aid to development of theoretical constructs in a poorly understood domain, or deductively, as a means of testing previously hypothesized constructs with new data. The intrinsic axis property of the PARAFAC model provides a plausible theoretical basis for such inferences when the solution can be replicated across samples. In this section, we explain the intrinsic axis property, discuss its mathematical basis and its limitations, and show why it often provides stronger grounds for attributing empirical reality to factors than the grounds provided by simple structure or other commonly used criteria.

Is there a "Most Valid" Rotation?

Some investigators have claimed that the particular orientation of axes in a factor-analytic solution is not important. Different rotations of a factor-analytic solution are said to correspond to different yet equally valid perspectives on the same complex phenomenon (for example, see Thurstone 1947, 332).

Such an attitude may be appropriate when one only uses factor analysis to provide a condensed description of the particular data set. However, if one uses it to obtain clues to the empirical processes responsible for the observed patterns, then alternative rotations cannot be considered "equally valid." They will lead to competing hypotheses concerning the underlying empirical reality, and these hypotheses will, in turn, lead to different and competing predictions of what would be expected in novel nonequivalent situations. It is this difference in implied predictions that gives empirical meaning to the claim that there is a "most valid" axis orientation.

A search for the most valid rotation or orientation of factor axes can be viewed as a part of the scientist's search for theoretical constructs that will prove the most generalizable (namely, those constructs not only successful at accounting for patterns of variation within the given data set, but those able to be used in explaining patterns in new and systematically different data sets, as well). Proportional profiles rotation seeks to maximize this sort of generalizability.

The argument can also be made on less pragmatic grounds: We believe that there are particular component processes that actually exist in the empirical situation and that they are likely to be better approximated by some factorial descriptions than others. If we could identify those axes that most closely correspond to the underlying component processes (which provide the best description of them), we would know which axis orientations should be considered most valid. Because of their better correspondence with empirical reality, we expect that such axes will lead to the most successful scientific constructs or theories, theories that will make the most accurate predictions in the widest range of situations.

Consider, as one example, the classic debate concerning the dimensions of intelligence. This is in part a debate over correct orientation of axes in the factor space of intellectual tests. In one position, we obtain a general factor and a set of uncorrelated group factors corresponding to specific intellectual abilities. In another orientation, we obtain several correlated broad intellectual abilities. These alternative solutions are not just different perspectives, since they lead to different predictions concerning the effect of drugs or brain damage on intellectual performance. Empirical findings can provide support for one set of factors as against the other. In fact, Bock (1973) has argued that recent discoveries about the functional specialization of the two hemispheres (based in part on effects of brain damage on cognitive abilities) have helped to confirm those factor-analytic theories that distinguish a broad class of verbal abilities from another broad class of spatial abilities.

To take another example, consider a recent disagreement between Eysenck and Guilford over the "real factor of Extroversion-Introversion" (Eysenck 1977; Guilford 1977). Guilford (1977) points out explicitly that the disagreement hinges on choice of axis orientation. In order to resolve such disputes, appeal must be made to empirical data beyond the particular two-way factor analysis, in the hope that different predictions of the two different factor-based constructs can be somehow tested empirically.

Limitations of Traditional Criteria for Orienting Axes

Since the two-way factor model provides a family of possible solutions—all of which fit the data equally well—investigators have had to devise additional criteria external to the model to aid in selecting a "best" solution from among this family. Confirmatory factor analysis uses rotation-to-target techniques and/or allows loadings to be constrained in certain ways. Exploratory factor analysis rotates to optimize some desired characteristic of the resulting factors; usually this is an index of simplicity, although other criteria have also been employed (see Comrey 1967; Eysenck 1950). In many situations, however, the factors that optimize these criteria may or may not be the "real" ones, that is, the ones that most closely approximate the empirical processes generating the observed relationships. For example, we often do not have strong empirical reasons for preferring the particular rotation of factors that maximizes simplicity, and hence we do not have strong grounds for attributing empirical validity to constructs that the simplest factors suggest. Furthermore, controversy can arise because different investigators prefer different rotation criteria and thus obtain differently oriented factors. This occurred in the Eysenck-Guilford dispute mentioned earlier.

We have discussed the limitations of traditional factor rotation procedures elsewhere (Harshman 1970, 8–14) and will therefore only briefly summarize some objections here. First of all, the likelihood of obtaining empirically valid dimensions by theoretically guided target rotation will depend on the validity of the theory used. Such rotations can provide evidence against a theory, if the target cannot be approximated by any rotation of the obtained factors. However, these rotations usually provide only weak evidence in favor of a theory; when factors can be found that are consistent with a given target, the theory is not discredited, but other rotations of the same axes may provide factors consistent with other theories. Thus, target rotation will not usually result in the discovery of theoretically novel dimensions.

A second objection is that rotation in search of meaningful interpretations lacks objectivity. In addition, it is limited by the investigator's notion of what is "meaningful"; a dimension may be difficult to interpret because it reveals an unfamiliar truth about the domain under investigation.

A final objection to current methods of factor rotation is that the likelihood of obtaining valid dimensions by rotation to simple structure (or its analytic approximations, such as Varimax) depends on one's assurance that maximizing the simplicity criterion is appropriate for the particular situation. Often, relationships between variables and underlying factors will not be maximally simple. As Comrey notes:

> If we sample at random from the entire universe of factors and use predominately factor-pure measures, simple structure will no doubt give results which are reasonably satisfactory. In many real-life factor analyses, however, where selection of variables is anything but random, and measures of considerable factor complexity sometimes predominate, one can only hope that the simple structure criterion is approximately applicable. . . . There is no particular reason why the variance of the squared factor loadings must be maximized, except that loadings are more easily interpreted if they are high or low rather than medium in absolute value. It may well be, however, that they should be in medium range, rather than high or low, depending on the data being analyzed. (Comrey 1967, 143)

The Principle of Parallel Proportional Profiles

Cattell (1944) reviewed seven principles for choice of rotation and concluded that the "most fundamental" was one that he called the "principle of parallel proportional profiles." This principle may be thought of as an application of the method of cross-validation to the problem of factor rotation, in order to determine empirically which rotation corresponds to "real functional unities in the psychological situation." However, Cattell recognized that "to require agreement in factors and factor loadings among correlation matrices derived from the same or similar test variables on the same or similar population samples, is an empty challenge. No new source of rotation determination is introduced, for such matrices will differ only by sampling errors and there will be an infinite series of possible parallel rotations in the two or more analyses." What is needed is a method for *meaningful* validation across samples, demonstrating a consistent empirical pattern that would not be found if the factors were a mere "mathematical artifact." How could this be obtained?

> The special and novel required condition is that any two matrices should contain the same factors, but *that in the second matrix each factor should be accentuated or reduced in influence by the experimental or situational design,* so that all its loadings are proportionately changed, thereby producing, from the beginning, an actual correlation matrix different from the first. (Cattell 1944, 274 [his italics])

Cattell states the principle even more clearly in a later article:

> The basic assumption is that, *if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study.* No inorganic factor, a mere mathematical abstraction, would behave in this way. . . . This principle suggests that every factor analytic investigation should be carried out on at least two samples, under conditions differing in the *extent* to which the same psychological factors . . . might be expected to be involved. We could then anticipate finding the "true" factors by locating the unique rotational position (simultaneously in both studies) in which each factor in the first study is found to have loadings which are proportional to (or some simple function of) those in the second: that is to say, a position should be discoverable in which the factor in the second study will have a pattern which is the same as the first, but stepped up or down. (Cattell and Cattell 1955, 84 [their italics])

Although he was convinced of the potential importance of this idea, Cattell had problems implementing it as a practical rotation procedure. He developed an algebraic method for orthogonal proportional profiles rotation of two sets of factor loadings (long before a similar method was proposed by Schönemann [1972] for MDS), but was unable to formulate a solution that allowed oblique axes (Cattell and Cattell 1955). Even more discouraging, his attempt at demonstrating a practical application of the orthogonal method to real data was not successful. Finally, he realized that his expectation of proportionality across occasions was an oversimplification: The relationship between two solutions obtained by conventional factor analysis of correlation matrices would be more complicated. In 1955, he did not see how to overcome this problem (Cattell and Cattell 1955), but it was subsequently resolved by Meredith (1964), who pointed out that strict proportionality could be obtained by use of covariances rather than correlations. More recently, Cattell refers to continued work on the method, which he now calls "confactor rotation." He reports encouraging results with synthetic data, but has yet (as far as we know) to report a successful application to real data.[3]

PARAFAC overcomes the problems that plagued Cattell in several ways. First, Cattell's idea is generalized from a method of comparing separate factor analyses into a true three-mode factor-analytic model. Second, the basic model is formulated in terms of the direct fitting of profile data, and so avoids the restriction to orthogonal axes implied by the indirect fitting case. Finally, the indirect fitting case is formulated in terms of summed cross-product or covariance matrices, overcoming the problem of nonproportionality, which occurs with correlation matrices. Also, there is an oblique axis generalization for indirect fitting (PARAFAC2; see Harshman 1972b), which has been applied successfully to at least one data set (Terbeek and Harshman 1972). Despite these modifications, however, the basic idea behind PARAFAC is still the insight by Cattell that a search for proportional profiles across several nonequivalent two-way data sets would determine the most empirically meaningful factor axes. Thus, the recent success of PARAFAC with many kinds of real data is a confirmation of what he has been proposing for more than thirty years.

Nature of the Intrinsic Axis Property

What is Unique?

Suppose we perform a PARAFAC-CANDECOMP three-way factor analysis (or multidimensional scaling) of a set of matrices $\{\mathbf{X}_k\}$. As described previously, we obtain a solution of the form

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{D}_k\,\mathbf{B}' + \mathbf{E}_k\,. \tag{5–15}$$

We now ask: Which characteristics of this representation are unique? That is, among the alternative representations (of the same structural form in the same number of dimensions) that would fit the data as well, which characteristics must remain the same, and which characteristics are free to vary?

It has been proven elsewhere (by Jennrich, reported in Harshman 1970; and by Harshman 1972a; Kruskal 1976, 1977) that with "adequate" data, the only alternative solutions involve changing the order of columns and/or stepping all the entries in a given column up or down by a constant multiplier. ("Adequate" data will be defined below.) To express this algebraically, let an alternative solution be represented as

$$\mathbf{X}_k = \overset{*}{\mathbf{A}}\,\overset{*}{\mathbf{D}}_k\,\overset{*}{\mathbf{B}}{}' + \mathbf{E}_k\,, \tag{5–16}$$

where $\overset{*}{\mathbf{A}}$ is an alternative version of the factor-loading matrix $\mathbf{A}$, and similarly $\overset{*}{\mathbf{B}}$ and $\overset{*}{\mathbf{D}}_k$ are alternative versions of $\mathbf{B}$ and $\mathbf{D}_k$, respectively. The uniqueness or intrinsic axis property of PARAFAC-CANDECOMP insures that $\overset{*}{\mathbf{A}}$ can only differ from $\mathbf{A}$ by a rearrangement of columns (which can be represented by postmultiplying $\mathbf{A}$ by a permutation matrix $\mathbf{P}$) and/or a multiplicative rescaling of its columns (which can be represented by postmultiplication of $\mathbf{A}$ by a diagonal matrix $\overset{\circ}{\mathbf{D}}_a$), which is compensated for by an inverse rescaling of the columns of $\mathbf{B}$ and/or $\mathbf{C}$. Similar statements hold for the $\overset{*}{\mathbf{B}}$ and $\overset{*}{\mathbf{C}}$ matrices. (The permutation matrix $\mathbf{P}$ must both pre- and postmultiply $\mathbf{D}_k$ so that while the order of the diagonal elements is changed, it remains diagonal.) Thus, the alternative $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}_k$ matrices must be related to the corresponding "original" versions of these matrices as follows:

$$\overset{*}{\mathbf{A}} = \mathbf{A}\,\overset{\circ}{\mathbf{D}}_a\,\mathbf{P}\ , \quad \overset{*}{\mathbf{B}} = \mathbf{B}\,\overset{\circ}{\mathbf{D}}_b\,\mathbf{P}\ , \quad \overset{*}{\mathbf{D}}_k = \mathbf{P}'\,\mathbf{D}_k\,\overset{\circ}{\mathbf{D}}_c\,\mathbf{P}\ , \tag{5–17}$$

and since the diagonal of $\overset{*}{\mathbf{D}}_k$ is the kth row of $\overset{*}{\mathbf{C}}$, we can also write:

$$\overset{*}{\mathbf{C}} = \mathbf{C}\,\overset{\circ}{\mathbf{D}}_c\,\mathbf{P}\ . \tag{5–18}$$

In addition, since the effects of any internal rescalings (described by the $\overset{\circ}{\mathbf{D}}$ matrices) must "cancel out," the $\overset{\circ}{\mathbf{D}}$ must conform to the requirement that

$$\overset{\circ}{\mathbf{D}}_a\,\overset{\circ}{\mathbf{D}}_b\,\overset{\circ}{\mathbf{D}}_c = \mathbf{I}\ . \tag{5–19}$$

These indeterminacies can be considered trivial because they do not affect the interpretation of the solution. Geometrically, the effect of the permutation matrix $\mathbf{P}$ is simply to "renumber" the axes without moving them, and the effect of the rescaling $\overset{\circ}{\mathbf{D}}_a$ and so forth is simply to stretch or contract the axis in one space, with compensatory contraction or stretch of the corresponding axis in the space of another mode. None of these changes affect the *orientation* of the axes in any of the three spaces. Since it is the orientation of axes that determines their meaning (by determining the pattern of loadings or projections of points onto the axes), the intrinsic axis property insures that the characteristics crucial for *interpretation* of dimensions are uniquely determined as a consequence of simply minimizing the error of fit (provided, as mentioned earlier, that the data are "adequate").

Although they do not affect interpretation, the indeterminacies can be distracting when comparing different solutions; thus, they are often removed by adopting simple conventions. The convention of ordering the factors from the largest to the smallest variance-accounted-for will generally fix the columnar order, and the convention of scaling two of the three loading matrices so that each column has a mean-square of 1.0 will determine two of the three $\overset{\circ}{\mathbf{D}}$ matrices in (5–19). (Motivation for this second convention will be provided below when PARAFAC is compared to traditional two-way factor analysis.) The third $\overset{\circ}{\mathbf{D}}$ matrix will then be determined by the requirement that the triple product of all three internal rescalings be the identity matrix. To fix sign patterns, dimensions in Modes *C* and *A* are reflected (multiplied by -1) when needed so that their mean cubed value is positive, with compensatory reflection(s) applied to Mode *B*.

Why is it Unique?

There are several perspectives from which the PARAFAC-CANDECOMP intrinsic axis property can be explained: (a) in terms of the algebra; (b) in terms of the geometry; (c) in terms of the added information used; and (d) as a function of "implicit constraints" imposed.

Algebraic Perspective. The proofs of uniqueness (Jennrich, in Harshman 1970; Harshman 1972a; Kruskal 1976, 1977) are too long to recount here, but with the algebraic notation defined above, we can quickly convey some idea of what is behind the uniqueness, based on the approach taken in the Harshman (1972a) proof.

First, let us recall the algebraic basis for the indeterminacy in the two-way case. The rotation problem arises because we can start with any given factorial representation

$$\mathbf{X} = \mathbf{A}\,\mathbf{B}' + \mathbf{E} \tag{5–20}$$

and generate an alternative representation by applying an arbitrary nonsingular linear transformation $\mathbf{T}$ to the loading matrix $\mathbf{A}$, and the compensatory transformation $\mathbf{T}'^{-1}$ to the "factor score" matrix $\mathbf{B}$, as follows:

$$\overset{*}{\mathbf{A}} = \mathbf{A}\,\mathbf{T}, \qquad \overset{*}{\mathbf{B}} = \mathbf{B}\,\mathbf{T}'^{-1}\,. \tag{5–21}$$

In the q-factor case, $\mathbf{A}$ is an n by q matrix whose columns give the projections of the n variables onto the q factor axes. $\mathbf{T}$ is a nonsingular q by q transformation matrix, the columns of which give the projections of the original axes of $\mathbf{A}$ onto the new axes of $\overset{*}{\mathbf{A}}$, and so $\overset{*}{\mathbf{A}}$ is an n by q matrix whose columns give the projections of the variables onto the new axes.

It is easy to show that the resulting alternative representation of $\mathbf{X}$, that is,

$$\mathbf{X} = \overset{*}{\mathbf{A}}\,\overset{*}{\mathbf{B}}{}' + \mathbf{E}\,, \tag{5–22}$$

is equivalent to our original representation (5–20). Simply substitute from (5–21) into (5–22) to obtain:

$$\begin{aligned} \mathbf{X} &= (\mathbf{A}\,\mathbf{T})\,(\mathbf{B}\,\mathbf{T}'^{-1})' + \mathbf{E} = \mathbf{A}(\mathbf{T}\,\mathbf{T}^{-1})\mathbf{B}' + \mathbf{E} \\ &= \mathbf{A}\,\mathbf{B}' + \mathbf{E}\,. \end{aligned} \tag{5–23}$$

In the three-way case, we can use a similar approach to examine the question of permissible transformations of the $\mathbf{A}$ matrices. Let us consider two alternative representations of a given set of $\mathbf{X}_k$ matrices, such as (5–15) and (5–16), again defining the transformation generating the second set of factor loadings and factor scores in general fashion, as in (5–21). (We will allow our alternative occasion-weight matrices, $\overset{*}{\mathbf{D}}_k$, to be any nonsingular diagonal matrices.)

By substituting from (5–21) into (5–16), we obtain the expression for $\mathbf{X}$ in terms of our original $\mathbf{A}$ and $\mathbf{B}$ matrices, and then determine the conditions under which it is equivalent to (5–15). The substitution of (5–21) into (5–16) yields

$$\mathbf{X}_k = (\mathbf{A}\,\mathbf{T})\,\overset{*}{\mathbf{D}}_k(\mathbf{B}\,\mathbf{T}'^{-1})' + \mathbf{E}\ ,$$

which, in general, simplifies to

$$\mathbf{X}_k = \mathbf{A}(\,\mathbf{T}\,\overset{*}{\mathbf{D}}_k\,\mathbf{T}^{-1})\mathbf{B}' + \mathbf{E}\ .$$

If we define

$$\mathbf{H}_k = \mathbf{T}\,\overset{*}{\mathbf{D}}_k\,\mathbf{T}^{-1}\ , \tag{5–24}$$

our expression becomes

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{H}_k\mathbf{B}' + \mathbf{E}\ . \tag{5–25}$$

The permissible $\mathbf{T}$ matrices—that is, the ones that produce alternative PARAFAC representations—are restricted to those for which $\mathbf{H}_k$ is diagonal; otherwise, (5–25) will not have the appropriate PARAFAC form. Furthermore, $\mathbf{H}_k$ must equal $\mathbf{D}_k$ if (5–25) is to be valid.

Unlike the two-way case, we find that not all axis transformations $\mathbf{T}$ will work. If $\mathbf{T}$ is a general orthogonal or oblique transformation, then $\mathbf{H}_k$ will not in general remain diagonal for different $\overset{*}{\mathbf{D}}_k$, as required by the PARAFAC model. In fact, it can be proven that this will occur only if $\mathbf{T}$ is a diagonal or a permutation matrix, or some product of these (Harshman 1972a). This corresponds to the case in which $\mathbf{T} = \mathring{\mathbf{D}}_a\mathbf{P}$, so that the relationship between $\mathbf{A}$ and $\overset{*}{\mathbf{A}}$ is as stated in (5–17).

It is easy to verify that when $\mathbf{T}$ has this form, the $\mathbf{X}_k$ represented by the alternative solution is equal to the original $\mathbf{X}_k$. By substituting from (5–17) into the right-hand side of (5–16), we obtain:

$$\overset{*}{\mathbf{A}}\ \overset{*}{\mathbf{D}}_k\ \overset{*}{\mathbf{B}}' = (\mathbf{A}\,\mathring{\mathbf{D}}_a\mathbf{P})(\mathbf{P}'\,\mathbf{D}_k\,\mathring{\mathbf{D}}_c\,\mathbf{P})(\mathbf{B}\,\mathring{\mathbf{D}}_b\mathbf{P})' \tag{5–26}$$

$$= \mathbf{A}\,\mathring{\mathbf{D}}_a\,(\mathbf{P}\,\mathbf{P}')(\,\mathbf{D}_k\mathring{\mathbf{D}}_c)(\mathbf{P}\,\mathbf{P}')(\,\mathring{\mathbf{D}}_b\mathbf{B}')\ . \tag{5–27}$$

And because $\mathbf{P}\,\mathbf{P}' = \mathbf{I}$ for any permutation matrix $\mathbf{P}$, these terms in (5–27) vanish. Further, since the diagonal matrices commute, $\mathring{\mathbf{D}}_a\mathbf{D}_k = \mathbf{D}_k\mathring{\mathbf{D}}_a$, and by using (5–19), we obtain:

$$\overset{*}{\mathbf{A}}\,\overset{*}{\mathbf{D}}_k\,\overset{*}{\mathbf{B}} = \mathbf{A}\,\mathbf{D}_k\,(\,\mathring{\mathbf{D}}_a\mathring{\mathbf{D}}_c\mathring{\mathbf{D}}_b\,)\mathbf{B}' = \mathbf{A}\,\mathbf{D}_k\,\mathbf{B}' = \mathbf{X}_k\ . \tag{5–28}$$

If the foregoing discussion seems involved, it might be useful to consider a simple algebraic analogy that makes the uniqueness seem less surprising and which will also be useful when we consider data "adequacy" below. PARAFAC-CANDECOMP performs what might be thought of as the analog of using simultaneous equations to obtain a unique solution: It performs simultaneous factor analysis of several two-way matrices. While the equations underlying a single two-way matrix are not sufficiently constrained to provide a unique solution, the PARAFAC-CANDECOMP equations used for the simultaneous factoring of a set of two-way arrays do constitute a sufficiently constrained set of relations to provide a unique solution, when the data are "adequate." To borrow the example given in Harshman, Ladefoged, and Goldstein (1977), the equation $x + y = 20$ has no unique solution, since an infinite number of x, y pairs will satisfy the weak constraints imposed by the specified relationship. Similarly, the matrix equation $\mathbf{A B}' = \mathbf{X}$ has no unique solution. However, if we require that our values of x and y also satisfy a second equation, such as $2x + 3y = 55$, then a unique solution is obtained. Likewise, if we require that the same loading matrix $\mathbf{A}$ satisfy two different equations, $\mathbf{A D}_1 \mathbf{B}' = \mathbf{X}_1$ and $\mathbf{A D}_2 \mathbf{B}' = \mathbf{X}_2$, then with adequate data the $\mathbf{A}$ (and $\mathbf{B}$, $\mathbf{D}_k$ matrices) become determined uniquely, except for the trivial indeterminacies of columnar order and scale noted above.

