

6

Data Preprocessing and the Extended PARAFAC Model

Richard A. Harshman and Margaret E. Lundy

DATA PREPROCESSING

Definitions and Objectives

Definitions

Two Kinds of Preprocessing. By *data preprocessing* we mean any transformations of the data values that are performed before fitting the main analysis model. For convenience, we define two broad classes of preprocessing: data *conversion* and data *adjustment*. This distinction is somewhat arbitrary, but, if not taken too seriously, it will simplify discussion.

Data conversions take one kind of data and transform it into a different kind, usually entailing a change in the form of the model that is subsequently fit to the data. In chapter 5, we discussed two types of data conversions and their implications: (a) the conversion of profile data into covariances (or sums of cross-products) prior to indirect fitting, and (b) conversion of proximity data into scalar products as part of multidimensional scaling. Data adjustments, on the other hand, take data of a given kind and transform it into an improved or more standardized version of data of the same kind. Adjustments do not require a change in the form of the model, although they may implicitly extend the model being fit, as we shall see later. In this section, we will only consider data adjustments.

There are two basic types of data adjustments used before direct fitting of profile data: (a) additive adjustments—for example, to "center" the data; and (b) multiplicative adjustments—for example, to "rescale" or "normalize" the data. Occasionally, a third type of adjustment might also be useful: (c) nonlinear adjustments—such as log transformation to "linearize" the data, or the Weeks and Bentler (1979) rank transformation to deal with nonmetric data. (We will not consider this third case in this chapter.)

This article has been reprinted with the permission of the authors, Richard A. Harshman and Margaret E. Lundy (Copyright 1983). See page 122 for acknowledgments.

Additive Adjustments. In traditional two-way factor analysis, the raw score matrix is often transformed so that the scores for each variable have zero-mean. This involves a simple additive adjustment—the mean for each variable is subtracted from each score involving that variable. At other times, score matrices are transformed so that the mean for each individual is zero, producing what is called *ipsative* data. Both kinds of additive adjustments are commonly called *centering*, presumably because they place the origin or zero-point of the measurement scale at the mean of a set of data values. There are also other kinds of additive adjustments that are sometimes used. For example, in multidimensional scaling, an *additive constant* is estimated for each subject, which transforms that subject's interval-scale dissimilarities into ratio-scale distances. Such additive adjustment does not produce data with zero-mean, in part because negative distances are not permitted. (These additive constants were mentioned briefly in chapter 5.)

In this section, we will restrict our discussion of additive adjustments to consideration of various kinds of centering. There are many ways of centering a three-way array, but we will find that only some of these are appropriate methods of preprocessing for three-way factor analysis or multidimensional scaling. In particular, we will demonstrate that, to be appropriate, centering must be performed over one-way subarrays, such as rows, columns, or "tubes."

Multiplicative Adjustments. In two-way factor analysis, the data are often adjusted so that the variance of each variable is unity. This is accomplished by multiplying all the data points for each variable by the reciprocal of the standard deviation for that variable. In other applications (such as multidimensional scaling), the data may be adjusted so that the scores for each subject have a sum of squares equal to unity. In general, such multiplicative adjustments are accomplished by considering the data in subsets and multiplying the values in each subset by a constant selected so that the size of observations in the subset will have some desired property (for example, mean-square 1.0). When used to equate the size of observations across subsets, such multiplicative adjustments are often called *standardizing* or *normalizing* (although *standardizing* is also used to refer to the combination of additive and multiplicative adjustment that transforms a set of values to z-scores). Multiplicative adjustments can also be used to vary the overall data size across different subsets (for example, as a means of accomplishing a weighted least-squares analysis, to be described below).

Multiplicative adjustments are often useful in three-way applications, but there are many possible ways of performing such adjustments, and, once again, we shall find that only some of these are appropriate for three-way factor analysis or multidimensional scaling. It will be demonstrated that multiplicative adjustments, unlike centering, must be performed over two-way subarrays (that is, "slices" of the three-way array).