Geometric Perspective. Suppose we plotted the point configurations for each of the two-way slices $\mathbf{X}_k$ of our three-way array, but left out any axes. (These plots are easiest to visualize in a two-dimensional case, so we might imagine that our plots are based on the first two principal components of each slice.) When the PARAFAC-CANDECOMP model is appropriate, we would observe a very similar configuration of points across the various slices. However, there would be certain systematic differences between the plots. When we compared one plot to another, all the points on one might be displaced outward from the origin along a certain line by an amount proportional to their distance from the origin along that line. This would correspond to a "stretching" of the space in a certain direction. In other directions, the space might be "contracted." By comparing the several spaces, one can detect such stretching and contraction and identify the directions along which the space is recurrently stretched or contracted.

Now, to explain these systematic variations in the configurations in terms of simple variations of a common set of factors, we must select an orientation of factor axes in the spaces that "line up with" the directions of stretch or contraction. Then, we can account for the systematic displacements of the point locations (the "stretches" or the configurations) as simply due to the increase or decrease in the importance of particular factors. This geometric interpretation bears a simple relation to the algebraic representation: The $\mathbf{A}$ (or $\mathbf{B}$) matrix would describe the common pattern of projections of points onto factor axes, and the $\mathbf{D}_k$ matrix would give the stretches or contractions in the kth space; that is, the rth diagonal cell of the $\mathbf{D}_k$ matrix would describe the proportion by which factor r was stretched on occasion k.

When one imagines looking at plots of two or more such spaces, each resembling the others but stretched or contracted in certain directions, it is easy to see how lining up the factor axes with the directions of stretch will force a particular unique orientation for the axes. Furthermore, to the extent that the axes deviate from proper alignment with these directions of stretch, the resulting factor solution will deviate from reproducing the actual stretches in the configurations across different slices of the three-way array, and so the goodness-of-fit will drop from its maximum value. Thus, by simply maximizing the fit of the PARA-

FAC-CANDECOMP model, one necessarily determines the axis orientation, without any recourse to outside criteria such as simple structure.

Use of Richer Information. From both the geometric and algebraic explanations, it becomes apparent that the PARAFAC-CANDECOMP model takes advantage of the extra information present in a three-way array to determine the "true" factor axis orientations. It is not surprising that proper use of this extra information can resolve the rotational indeterminacy of two-way factor analysis. Yet other three-way models, with access to the same type of information, do not provide a unique orientation for the factor axes because they do not use the information in the same way. Tucker's three-way model, for example, allows representations in which the axes can be oriented in any position, and the Mode *A* axes can be oriented independently of the Mode *B* axes, which in turn can be oriented independently of those in Mode *C*. The model compensates for the effects of these variations by adopting a more complex representation of the relationships among dimensions. Not only are there contributions from the Mode *A* loadings of dimension one multiplied by the Mode *B* and Mode *C* loadings for that dimension, but also contributions from the Mode *A* loadings of dimension one multiplied by the Mode *B* loadings of dimension two, times the Mode *C* loadings of dimension three, and so on for all possible combinations of dimensions in the different modes. This model produces a family of more complex representations that can generate the same patterns of changes across the three-mode data set as the PARAFAC-CANDECOMP model (and which can generate more complex patterns, as well). However, when the Tucker and PARAFAC models have approximately the same goodness-of-fit, the PARAFAC representation would probably be preferred because of its simplicity and straightforward empirical interpretation. In fact, some implementations of Tucker's model have methods of transforming the solution so as to approximate PARAFAC-CANDECOMP-type representations (see Kroonenberg and de Leeuw 1980).

Use of "Implicit Constraints." It might be argued that fitting the PARAFAC-CANDECOMP model is equivalent to imposing extra constraints on the more general Tucker-type of representation. (The exact form of the "constraints" that make the two models equivalent will be discussed below.) From this perspective, three-way analysis with a model "constrained" to optimize proportional profile form is like two-way analysis with the solution "constrained" to optimize simple structure or some other special form. Using this analogy, one might argue that it is these extra constraints of factor proportionality across levels of Mode *C* that remove the rotational indeterminacy, just as "simple structure" constraints remove the indeterminacy in the two-way case.

Even if one adopts this perspective, however, it is interesting to note that the PARAFAC "constraints" do not specify anything about the *form* of the factor-loading matrices **A** , **B** , or **C** . The factors may be simple or complex in structure, and they may be orthogonal or oblique (in the direct-fit case). PARAFAC-CANDECOMP does not specify preferred loading relationships *within* a mode, as simple structure does, but rather constrains the way a factor varies *across* different slices of the three-way array (ie. $\mathbf{A}_k = \mathbf{A}\mathbf{D}_k$). We will argue below that the PARAFAC-CANDECOMP-type of "constraint" differs from other two- and three-way constraints in that it is intrinsic to a plausible model of factor variation, and that comparison of split-half solutions can be used to test its appropriateness. Furthermore, if the model is found to be appropriate, this result can be used to provide empirical evidence in support of particular theoretical constructs being inferred from, or tested by means of, the given solution.

When is it Unique?

Not all sets of three-way data contain the information necessary to uniquely determine the orientation of factor axes; the data set must be "adequate." If we think of three-way data as composed of several two-way "slices," then the conditions of data adequacy can be briefly summarized as follows: The slices must involve the same set of factors *and* differ in the relative importance or contribution of these factors.

Same Factors across Slices. All current three-way factor analysis models (with the exception of Corballis and Traub 1970; and Swaminathan, chapter 8) assume that there is a common set of factors that generates data at all different levels of the three-way array. The models only differ in terms of how the weighting or pattern of combination of the factors is allowed to vary across the different levels in order to account for systematic differences between successive two-way arrays.

This assumption of a common set of factors may not always be strictly appropriate. In a long-term longitudinal study, for example, the pattern of loadings on a given factor may shift somewhat, because the meaning of test or personality items may not be the same when administered to children as when administered to adults, even if the underlying trait or process is the same (Harshman and Berenbaum 1981). It is also possible that the empirical entity represented by a particular factor will change somewhat in "quality" or nature across time.

To take another example, suppose we are analyzing an array of stimulus ratings, where each of several stimuli are rated on each of several scales by several raters. The cognitive or perceptual dimensions underlying the ratings may not have exactly the same quality from one rater to another, even if they are constant within a given rater across stimuli and rating scales. Suppose two raters both use the dimension of "valuable-worthless." One rater may think of the dimension in a slightly different way than another, which would imply that one person's pattern of factor loadings might differ from the other by more than a simple proportional reweighting.

Nonetheless, the assumption of common factors but different weighting or combination rules across the levels of the array is a useful approximation that greatly reduces the number of parameters in the solution and usually leads to an interpretable summary account of what is going on. (For a successful application of PARAFAC to lifespan longitudinal data, see Haan 1981.) Indeed, the same kind of common factor assumption is basic to the use of two-way factor analysis or multidimensional scaling.

The PARAFAC-CANDECOMP model actually allows considerable flexibility in the dimensional representation across levels, because it does not insist that every factor be involved at every level of the data. It is possible that some factors have zero loadings on some variables, or on some people, or on some occasions. In fact, one can imagine a complete change in the factor structure of the slices over successive occasions, as one set of factors comes to take on zero loadings and another begins to take on nonzero loadings. The only "sameness" that must be assumed is the qualitative sameness of a given factor across slices; that is, if a factor is present on two or more occasions, it is assumed to have the same pattern of factor loadings on those occasions, except that it may be stepped up or down by the occasion weights.

Differences in Factor Importance across Slices. As Cattell noted in the quote cited earlier, comparison of two-way arrays that are merely replications of one another—differing only by random sampling—will provide no added information over a single two-way array; hence, it will not reduce the indeterminacy of the solution. There must be systematic differences between at least some of the slices of the three-way array. In particular, the slices must differ in terms of relationships among variables, and these differences must be attributable to shifts in the relative importance of underlying factors. Furthermore, the pattern of shifts must be distinct for each factor; if two factors shift in importance in the same way across levels, they will not be uniquely determined. In other words, each factor that is to be uniquely recovered must have a pattern of effects in each mode that is distinct from all other factors.

We can gain an intuitive understanding of the necessary conditions for uniqueness by referring back to the same algebraic and geometric analogies that we employed earlier.

Algebraic Interpretation. We have compared the simultaneous factorization of several two-way arrays to the solution of simultaneous equations. As a simple example, we took the equations $x + y = 20$ and $2x + 3y = 55$; in this case, there was only one value of x and y that would satisfy both equations (namely, $x = 5$, $y = 15$). However, if our second equation were $2x + 2y = 40$, there would be no unique solution. In order for x and y to be determined uniquely, the ratio of the x coefficients in the two equations must be different from the ratio of the y coefficients. Similarly, for two factors to be distinguished and their loadings determined uniquely, there must be at least two levels of Mode C for which the ratio of the coefficients for the first factor differs from the ratio for the second factor. If two factors always change by the same percentage across levels of Mode C, then in the solution these two factors will not be uniquely determined.

Because all three modes have the same status in the basic PARAFAC model (5–3), the requirement that factors have distinct patterns of variation applies in each mode. If two factors have proportional loadings in any one of the three modes, but show distinct loading patterns in the other two, then there will be a family of possible solutions in which different linear combinations of these two factors occur (for example, see Harshman 1970, 41). (We note in passing that if two factors have proportional loadings in *two* modes, they cease to be distinct factors; their contribution is represented by a single factor that has loadings proportional to theirs in the two modes where they were proportional, and has loadings equal to a weighted sum of their loadings in the third mode. The same thing happens if the two factors have proportional loadings in all three modes.)

Geometric Interpretation. As we pointed out earlier, the orientation of factor axes can be established geometrically by comparing the configuration of points derived from different two-way slices and determining the directions along which one configuration of points is stretched relative to another. But such comparisons will only be informative if the factors are stretched by different amounts. If two axes were each stretched by the same amount, we would observe only a uniform dilation of the plane in which those axes lie; all interpoint distances—that is, their projections in that plane—would be increased to the same degree. Since such a uniform dilation would be produced by an equal stretch of any set of axes in that plane, regardless of orientation, we could not determine the correct orientation of the two axes in that plane from a comparison of these two configurations. In order to distinguish and uniquely orient these two factors, we would have to compare other spaces (using other slices of our three-way array) until we found two spaces for which the factors in that plane did not increase or decrease by the same proportion.

Partial Uniqueness. The geometric interpretation of the necessary conditions for uniqueness permits us to easily grasp how the uniqueness need not break down entirely if the conditions are not met for all factors. Those factors that have distinct patterns of changes across three modes will be uniquely determined, but those that do not will show rotational indeterminacy within the subspace that they span. For a concrete example, consider the comparison of two three-dimensional spaces in which factors 1 and 2 each increase by 50%, but factor 3 increases by 75%. The second space will look like one that is uniformly expanded by 50% and then stretched by an additional 25% along one direction. This extra stretch will allow us to identify the axis orientation for factor 3. But the plane in which factors 1 and 2 reside is uniformly expanded, and so the axes can be rotated to any arbitrary position within this plane and still be able to reproduce the relationship between the two spaces.

Theoretically, a partial breakdown of uniqueness should only create ambiguities of interpretation within the subspace(s) spanned by the inadequately distinct factors. In practice, however, if major factors are nonunique, it may seem as though no interpretable solution can be obtained with a given data set. This is because PARAFAC solutions are usually obtained in low dimensionalities first, to determine if any sensible dimensions can be extracted, before proceeding to higher dimensionalities. If nonunique factors account for a substantial proportion of the variance, these factors may emerge in the lower-dimensional solutions, producing uninterpretable results that discourage attempts at further analysis.

When several factors are nonunique, it is commonly because they do not change across one of the three modes. Such factors can be removed from the data by appropriate PARAFAC preprocessing (by centering across the mode in which they are constant). This allows the remaining factors to be analyzed successfully.

Minimum Conditions for Uniqueness. Successive proofs of uniqueness have established successively less stringent requirements for determination of unique axes. Jennrich's original proof (in Harshman 1970) showed that q factors could be determined uniquely if the **A**, **B**, and **C** matrices all have rank q. But empirical results (Harshman 1970, 39–44) suggested that this condition was stronger than necessary. Also, Cattell and Cattell (1955) had presented an algorithm for orthogonal rotation of two factor-pattern matrices to proportional profiles that would provide a solution for any number of factors. In 1972, a PARAFAC uniqueness proof was formulated that showed that if the factor-loading matrices for two modes had rank q, the third mode need only have two levels (and hence a maximum rank of 2), even if there were more than two factors in the solution (Harshman 1972a). Uniqueness was obtained so long as each factor had a distinct ratio of change across the two levels. Although it was not stressed at the time, the proof also implied that even easier conditions obtained when there were more than two levels. In that case, not all factors need show distinct amounts of change across any particular pair of levels; some factors might be distinguished by means of distinct changes between levels 1 and 2, others by distinct changes between levels 2 and 4, and so on.

Some surprising uniqueness properties of three-way arrays have been discovered empirically, one of which was reported in the original PARAFAC monograph (Harshman 1970, 44). It was found that more factors could be uniquely determined by a three-way array than would seem possible from a conventional two-way perspective. In particular, it was found that more factors could be determined uniquely than there were levels in any of the three modes of the data array! In the example cited, 10 factors with random loadings were used to construct an $8 \times 8 \times 8$ data array, and PARAFAC correctly and uniquely recovered these 10 factors upon analysis of the array. Since no two-way slice of such an array could have more than rank 8, the unique recovery of 10 factors showed the remarkable degree to which the extra richness of relationships in a three-way array permits recovery of information that would not be possible in a two-way array.

Kruskal, intrigued by this empirical finding, undertook a rigorous investigation of the mathematical issues of uniqueness and generalized rank of three-way arrays. His results (Kruskal 1976, 1977) constitute by far the most complete treatment of this question to date. Among other things, he shows mathematically why 10 and even 11 PARAFAC factors can be uniquely determined by an $8 \times 8 \times 8$ array. He proves a number of important theorems concerning uniqueness and generalized rank. However, in this chapter, we will sketch only one of his basic results concerning uniqueness.

To lay the groundwork for Kruskal's result, we first restate a familiar fact concerning rank and then, in parallel terms, explain a stronger property related to rank. Consider the $n \times q$ matrix **A**. The rank of **A** is equal to r_a if the columns are linearly independent in at least one set of r_a columns from **A**, and if there is no set of $(r_a + 1)$ columns from **A** that consists of linearly independent columns. Now let us consider a stronger property, which Kruskal defines but does not name, but which we shall call *k*-rank ("Kruskal-rank"). **A** has *k*-rank of k_a if the columns are linearly independent in every set of k_a columns from **A**, and if there is at least one set of $(k_a + 1)$ columns of **A** that includes linearly dependent columns. For example, if **A** has six columns, all of which are linearly independent except one (which is a linear combination of three others), then **A** has rank of five but *k*-rank of three.

A PARAFAC solution consists of three factor-loading matrices, each with its corresponding *k*-rank. We have used k_a to represent the *k*-rank of **A**; in a similar fashion, we can let k_b represent the *k*-rank of **B** and k_c the *k*-rank of **C**. Kruskal shows that if the factor-loading matrices **A**, **B**, and **C** provide an exact *q*-factor PARAFAC representation of the three-way array **X**, this representation is unique (up to trivial permutations and rescalings, as discussed earlier) whenever

$$(k_a + k_b + k_c) \geqq (2q + 2) \ . \tag{5–29}$$

(Most often, what we called **X** will in fact be $\hat{\mathbf{X}}$, the *q*-dimensional fitted part of a data array of higher dimensionality.)

Not only does Kruskal's result prove that it is possible to extract more factors than variables and still obtain a unique solution, but it also confirms another surprising property that was indicated by earlier results obtained with synthetic data: PARAFAC can uniquely recover factors that are linearly dependent on other factors. In unpublished research, we have found that if a factor is a linear combination of three other factors in two modes, but has linearly independent patterns of change in a third mode, this factor can be uniquely recovered by PARAFAC analysis. However, if it is only a linear combination of two other factors, then it will not be uniquely recovered. While Kruskal's articles do not make this consequence explicit, his result implies that this will happen and further implies an even stronger result. It follows from Kruskal's theorems that a factor can be a linear combination of other factors in *all three modes* and still be uniquely recovered, provided that it is a combination of a sufficient number of other factors. (A factor that is a linear combination of three other factors in all three modes cannot be uniquely recovered, but one that is a linear combination of four other factors can be uniquely recovered.) In further synthetic data tests, we have recently verified that uniqueness holds under these conditions as well.

Aside from their mathematical interest, such results have potential bearing on the question of "higher-order" factors, which are sometimes thought of as linear combinations of "lower-order" factors. They are also relevant to the question of an oblique axis generalization of the indirect fitting model (PARAFAC2), since data computed from oblique axes can be reformulated as data generated by orthogonal axes plus appropriate linearly dependent extra dimensions (Harshman 1973). The surprising strength of the intrinsic axis property suggests that all the interesting characteristics and potentials of PARAFAC-CANDECOMP trilinear representation of three-way arrays have yet to be uncovered.

Empirical/Inferential Significance of the Intrinsic Axis Property

We have argued that the intrinsic axis property of PARAFAC-CANDECOMP gives the factors obtained by this method a special empirical significance. But some authors have questioned this, pointing out that other matrix decomposition methods, such as principal components, can also be interpreted as providing a unique set of dimensions (for instance, see Schönemann 1972). In fact, even two-way factor extraction followed by analytic rotation (using Varimax or some similar criterion) can be interpreted as a two-step procedure that provides a unique solution. Why, then, should the "unique axes" provided by PARAFAC-CANDECOMP be any more meaningful than the "unique axes" provided by these other methods? We suggest that there are two basic reasons: (a) empirical plausibility; and (b) empirical confirmability.

Plausibility of Assumptions

The first basic consideration is the empirical plausibility of the assumptions required to obtain a unique solution. With intrinsic axis methods, one makes no additional assumptions beyond the appropriateness of a particular three-way factor model. Since for this model different axis orientations yield different fit values, the "correct" orientation of factor axes is established as an intrinsic part of fitting the model to data. With the other methods, in addition to the plausibility of the factor model, one must also consider the plausibility of the additional assumptions involved in selecting axis orientation; as noted earlier, these assumptions have little or no empirical rationale in many applications.

In principal components analyses, for example, the axis locations are fixed by the additional requirement that the first dimension explain the maximum amount of variance possible and similarly that each successive dimension explain as much remaining variance as possible. But if there are several factors underlying a given data set, maximizing the variance explained by the first dimension usually makes the first dimension represent a *combination* of the underlying influences operating in the situation, rather than any one of them. It is not likely that real influences or processes will correspond to axes which have the successive variance-maximizing characteristics assumed by principal components analysis. Likewise, as noted earlier, the assumptions underlying simple structure rotation are often debatable, particularly in the many common applications of factor analysis in which the variables are not selected with an eye to eventual factor rotation but are determined by other characteristics of the study. With such data, the variables are likely to be factorially complex and concentrated in a particular domain in a way inconsistent with simple structure.

It would be misleading, however, to say that PARAFAC involves no new assumptions. The assumptions that determine axis orientation are external to the factor model in the two-way case and intrinsic to it in the PARAFAC-CANDECOMP three-way case. As was pointed out earlier, the direct application of PARAFAC to profile data requires the assumption of system variation, although indirect fitting by means of covariance analysis allows weaker assumptions. We argue that these assumptions are "natural" or empirically plausible for many types of three-way data (so much so that they constitute a straightforward generalization of the factor model to three-way arrays). They are also, in a certain sense, less restricting; there are no assumptions about the pattern of factor loadings *within* a given occasion but only assumptions about the form of factor variation *between* occasions. Thus, the factors can have simple or complex structure and can be orthogonal or oblique (in the profile data case), whichever best fits the data.

Confirmability

Since PARAFAC-CANDECOMP is not a completely general three-way model, we need to make an argument for its appropriateness when applying it to a given data set. When a PARAFAC solution is to be used as a starting point for scientific inferences, this argument needs to be particularly persuasive. Hopefully, it would consist of more than pointing out the conceptual plausibility of the model in a particular application. It would consist of empirical *evidence* that two essential preconditions are fulfilled: (a) the PARAFAC-CANDECOMP model is appropriate for these data (at least as a first approximation); and (b) the pattern of factor variation within the data is adequate to determine reliably the orientation of axes. Only then can the orientation of axes in the solution be taken as substantial evidence for or against specific empirical hypotheses. (In the same sense, factors determined by rotation of a two-way analysis to some target orientation or to the position that maximizes some "simplicity" function do not provide substantial evidence for any empirical hypothesis until one is persuaded of the likely appropriateness of the rotation criterion.)

Fortunately, one can test whether the two essential preconditions are fulfilled. It is an important consequence of the intrinsic axis property that analyses can be performed in such a way as to provide *confirmatory evidence* for both the appropriateness (or partial appropriateness) of the model and the adequacy of the data to determine a unique axis orientation. As far as we are aware, it is not generally possible to obtain comparable internal evidence for the validity of axes when their orientation is determined by other methods, such as simple structure or principal components.

Empirical Tests of PARAFAC Axes. The simplest and most common approach to testing a given PARAFAC solution has been to use a split-half or double-split-half technique. Suppose we wish to demonstrate that the pattern of factor variation across all three modes of a given data set is appropriate and adequate to determine a stable orientation of factor axes. If we split our data set into two halves (for example, by randomly assigning the subjects to two different subsamples) and perform a separate PARAFAC analysis of each half, we can compare the solutions to assess whether or not the necessary pattern of three-way variation was present in the data. If the factors did not show adequate proportional changes in relative importance across levels of Mode *C* (as well as *A* and *B*), then the same set of axes will not be obtained in the two solutions. Instead, arbitrary orientations of factor axes will be obtained, and these will differ from one split-half to the other. If certain factors are well determined, but other are not, this too will be revealed. *Thus, the observation of similar factors in two different split-halves provides confirmatory evidence that the data contains enough systematic variance of the appropriate kind to determine a unique, stable set of factor axes.*

The split-half procedure is an approximate test of stability, but the results obtained will depend to some extent on the particular random split that is used for the test. More sophisticated procedures allow one to minimize this dependence on a particular split. The simplest approach is to use a double-split-half procedure. "Orthogonal" splits can be generated by dividing the data into four subsamples—call these S_1, S_2, S_3, and S_4—and defining the two split-half comparisons as follows: Split One is ($S_1 + S_2$) versus ($S_3 + S_4$), and Split Two is ($S_1 + S_3$) versus ($S_2 + S_4$). With such a procedure, an accidental inequality between the halves of one split is likely to be rectified in the other orthogonal split. Validation of a given dimension in either of the two orthogonal splits is sufficient to demonstrate its "reality."

More sophisticated tests are possible using resampling methods, such as "jackknifing" (Mosteller and Tukey 1977; chapter 8) and "bootstrapping" (Efron 1982). In these methods, one takes multiple repeated samples from the original data set and performs independent analyses of these samples. By examining the variation in the solutions that result, inferences can be made about the reliability of different aspects of the solution. A brief discussion of these techniques can be found in Gifi (1981, section 12) and a very simple introduction to bootstrapping is given in Diaconis and Efron (1983). Weinberg, Carroll, and Cohen (1983) have recently applied these methods successfully to INDSCAL solutions.

It is also possible to perform significance tests for particular structural characteristics in three-way data by applying the methodology of permutation and randomization tests (Edgington 1969, 1980). For example, one can test whether there is significant "system variation" in a three-way array by proceeding as follows: Fit a PARAFAC model in q dimensions and note the goodness-of-fit value (Stress or r-squared). Then randomly permute the data observations across levels of Mode *C* but within the same levels of Modes *A* and *B* (permuting the values of subscript k, but keeping the i and j subscripts unchanged). By applying a different random permutation within each Mode *C* "tube" of the data (a different permutation of the k for each i,j pair), we can "scramble" the systematic variation across the third mode while preserving the systematic variation across Modes *A* and *B*. The resulting data will still contain q factors, but they will not show the systematic variation across Mode *C* necessary to determine a reliable unique solution. If the null hypothesis of no Mode *C* system variation is true, then the original data should not differ systematically from these permuted data sets. If the alternate hypothesis of system variation is true, the fit of the PARAFAC model to the original data should be better than to the permuted data.

Suppose we perform the permutation process 19 times and fit the *q*-factor PARAFAC model to each of the permuted data sets. We then rank the 20 fit values (19 permuted plus one nonpermuted). If there is no system variation, the fit for the unpermuted data is equally likely to be at any of the 20 possible ranks. Thus, if the fit value obtained for all 19 permuted data sets is less than that for the original nonpermuted data, we have observed a ranking that would happen only one time out of 20 under the null hypothesis. Thus, we can reject at the .05 level the null hypothesis of no system variation; there appears to be systematic variation across Mode *C* that is fit by the PARAFAC model. (Similar applications of permutation tests to multilinear models are described in Harshman, Green, Wind, and Lundy 1982; Harshman and Reddon 1983; Hubert 1983).

No Similar Tests for Two-Way Solutions. Note the contrast between intrinsic axis solutions and principal components or simple structure solutions. With either of these latter methods, the finding of consistent axis orientations across two split-halves (or by other means, such as the bootstrap) does *not* constitute evidence for their empirical validity. Similar axis orientations necessarily occur, so long as the configurations of points in the factor spaces are similar across split-halves, since these two-way methods determine axis orientation by finding directions in the configuration that maximize some simplicity or variance criterion. Indeed, so long as the split-half configurations are similar, any arbitrary rotation principle based on relations of axes to points in the space (for instance, a "most-complex-structure" criterion) would show similar results in the two split-halves. Obviously, no evidence for the correctness of a particular rotation criterion is provided by such a replication. For intrinsic axis methods, however, a consistent configuration is not sufficient to ensure consistent axis orientations. In each split-half there must be systematic stretches and contractions of the configuration as one proceeds across levels of the third mode, and these stretches must be in consistent directions in the two split-halves. Thus, replication of an intrinsic axis solution demonstrates the reliable presence of those characteristics of the data postulated by the three-way model and required to orient factor axes. Hence, in some sense, it validates the criterion used for orienting axes.

Cattell (1978) and others have sometimes argued for the confirmability of simple structure rotation. They have suggested that the occurrence of clear simple structure in the rotated solution will provide evidence for the "reality" of the hyperplanes and thus the appropriateness of seeking simple structure in the first place. While there is some logic to this argument, the conditions under which it might actually be persuasive seem quite limited. The presence of "clear hyperplanes," with many near-zero loadings on each factor, may often be an artifact of how the variables were selected (for example, if they were selected in clusters of related items) and may not indicate any structure in nature. But even if we could obtain "neutral" or representative selections of variables, there is little distributional or Monte Carlo data on the likelihood of obtaining various degrees of simplicity "by chance" when the variables are selected "randomly." (However, see Cattell 1978, appendix A.6.) Furthermore, it is not clear how to go about getting better data of this kind, since such Monte Carlo work would very likely have to be based on questionable assumptions concerning distributions of variables and of factor loadings on variables. Thus, although there might be a theoretical argument for the confirmability of simple structure, based on clarity of hyperplanes when the variables are selected in a certain way, this approach is much more difficult to apply in practice than the one for confirmability of intrinsic axis solutions by split-half or related methods. Consequently, the evidence provided by split-half confirmation of an intrinsic axis solution is usually much stronger.

Empirical Implications of Confirmation

Split-half or bootstrap confirmation of intrinsic axis factors has empirical significance beyond simply verifying that the model was (at least partly) appropriate and the data adequate. By demonstrating the occurrence of a particular kind of systematic variation within the data, it presents us with a fact that is hard to explain, except in terms of the conceptual framework that underlies the proportional profiles idea (that is, in terms of variation of contributions of "empirical unities" of the sort conceptualized by the PARAFAC three-way factor model).

To put it geometrically, suppose separate factor analyses of individual slices of a three-way array reveal a series of parallel configurations in which variables show basically the same relationships, but in which the points in some configurations are displaced outward from the origin in a particular direction by an amount proportional to their distance from the origin and contracted inward in other directions in a similar proportional way. Suppose, in other words, that the configurations show the sort of coordinated shifts of points describable as systematic stretches and contractions, as postulated by the PARAFAC-CANDECOMP model. This systematic pattern of relationships constitutes a striking empirical fact about the data that is very hard to explain except as the result of variations in the strength of a few underlying influences or processes that affect the relationships among variables. The systematic displacement of points in a configuration along a certain direction seems to imply some common influence on the several variables that get displaced, a common influence that has increased or decreased in magnitude from one configuration to the next. In other words, there is an empirical reality that corresponds (at least to some extent) to the abstract mathematical "factor" oriented along the direction of stretch.

Inductive versus Deductive Use of Factor Analysis. PARAFAC and related intrinsic axis methods significantly strengthen the inductive or hypothesis generating ability of factor analysis. When split-half testing (or similar methods, such as bootstrapping) confirms a set of dimensions, one can proceed with considerably stronger confidence to the construction of empirical hypotheses based on these dimensions. However, the significance of the split-half evidence can be both inductive and deductive. In one sense, the appeal to split-half methods is itself a test of a hypothesis; that is, in each half, one generates a hypothesis about the orientation of axes and tests it by replication in the other half. However, PARAFAC can also be used in a purely deductive mode to test hypotheses generated by other means, such as theory concerning configurations and axis orientations. In this mode, it will usually differ from other "deductive" or hypothesis-testing applications of factor analysis in that the test would not involve rotation-to-target to see how close a given matrix can be approximated; instead, the solution would be compared to the theoretical matrix without rotation (but possibly with rescaling of columns) to see how close both the configurational part of the hypothesis and the axis orientation part of the hypothesis agree with the data.

The configuration and axis orientation aspects of theory can be tested separately, however. If the fit value does not decrease much when the PARAFAC solution is rotated from the observed position to the position closest to the theoretical hypothesis, and if the configuration provides a good match to theory, then the evidence against the axes predicted by theory might be considered weak. However, if a PARAFAC solution in several different data sets shows an axis orientation that diverges from theory in a consistent manner, then the theory would be cast into doubt even if the theoretical orientation of axes had only slightly poorer fit. This is because the theory would presumably be unable to account for the consistency of the divergent PARAFAC results.

Limitations on Empirical Interpretation. Replication of intrinsic axis solutions—across split-halves or different experiments—implies that something meaningful and systematic is going on that can, at least in part, be captured by the PARAFAC-CANDECOMP model. But we must remember, of course, that this evidence does not imply that the model is perfectly appropriate, just that it captures enough variance to be a generalizable approximation to a more complex world. As noted earlier, for example, the model's assumption that common factors are present at the different levels of the data may not always be correct. However, the nature of the model is such that the axes established by such an approximation are likely to retain empirical meaning, even when the reality is considerably more complex.

When the reality is too complex for such an approximation to capture the bulk of the systematic variance, analysis of both real and simulated data suggests that the PARAFAC model will be much more likely to give degenerate solutions, with uninterpretable highly correlated dimensions, than misleading interpretable ones. And when these degenerate solutions arise, there are ways of constraining the PARAFAC solution so that a meaningful subset of the variance can often be captured and interpreted. With much more complex data, one might want to go to a more general model, such as Tucker's three-mode model, but one then loses the intrinsic axis property. To compensate for this loss, a constrained PARAFAC solution might be attractive as a supplement; it could provide an intrinsic axis reference point to aid in interpretation and rotation of the Tucker solution. Another intermediate approach is to rotate Tucker's core matrix to approximate diagonal form (McCallum 1976; Cohen 1974; Kroonenberg 1981b).

COMPARISON WITH TUCKER'S AND OTHER MODELS

Comparison with Tucker's Three-Mode Model

Tucker's Model "T3"

Tucker was the first to extend factor analysis to three-way arrays (Tucker 1963, 1964, 1966), and his model is still the basic reference point to which all other three-way procedures are compared. It is a very general model and incorporates most of the other three-way models as special cases. Kroonenberg and de Leeuw (1980) distinguish two versions of the model: (a) Tucker's original version, which reduces dimensionality of all three modes; and (b) a variant, which only reduces dimensionality in two modes (that is, it does not reduce the dimensionality of Mode C). They call the first model "T3" and the second "T2." We will adopt this terminology and concentrate our initial discussion on T3.

Our objective in this discussion is to compare and contrast the general T3 model and PARAFAC-CANDECOMP. For more details on T3 and its applications, the reader is referred to the works by Tucker (cited above) and to other chapters in this volume, including those by Bloxom and Kroonenberg.

Tucker's T3 in Scalar Notation. If x_{ijk} is an entry in a three-way data array, the model T3 can be written in scalar notation as follows:

$$x_{ijk} = \sum_{r=1}^{q_a} \sum_{s=1}^{q_b} \sum_{t=1}^{q_c} a_{ir} b_{js} c_{kt} g_{rst} \ . \qquad (5\text{–}30)$$

Here the a, b, and c coefficients have the same meaning as in the PARAFAC model (equation [5–3]); they are the factor loadings or weights for Modes A, B, and C, respectively. Tucker calls the small three-way array of g_{rst} coefficients the "core matrix." The g coefficients describe the sizes of interactions among factors across modes, as will be explained below. Instead of a single summation, there is a triple summation that runs over all combinations of entries in the three modes. The number of factors in one mode does not necessarily equal the number in another, and so in (5–30), q_a is used to represent the number of factors in Mode A, q_b the number in Mode B, and q_c the number in Mode C.

Since we have already discussed PARAFAC in some detail, we begin our discussion of T3 by noting the basic characteristics that the two models share. First, they are both designed primarily for direct fitting of the data matrix rather than indirect fitting of covariances, although, like PARAFAC-CANDECOMP, T3 can be applied to covariances to perform indirect factor analysis or to scalar products in order to perform multidimensional scaling (for example, see Tucker 1972). Second, both T3 and PARAFAC-CANDECOMP extend the idea of proportional changes in factor contributions to include the third mode; the Mode *C* weights are analogous to Mode *A* and *B* weights in that they simply multiply the contributions of a given factor by a coefficient representing the importance of that factor at a particular level of the given mode. Third, both T3 and PARAFAC-CANDECOMP are even-handed with respect to the role of different modes; all modes have the same formal properties—none has a preferred or special status.

There are, however, five main differences between Tucker's model and PARAFAC. The first is conceptual; the other four are algebraic and follow from the first one:

a. T3 is based on a different conception of "factor" than PARAFAC, although reconciliation is possible, as we shall see below.
b. In T3, all between-mode combinations of loadings are permitted, not only $a_{i1}b_{j1}c_{k1}$ but $a_{i1}b_{j3}c_{k2}$, and so forth; hence, each factor in Mode *A* potentially "interacts with" every one of the factors in Modes *B* and *C*, in all possible combinations; similarly, all factor combinations are possible for the other modes.
c. T3 incorporates a fourth set of coefficients, the g_{rst}, which describe the size of the factor "interactions" across modes; hence, T3 is a quadrilinear rather than trilinear model.
d. T3 allows the number of factors in Mode *A* (the number of columns of **A**) to differ from the number of factors in Mode *B*, which in turn may differ from the number in Mode *C*.
e. T3 does not have the intrinsic axis property but instead has a greater transformational indeterminacy than two-way factor analysis.

Differences in the Conceptual Model. As noted above, the Tucker T3 model was derived from a different basic conceptual idea of "factor" than that used by PARAFAC. Briefly, the difference is that a PARAFAC factor is more global than a T3 factor. For PARAFAC, a factor represents an empirical entity, process, or influence that is observed as a result of the situation being measured, rather than as a result of any particular mode of data classification or measurement. Thus, a single PARAFAC factor would be expected to have influence across the levels of all three modes of measurement. In other words, a PARAFAC factor is not "in a mode" but in a situation itself, with influences on or relationships with each of the three modes. In contrast, a T3 factor is conceptualized as an aspect, characteristic, or idealized type of level of a given mode. From Tucker's perspective, it is not the factors by themselves that generate variation in the data but rather the *interaction* of factors, that is, the interaction of particular aspects (factors) of Mode *A* with particular aspects (factors) of Modes *B* and *C*.

The distinction may be clarified by an example. Suppose we have a set of "semantic differential" data, a 25 × 30 × 200 array consisting of ratings of 25 stimuli on 30 rating scales, as made by 200 raters. From the PARAFAC perspective, the factors that we are seeking are global "dimensions of affective meaning." In the terms used by Osgood, Suci, and Tannenbaum (1957), these semantic dimensions might correspond to "Evaluation," "Potency," and "Activity." Such dimensions are not "in" any of the modes but are thought to be basic to the nature of affective meaning and are thus a natural part of the data-generating situation; through the particular stimulus-rating procedure, these global dimensions have patterns of relative influence or importance (given by loadings) for the several levels of each of the data modes. Thus, the first column of the Mode *A* loading matrix might describe the relationship between "Evaluation" and each of the stimuli (that is, which stimuli were perceived as strongly "good" or "bad"). The first column of **B** would then describe the relationship between "Evaluation" and the rating scales (that is, which rating scales were strongly expressive of "goodness" or "badness"). Finally, the Mode *C* loadings would show which raters stressed evaluative aspects strongly in making their ratings.

From Tucker's perspective, the semantic differential rating scale data is described differently. Just as each original data value is the result of the interaction of a particular rater with a particular stimulus and a particular rating scale, so the underlying patterns in the data are described as the result of interactions of features of stimuli, aspects of rating scales, and traits of raters. The objective of three-mode factor analysis is to discover those few different *features* or *types* of the stimuli (the Mode *A* factors), *kinds of components* of the rating scales (the Mode *B* factors), and *traits* or *types* of raters (the Mode *C* factors) that permit an adequate description of the patterns underlying the observed data variation. Tucker would call these factors "idealized" stimuli, "idealized" rating scales, and "idealized" raters. The core matrix **G** gives the pattern of interactions among these idealized entities and hence might be thought of as a miniature idealized image of the original data set.

A T3 factor-loading matrix for stimuli would then describe how the patterns for each actual stimulus can be approximated by some linear combination of the patterns for these idealized stimuli, implicitly telling how much the stimulus partakes of each of the idealized stimulus features. Likewise, the factor loadings for scales describe how the pattern of scores for each rating scale can be represented as some weighted linear combination of the idealized scale characteristics; similarly, each rater would be represented as some combination of the idealized raters or traits. When a given rater makes a rating, each of his idealized traits reacts to the various features (idealized stimulus components) in the particular stimulus and to the characteristics of the particular rating scale (the mixture of idealized scales in the actual scale) to produce the rating.

Tucker's conception of "factor" is not the same as the conventional notion most commonly used in two-way factor analysis. Thus, while PARAFAC reduces to two-way factor analysis when there is only one level to Mode *C*, this is not true of T3, unless the core is an identity matrix. This point is made particularly clear when one reads the Levin (1965) explication of the Tucker model.

Resulting Algebraic Differences. It is easy to see how these two contrasting notions of "factor" lead to the algebraic differences between T3 and PARAFAC. For T3, the Mode *A* factors are different in kind from the Mode *B* and Mode *C* factors, each arising from a different source; thus, there is no reason to require the same number of factors in each mode. However, for PARAFAC, if a factor is present in the situation, it is reflected in all three modes; hence, the dimensionality of the three modes should be the same (except when a factor fails to vary or to vary independently in a given mode—see below). For T3, interactions between all factors of one mode and all factors of another is what generates the data variance. Therefore, computation of all combinations of loadings across modes and the use of a core matrix to weight these combinations is natural and essential. But for PARAFAC-type factors, such interaction between modes does not have any meaningful interpretation and hence is not a part of the model. (However, an interpretation of the core matrix in slightly different terms is possible, as will be explained below.)

Lack of Uniqueness. Because of the complexity of relations permitted by the core matrix, the Tucker three-mode model permits free rotation or linear transformation of any three of its four sets of component parameters, with no loss of fit to the data (provided the compensatory transformation is applied to the fourth set of parameters). This means, for example, that each of the three loading matrices can be subjected to an arbitrary nonsingular linear transformation, and the inverses of these transformations can all be absorbed into the core matrix. Even system-variation data, which has only one PARAFAC representation, can be represented by T3 in terms of many alternative sets of factors, because each alternative set also has a different pattern of interactions between modes.

Since the T3 model allows a far greater family of possible solutions than were available for two-way factor analysis, it places even stronger demands on the user for intelligent application of rotation or transformation criteria; one might, for example, use simple structure when appropriate, or target rotation, or even rotation to transform T3 into a simpler model, perhaps based on some external theory (as in Dunn and Harshman 1982). Such "freedom of rotation" can prove useful for exploration purposes, but many investigators regret the loss of the empirical information which would have been provided by the intrinsic axis property. As a compromise, some programs that fit the Tucker model include options to approximate the simpler PARAFAC-CANDECOMP model by a special rotation to approximately "diagonalize" the core matrix and thus perhaps allow some of the information about factors provided by intrinsic axis methods to be recovered (see below).

Both PARAFAC and T3 models seem to make sense. Both can be applied to the same data and yield useful insights. Are they really, then, as different as they seem? As we shall see, the two models can be interpreted in such a way as to make them much more closely related. We will first see how the PARAFAC model can be considered as a special case of the Tucker T3 model and how in this case the interpretations can also be parallel. We will then look at the reverse relationship, constructing the T3 model as a special case of PARAFAC-CANDECOMP. Finally, a particular formulation of the T3 model in matrix terms permits it to be interpreted in terms of "global" or PARAFAC-like factors that vary in a more complex fashion than provided for in the PARAFAC-CANDECOMP model.

PARAFAC-CANDECOMP as a Special Case of T3 and Vice Versa

Embedding PARAFAC-CANDECOMP in T3. If we start with the T3 model but eliminate multiple interactions of factors across modes, we obtain a special case of T3 that is structurally equivalent to PARAFAC-CANDECOMP. This can be accomplished by simply requiring that the core matrix elements g_{rst} be zero except when $r = s = t$, in which case we specify that they are equal to 1.0. (This equivalence is well known and has been pointed out by many authors, including Carroll and Chang 1970; Harshman 1970; Tucker 1972.) These restrictions turn the core matrix into a three-way analog of an identity matrix, with zeros in all cells except for the "superdiagonal" cells running diagonally through the body of the cube from the upper left cell of the front slice of the cube to the lower right cell in the back slice.

Now, suppose instead that we fit the more general T3 model but after suitable rotation find that the core matrix has large entries on the superdiagonal and small entries elsewhere. How would we interpret this solution in terms of Tucker's conception of "factor"? Suppose for the moment that we are analyzing the semantic differential data described earlier and that three large superdiagonal cells are identified, so the core matrix is $3 \times 3 \times 3$. A Tucker interpretation might proceed as follows: There are three different basic factors in the rating scale mode, which, under suitable rotation, correspond to "Evaluation," "Activity," and "Potency" (that is, all the rating scales employing highly evaluative terms load on the Evaluation factor, and so forth). There are also three major features of the stimuli. In the rotation that "diagonalizes" the core matrix, these three stimulus mode factors seem to group the stimuli that are highly evaluative into one group, those that are highly active into another, and those that are highly potent into a third, with stimuli showing a mixture of several of these characteristics loading on several of the factors. Thus, it might be appropriate to give the same names to the three stimulus dimensions as to the rating scale dimensions.

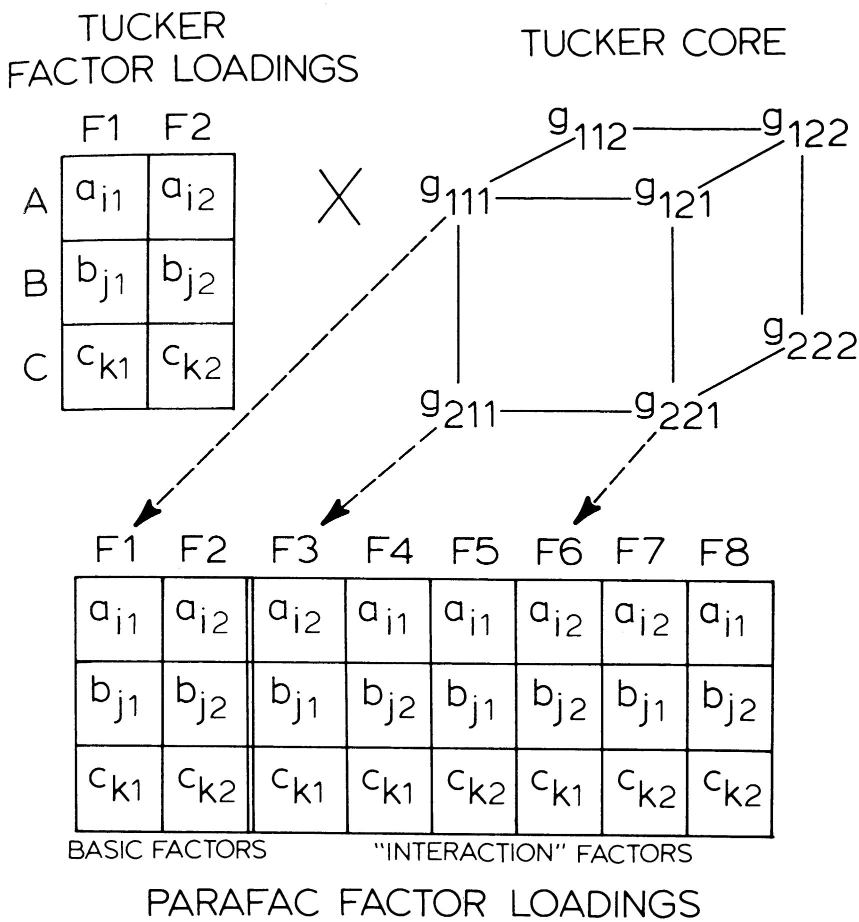
It is also significant (so the interpretation might proceed) that the Evaluation feature of the rating scales interacts almost exclusively with the "good-bad" feature of the stimuli and very little with the Potency or Activity aspects of the stimuli (a consequence of the superdiagonal core matrix). Likewise, the Activity and Potency factors of the scale mode interact mainly with the corresponding characteristics of the stimulus mode. Finally, the patterns of ratings of different individuals seem to vary mainly in terms of their relative emphasis on the three factors of the other two modes. In such a state of affairs, it would seem plausible to suggest that Evaluation, Activity, and Potency are "global" or general cognitive-affective dimensions, determined perhaps by culture or human psychology, but in any case more general than any particular mode of measurement by which the rating data were collected. It appears that the same cognitive-affective dimensions influenced raters' perceptions of both the rating scales and the stimuli. In this sense, we might say that the first factor of Mode A is the *same factor* as the first factor of Modes B and C and likewise for the second and third factors. We would conclude that a common set of three factors appears in all three modes of the Tucker analysis, thus arriving at a very PARAFAC-like interpretation of the data.

Embedding T3 in PARAFAC-CANDECOMP: Method I. It is instructive to see how the structure and interpretive framework of the two models can be made to correspond in a complementary way by representing a general T3 data structure in PARAFAC-CANDECOMP terms. In the one-dimensional case, this correspondence is trivial, since both models are directly equivalent. In the multidimensional case, however, we need some way to represent in the PARAFAC-CANDECOMP model all the interactions provided for by the T3 core matrix. The most direct but cumbersome way to do this is to provide a distinct PARAFAC factor for each additive component in the T3 model, that is, for the additive contribution corresponding to each cell in the core matrix.[4]

For simplicity, let us consider the case in which T3 has two factors in each mode—and thus a $2 \times 2 \times 2$ core matrix. To represent data with this T3 structure in PARAFAC-CANDECOMP terms, we could construct a PARAFAC model with eight factors, one to represent each possible interaction of a T3 Mode A factor with a T3 Mode B factor and a T3 Mode C factor. The PARAFAC factor representing a given interaction would have loadings in each mode corresponding to the T3 factor in that mode that was involved in the interaction (see Figure 5–1). For example, the PARAFAC factor corresponding to the g_{212} term would have loadings in Mode A corresponding to the second Mode A factor of T3, loadings in Mode B corresponding to the first Mode B factor of T3, and loadings in Mode C corresponding to the second Mode C factor of T3. In order that these interactions have the correct relative sizes, one mode, say Mode C, should have each set of factor loadings multiplied by the corresponding core matrix g value. There would appear to be considerable redundancy in such a PARAFAC representation. In this example, the same set of Mode A loadings would appear four times, each time linked with a different set of Mode B and/or C loadings; similar redundancies would occur in the other two modes (Figure 5–1).

What might we make of such a solution if it were obtained with real data? Let us assume that the two factors involved are Activity and Evaluation. Our first PARAFAC factor might have Activity loadings in both Modes A and B and so we would interpret it in the normal PARAFAC fashion as a factor representing the connotative dimension of activity. But the second factor might have Activity loadings in Mode A and Evaluation loadings in Mode B, representing the contribution of the g_{121} cell of the T3 core matrix. We would be forced to conclude that somehow the raters were adding a component to their responses on evaluative scales that was determined by the activity feature of the stimuli. While at first this might seem peculiar, on reflection, we might

Figure 5–1. Carroll's Method of Transforming a Tucker Representation into the Corresponding PARAFAC-CANDECOMP Representation



conclude that for these raters, *activity itself* might seem good. Thus, this factor, expressing an Activity-Evaluation interaction, might be said to arise from a nonindependence of the two concepts in the minds of the raters. Similar interpretations could be made about the other "interaction" factors; variations between subjects in the weights assigned to these factors would correspond to variations in overlap of Activity and Evaluation. Note that here we are still talking of "global" factors of the PARAFAC kind but interpreting the interaction components as arising from dependent effects of these global factors—a kind of nonorthogonality.

Since PARAFAC already permits nonorthogonality of factors, it might seem puzzling that additional factors would be needed to express such overlap of meaning. If activity has overtones of goodness in the minds of our raters, then this could be represented by incorporating some loadings on evaluative scales into the basic Activity dimension. Such a nonorthogonality in the basic factor, however, would produce the same pattern of correlated overtones for all subjects. The addition of interaction factors allows *individual differences* in the degree of obliqueness, as determined by the Mode *C* weights. Tucker introduces this kind of interpretation of core matrix interaction terms, for example, when discussing multidimensional scaling applications of T3 (Tucker 1972). The nondiagonal cells of the first frontal plane of the core matrix allow the model to represent one pattern of nonorthogonality between dimensions. Likewise, the nondiagonal cells of the second plane represent a different pattern of nonorthogonality. Each subject's individual pattern of obliqueness is some combination of these core patterns. It is with such reinterpretations that we can bridge the gap between the two models.

By embedding T3 in PARAFAC, it is possible to describe T3 as a "special case" of the PARAFAC-CANDECOMP model, one in which the factors do not all have distinct loading patterns within a given mode but rather repeat their loading patterns to produce all possible combinations of factors across the three modes. Unfortunately, our chance of obtaining such a recognizable pattern in a PARAFAC solution with real data is practically nil (unless we employ special rotation methods), because the axis orientation would not be unique. Recall that uniqueness requires distinct patterns of factor variation across the levels of each mode. In the PARAFAC solution that represents T3, there are two sets of four factors in each mode that have identical loading patterns (although these are different groups of four in each mode). Clearly, the requirement that factors show distinct patterns of variation is grossly violated, and so this kind of PARAFAC representation of data with general T3 structure will show as much indeterminacy as the T3 representation.

While Method I provides the most straightforward way of embedding T3 in PARAFAC-CANDECOMP, there are other ways that require fewer PARAFAC dimensions. However, the less mathematically inclined reader can skip the next two sections and go directly to the section "Reconciliation of T3 and PARAFAC Perspective" with no loss of continuity.

Simplifying the Representation: Method II. There are more compact ways of embedding T3 in PARAFAC-CANDECOMP. A first simplification allows us to reduce the number of PARAFAC dimensions needed from one per cell of the core to one per cell of the smallest face of the core. This is possible because, as noted earlier, any two factors that share the identical pattern of loadings in two of three modes can be combined into a single factor. Our initial method of embedding T3 in PARAFAC-CANDECOMP systematically creates factors with identical loading patterns in two modes. All factors, for example, that correspond to the interaction of a particular Mode *A* factor with a particular Mode *B* factor differ only in the Mode *C* loadings. Such factors can be combined so that a more compact representation is obtained, with one PARAFAC factor for each combination of Mode *A* and *B* factors from T3, rather than one for each combination of Mode *A*, *B*, and *C* factors.

Suppose we take two factors with identical loading patterns in Modes *A* and *B* but different patterns in Mode *C*. These factors can be combined as follows:

$$
\begin{aligned}
& a_{i1}b_{j1}c_{k1}g_{111} + a_{i1}b_{j1}c_{k2}g_{112} \\
& = a_{i1}b_{j1}(c_{k1}g_{111} + c_{k2}g_{112}) \; .
\end{aligned}
\tag{5–31}
$$

In terms of our example, the two factors that have the Evaluation pattern of Mode A and B loadings can be combined into one, and the two that have the Activity pattern of Mode A and B loadings can also be combined into one. Similarly, the two factors with Activity loadings in Mode A and Evaluation loadings in Mode B (or vice versa) can be combined. The result is a four-dimensional PARAFAC representation, and a new set of Mode C weights.

In general, we can use this technique to reduce the PARAFAC representation of an arbitrary q_a by q_b by q_c T3 model from $q_a q_b q_c$ PARAFAC dimensions to $q_a q_b$ PARAFAC dimensions, or (by collapsing across Mode A or B) $q_b q_c$ or $q_a q_c$ dimensions if this product is smaller. This procedure will work regardless of the number of levels of the mode that collapses. For example, if Mode C has the highest dimensionality, then we can define our representation as

$$
a_{iu}b_{ju}c_{ku} = a_{ir}b_{js} \left(\sum_{t=1}^{q_c} c_{kt}g_{rst} \right) \tag{5–32}
$$

with

$$
u = r + (s - 1)(q_a) \; ,
$$

where u is the index of the PARAFAC factor, r, s, and t are indices of the T3 factors in Modes A, B, and C, respectively, and there is a distinct u for each a,b combination. (An alternative way of demonstrating that no more than q_a times q_b dimensions are needed is given below, where a matrix formulation of the model is considered.)

Still More Compact Representations: Method III. By simplifying the core matrix before applying (5–32), it is often possible to achieve an even more condensed PARAFAC representation of a T3 structure. This simplification, interestingly enough, can be accomplished by applying PARAFAC to the core itself. To summarize briefly, by using enough PARAFAC dimensions, it is always possible to find an exact PARAFAC representation of an arbitrary core. In general this will require more PARAFAC factors than Tucker factors—and thus more PARAFAC factors than rows or columns of the core matrix (see Kruskal 1976, 1977)—but still relatively few dimensions compared to the number required by Method II. This exact-fitting PARAFAC solution provides a reexpression of the core in terms of a set of diagonal matrices that are pre- and postmultiplied by particular transformations (in other words, in terms of an equation of the form [5–4]). These transformations are then absorbed into the Tucker A and B matrices to produce the new T3 representation. Because of its intrinsic interest, as well as its relationship to previous proposals for simplifying the T3 core (McCallum 1976; Carroll and Pruzansky 1979; Cohen 1974), this procedure is described in more detail in appendix 5–2.

Method III often compresses our PARAFAC representation of a given T3 structure into considerably fewer dimensions than even

Method II requires. For example, it is possible to represent two-dimensional T3 data by means of a 2- or 3-factor PARAFAC model.[5] In the case of larger core matrices, our Monte Carlo tests indicate that the number of PARAFAC dimensions needed to embed the T3 model is considerably less than the $q_a q_b$ upper bound give by Method II. In the $3 \times 3 \times 3$ case, it appears that the PARAFAC model can be reduced from 9 to 5 factors. In the $2 \times 3 \times 4$ case, the representation can be reduced from 6 to 4 factors. Examples such as this demonstrate that often many of the cells in the core matrix are "nonessential" in the sense that they could be constrained to be zero with no loss in fit.

Because one can transform a T3 representation into a PARAFAC representation by applying PARAFAC to the core, as described in appendix 5–2, the question of how many PARAFAC dimensions are required to embed a given T3 model reduces to the question of how many PARAFAC dimensions are required to represent its core matrix. In the most general case, the core matrix is simply an arbitrary three-way array, and so the number of PARAFAC dimensions needed to represent T3 data that has q_a, q_b, and q_c dimensions in Modes *A*, *B*, and *C*, respectively, is formally equivalent to the generalized or trilinear rank of a q_a by q_b by q_c three-way array. The issue of trilinear rank is considered in some detail by Kruskal (1976, 1977).[6]

Reconciliation of T3 and PARAFAC Perspective

T3 in Matrix Notation. A useful alternative perspective on T3 can be gained by considering the following matrix representation of the model

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{H}_k\,\mathbf{B}' + \mathbf{E}_k\,, \tag{5–33}$$

where

$$\mathbf{H}_k = \sum_{t=1}^{q_c} c_{kt}\,\mathbf{G}_t. \tag{5–34}$$

In (5–33), $\mathbf{X}_k$ is the *k*th slice of the three-way data array, $\mathbf{A}$ is an n by q_a matrix of Mode *A* factor loadings, $\mathbf{B}$ is an m by q_b matrix of Mode *B* factor loadings, and $\mathbf{H}_k$ is a q_a by q_b matrix giving the relationships between the Mode *A* and Mode *B* factors for the *k*th level of Mode *C* (for instance, for the *k*th person if Mode *C* represents individuals). For each *k*, the $\mathbf{H}_k$ matrix is obtained by taking a weighted sum of q_c slices of the core matrix $\mathbf{G}$, with the weights given by the *k*th row of the Mode *C* factor-loading matrix.[7]

Let us return once again to our example of the rating scale data. On the one hand, (5–33) can be interpreted in terms of "factors" in the sense meant by Tucker—as aspects, features, or idealized levels of a given mode. Treating our semantic differential example from this perspective, $\mathbf{H}_k$ gives the *k*th individual's pattern of interactions of the stimulus features with the rating scale attributes. On the other hand, it is possible to reinterpret (5–33) in terms of "factors" in the sense meant by PARAFAC—as global influences acting in the situation and expressed in all three modes; in our example, the factors would be basic semantic or affective features, with Mode A giving their relation to the stimuli and Mode B giving their relation to the rating scales. In this second interpretation, $\mathbf{H}_k$ would describe the way that the kth individual modified the use of, or shifted the overtones associated with, these global factors.

In order to reinterpret $\mathbf{H}_k$ as describing individual differences in use of global dimensions, we need to decide how to apply it to the $\mathbf{A}$ and $\mathbf{B}$ matrices. In terms of our example, we need to decide whether $\mathbf{H}_k$ describes individual differences in the perceptions of the stimuli, in the semantic dimensions underlying the scales, or in both.

Individual Differences in Mode A. One possibility is to apply the $\mathbf{H}_k$ solely to the stimulus mode (which in our example we assume to be Mode A), and so conceptualize the model as follows:

$$\mathbf{X}_k = (\mathbf{A}\,\mathbf{H}_k)\,\mathbf{B}' \tag{5–35}$$

or

$$\mathbf{X}_k = \mathbf{A}_k\,\mathbf{B}' \,. \tag{5–36}$$

This approach attributes the variations in patterns of judgments across different raters to variations in the perceptions of the stimuli. The perceptual dimensions would be assumed constant —an unchanging pattern of rating-scale loadings would reflect a fixed notion of Evaluation, Activity, and Potency—but the degree to which these properties were perceived as characteristic of different stimuli would change from subject to subject.

Note, however, that the changes in perceived stimulus properties would only be of certain particular kinds. The model cannot represent the fact that an individual point might be the only one to shift position relative to other points in one rater's space. The $\mathbf{H}_k$ produces a linear transformation of the stimulus space and so describes coordinated changes in the positions of all the stimuli. In fact, in the T3 model (as contrasted with the T2 model, to be discussed briefly below), the patterns of individual variations are even more restricted: Every $\mathbf{H}_k$ is made up of a weighted combination of slices $\mathbf{G}_t$ of the core matrix, $\mathbf{G}$, with each slice describing a characteristic pattern of transformation of the stimulus space. Thus, individual raters differ only in the degree to which they incorporate one or another of these few basic patterns of individual variation into their personal perception of the stimuli.

As with PARAFAC-CANDECOMP, the stimulus space for subject k might be stretched or contracted along the axes, but in addition it might also be subjected to rotation, shear, or other linear transformations. If, for example, the space were compressed in such a way that certain "active" stimuli were located closer to the "good" ones and no longer at right angles to them, then we might interpret the solution as indicating that for person k those stimuli were perceived as having a mixture of activity and goodness, not simply activity alone.

Individual Differences in Mode B. Now suppose that we group terms in the opposite way and interpret the model as

$$\mathbf{X}_k = \mathbf{A}\,(\mathbf{H}_k\,\mathbf{B}') \tag{5–37}$$

or

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{B}_k' \;. \tag{5–38}$$

With this grouping, the model could be interpreted as describing differences in the nature of the semantic dimensions across individuals. For example, the *k*th rater might have an Activity dimension (column two of his $\mathbf{B}_k$) that was a linear combination of the basic Activity dimension (described by column two of $\mathbf{B}$) and Evaluation (column one of $\mathbf{B}$); this could be interpreted as indicating that the notions of activity and goodness were not as independent for this person as they were in the group space described by $\mathbf{B}$.[8]

Individual Differences in Both Modes. The third possible interpretation of (5–33) involves splitting $\mathbf{H}_k$ into two or three pieces. Here $\mathbf{H}_k$ is considered to be the product of several matrices that represent transformations of both the $\mathbf{A}$ and $\mathbf{B}$ matrices and possibly also a diagonal matrix of weights on the dimensions. For example, we can let

$$\mathbf{H}_k = \mathbf{T}_{ak}\,\mathbf{T}'_{bk} \qquad \text{or} \qquad \mathbf{H}_k = \mathbf{T}_{ak}\,\mathbf{D}_k\,\mathbf{T}'_{bk} \;, \tag{5–39}$$

so that

$$\mathbf{X}_k = (\mathbf{A}\,\mathbf{T}_{ak})(\mathbf{T}'_{bk}\,\mathbf{B}') \text{ or } \mathbf{X}_k = (\mathbf{A}\,\mathbf{T}_{ak})\,\mathbf{D}_k\,(\mathbf{T}'_{bk}\,\mathbf{B}') \;.$$

Here we would interpret the differences in judgments made by rater *k* as due to shifts both in the perception of the stimuli and in the overtones of the fundamental semantic dimensions used in making the evaluations. The more detailed representation would insert a $\mathbf{D}_k$ matrix of dimension weights between $\mathbf{T}_{ak}$ and $\mathbf{T}'_{bk}$ to represent general dimension weights for occasion *k*. Alternatively, we can view the dimensions as fixed except for differences in weight or salience and the interactions among the dimensions as changing across levels of Mode *C*. In this case, the $\mathbf{H}_k$ matrix would be written

$$\mathbf{H}_k = \mathbf{D}_k\,\mathbf{T}_k\,\mathbf{D}_k \;, \tag{5–40}$$

so that

$$\mathbf{X}_k = (\mathbf{A}\,\mathbf{D}_k)\;\mathbf{T}_k\;(\mathbf{D}_k\,\mathbf{B}') \;.$$

This latter approach is sometimes taken when $\mathbf{X}_k$ is symmetric, such as when the data are covariances. Then, $\mathbf{T}_k$ expresses obliqueness of axes and $\mathbf{D}_k$ dimension weights or saliences.

We can see that by reinterpreting the $\mathbf{H}_k$ matrices, we can formulate an alternative conceptual framework for Tucker's three-way algebraic model, one which involves the same sort of global factors as underlie the conceptual framework for PARAFAC. That is, loadings for a given column in all three modes can be interpreted as referring to the same factor. The greater complexities that can be represented by the Tucker model are seen as referring to more complex patterns of individual variations. In particular, these complexities are seen as referring to changes across individuals in the perceived grouping of stimuli (5–36), in the overtone content of semantic dimensions (5–38), or both (5–39). Alternatively, the variations can be attributed to individual differences in the interactions or obliqueness among dimensions (5–40).

Different Dimensionality in Modes A and B. While reconciliation of Tucker and PARAFAC schemes of interpretation seems feasible if there are the same number of factors in each of the three modes, what can we make of the T3 models in which there are a different number of factors in one mode compared to another? At first this would seem to represent a serious block to application of our scheme of reinterpretation. There is, however, a simple solution: Situations in which T3 modes have differing dimensionality correspond to situations in which two or more PARAFAC factors do not show linearly independent patterns of variation in a particular mode but are independent in the other two modes. In the PARAFAC representation of this situation, all three factor-loading matrices will have the same number of columns, but the rank of the matrices would differ. With T3, however, all loading matrices are *by convention* required to have full column rank; thus, T3 would typically represent this situation by having fewer factors in some modes than others. With fallible data, true linear dependence would normally not occur, even if we filled out the T3 loading matrices for all three modes to the dimensionality of the mode with the most factors; but sometimes when this is attempted, the eigenvalues corresponding to the last few factors are very small in certain modes and so the corresponding dimensions are typically considered to make trivial contributions in those modes. By convention, the factors are eliminated from the modes to which they do not make substantial *independent* contributions.

Suppose that for a particular stimulus set, two factors happen to show a more or less identical pattern of variation across stimuli (levels of Mode A), although they are clearly distinguished in terms of their weights on rating scales and people. If we adopt the usual practice of requiring that all the Tucker factor-loading matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ have full column rank, then a Tucker solution will be obtained in which the Mode A loading matrix has one less dimension than Modes B and C. The "missing" linearly dependent factor loadings in Mode A will be recreated, however, if the rectangular $\mathbf{H}_k$ matrix is used to transform the Mode A loadings, as in (5–36). There will be an extra column of elements in $\mathbf{H}_k$ that will describe how to regenerate the "missing" loadings in Mode A; that is, $(\mathbf{AH}_k)$ or $\mathbf{A}_k$ will have as many columns as $\mathbf{B}$ and will contain the column of linearly dependent loadings needed to go with the additional column of the Mode B and C factor-loading matrices. By thinking of T3 as in (5–36), for example, we can maintain our alternative interpretation in terms of "global" PARAFAC-like dimensions, even when the dimensionality of the formal T3 model varies across modes.

Transforming the Core to Approximate PARAFAC-CANDECOMP. Once the T3 model is interpreted in a way consistent with the

PARAFAC notion of "factor," it becomes possible to use proportional profiles to help deal with the transformational indeterminacy of the T3 solution. By rotating the solution into an approximation of PARAFAC form, we might be able to simplify the interpretation of the T3 dimensions, while testing the feasibility of going to a strict PARAFAC analysis (by observing how much the best approximation still diverged from a true superdiagonal core matrix). If a T3 solution can be rotated or transformed so that the core is very close to being diagonal (or superdiagonal), then the resulting axes can be interpreted in terms of the PARAFAC-CANDECOMP model. An approximate intrinsic axis property will apply, as well. In such a case, if the deviations from "pure" PARAFAC-CANDECOMP structure are attributed to chance (the off-diagonal core cells are considered to be small deviations from zero due to fitting error in the data), then it might be appropriate to replace the T3 solution with a true least-squares PARAFAC-CANDECOMP solution. If, on the other hand, there are a few large off-diagonal cells in the core matrix, even after transformation to achieve closest approximation to diagonal form, then by studying the core, the analyst may be able to gain further insight into the patterns of individual difference in the data. (Current research on "degenerate" PARAFAC solutions, discussed in chapter 6, suggests that strict minimization of all off-diagonal core cells may not lead to the most interpretable rotation when certain kinds of T3 structure are present. It may be better to retain a few large off-diagonal cells when this will cause the others to be much closer to zero and/or the loadings matrices to be less correlated and more interpretable. Such a rotation of the T3 core might produce factor loadings that correspond to an orthogonally constrained PARAFAC analysis—see chapter 6.)

This type of cross-model comparison was originally suggested in the slightly more restricted context of three-way multidimensional scaling; Harshman (1972b) suggested transformation of IDIOSCAL (symmetric T2) to maximize agreement with PARAFAC2, and Cohen (1974) suggested transformation of T3 or T2 for comparison with INDSCAL. The first successful demonstration of this idea was by Cohen (1974), who implemented it by the ingenious method of applying CANDECOMP directly to the core matrix to obtain an approximation in the same dimensionality as the Tucker solution, then using the factor-loading matrices of the CANDECOMP solution to transform the Tucker dimensions. Similar procedures involving transformation of the Tucker core to INDSCAL form were subsequently suggested by McCallum (1976), and Carroll, Pruzansky, and Kruskal (1980). The latter authors developed an application of their linearly constrained three-way model CANDELINC, which turned out to be precisely equivalent to Cohen's procedure and suggested that the INDSCAL-like solution that they could obtain in this way would be useful as a starting position for true least-squares INDSCAL analysis (see Carroll and Pruzansky 1979). Kroonenberg and de Leeuw (1980) (see also Kroonenberg 1981b) have developed a program called TUCKALS2 for fitting the T2 model to data by the method of alternating least-squares. They have incorporated into this procedure an option for approximate diagonalization of the core matrix, based on an algorithm described by de Leeuw and Pruzansky (1978).

The "Extended Core" Tucker Model T2

Tucker's original model places fairly strong constraints on the patterns of individual differences in use of dimensions. The $\mathbf{H}_k$ for each level of Mode *C* must be a weighted combination of the slices of the core matrix. A more general model would be to allow each individual to have an arbitrary $\mathbf{H}_k$. This was suggested in the context of analysis of symmetric scalar product data for MDS by Carroll and Chang (1970), who subsequently called the model IDIOSCAL, for IDIOsyncratic SCALing model (Carroll and Chang 1972). The same model, given a slightly different interpretation, was independently proposed by Jennrich (1972). Tucker (1972) incorporated the possibility for this kind of generality by allowing the third dimension of the core matrix to become "extended" until it had as many levels as there were levels of Mode *C* of the data array itself. The extended core also was useful in interpretation of Tucker's model in the context of MDS. The Tucker model with this extended core is called T2 by Kroonenberg and de Leeuw (1980) because it provides a reduced-dimensionality representation for only two of the three modes of the data. The T2 algebraic model would be given by (5–33), in which each $\mathbf{H}_k$ would be one slice (the kth frontal plane) of the core.

Most of the points that were made above about the T3 model also apply to T2. The same alternative interpretations of "factor" can be applied, and the same reinterpretations of the model can be used to place it and PARAFAC-CANDECOMP in a common conceptual framework. Interestingly, much the same limits on dimensionality also apply. The earlier demonstration that it is possible to combine all the dimensions that have the same Mode *A* and Mode *B* loadings but different Mode *C* loadings applies regardless of the number of levels of Mode *C*. Hence, an upper bound for the number of PARAFAC dimensions needed to "embed" or represent a given T2 structure is still the product of the two smaller ways of the core matrix—in this case, $q_a q_b$—regardless of the number of distinct levels of Mode *C*. For example, a T2 model with two dimensions in Mode *A*, three in Mode *B*, and 80 in Mode *C* (for 80 subjects) would be equivalent to a six-dimensional PARAFAC model or a T3 model with a 2 × 3 × 6-dimensional core.

Comparison with Other Three-Way Models

Several other three-way models have been proposed, but space limitations prevent us from discussing them in any detail here. Table 5–1 presents a summary of one family of related models for three-way profile data. Some of these models have been proposed and programmed by one or more groups of investigators, others have been proposed but not yet incorporated into a working program, and still others are included simply for comparison or theoretical interest. The models range from the most general, which allows a completely different set of dimensions for each two-way slice of the three-way array (T1–1), to the most restricted, which assumes exactly the same structure in all slices (T1–9). Tucker's T3 and T2 and the PARAFAC-CANDECOMP

TABLE 5-1. Some Three-Way Models for Profile Data

(T1-1)	$\underline{X}_k = \underline{A}_k\underline{D}_k\underline{B}'_k$	Independent spaces
(T1-2)	$\underline{X}_k = \underline{AD}_k\underline{B}'_k$ or $\underline{AB}'_k$	Unconstrained object variation (equals two-mode analysis of the "string out" data)
(T1-3)	$\underline{X}_k = \underline{A}(\underline{B} + \underline{V}_k)'$	("Continuous") Object variation
(T1-4)	$\underline{X}_k = \underline{AD}_k(\underline{B} + \underline{V}_k)'$	Mixed variation
(T1-5)	$\underline{X}_k = \underline{AH}_k\underline{B}'$	Tucker's T2
(T1-6)	$\underline{X}_k = \underline{A}(\sum_{t=1}^{q_c} c_{kt}\underline{G}_t)\underline{B}'$	Tucker's T3
(T1-7)	$\underline{X}_{kf} = \underline{AD}_k\underline{HD}_f\underline{B}'$	PARAFAC3
(T1-8)	$\underline{X}_k = \underline{AD}_k\underline{B}'$	PARAFAC1
(T1-9)	$\underline{X}_k = \underline{AB}'$	Two-mode analysis of "collapsed" (averaged) data

models will be found at intermediate positions. This table provides an idea of some of the different variations of three-way factor analysis that might be of interest. It focuses on versions of the models suitable for direct fitting of profile data and does not include special forms that arise from indirect fitting.

Corballis' Three-Way Model

A different kind of three-way model not included in Table 5–1 has been presented by Corballis (1973) as an extension of a factor-analytic model for change proposed by Corballis and Traub (1970). It is defined in terms of correlation matrices rather than profile data matrices. Even though it has been neglected in the three-mode literature, it is mentioned here because it provides an interesting complement to the more familiar models given in Table 5–1. It has the form

$$\mathbf{C}_{kk^*} = \mathbf{A}_k \mathbf{D}_{kk^*} \mathbf{A}'_{k^*} + \mathbf{E}_{kk^*} \, , \tag{5–41}$$

where $\mathbf{C}_{kk^*}$ is the matrix of correlations or covariances between the values obtained for a set of variables when they were measured on a set of cases on occasion k and the values for the same variables on the same cases but measured on occasion k^*. $\mathbf{A}_k$ is the factor-loading matrix for occasion k; $\mathbf{A}_{k^*}$ is the corresponding factor-loading matrix for occasion k^*; and $\mathbf{D}_{kk^*}$ is a diagonal matrix of weights, the product of the diagonal scaling matrices for occasions k and k^*. Thus the model considers both the matrices of covariances *within* occasions (when $k = k^*$) and *between* occasions (when $k \neq k^*$). Note that this model differs from any we have considered previously in that the factor-loading matrix **A** is subscripted. The model incorporates the idea that the precise pattern of factor loadings and factor scores for a given factor might change between occasions (such as when a given test is not measuring exactly the same thing when administered to subjects at different ages). These changes are not assumed to be proportional or to have any other particular form; in fact, they are not, in general, representable as a linear transformation of some common **A**. In this respect, the model is more flexible than any other three-way models that we have considered.

It might be thought that this model reduces to the overly general model given in the first line of Table 5–1, where there is no constraint on how the dimensions for one value of k (one occasion or individual) are related to those for another. This is not true, however, because (5–41) requires an assumption that each factor is orthogonal to all other different factors not only within occasions, but also between occasions. This requirement has the further consequence that all patterns of factor *change* across occasions must also be orthogonal.

Corballis and Traub (1970) and Corballis (1973) point out that the strong orthogonality assumptions of this model are generally sufficient to insure a *unique* solution for all the $\mathbf{A}_k$ loading matrices and for the other parameters, as well. Is this uniqueness another possible source of empirically meaningful axes? Once again, we would suggest that one's confidence in the orientation of axes provided by any procedure is a function of the plausibility of the assumptions invoked to obtain the unique solution. To many investigators, the orthogonality assumptions might seem too strong to be plausible. But it is an interesting question: How plausible are they *as an approximation,* and how distorted would the recovered factors be when the "true" factors are moderately correlated? This is presumably a problem for a Monte Carlo study. Corballis (1973) cautions us that "a potential difficulty is that the factor rotations specified by the model may not make sense, [that is], the factors may be uninterpretable." To our knowledge, there have been no applications of this model other than the demonstration application in the Corballis article (1973).

PARAFAC3 and PARAFAC2

It might be useful to note briefly some related intermediate versions of three-way factor-analytic models that form a bridge between PARAFAC1 and T3 or that have been considered elsewhere as interesting variants. However, this somewhat technical discussion is not essential for understanding the subsequent sections of this article; therefore, the less theoretically inclined reader may wish to skip to the discussion of ALSCOMP, below.

PARAFAC3 in General Form. In Table 5–1, a model called PARAFAC3 is placed between T3 and PARAFAC1. If we consider this model in its most general form, where the left-hand matrix of weights $\mathbf{D}_k$ need not be the same as the right-hand matrix $\mathbf{D}_f$, PARAFAC3 can be interpreted as a four-mode or fourway model. Applied to the analysis of cross-product or covariance matrices, it resembles the Corballis (1973) model in that it fits both the between-occasion and within-occasion cross-products. In this application, $\mathbf{X}_{kf}$ would represent the matrix $(\mathbf{X}'_k \mathbf{X}_f)$, in other words, the matrix of cross-products between data on occasions k and f. In this application, PARAFAC3 differs from Corballis' model in that it assumes constant $\mathbf{A}$ and $\mathbf{B}$ matrices rather than ones that vary across occasions. It also incorporates a central matrix $\mathbf{H}$, which allows for oblique relationships among dimensions.

The PARAFAC3 model is not restricted to analysis of cross-products or covariances; it can be applied to rectangular matrices of raw-score or profile data or to square matrices of asymmetric relationships. Such data might be obtained in some kind of four-way experiment. For example, $\mathbf{X}_{kf}$ might represent the ratings of stimulus k in context or situation f, and all $\mathbf{X}_{kf}$ matrices would contain ratings made by the same common set of raters (represented by the rows of $\mathbf{X}$) on a common set of rating scales (represented by the columns of $\mathbf{X}$). Or, it might be applied to transition matrices, such as a matrix whose rows and columns correspond to the same categories of automobiles and in which the cell x_{ij} represents the trade-in of an automobile in category i for a new one in category j. The matrix $\mathbf{X}_{kf}$ might then represent the number of people involved in such transitions who are trading in their old automobiles manufactured in year k for new automobiles manufactured in year f. In this application, the $\mathbf{H}$ matrix would represent an asymmetric pattern of transitions between categories. Such an application would be a four-way example of DEDICOM (see below).

The PARAFAC3 model has not been incorporated into a computer program, but it has been subjected to some mathematical analysis. An important reason for interest in this model is that it has been proven (Harshman 1981) that PARAFAC3 has the intrinsic axis property under reasonable conditions of data adequacy that are similar, in some respects, to those for PARAFAC1. It provides a unique solution without requiring the orthogonality assumptions of the Corballis model.

PARAFAC3 in Three-Mode Form. If we take the special case of PARAFAC3, where $k = f$, we obtain a three-mode model with intermediate generality between T3 and PARAFAC1. This version can be interpreted in the same terms as the general Tucker T3 model, in which there are "interactions" between the Mode A and Mode B dimensions; but in PARAFAC3, these interactions are fixed and described for all subjects by $\mathbf{H}$. It can also be reinterpreted in terms of more "global" dimensions by the methods noted earlier for T3. In either interpretation, individual differences between subjects (or whatever sources of variation are represented by the levels of Mode C) are of the simpler kind assumed by all the PARAFAC-series models: Differences in the importance or salience of the dimensions for different levels of Mode C. The weights defining the significance or contribution of dimensions on the kth occasion are given by the diagonal elements of the diagonal matrices $\mathbf{D}_k$.

PARAFAC2 and DEDICOM. When the three-mode version of PARAFAC3 is applied to symmetric matrices of scalar products or covariances, we obtain the PARAFAC2 model. This model is useful for oblique axis indirect fitting of PARAFAC and as a multidimensional scaling model that allows individual differences in the saliences of *oblique* dimensions (Harshman 1972b). While PARAFAC2 is more general than the PARAFAC1 model for covariances, since the indirectly fit axes can be oblique, it requires a constant set of angles between dimensions across levels of Mode C (for instance, across occasions) and thus is more restricted than T3.

PARAFAC2 involves an $\mathbf{H}$ that is square and symmetric. (The symmetry is not required, however.) When PARAFAC3 or PARAFAC2 is applied to a set of square n by n matrices describing relationships among n things—such as stimulus confusion matrices (see Dawson and Harshman 1983), matrices describing the number of people making telephone calls or perhaps migrating from the row place to the column place, the number of people switching from the row product to the column product, and so forth—we obtain the three-way DEDICOM models (see Harshman, Green, Wind, and Lundy 1982; Harshman 1978). If we allow $\mathbf{H}$ to be asymmetric but require $\mathbf{B} = \mathbf{A}$, we obtain the three-way "single domain" DEDICOM model; if we relax this requirement, we obtain the three-way "dual domain" model.

Despite its intermediate position in Table 5–1, PARAFAC2 (and the three-mode case of PARAFAC3) is not easily obtained as a special case of T3. This is because PARAFAC2 requires that the individuals' $\mathbf{H}_k$ matrices be obtained by weighting a fixed cosine matrix. In other words,

$$\mathbf{H}_k = \mathbf{D}_k \mathbf{W} \mathbf{D}_k \quad , \tag{5–42}$$

where $\mathbf{D}_k$ is a diagonal matrix giving the weights of the dimensions for the kth individual (or on the kth occasion). It is not obvious that a set of matrices of this kind can be obtained by taking different linear combinations of a few slices of any core matrix. Nonetheless, because any "extended" T2 core can be replaced by a moderate-sized diagonal or superdiagonal T3 core (as shown earlier), it is possible to represent PARAFAC2 and PARAFAC3 in terms of a T3 model. First, we set up a T2 model for which the slices of the extended core matrix have the form given in (5–42); then, we "collapse" the extended core to a T3 core with at most $q_a q_b$ slices.

Given that such a T3 representation is possible for any PARAFAC2 and PARAFAC3 structure, should we call these models special cases of T3? This may straining the sense of "special case." Even though there exist T3 solutions in moderately higher dimensionalities that would provide the same fitted values $\hat{x}_{ijk}$ as these PARAFAC models, one cannot obtain these models from T3 (or T2 or IDIOSCAL) by simply constraining certain terms to be zero, or fixed, or equal, and so forth. In order to find an "extended core" that has the special form required by PARAFAC2 and PARAFAC3, one must solve what is really the same problem as fitting these PARAFAC models to any other data array. Thus, it is perhaps better to treat PARAFAC2 and PARAFAC3 as separate models with their own interesting properties, such as intrinsic axes in certain circumstances (Harshman 1981).

The ALSCOMP Procedure

The Basic Idea of ALSCOMP: Nonmetric PARAFAC-CANDECOMP. Sands and Young (1980) have developed an algorithm and associated computer program called ALSCOMP (for Alternating Least-Squares COMPonents analysis). ALSCOMP is not really a new structural model but rather a new and more flexible estimation procedure, one that allows *nonmetric* fitting of the basic PARAFAC-CANDECOMP model (2–3) and (2–4) to a wide class of data. With their procedure, the data can be treated as ratio, interval, ordinal, or nominal scale, or some combination of these, and can be considered as "subject conditional," or "variable conditional," or the like (that is, having a measurement scale that differs from subject to subject, variable to variable, and so on).

The authors compare their procedure to other recently developed three-way procedures (including those for PARAFAC, CANDECOMP, and the Tucker three-mode model) and stress that "because all of these [other] . . . procedures place stringent requirements on the measurement characteristics of the data, none of them are applicable to most of the data types usually encountered in psychological research" (Sands and Young 1980). They suggest, therefore, that the ALSCOMP algorithm provides the only suitable method of fitting the PARAFAC-CANDECOMP model to most social sciences data.

How Essential is Nonmetric Fitting? While it is true that early applications of PARAFAC and CANDECOMP analysis often met with discouraging results, this probably occurred because the intrinsic axis property was interfered with by problems with the data characteristics and/or model appropriateness. PARAFAC has performed well in recent years, providing meaningful solutions to a wide range of data. How are we to explain this fact? The answer probably lies in the modifications and improvements to the analysis and particularly to the data preprocessing procedures that have been developed more recently. These improvements (discussed in chapter 6) have resulted in much greater flexibility in application of the model to real data; in fact, they have extended the flexibility or applicability of PARAFAC in some ways beyond that of ALSCOMP (although in other ways ALSCOMP is still more general).

Where the more sensitive intrinsic axis property is not part of the solution, metric procedures have been successful for some time. Sands and Young do not comment on the apparently successful application of Tucker's model to a number of different data sets (as reviewed elsewhere in this volume), nor do they mention the much longer history of the successful social sciences applications of conventional two-way factor analysis, which is also a metric procedure and thus involves the same "stringent" theoretical requirements on the measurement characteristics of the data.

Comparison of ALSCOMP with PARAFAC. What are the advantages and disadvantages of ALSCOMP versus PARAFAC algorithms for fitting (5–3) and (5–4)? We do not have space to discuss this issue in detail here, but there are several points that should be mentioned. PARAFAC preprocessing methods (discussed in the next section) eliminate the requirement for ratio-scale data and permit several kinds of data conditionality, so the basic criticism of Sands and Young is considerably weakened. Nonetheless, PARAFAC is still a metric analysis procedure. As we noted earlier, however, almost all factor analysis and many MDS procedures are metric; yet, they have been shown to be generally quite robust, perhaps because the solution is highly overdetermined by the data. The differences between the results obtained by metric and nonmetric fitting are typically modest, even when the data are known to be only ordinal scale, as Weeks and Bentler (1979) concluded from a Monte Carlo study using two-way data and as Harshman and Howe (1979) found in a Monte Carlo study using three-way MDS data that violated the Euclidean distance formula. Carroll and Chang (1970) also developed a "quasi-nonmetric version of INDSCAL," but in practice it gave results that were not appreciably different from the metric version. Hence, it was not considered to justify the extra analysis cost (Carroll and Chang 1970). (They did not develop a fully nonmetric version, because they doubted that it would provide any greater advantages.)

The distinction between continuous and discrete data is also not crucial for PARAFAC, as demonstrated by recent Monte Carlo results of Sentis, Harshman, and Stangor (1983), who found that PARAFAC quite successfully recovered continuous latent structure from binary data. Thus, it appears that in the vast majority of cases, nonmetric fitting may provide at best only a subtle improvement over the solution obtained by metric procedures.

Despite all the reservations listed above, there are no doubt certain cases involving systematic and extreme violations of interval-scale properties in which nonmetric procedures provide significantly better representations of the data than metric procedures. It is not clear, however, whether such extreme cases occur in real data except on rare occasions. Furthermore, in those special cases in which nonmetric analysis would make a substantial improvement, there may be special approaches to metric analysis that would provide similar results. Weeks and Bentler (1979) suggest that preprocessing the data by conversion to ranks (their "rank-linear" procedure) will provide most of the benefits of nonmetric analysis in such cases. We have not yet explored the effects of this kind of preprocessing in conjunction with PARAFAC. However, for cases in which nonlinearity is the problem, it has been demonstrated earlier (Harshman 1970, chapter 6; Terbeek and Harshman 1972) that the intrinsic axis properties of PARAFAC may often make it feasible to explicitly represent nonlinearities as extra dimensions and thus not only fit such data but quantify their nonlinear structure. Similar representations of nonlinearities are possible with intrinsic axis multidimensional scaling methods such as INDSCAL or the MDS application of PARAFAC. In fact, Chang and Carroll (1978) in an INDSCAL study of color perception, found "extra" dimensions beyond what would be theoretically expected; the form of the dimensions suggested that they might be due to nonlinear components involved in the reported similarities of colors.

Since nonmetric analysis involves fewer assumptions than metric analysis, it might seem advisable to use nonmetric approaches "just to be safe." However, one must balance potential benefits of nonmetric analysis against potential disadvantages. Sometimes the nonmetric procedures provide worse solutions than metric procedures, because they are subject to certain degeneracies—such as a number of points collapsing into a single location in the space—that the metric procedures avoid. Study of the Sands and Young article also suggests that, with some data, ALSCOMP may also be more subject to local optimum problems than metric ALS (Alternating Least-Squares) procedures such as PARAFAC. A nonmetric analysis fits far more parameters to a given data set than does a metric analysis; sometimes the data set is not large enough to determine all these parameters adequately. Also, ALSCOMP requires the user to specify the level of measurement of the data. The Monte Carlo results reported by Sands and Young (1980) indicated that incorrect specification could lead to considerably poorer recovery of the latent structure.

Finally, it should be pointed out that some of the most serious problems with fitting the PARAFAC-CANDECOMP model may be due to other things than violation of metric assumptions. Two cases, in particular, are worth noting:

a. Often the problem is that the data include certain unwanted components that interfere with intrinsic axis solutions (for instance, multiple factors constant in one mode but not others). It takes particular kinds of preprocessing to remove such contaminants and permit meaningful intrinsic axes to be defined by the part of the data that remains.
b. At other times, the data is technically inappropriate for either PARAFAC or ALSCOMP because it is generated by a process with a more complex structure (similar to the structure of Tucker's model). As we note in chapter 6, uninterpretable "degenerate" solutions with very highly correlated factors sometimes provide better fit to such data than interpretable solutions with factors resembling the "true" latent axes. The problems caused by such complex data should interfere equally with metric and nonmetric analysis. To cope with this situation, extended PARAFAC incorporates options for special analysis constraints that can block such degenerate solutions. Analysis with these constraints will permit recovery of reasonable approximations of the true dimensions, even though the fit to the data is lower than for the degenerate solution.

For these reasons, the extended PARAFAC model—made possible by the current PARAFAC procedure that incorporates special preprocessing and constraint options—provides certain kinds of generality that the ALSCOMP model does not.

In summary, then, both the ALSCOMP and PARAFAC programs have their particular strengths and weaknesses. Although there may still be situations in which the ALSCOMP nonmetric procedure could reveal more than PARAFAC (such as nominal data), there appear to be other circumstances in which ALSCOMP could not deal with the data as effectively as PARAFAC preprocessing and the extended PARAFAC model. We do not yet know the relative frequency of the two classes of situations "in the real world." Overall, it may be that PARAFAC and ALSCOMP will frequently give similar results, but such a statement is speculative, since to our knowledge no systematic comparison of the two procedures has yet been undertaken.

Other Models and Procedures. Space limitations prevent us from discussing a number of other models and procedures for three-way factor analysis that have been developed. For example, Bloxom (1968) has developed a version of Tucker's model in which the subject weights and errors are treated more specifically as random variables. Bentler and Lee (1978, 1979) have carried the statistical development of T3 even further; their approach also links three-mode factor analysis with structural equation modeling and permits confirmatory factor analysis, estimates of standard errors, and so on. Finally, the important models for the analysis of covariance structures proposed by Jöreskog (1971) and others are certainly relevant in the broader context of three-way data analysis but had to be omitted from this discussion in order to permit more detailed development of ideas within the scope of this chapter.

Comparison with Two-Way Factor Analysis

Of the various three-way models, PARAFAC is probably the one most directly related to traditional two-way factor analysis. When we apply PARAFAC to a three-mode array with only one level to the third mode (namely, the two-way special case), it reduces directly to traditional two-way factor or component analysis. In contrast, the Tucker model reduces to a less conventional but interesting representation in which there is a core matrix that has only one slice. (The two-way version of Tucker's model is discussed in some detail by Levin 1965.)

On the one hand, with appropriate preprocessing and scaling of the output loadings (as noted briefly earlier and described in more detail in appendix 5-1), PARAFAC gives results identical to those obtained with traditional programs for two-way singular value decomposition or principal component analysis. This property was exploited by Reddon, Marceau, and Jackson (1982), who used PARAFAC as an efficient way to obtain the first few vectors of the singular value decomposition of the more than five hundred items of the MMPI; their solution was then rotated by Varimax to obtain an interpretable principal axis solution. On the other hand, by fitting covariances or correlations and choosing the option to ignore the diagonal, PARAFAC can be used to perform common factor analysis; the solution that results is equivalent to that produced by the MINRES procedure (Harman and Jones 1966) or by the more common principal factor method based on iterating on the diagonal (when iteration on the diagonal is allowed to reach true convergence).

Of course, when PARAFAC is applied to two-way data, the intrinsic axis property is not obtained; the solution shows the classical rotational indeterminacy of two-way factor analysis. However, by exercising the option to require orthogonal loading matrices in both modes simultaneously, a unique solution can be obtained with axes oriented as in unrotated principal components analysis or principal factor analysis. This solution can then be rescaled and rotated by some suitable analytic or graphical method to obtain the final desired result.

The fact that PARAFAC bears this very direct relationship to two-way factor analysis facilitates interpretation of the loadings in the three-way case. We saw earlier how interpretive conventions can be carried over directly from the two-way case. For example, when doing direct fitting, weights for one of the three modes can be given a conventional interpretation as "factor loadings" and weights for the other two can be interpreted as defining "factor scores" (or "factor score estimates") of the traditional kind.

In this chapter, we have discussed the PARAFAC models for factor analysis (both direct and indirect fitting) and multidimensional scaling. We have examined the intrinsic axis property and compared PARAFAC with other models, particularly Tucker's three-mode factor analysis models.

In chapter 6, we will describe the results of our efforts at extending the domain of PARAFAC analysis to a wider range of data types by means of three-way data preprocessing and special analysis procedures.[9]

APPENDIX 5-1: SCALING AND INTERPRETATION OF PARAFAC LOADINGS

Size-Standardization of Loadings

Size Indeterminacy

There is a multiplicative indeterminacy in the scaling of factor loadings for PARAFAC, as there is with all factor-analytic procedures. The loadings for a given factor can be scaled upward or downward in one mode, provided that compensatory adjustments are made to the size of loadings in the other mode(s). For example, all the loadings for a given factor can be doubled in one mode and halved in another, and the resulting factor contributions (triple products) remain unaffected. In general, we can rescale the a_{ir}, b_{jr}, and c_{kr} loadings for factor r by any constant multipliers $k_{r(a)}$ for Mode A, $k_{r(b)}$ for Mode B, and $k_{r(c)}$ for Mode C, so long as $(k_{r(a)})*(k_{r(b)})*(k_{r(c)}) = 1$. The new loadings for factor r would be defined as follows:

$$\overset{*}{a}_{ir} = (k_{r(a)})a_{ir} \ , \quad \overset{*}{b}_{jr} = (k_{r(b)})b_{jr} \ , \quad \overset{*}{c}_{kr} = (k_{r(c)})c_{kr} \ .$$

These rescaled loadings would provide exactly the same fitted values, residuals, and so forth as the prior loadings.

Since within any mode such rescaling multiplies all the loadings for a given factor by the same constant, this size indeterminacy does not affect the pattern of relationships used for identification and interpretation of a given factor. It *can* affect interpretation, however, if loadings below a given size are considered "insignificant" and are disregarded; a factor scaled up or down will appear to have more or fewer large-sized loadings. Furthermore, if different k_r values are used for each r (that is, for each factor), then rescaling will change the relative sizes of loadings in different columns and will therefore complicate the comparison of relative contributions of factors to a given variable or occasion.

Standardization Conventions

A similar indeterminacy exists in two-way factor and principal component analysis of profile data, where it has been resolved by adopting a convention in which the factor weights for one mode are set to unit variance and called "factor scores," "component scores," or "factor score estimates." As a result, weights in the other mode reflect the scale of the data and the "absolute" size of factor contributions and are called "loadings." For example, when the data consist of variables measured on cases, the standardization is usually applied to the person weights while the size of the factor contributions is expressed in the variable weights.

We have adopted a similar convention for PARAFAC analysis of three-way profile data: Two of the three loading matrices are typically size-standardized so that the mean-squared factor loading for each factor in each mode is equal to 1.0; as a result, the size of loadings in the remaining factor-loading matrix is determined by the size of factor contributions to the data. (In contrast to the two-way case, we often use the term "loadings" to refer to weights in any of the three modes, since we treat the modes more even-handedly.) The loadings in this nonstandardized matrix take on the same units as the original data measurements; thus, they have some sort of "absolute" meaning that allows one to compare the size of loadings across columns, and to evaluate the size of a given loading against some external criterion to determine whether it reflects a substantial or trivial relationship. With standardized modes, on the other hand, loadings are evaluated in relative terms; a large loading means that the factor has a relatively strong relationship to that level of the mode, compared to other levels. Thus, the loadings are interpreted in the same way as any other z-score standardized variables.

Any one of the three matrices resulting from analysis of a given data set could be taken as "the" primary loadings matrix, provided the k_r values used to standardize the solution had been selected so that the columns of the other two matrices were set to unit mean-squares. Indeed, the results of a given analysis can be rescaled several different ways to change the mode that is interpreted as "loadings" in the sense used in two-way factor analysis. The investigator could therefore initially look at "loadings" on variables and factor or component "scores" on people and occasions but then rescale the variable and occasion entries and look at occasion "loadings" and variable and person "scores." No new analysis or estimation procedures would be required to get values in any mode that would reflect the scale of the data and provide cross-factor comparability. However, if the "loadings" were to be given additional interpretations—for example, as beta weights or variance components—then particular data-centering and size-standardization would be required, as described below. Since in some data sets these prerequisite characteristics might not be true of every mode, the special interpretations of loadings that they permit might not be possible for every mode. (Details on the required conditions are discussed in the following two parts of this appendix.)

Interpretations of Loadings

In two-way factor analysis, factor loadings are often given special interpretations, for example, as beta weights. If the factors are orthogonal, loadings are often interpreted as the correlation between a variable and a factor, and the sum of the squared loadings for a given variable is the proportion of the variance predicted by the factors, also known as the "communality." Before we can invest PARAFAC loadings with such additional meaning, certain conditions must be met. The data must be size-standardized in certain ways, and the loadings must then be scaled in a coordinated fashion. In the following discussion, we will first consider the more general interpretations of PARAFAC loadings that are possible, regardless of data standardization, and then develop the additional meanings that can be attributed to the loadings in certain special cases.

General Interpretation of Loadings as Regression Weights

The trilinear PARAFAC and bilinear factor analysis or principal component models can be thought of as equivalent to multiple regression models, except that the data is being predicted from latent factors rather than observed variables. Recall that the basic multiple regression model can be written (in the q predictor case) as

$$y_j = a_1 x_{j1} + a_2 x_{j2} + \ldots + a_q x_{jq} + e_j \;, \tag{5–43}$$

where the y_j is the value of the dependent variable for the jth case, and x_{j1}, x_{j2}, . . . are the values of the independent or predictor variables for the jth case; the a_1, a_2, . . . are the regression weights; and e_j is the residual or error of prediction. Typically, there is also an intercept or constant term included, but we omit this and consider regression "through the origin." The intercept term would permit the regression model to account for an additive constant in y. We consider such an additive constant as one of the h-terms in our model of interval-scale conditional-origin data (chapter 6, equation [6–3]) and assume that it has been removed by an appropriate preprocessing stage.

The regression model (5–43) can be transformed into the two-way factor-analytic model by letting $x_{j1} = b_{j1}$, that is, by taking the predictor variables to be the scores of each case on latent factors rather than on observed variables. To complete the transformation into a factor-analytic model, we simply consider n

different y-variables at the same time, so that we have n different regression equations predicting the n variables from the q latent factors. For the ith such variable, the score of the jth case would then be written y_{ij}, and the equation predicting that variable would have the regression weights a_{i1}, a_{i2}, and so forth, and an error term e_{ij}.

To generalize this to three-way factor analysis, we simply take a j^* that ranges over two modes rather than one—for example, over occasions as well as cases. If there are m cases (that is, if the maximum value of j is m) and p occasions (that is, the maximum value of k is p), then $j^* = j + (k - 1)m$, and the maximum value of j^* is (mp). To obtain a regression of the same form as (5–43), we replace x_j with x_{j^*}, where $x_{j^*} = (b_j c_k)$. Thus, we obtain a regression in which the data values are predicted from factor or component scores for each case on each occasion, and these scores are given by the product of Mode B and Mode C weights. By considering i variables simultaneously, we obtain a set of regression equations for which the ith equation is

$$y_{ij^*} = a_{i1}x_{j^*1} + a_{i2}x_{j^*2} + \ldots + a_{iq}x_{j^*q} + e_{ij^*} , \tag{5–44}$$

or

$$y_{i(jk)} = a_{i1}(b_{j1}c_{k1}) + a_{i2}(b_{j2}c_{k2}) + \ldots \tag{5–45}$$

$$+ a_{iq}(b_{jq}c_{kq}) + e_{i(jk)} .$$

We see that the factor loadings a_{i1}, a_{i2}, and so on are simply regression weights in a multiple regression equation predicting the y values of the ith variable from scores on the factors, as before; only now we consider the scores to be a function of the occasion as well as the case. Carroll and Chang (1980) use similar reasoning in their development of the alternating least-squares method for fitting CANDECOMP.

Interpreting Regression and Factor Weights

In the most general case, application of regression involves no standardization of predictor or predicted variables. In such applications, the B weights (as nonstandardized regression weights are often called) have no special interpretation except that they give the amount of each predictor variable needed to generate the composite that best approximates the predicted variable. Similarly, when PARAFAC is applied to raw data with no special scaling, the sizes of the factor loadings have no special interpretation except that they give the amount that must be contributed by each factor in each mode for the q factors to best predict the data array. Additional meaning can be attributed to regression weights or factor loadings only when the data and the factors have special properties, such as zero-mean, unit variance, and/or orthogonality.

For example, recall the effects of variable standardization in regression. Size-standardizing the x values permits straightforward comparison of the sizes of regression weights. Differences in weights now directly reflect differences in size of the contribution of the associated predictors to y, without being influenced by the overall size (mean-square) of each predictor variable.

Size-standardizing y permits the size of regression weights to be compared across different variables, data sets, and so forth. When both the predicted and predictor variables are standardized to be z-score variables (with mean zero and variance 1.0), then the B weights of the regression become fully standardized and are called "beta weights." In addition, if the predictor variables are mutually uncorrelated, the transformation of both x and y to z-scores permits each beta weight to be interpreted as the simple product-moment correlation between the predictor and the predicted variable.

Similarly, in PARAFAC analysis, standardization of y (the data) and x (the predictor "factor scores") allows the loadings to take on the additional interpretations possible with z-score regression. Suppose, for example, that we want Mode A weights to take on this extra meaning. We need to standardize the data and the output factor loadings so that Mode A factor weights are interpretable as the beta weight "loadings" and the products of Mode B and C weights give the predictor variable "factor scores" in z-score form. To do this, the data should be centered across Mode B and/or C, so that there is a mean of zero within each level of Mode A. (Recall that fiber-centering a given mode will slab-center the other two modes, as is apparent from inspection of Figure 6–1 in chapter 6.) The data should also be size-standardized on Mode A, so that within each level of Mode A the data have a mean-square of 1.0 (computed across all levels of the other two modes). This would transform the data at each level of A into z-scores and thus provide the necessary standardization of y in our corresponding regression equation.

In addition, the output *loadings* should be standardized so that the b and c weights have a mean-square of 1.0 for each factor (and thus the Mode A loadings reflect the scale of the data); then the x_{j*r} factor scores or factor contributions will also be z-scores. To show this, we first show that the x_{j*r} will have a mean of zero and then that they will have a variance of one.

It is easy to see why the x_{j*r} predictor variables must have a mean of zero. Since we are assuming that either Mode B or C (or both) has been centered, it follows that either the Mode B or Mode C factor-loading table is column-centered (as proven in chapter 6). As a result, the total set of "factor scores," across all levels of both Modes B and C—that is, the products $b_{jr}c_{kr}$ for all jk combinations—has the zero-mean property of z-scores. This is because

$$\sum_j \sum_k (b_{jr}c_{kr}) = \sum_j b_{jr} \left(\sum_k c_{kr} \right) = \sum_k c_{kr} \left(\sum_j b_{jr} \right) ,$$

which implies that the sum of the bc products are zero whenever either the sum of the b or of the c terms is zero.

It is also easy to establish that the bc products for any factor have unit variance whenever both the b and the c weights for that factor have unit variance, since the mean-square of the bc products is equal to the product of their respective mean-squares. In other words,

$$\frac{1}{mp} \sum_j \sum_k (b_{jr}\, c_{kr})^2 = \frac{1}{mp} \sum_j \sum_k b_{jr}^2\, c_{kr}^2$$

$$= \frac{1}{m} \sum_j b_{jr}^2 \left(\frac{1}{p} \sum_k c_{kr}^2\right) .$$

Thus, the bc terms, the x_{j*} of equation (5–44), have unit variance and zero means and hence are z-scores.

We have now shown how to set up a PARAFAC analysis so that both the y and x terms in the corresponding regression equation (5–44) will be z-scores. When this is done, the factor loadings in the nonstandardized mode (in the example, Mode A) of the PARAFAC output can be considered beta weights. This makes them strictly comparable to loadings obtained by conventional two-way factor analysis.

If, in addition, the factors are orthogonal across Mode B or C (or simply within each B by C slice), the Mode A factor loadings can be interpreted as simple product-moment correlations between the data (at that level) and the contributions of the factor (at that level). Additional interpretations in terms of variance components are also possible, but they will be discussed later in this appendix.

Of course, when analyses are performed without imposition of an orthogonality constraint in Mode B or C, exact orthogonality of the resulting loadings will be very unlikely. However, if for each pair of factors the product of the Mode B cosine and the Mode C cosine is small, the above relationships will be closely approximated.

If the factors are not orthogonal across levels of Mode B or C (or at least within the Mode B by C slices at each level of Mode A), standardizing the data and the output to obtain a loading matrix interpretable as beta weights will produce a factor-loading matrix of the kind obtained in traditional *oblique* two-way factor analysis. In terms of the above example, the Mode A loadings matrix would be the analog of the factor pattern matrix in an obliquely rotated two-way solution. The $b_{jr}\, c_{kr}$ products themselves would be the estimated factor scores. Thus, the matrix of correlations between the $b_{jr}\, c_{kr}$ scores for the factors would be the phi matrix of "correlations among the factors," as traditionally interpreted. This correlation could be quickly computed for any pair of factors by computing the cosine of the angle between the two factors in Mode B and similarly in Mode C and then taking the product of these two cosines. (The factor cosines are computed as the inner product of the factor-loading vectors after the vectors have been scaled to unit length.)

Other interpretations of loadings from two-way factor analysis can also be carried over into the three-way domain. Thus, for those accustomed to looking at traditional factor loadings and interpreting them in particular ways, a method is available for obtaining interpretations of this kind from three-way PARAFAC analysis.

Interpretation of Loadings as Variance (or Mean-Square) Components: Orthogonal Case

If the factors are orthogonal, it is possible to interpret the loadings as describing components of the total variance (or mean-square) contributed by particular factors at particular levels of a mode, even when the data have not been size-standardized and centered in the way required for interpretation of loadings as beta weights. In the following discussion, we will prove such relationships for Mode A factor loadings only, since the arguments for Mode B and Mode C factors are parallel.

We start with an expression for MSQ, the mean-square data value at level i of Mode A (if the data has zero-mean at each level of A, MSQ will be the variance of level i):

$$\frac{1}{mp} \sum_j \sum_k (x_{ijk})^2 = MSQ_i \ . \tag{5–46}$$

By substituting the PARAFAC model for x_{ijk}, we obtain:

$$\frac{1}{mp} \sum_j \sum_k \left(\sum_r (a_{ir} b_{jr} c_{kr}) + e_{ijk} \right)^2 = MSQ_i \ . \tag{5–47}$$

Now, because the error part is orthogonal to the systematic factor part, we have:

$$\frac{1}{mp} \sum_j \sum_k \left(\left(\sum_r (a_{ir} b_{jr} c_{kr}) \right)^2 + (e_{ijk})^2 \right) = MSQ_i \ . \tag{5–48}$$

If we assume that the factors are orthogonal to one another, their cross-products vanish, so we can write:

$$\frac{1}{mp} \sum_j \sum_k \left(\sum_r (a_{ir}^2 b_{jr}^2 c_{kr}^2) + e_{ijk}^2 \right) = MSQ_i \ . \tag{5–49}$$

By rearranging the order of summation, pulling the constant a-coefficient outside the summation over j and k, and moving the constant divisors m and p inside where appropriate, we obtain:

$$\frac{1}{mp} \sum_r \left(a_{ir}^2 \left(\sum_j \sum_k (b_{jr}^2 c_{kr}^2) \right) \right) + \frac{1}{mp} \sum_j \sum_k e_{ijk}^2 = MSQ_i \tag{5–50}$$

or

$$\sum_r \left(a_{ir}^2 \left(\frac{1}{m} \sum_j b_{jr}^2 \right) \left(\frac{1}{p} \sum_k c_{kr}^2 \right) \right) + \frac{1}{mp} \sum_j \sum_k e_{ijk}^2 = MSQ_i \ .$$

If we have size-standardized the PARAFAC *loadings* (as opposed to standardizing the *data*) so that they have a mean-square of 1.0 in Mode B and Mode C, we obtain:

$$\sum_r (a_{ir}^2)(1)(1) + \frac{1}{mp} \sum_j \sum_k e_{ijk}^2 = MSQ_i$$

or, subtracting the error mean-square from both sides,

$$\sum_r (a_{ir}^2) = MSQ_i - \frac{1}{mp} \sum_j \sum_k e_{ijk}^2 \quad . \tag{5–51}$$

In other words, when the factors are orthogonal in Mode B or C (or simply across all the points in the B-C slice), the sum of squared loadings in a row of the Mode A matrix gives the mean-square accounted for by the q-factor model at that level of Mode A (namely, the total mean-square minus the error mean-square at that level of Mode A). Furthermore, it follows that: (a) each squared loading equals the mean-square contribution of a particular factor at that level; (b) by computing the average squared loading for each column of the Mode A matrix, one can obtain the mean-square contribution of each factor to the total data; and (c) the sum of these quantities across the r factors equals the total mean square predicted by the PARAFAC model at that dimensionality.

Interpretation for Nonorthogonal (or Orthogonal) Factors

When factors are not orthogonal, then the step from (5–48) to (5–49) does not follow. However, if we have size-standardized Modes B and C so that their mean-squared loading is 1.0, there is still a straightforward interpretation for the size of a_{ir} (disregarding sign) that holds whether or not factors are orthogonal: a_{ir} gives the root-mean-square average size of the contribution of factor r to level i of the data. (Recall that by a factor's "contribution" we mean the triple product $a_{ir}b_{jr}c_{kr}$, one of the q additive components of the predicted value for x_{ijk}.) The size of factor contributions is expressed in the same units as the data; for example, if the original measurements were in centimeters displacement, the factor contributions will of necessity be in terms of centimeters displacement, and so a_{ir} can be interpreted as giving the (RMS) average centimeters of displacement at level i due to factor r (see, for example, Harshman, Ladefoged, and Goldstein 1977). This property of loadings—sharing the same units as the data—is sometimes quite useful for application of the results of the factor analysis (see, for example, Ladefoged, Harshman, Goldstein, and Rice 1978).

It is easy to establish the validity of this general interpretation. When the Mode B and C loadings for each factor have a mean-square of 1.0, then, as we proved above, the bc products also have a mean-square of 1.0. But whenever a set of numbers with unit mean-square are multiplied by a constant (such as a_{ir}), their mean-square becomes the square of that constant (a_{ir}^2) and so their root-mean-square must be the absolute value of that constant ($|a_{ir}|$). The algebraic argument for this is obtained by following the same steps as those that took us from equation (5–50) to (5–51) but considering only the one-factor case, so that the sum over r vanishes.

Since the mean-square of the contributions of factor r at level i of Mode A is $(a_{ir})^2$, the average squared value in column r of A gives the mean-square contribution of factor r in the data as a whole. When the factors are not orthogonal, the sum of the column mean-squares will not equal the total mean-square predicted by the model, because some of the mean-squares will be overlapping. But as the factors become less and less oblique, individual factor mean-squares overlap less and less, and their sum approaches the mean-square of $\hat{x}_{ijk}$, the part of the data fit by the model. This is true not only in the data as a whole, but also at each level of A.

Special Cases

If the data are centered across Mode B or C and thus have a mean of zero at each level of Mode A, then the factor contribution mean-squares discussed above become equal to factor contribution *variances,* and the equations above can be interpreted in terms of *variance components*. For example, $(a_{ir})^2$ would give the variance of the contributions of factor r to level i of the data; thus, the absolute value of the loading would give the standard deviation of the factor contributions, a reasonable measure of factor influence and one which (as noted earlier) is in the same scale as the data itself. To take another useful example, (5–51) would imply that for orthogonal factors, the squared loadings in a given row of the Mode A matrix sum to the variance predicted by the factors (the "variance accounted for") at that level.

When the data are also size-standardized so that their overall mean-square is 1.0, then, for orthogonal factors, each squared factor loading can be interpreted as the *proportion* of variance contributed by that factor at that level. Thus, when the squared loadings are summed across rows, the sums can be interpreted as the proportion of data variance accounted for by the q-factor model at each level of Mode A (the "communality" of each level). When averaged down columns, the mean-squared loading of each factor can be interpreted as the proportion of the total data variance accounted for by that factor. (In this latter case, taking the column *sums* of squares, rather than mean-squares, will give the three-way equivalent of eigenvalues or squared singular values.)

If the preprocessing of a given data set had included double size-standardization—for example, both within levels of Mode A and within levels of Mode B, when the data were centered on Mode C—then it would be possible to consider two different sets of alternative standardized loadings for the data. One might scale the output so that the Mode B and C loadings have sums of squares equal to 1.0 (in order to interpret the Mode A loadings in the traditional way), and then rescale the output so that Modes A and C have mean-squares of 1.0 (to interpret the Mode B loadings in the traditional way).

Before we invest too much effort in such procedures, however, we should remind ourselves that a given set of factor loadings can be examined to determine the meaning of the factor or its pattern of effects in a given mode, regardless of the scaling of the loadings matrix. Such basic interpretation is accomplished by looking at the relative sizes of different loadings within the factor and determining what might best distinguish those items lying at one pole of the dimension from those lying at the opposite pole. Since such comparisons within a factor are not affected by columnar scale adjustment of the loadings matrix, it is only for special interpretation in terms of factor-variable correlations or variance-accounted-for that one need worry about the special standardizations discussed above.

"Factor (Component) Scores"

Give the strong parallel that can be established between one of the PARAFAC loadings matrices (in the example above, the Mode *A* matrix) and the traditional two-way "factor loadings" matrix, what can we make of the PARAFAC loadings matrices for the other two modes (in the example, the Mode *B* and *C* matrices)? As we have already noted, these matrices can be compared to the traditional factor score or component score matrices from a two-way analysis. A slightly different perspective, which we have adopted in this chapter, interprets the entries in the Mode *B* and Mode *C* loading matrices as expressing the *average* (that is, root-mean-square) size of factor or component scores for each factor, within particular levels of the mode in question. The individual factor score estimates, or principal component scores, are given by the direct or Kronecker product of the Mode *B* and Mode *C* loadings vectors for each factor (that is, the $b_{jr}c_{kr}$ double-product terms corresponding to each jk combination).

It should be noted, however, that factor scores computed in this fashion strictly conform to the "strong" restrictions imposed by the PARAFAC "system variation" model, as explained earlier. In contrast, indirect estimation of factor scores by regression methods (as described by Harshman and Berenbaum 1981) would allow estimation of variance components of the factor scores that might follow more general patterns of variation.

As we noted earlier, some contemporary psychometricians have adopted conventions regarding the use of the terms "factor" versus "component" that are more restricted than our usage here. They would refer to the $(b_{jr}c_{kr})$ products as "component scores," since they are based on loadings obtained by direct fitting. They would also use "component scores" to refer to any scores estimated after indirect fitting, so long as the diagonals of the covariance matrices used in the indirect fitting contained unaltered variances. They would reserve the term "factor score estimates" for the scores obtained (by regression or other methods) after fitting the "common factor model" (namely, when indirect fitting is performed in which the covariance matrix diagonals are replaced by estimates of common variance or "communality"; this is accomplished in PARAFAC when the diagonals are "ignored" by iterative reestimation during the ALS fitting procedure).

Loadings Standardization when Doing Indirect Fitting

If the data being analyzed are summed cross-products, or covariances, a *modified* standardization convention is required to maintain comparability between the results of three-way and two-way analysis and between direct and indirect fitting methods. Basically, the Mode *C* loadings must be scaled so that their mean is 1.0 and the scale of the data is jointly reflected in Modes *A* and *B*.

For solutions obtained by analysis of covariance matrices, the Mode *A* and Mode *B* loadings tables are identical, since the covariance data is symmetric across Modes *A* and *B*. Both sets of loadings correspond to the Mode *A* table of the direct fit solution. However, the Mode *C* table contains entries that correspond to the *squares* of the entries obtained with direct fitting. This is because the data variances and covariances are averages of cross-products, and in each cross-product, the Mode *C* loadings occur twice. For this reason, we scale the Mode *C* weights so that the average *first power* of the entries in each column is equal to one; it is not appropriate to set their average *square* to one, since they are already squared quantities. As a result, we obtain Mode *C* weights that directly equal the factor (or "factor score") variances at each level of Mode *C*; the average factor score variance over all levels of Mode *C* is 1.0, as is appropriate for *z*-scores. The scale of the data is then jointly reflected in Modes *A* and *B*. (Both Mode *A* and *B* loadings are multiplied by the square root of the scale factors that would otherwise be applied to a single mode—these scale factors are the square roots of our three-mode generalization of eigenvalues.)

When the data have been suitably size-standardized so that the average covariance matrix is a correlation matrix[10] (we call this "Equal Average Diagonal" or EAD standardization), then the resulting Mode *A* or *B* weights will have the characteristics of traditional "loadings." They can be interpreted as beta weights and have the same properties as the loadings obtained from two-way factor analysis of correlations or from factor analysis by direct fitting of *z*-score standardized data. In fact, with error-free data, when extracting the correct number of factors for perfect fit, the two kinds of loadings will be identical. That is, the Mode *A* loadings obtained by indirect fitting, when scaled in this modified fashion, are identical to the Mode *A* loadings that would be obtained by direct fitting of the profile data, when appropriately scaled and centered, as described earlier. In the more general case of fallible data with "true" and "error" factors that are not strictly orthogonal in any mode, the result of indirect fitting will be similar but not identical to the result of direct fitting.

When doing indirect fitting, one can compute measures of the fit to the covariances (for example, mean-square error, Stress, *R*-squared), but one often wants to obtain an index of the implied fit of the factors to the original data from which the covariances were computed. This is what is normally reported in two-way factor analysis. Such a fit value is easily obtained by simply averaging the squared loadings in each column of the Mode *A* matrix, to obtain the variance accounted for by that factor (or summing the squared loadings to obtain the generalized eigenvalue for each factor). By summing variance estimates across factors, we obtain the total variance accounted for by the solution. The latter value could also have been obtained by summing squared loadings in each row of **A** to get the predicted diagonal elements of all the covariance matrices fit by the analysis. The average of these diagonals gives the desired quantity, namely, the variance of the original profile data that is indirectly fit by the model; since the original diagonal elements of the covariance matrices represent the variances of the variables in the profile data, the fitted part of these diagonals represents the fitted part of the original data variance.

When EAD normalization is used, the average diagonal in the data is 1.0, and so the average predicted diagonal gives the proportion of variance accounted for by the solution, and the mean-squared loading for a given factor gives the proportion of the total variance contributed by that factor. The proportion of the common variance contributed by each factor is simply the mean-squared Mode *A* loading for that factor divided by the total of mean-squared Mode *A* loadings for all factors in the solution.

APPENDIX 5–2: A METHOD OF TRANSFORMING ANY TUCKER REPRESENTATION INTO A COMPACT PARAFAC-CANDECOMP REPRESENTATION

Suppose a three-way array has an exact Tucker representation and we wish to find the corresponding PARAFAC-CANDECOMP representation that provides perfect fit in the lowest possible dimensionality. The following is a method of deriving the desired PARAFAC-CANDECOMP representation by operations on the Tucker representation. We will describe the method in terms of operations on a T3 model; however, the same procedure will work with a T2 model, as noted at the end of this appendix.

This procedure demonstrates that the trilinear rank of an array—that is, the number of "triads" or PARAFAC dimensions needed to provide exact fit to the array (Kruskal 1977)—is determined by the trilinear rank of the core matrix, which is the number of PARAFAC dimensions needed to exactly fit the core matrix.

The T3 Starting Point

Standard Form

Let $\mathbf{X}$ be an n by m by p three-way array. For notational convenience, we conceptualize it as composed of p successive n by m "slices" or two-way arrays; the kth such slice is called $\mathbf{X}_k$. To express the T3 representation of $\mathbf{X}$ in matrix terms, we use the same convention as in (5–4); that is, we represent the three-way array by providing a general expression for the kth slice. We call the initial form of our T3 model $(\mathrm{T3})_1$ and write it as follows:

$$(\mathrm{T3})_1: \quad \mathbf{X}_k = \mathbf{A}\ [\ \sum_{t=1}^{q_c} c_{kt}\mathbf{G}_t\ ]\ \mathbf{B}' , \tag{5-52}$$

where the matrices **A**, **B**, and **C** are factor-loading matrices for Modes *A*, *B*, and *C*, respectively. **A** is n by q_a with arbitrary element a_{ir}, **B** is m by q_b with arbitrary element b_{js}, and **C** is p by q_c with arbitrary element c_{kt}. The "core matrix" **G** is a three-way array considered for convenience as a set of slices $\mathbf{G}_t$; there are q_c such slices, each slice being q_a by q_b.

Individualized Form

If we let $\mathbf{H}_k$ represent the matrix of interactions between Mode *A* and *B* dimensions for the *k*th level of Mode *C* (for instance, for the *k*th person), then we can write an "individualized" form of the model,

$$\mathbf{X}_k = \mathbf{A}\,\mathbf{H}_k\,\mathbf{B}' , \tag{5-53}$$

where

$$\mathbf{H}_k = \sum_{t=1}^{q_c} c_{kt}\,\mathbf{G}_t \tag{5-54}$$

In this "individualized" form of the model, each level of Mode *C* has a matrix $\mathbf{H}_k$, which gives that level's interactions between the Mode *A* and *B* dimensions. For example, if the levels of Mode *C* represent persons, then each $\mathbf{H}_k$ gives the idiosyncratic changes in Mode *A* and/or *B* axis orientations and weights for the *k*th person. As shown in (5-54), each $\mathbf{H}_k$ is a weighted combination of the slices of the core matrix, with the *t*th slice weighted by the *k*th person's loading on the *t*th dimension of Mode *C*.

The Transformation

To find the most compact PARAFAC-CANDECOMP model that represents this same data structure, we need to perfectly "diagonalize" the slices of the core array. We can do this by applying PARAFAC to the core, although other related methods, such as that of de Leeuw and Pruzansky (1978), should also work. Note that we do not seek an approximation but rather a perfectly fitting PARAFAC representation. To find this exact representation, we take advantage of the fact that PARAFAC can fit more dimensions than there are levels to any way of the array (Harshman 1970; Kruskal 1976). For an arbitrary core that is roughly cubical, the required diagonalized equivalent will in general be slightly larger than the original core, but not by much. As Kruskal has pointed out (personal communication, April 1983), in all such cases that we have examined, the PARAFAC dimensionality q does not exceed $(q_a + q_b - 1)$, where q_a and q_b are the number of levels of the two smaller ways of the core.

Thus, we begin by applying PARAFAC to the core. To describe the PARAFAC representation of the core, we employ a matrix formulation of the same kind as used earlier in this chapter. The PARAFAC representation of the three-way core array **G** is specified by giving a general matrix representation of an arbitrary tth slice $\mathbf{G}_t$:

$$\mathbf{G}_t = \bar{\mathbf{A}}\,\bar{\mathbf{D}}_t\,\bar{\mathbf{B}}' \,, \tag{5–55}$$

where (if an exact PARAFAC solution requires q factors) $\bar{\mathbf{A}}$ would be a q_a by q matrix of PARAFAC loadings on Mode A of the core, and $\bar{\mathbf{B}}$ would be a q_b by q matrix of PARAFAC loadings on Mode B of the core, with $\bar{\mathbf{D}}_t$ being a diagonal q by q matrix, whose diagonal elements are the PARAFAC loadings on Mode C of the core. To provide an exact representation, the $\bar{\mathbf{D}}_t$ will sometimes be larger than the corresponding $\mathbf{G}_t$. For example, in the case of a $2 \times 3 \times 5$ core matrix, where each $\mathbf{G}_t$ slice is 2×3, the PARAFAC representation would require at most four dimensions to perfectly fit $\mathbf{G}$. (For some $2 \times 3 \times 5$ cores, it would require less.) In the four-dimensional case, $\bar{\mathbf{A}}$ would be 2×4, $\bar{\mathbf{B}}$ would be 3×4 (and so $\bar{\mathbf{B}}'$ would be 4×3), and $\bar{\mathbf{C}}$ would be 5×4. Each $\mathbf{D}_t$ would be a diagonal 4×4 matrix.

We now need to find the factor loadings for the data array **X** that go along with the "diagonalized" form of its core array. These will be obtained by appropriately transforming the original Tucker loadings. Substituting our PARAFAC-CANDECOMP representation of the core (5–55) back into our basic model $(T3)_1$ (given by [5–52]), we obtain:

$$\mathbf{X}_k = \mathbf{A}\left[\sum_{t=1}^{q_c} c_{kt}\,(\bar{\mathbf{A}}\,\bar{\mathbf{D}}_t\bar{\mathbf{B}}')\right]\mathbf{B}' \,. \tag{5–56}$$

Since $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ do not change with t, we can move these constant terms outside the summation over t, and obtain:

$$\mathbf{X}_k = \mathbf{A}\left[\bar{\mathbf{A}}\left(\sum_{t=1}^{q_c} c_{kt}\bar{\mathbf{D}}_t\right)\bar{\mathbf{B}}'\right]\mathbf{B}' \,. \tag{5–57}$$

Regrouping terms, we obtain:

$$\mathbf{X}_k = (\mathbf{A}\,\bar{\mathbf{A}})\left(\sum_{t=1}^{q_c} c_{kt}\bar{\mathbf{D}}_t\right)(\bar{\mathbf{B}}'\,\mathbf{B}') \,. \tag{5–58}$$

We can interpret the $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ matrices as defining linear transformations of the original $(T3)_1$ loading matrices for Modes A and B. So, if we define the transformed loading matrices as

$$\overset{*}{\mathbf{A}} = \mathbf{A}\,\bar{\mathbf{A}}\,, \qquad \overset{*}{\mathbf{B}} = \mathbf{B}\,\bar{\mathbf{B}} \,, \tag{5–59}$$

and if we define a new "diagonalized" core matrix $\bar{\mathbf{G}}$ composed of q_c diagonal slices $\bar{\mathbf{G}}_t$ such that

$$\bar{\mathbf{G}}_t = \bar{\mathbf{D}}_t \,,$$

then we can write an expression for a partially transformed T3 model, which we call $(T3)_2$:

$$(T3)_2: \quad \mathbf{X}_k = \overset{*}{\mathbf{A}} \left(\sum_{t=1}^{q_c} c_{kt} \overline{\mathbf{G}}_t \right) \overset{*}{\mathbf{B}}' \, . \tag{5–60}$$

This model has the same structural form as $(T3)_1$ given above in (5–52) but uses transformed component matrices $\overset{*}{\mathbf{A}}$ and $\overset{*}{\mathbf{B}}$, which are n by q and m by q, rather than $\mathbf{A}$ and $\mathbf{B}$, which are n by q_a and m by q_b. It also has a modified core $\overline{\mathbf{G}}$, which is q by q by q_c, instead of $\mathbf{G}$, which is q_a by q_b by q_c. It is not fully transformed into a PARAFAC-CANDECOMP form, however, since it has the same matrix of Mode C loadings as $(T3)_1$, and the core matrix is not superdiagonal.

Nonetheless, model $(T3)_2$ gives rise directly to a PARAFAC-CANDECOMP representation when we write it in "individualized" form, as was done for $(T3)_1$ in (5–53). We simply define

$$\overline{\mathbf{H}}_k = \sum_{t=1}^{q_c} c_{kt} \overline{\mathbf{G}}_t = \sum_{t=1}^{q_c} c_{kt} \overline{\mathbf{D}}_t \, , \tag{5–61}$$

which allows us to write $(T3)_2$ as

$$\mathbf{X}_k = \overset{*}{\mathbf{A}} \, \overline{\mathbf{H}}_k \, \overset{*}{\mathbf{B}}' \, . \tag{5–62}$$

Now it is apparent from (5–61) that all the matrices that are summed to produce $\overline{\mathbf{H}}_k$ are diagonal, and so $\overline{\mathbf{H}}_k$ is itself diagonal. Hence, (5–62) can be considered a PARAFAC model. For uniformity of notation, we can let $\overline{\mathbf{H}}_k = \overset{*}{\mathbf{D}}_k$ and rewrite (5–62) as the PARAFAC-CANDECOMP model

$$\mathbf{X}_k = \overset{*}{\mathbf{A}} \, \overset{*}{\mathbf{D}}_k \, \overset{*}{\mathbf{B}}' \, , \tag{5–63}$$

where $\overset{*}{\mathbf{A}}$ is an n by q matrix of Mode A loadings, $\overset{*}{\mathbf{B}}$ is an m by q matrix of Mode B loadings, and $\overset{*}{\mathbf{D}}_k$ is a q by q diagonal matrix whose diagonal elements constitute the kth row of $\overset{*}{\mathbf{C}}$, a p by q matrix of Mode C loadings. This gives the PARAFAC-CANDECOMP representation of the three-way array $\mathbf{X}$, which is equivalent to the Tucker representation $(T3)_1$.

T3 in "Superdiagonal" (PARAFAC-CANDECOMP) Form

From (5–63) we can directly obtain a T3 representation in which the core matrix is superdiagonal (that is, has nonzero entries only when $r = s = t$). We simply define our transformed Mode C loading matrix for T3 as a matrix that has the diagonals of the $\overset{*}{\mathbf{D}}$ matrices as its rows. That is, $\overset{*}{\mathbf{C}}$ is a p by q matrix for which

$$\overset{*}{c}_{kt} = \overset{*}{d}_{tt(k)} \, ,$$

where $\overset{*}{d}_{tt(k)}$ is the tth diagonal element of $\overset{*}{\mathbf{D}}_k$. With these revised Mode C loadings, we can define the core matrix as the three-mode equivalent of an identity matrix. Thus, if $\overset{*}{g}_{uut}$ is the element in the uth diagonal cell of the tth slice of the revised core matrix $\overset{*}{\mathbf{G}}$, then

$$\overset{*}{g}_{uut} = 1 \quad \text{if} \quad t = u, \quad 0 \text{ otherwise.}$$

We can now write the T3 model in explicit superdiagonal or PARAFAC-CANDECOMP form, as follows:

$$(\mathrm{T3})_3\text{:} \quad \mathbf{X}_k = \overset{*}{\mathbf{A}} \left[\sum_{t=1}^{q} \overset{*}{c}_{kt} \overset{*}{\mathbf{G}}_t \right] \overset{*}{\mathbf{B}}' . \qquad (5\text{–}64)$$

Transforming T2

The same procedure can be applied to find the PARAFAC-CANDECOMP representation for a Tucker T2 model, in which the core matrix has as many Mode C levels as the data array $\mathbf{X}$. We simply apply PARAFAC to the extended core. In the argument above, this would mean that $q_c = p$. This presents no new difficulties for the PARAFAC decomposition in (5–55), and the rest of the transformation follows as before.

The fact that corresponding dimensions across Mode C can always be "collapsed" (as described earlier in this chapter) places the absolute upper bound on the number of needed PARAFAC dimensions as $(q_a q_b)$, or the product of whichever two dimensionalities are the smallest. (Kruskal [1977] gives the same upper bound.) Consideration of parameter counts or "degrees of freedom" of the core and corresponding PARAFAC representation shows that for T2, "extended cores" that have many levels in one mode, $(q_a q_b)$ will often be both a lower and upper bound. On the other hand, for small cores in which q_a, q_b, and q_c are similar, our experience suggests that the actual upper bound is much lower, perhaps $(q_a + q_b - 1)$. In either case, the transformed representation would be relatively more compact than the original.

As noted earlier in this chapter, the idea of approximately "diagonalizing" the Tucker T2 core array has been proposed previously by several authors (the first of which, to our knowledge, was Cohen [1974]). Thus, the transformations presented here should be compared with the proposals of Carroll and Pruzansky (1979), Cohen (1974), de Leeuw and Pruzansky (1978), McCallum (1976), and the application of this idea in Kroonenberg (1981b).

NOTES

1. A matrix approach that clearly displays the three-way symmetry of the PARAFAC model can be implemented by means of the Kronecker product notation. This is the approach used by Jennrich in Harshman (1970, chapter 5). (Tucker [1966] also

employed the Kronecker product for discussion of three-mode factor analysis.) The data would be represented as the sum of q different three-way arrays, each array corresponding to the contribution of one factor. The array corresponding to the contribution of the rth factor would be represented as the Kronecker product of the rth column of **A** with the rth column of **B** and **C**. If we let **X** equal the three-way array $\{x_{ijk}\}$ and **E** the three-way array of error terms, and if we use $\mathbf{a}_r$ to represent the rth column of **A**, and similarly for **B** and **C**, then, by using $\otimes$ to represent the Kronecker product of two vectors, we can write:

$$\mathbf{X} = \sum_r \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r + \mathbf{E} \ .$$

While elegant, this representation would involve us with notation and mathematics less familiar to many who work in this area; hence, it has not been adopted here.

2. Some would insist on the term "component scores" rather than "factor scores," as noted earlier.

3. Cattell discusses proportional profiles in several places in a recent book (Cattell 1978), describing the "confactor rotation" method with some optimism. However, he still considers the implicit orthogonality constraint (which we saw earlier to be a consequence of the use of indirect fitting) to be an unsolved problem. He does not discuss direct fitting using PARAFAC1 or indirect fitting using the PARAFAC2 model, as possible ways around this problem.

4. The authors would like to thank J. D. Carroll for suggesting this simple method of embedding T3 in PARAFAC-CANDECOMP.

5. Kruskal (personal communication, March 1983) has proven that a $2 \times 2 \times 2$ core matrix will have a maximum rank of 3 and that a $3 \times 3 \times 3$ core will have a maximum rank of 5, consistent with our Monte Carlo results. The interesting thing is that under a wide range of plausible conditions, a $2 \times 2 \times 2$ core will have rank 2 or 3, depending on the relative size of different core elements. Kruskal provides algebraic conditions (inequalities based on products of particular elements of the core) that determine whether a $2 \times 2 \times 2$ core will have rank 3 or 2.

6. Some of our Monte Carlo results have led Kruskal to reexamine these proofs; he now cautions us (personal communication, March 1983) that the theorems 3a–3d in Kruskal (1977) require minor modification and are not correct as stated. Thus, these particular theorems concerning trilinear rank have been revised. However, theorems in Series 1, 2, and 4 of that paper still appear to be valid. Copies of corrected versions of Kruskal (1977) are available from J. Kruskal.

7. This expression for T3 leads to a simple demonstration of an upper bound on Tucker dimensionalities (closely related to the upper bound on the embedding PARAFAC representation, given in Method II and equation [5–31], above). Any Tucker representation can always be replaced by one in which the largest mode of the core has no more dimensions than the product of the dimensionalities of the other two modes. We show this as follows:

Suppose, without loss of generality, that in our model Mode C has the most dimensions (that is, q_c is larger than either q_a or q_b). We can construct an alternative core array and corresponding Mode C loading matrix that will generate the same data as the original model. First, we use the original q_a by q_b by q_c core and Mode C loadings to construct the p different $\mathbf{H}_k$ matrices, one for each level of Mode C, as defined by (5–34). (If we are starting with a T2 model, this first step is unnecessary.) Then we construct a new core matrix that is q_a by q_b by $q_a q_b$—that is, one that has q_a times q_b slices, rather than q_c. We construct the slices of the new core such that there is only one nonzero element in each slice, and this element occurs at a different location in each slice; furthermore, we set these nonzero elements equal to 1.0. This establishes the new core. Note that we now have one slice of the core for each cell of an $\mathbf{H}_k$ matrix, since $\mathbf{H}_k$ is q_a by q_b. This allows us to obtain the Mode C matrix corresponding to this new core by simply assigning to each person a set of Mode C loadings equal to the values in his $\mathbf{H}_k$ matrix. In particular, we obtain the kth person's loading for factor t from the cell of $\mathbf{H}_k$ that corresponds to the nonzero cell in the tth slice of the new core matrix.

8. The question might be raised as to whether this second interpretation (5–38) is really different from the first one (5–36). We would argue that they *could* describe different states of affairs. On the one hand, if a new set of stimuli, when judged by person k, failed to show the same close association between active and good objects, it would appear that person k's distinct perspective was specific to particular stimuli, and model (5–35/5–36) might be more appropriate. On the other hand, if *all* new stimuli would be judged by person k in a way more closely linking goodness and activity, then it might be argued that model (5–37/5–38) is more appropriate. (Even here, however, it would seem that one could conceptually distinguish two different cases. In the first case, the basic *meaning* of "activity" is unaltered, but associated with this meaning is a derivative evaluation of "goodness," perhaps because person k believes that activity has good consequences. In the second case, the basic semantic category itself is changed; for example, person k might not use "activity" in its pure sense but rather as some notion of "vitality" that incorporates both aspects of activity and goodness.) This is a subtle issue, but we believe that the distinction is "real" if it can lead to different predictions. One way in which this might occur would be in a four-mode situation.

It is possible to construct four-mode models in which the indeterminacies of the Tucker model are reduced or eliminated because different ways of resolving these indeterminacies make different predictions about how the data might vary across a fourth mode (see, for example, PARAFAC3, given in Table 5–1 and discussed above under Other Models). One might imagine such a four-mode model being applied to studies in which each rater evaluates a given set of stimuli, using a standard set of rating scales, but repeats the ratings for each stimulus with respect to several different situations or contexts.

9. General acknowledgments for chapters 5 and 6 are made at the beginning of this chapter. We would also like to thank

John Reddon for helpful comments on chapter 5.

10. The covariance matrices are uniformly rescaled as follows:

$$\overset{*}{c}_{ijk} = \frac{c_{ijk}}{[(\bar{c}_{ii.})^{1/2} \ (\bar{c}_{jj.})^{1/2}]} ,$$

where

$$\bar{c}_{ii.} = \frac{1}{P} \sum_{k=1}^{p} c_{iik} ,$$

and similarly for $\bar{c}_{jj.}$. This imposes the same rescaling on all levels of Mode *C* and produces the same covariances as would have been obtained if the raw data for each level of Mode *A* had been converted to *z*-scores before covariances were computed.

REFERENCES

Bentler, P. M., and S.-Y. Lee. 1978. Statistical aspects of a three-mode factor analysis model. *Psychometrika* 43:343–52.

———. 1979. A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology* 32:87–104.

Bloxom, B. 1968. A note on invariance in three-mode factor analysis. *Psychometrika* 33:347–50.

Bock, R. D. 1973. Word and image: Sources of the verbal and spatial factors in mental test scores. *Psychometrika* 38:437–57.

Carroll, J. D., and P. Arabie. 1980. Multidimensional scaling. *Annual Review of Psychology* 31:607–49.

Carroll, J. D., and J. J. Chang. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* 35:283–319.

———. 1972. IDIOSCAL (Individual DIfferences in Orientation SCALing): A generalization of INDSCAL allowing idiosyncratic reference systems as well as an analytic approximation to INDSCAL. Paper presented at the spring meeting of the Psychometric Society, Princeton, New Jersey, March.

Carroll, J. D., and S. Pruzansky. 1979. Use of LINCINDS as a rational starting configuration for INDSCAL. Manuscript, Bell Laboratories, Murray Hill, New Jersey.

Carroll, J. D., and M. Wish. 1974. Multidimensional perceptual models and measurement methods. In *Handbook of perception,* vol. 2, ed. E. C. Carterette and M. P. Friedman. New York: Academic Press.

Carroll, J. D., S. Pruzansky, and J. B. Kruskal. 1980. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* 45:3–24.

Cattell, R. B. 1944. "Parallel Proportional Profiles" and other principles for determining the choice of factors by rotation. *Psychometrika* 9:267–83.

———. 1978. *The scientific use of factor analysis in behavioral and life sciences.* New York: Plenum.

Cattell, R. B., and A. K. S. Cattell. 1955. Factor rotation for proportional profiles: Analytical solution and an example. *British Journal of Statistical Psychology* 8:83–92.

Chang, J. J., and J. D. Carroll. 1978. Three are not enough: An INDSCAL analysis suggesting that color space has seven (± one) dimensions. Manuscript, Bell Laboratories, Murray Hill, New Jersey.

Cohen, H. S. 1974. Three-mode rotation to approximate INDSCAL structure (TRIAS). Paper presented at the annual meeting of the Psychometric Society, Palo Alto, California, March.

Comrey, A. L. 1967. Tandem criteria for analytic rotation in factor analysis. *Psychometrika* 32:143–53.

Corballis, M. C. 1973. A factor model for analysing change. *British Journal of Mathematical and Statistical Psychology* 26:90–97.

Corballis, M. C., and R. E. Traub. 1970. Longitudinal factor analysis. *Psychometrika* 35:79–98.

Dawson, M. 1982. Multidimensional responses to metaphor. Master's thesis, University of Western Ontario.

Dawson, M., and R. A. Harshman. Forthcoming. Applications of an asymmetric multidimensional scaling model to alphabetic confusion matrices. Manuscript, University of Western Ontario.

De Leeuw, J., and S. Pruzansky. 1978. A new computational method to fit the weighted Euclidean distance model. *Psychometrika* 43:479–90.

Diaconis, P., and B. Efron. 1983. Computer-intensive methods in statistics. *Scientific American* May:116–26, 128, 130.

Dunn, T. R., and R. A. Harshman. 1982. A multidimensional scaling model for the size-weight illusion. *Psychometrika* 47:25–45.

Edgington, E. S. 1969. Approximate randomization tests. *Journal of Psychology* 72:143–49.

———. 1980. *Randomization tests.* New York: Marcel Dekker.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. *Society for Industrial and Applied Mathematics Monographs* 38.

Eysenck, H. J. 1950. Criterion analysis—an application of the hypothetico-deductive method to factor analysis. *Psychological Review* 57:38–53.

———. 1977. Personality and factor analysis: A reply to Guilford. *Psychological Bulletin* 84:405–11.

Gandour, J. T., and R. A. Harshman. 1978. Cross-language differences in tone perception: A multidimensional scaling investigation. *Language and Speech* 21:1–33.

Gifi, A. 1981. *Non-linear multivariate analysis.* Preliminary edition. Leiden, the Netherlands: University of Leiden, Department of Data Theory.

Green, P. E. 1978. *Analysing multivariate data.* Hinsdale, Illinois: The Dryden Press.

Guilford, J. P. 1977. Will the real factor of Extroversion-Intro-

version please stand up? A reply to Eysenck. *Psychological Bulletin* 84:412–16.
Haan, N. 1981. Common dimensions of personality development: Early adolescence to middle life. In *Present and past in middle life,* ed. D. H. Eichorn, J. A. Clausen, N. Haan, M. P. Honzik, and P. H. Mussen. New York: Academic Press.
Harman, H. H., and W. H. Jones. 1966. Factor analysis by minimizing residuals (MINRES). *Psychometrika* 31:351–68.
Harris, R. J. 1975. *A primer of multivariate statistics.* New York: Academic Press.
Harshman, R. A. 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics* 16:1–84 (University Microfilms No. 10,085).
————. 1972a. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics* 22:111–17 (University Microfilms No. 10,085).
————. 1972b. PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics* 22:30–44 (University Microfilms No. 10,085).
————. 1973. Oblique coordinate systems in multidimensional scaling: Theory and experimental test. Manuscript, University of California at Los Angeles.
————. 1978. Models for analysis of asymmetrical relationships among N objects or stimuli. Paper presented at the first joint meeting of the Psychometric Society and the Society for Mathematical Psychology, Hamilton, Canada, August.
————. 1981. Generalizing INDSCAL and PARAFAC1: Uniqueness of certain representations of three-way crossproduct arrays. Manuscript, University of Western Ontario.
Harshman, R. A., and S. A. Berenbaum. 1981. Basic concepts underlying the PARAFAC-CANDECOMP three-way factor analysis model and its application to longitudinal data. In *Present and past in middle life,* ed. D. H. Eichorn, J. A. Clausen, N. Haan, M. P. Honzik, and P. H. Mussen. New York: Academic Press.
Harshman, R. A., and M. L. Howe. 1979. Systematic effects of noneuclidean input on the recovered subject and stimulus spaces in three-way MDS. Paper presented at the annual meeting of the Psychometric Society, Monterey, California, June.
Harshman, R. A., and G. Papcun. 1976. Vowel normalization by linear transformation of each speaker's acoustic space. *Journal of the Acoustical Society of America* 59:S71 (abstract).
Harshman, R. A., and J. R. Reddon. 1983. Three-way (PARAFAC) factor analysis of WAIS and WAIS-R suggests different axes than simple structure. Paper presented at annual meeting of the Classification Society (North American Branch), Philadelphia, May.
Harshman, R. A., P. E. Green, Y. Wind, and M. E. Lundy. 1982. A model for the analysis of asymmetric data in marketing research. *Marketing Science* 1:205–42.
Harshman, R. A., P. Ladefoged, and L. Goldstein. 1977. Factor analysis of tongue shapes. *Journal of the Acoustical Society of America* 62:693–707.

Horan, C. B. 1969. Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika* 34:139–65.

Horst, P. 1965. *Factor analysis of data matrices.* New York: Holt, Rinehart, and Winston.

Hubert, L. J. 1983. Inference procedures for the evaluation and comparison of proximity matrices. In *Numerical Taxonomy,* ed. J. Felsenstein. New York: Springer Verlang.

Jennrich, R. 1972. A generalization of the multidimensional scaling model of Carroll and Chang. *UCLA Working Papers in Phonetics* 22:45–47.

Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. *Psychometrika* 36:409–26.

Kettenring, J. R. 1983. A case study in data analysis. In *Statistical data analysis,* ed. R. Gnanadesikan. Providence, R.I.: American Mathematical Society.

Kroonenberg, P. M. 1981a. *User's guide to TUCKALS3. A program for three-mode principal component analysis (WEP-Reeks, WR 81-6-RP).* Leiden, The Netherlands: University of Leiden.

————. 1981b. *User's guide to TUCKALS2. A program for three-mode principal component analysis with extended core matrix (WEP-Reeks, WR 81-35-RP).* Leiden, The Netherlands: University of Leiden.

Kroonenberg, P. M., and J. de Leeuw. 1980. Principal components analysis of three-mode data by means of alternating least-squares algorithms. *Psychometrika* 45:69–97.

Kruskal, J. B. 1976. More factors than subjects, tests than treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika* 41: 281–93.

————. 1977. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications* 18: 95–138.

————. 1978. Factor analysis and principal components: Bilinear methods. In *International encyclopedia of statistics,* ed. W. H. Kruskal and J. M. Tanur. New York: Free Press.

————. 1981. Multilinear models for data analysis. *Behaviormetrika* 10:1–20.

————. 1983. Multilinear methods. In *Statistical data analysis,* ed. R. Gnanadesikan. Providence, R.I.: American Mathematical Society.

Kruskal, J. B., and M. Wish. 1978. *Multidimensional scaling.* Beverly Hills: Sage.

Kruskal, J. B., R. A. Harshman, and M. E. Lundy. 1983. Relationships between Tucker's three-mode factor analysis, PARAFAC-CANDECOMP and CANDELINC. Paper presented at the annual meeting of the Psychometric Society, Los Angeles, June.

Ladefoged, P., R. A. Harshman, L. Goldstein, and L. Rice. 1978. Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America* 64:1027–35.

Levin, J. 1965. Three-mode factor analysis. *Psychological Bulletin* 64:442–52.

Linker, W. 1982. Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. *UCLA Working Papers in Phonetics* 56:1–134.

MacCallum, R. C. 1976. Transformation of a three-mode multidimensional scaling solution to INDSCAL form. *Psychometrika* 41:385–400.

McDonald, R. P. 1979. The simultaneous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology* 32:212–28.

Meredith, W. 1964. Notes on factorial invariance. *Psychometrika* 29:177–85.

Meyer, J. P. 1980. Causal attribution for success and failure: A multivariate investigation of dimensionality, formation and consequences. *Journal of Personality and Social Psychology* 38:704–18.

Mosteller, F., and J. W. Tukey. 1977. *Data analysis and regression: A second course in statistics.* Reading, Mass.: Addison-Wesley.

Osgood, C. E., G. J. Suci, and P. H. Tannenbaum. 1957. *The measurement of meaning.* Urbana: University of Illinois Press.

Reddon, J. R., R. Marceau, and D. N. Jackson. 1982. An application of singular value decomposition to the factor analysis of MMPI items. *Applied Psychological Measurement* 6: 275–83.

Sands, R., and F. W. Young. 1980. Component models for three-way data: An alternating least squares algorithm with optimal scaling features. *Psychometrika* 45:39–67.

Schönemann, P. H. 1972. An algebraic solution for a class of subjective metrics models. *Psychometrika* 37:441–51.

Sentis, K. P., R. A. Harshman, and C. Stangor. Forthcoming. PARAFAC three-way factor analysis of dichotomous data: A Monte Carlo study. Manuscript, Ted Bates Advertising/New York.

Snyder, C. W., Jr., W. D. Walsh, and P. R. Pamment. 1983. Three-mode PARAFAC factor analysis in applied research. *Journal of Applied Psychology* 68:572–83.

Spence, I. 1977. Multidimensional scaling. In *Quantitative ethology,* ed. P. Colgan. New York: Wiley-Interscience.

Terbeek, D. 1977. A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics* 37:1–271.

Terbeek, D., and R. A. Harshman. 1972. Is vowel perception non-Euclidean? *UCLA Working Papers in Phonetics* 22:13–29.

Thurstone, L. L. 1947. *Multiple-factor analysis: A development and expansion of the vectors of mind.* Chicago, Illinois: University of Chicago Press.

Torgerson, W. S. 1958. *Theory and methods of scaling.* New York: Wiley.

Trick, L. 1983. The effect of individual differences in analogic reasoning and perceived tenor-vehicle similarity of metaphor comprehension and appreciation. Master's thesis, University of Western Ontario.

Tucker, L. R. 1963. Implications of factor analysis of three-way matrices for measurement of change. In *Problems in measuring change,* ed. C. W. Harris. Madison, Wis.: University of

Wisconsin Press.
———. 1964. The extension of factor analysis to three-dimensional matrices. In *Contributions to mathematical psychology,* ed. N. Frederiksen and H. Gulliksen. New York: Holt, Rinehart, and Winston.
———. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31:279–311.
———. 1972. Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika* 37:3–27.
Velicer, W. F., A. C. Peacock, and D. N. Jackson. 1982. A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research* 17:371–88.
Weeks, D. G., and P. M. Bentler. 1979. A comparison of linear and monotone multidimensional scaling models. *Psychological Bulletin* 86:349–54.
Weinberg, S. L., and R. A. Harshman. 1980. The structure of semantic space: A reexamination. Paper presented at the annual meeting of the National Council on Measurement and Education, San Francisco, April.
Weinberg, S. L., J. D. Carroll, and H. S. Cohen. Forthcoming. Confidence regions for INDSCAL using the jackknife and bootstrap techniques. Manuscript, Bell Laboratories, Murray Hill, New Jersey.