Objectives of Preprocessing

Preprocessing plays a very important role in three-way analysis. We saw in the previous chapter how conversion of profile data to

covariances or scalar products allows us to use indirect fitting, which permits modified or alternative models of the data, thus extending the power and generality of the PARAFAC procedure. Preprocessing plays an equally important role in direct fitting of profile data. Once again, it will allow us to fit extended versions of our analysis model and thus increase the generality of the three-way procedure. We will explore this extended PARAFAC perspective in more detail later in this chapter. In the present section, however, we adopt a complementary perspective, where preprocessing is used to make the data appropriate for the PARAFAC model. We will also focus on several additional advantages of preprocessing data before analysis.

Eight Objectives. There are at least eight basic reasons for preprocessing data before performing a PARAFAC analysis: (a) to make the data appropriate for the PARAFAC model by removing unwanted constants and eliminating some kinds of *conditionality* (to be explained below); (b) to emphasize relationships among patterns of change in certain modes, rather than among baseline values; (c) to weight and thus emphasize or ignore particular subsets of the data during the analysis, or, conversely, to equate the influence that different variables, stimuli, subjects, and so forth have on the form of the final solution; (d) to equate the size of presumed error variance components across different subsets of the data; (e) to standardize the data so that comparison of loadings across levels is facilitated; (f) to standardize the overall scale so that comparisons across data sets is facilitated; (g) to standardize the data so that useful added interpretations of the factor loadings are possible (for example, as correlations, variance components, and so on); and (h) to permit fitting of an extended PARAFAC model.

Objectives (a) and (b) are accomplished by various additive adjustments, and objectives (c) and (d) are accomplished by multiplicative adjustments. Finally, objectives (e) through (h) are accomplished by means of both additive and multiplicative adjustments. Of these various objectives, (a) is probably the most general and important application of preprocessing, and it will be considered in detail in the following discussion. After that, we will consider (b) through (g). Interpretation in terms of the extended PARAFAC model (objective [h], which is closely related to objective [a]) will be taken up in the latter part of the chapter.

Making One's Data More Appropriate for the PARAFAC Model

Restrictiveness of the Formal Model

The form of the basic PARAFAC-CANDECOMP model (5-3) appears to require what Stevens (1946) would call *ratio-scale* measurements. That is, the data values must be strictly proportional to the underlying quantities being measured, so that a zero in the data represents a complete lack of the property being measured, and 4 units represents exactly twice as much of the property as 2 units. This strong level of measurement appears necessary because the factor model is formulated in terms of proportional changes in factor contributions from one level of a given mode to

the next. Computing such proportional changes implies estimation of a true zero for each factor, which implies knowledge of the true zero-point on the variables from which those factors are to be extracted. Yet with much of the data encountered in the social sciences and elsewhere, one does not know the true zero; there are arbitrary origins and unknown additive constants that make the measurements interval scale at best. Furthermore, social sciences data is often "row-conditional" or "column-conditional"; that is, there may be a different measurement scale from one row to the next or from one column to the next. For example, each person may have a different subjective origin to his scale, or each variable might involve a different additive constant. Since there are no explicit terms in the model and no features in the fitting procedure designed to take care of such arbitrary origins and constants, Sands and Young (1980) have suggested that the practical application of the model may be seriously limited.

However, the model's formal measurement level is not as crucial as one might expect. While it is true that PARAFAC treats the data as ratio scale, so do all other metric factor analysis models, including Tucker's three-mode model and almost all varieties of traditional two-way factor analysis. Despite this formal limitation, factor analysts have not had trouble coping with interval-scale data. In part, this is because factor analysis was traditionally conceptualized in terms of correlation matrices. The conversion of raw data to correlations before performing factor analysis eliminated the problem of origin and so allowed investigators to avoid the issue. The first stages of such conversion implicitly perform a centering operation that transforms an interval-scale data matrix into appropriate ratio-scale form. (Such centering operations will be discussed below.) But conversion to correlations is not necessary for interval-scale data to be successfully factor-analyzed. As we shall see, additive constants can be represented as extra factors with constant loadings in one or more modes, and thus, in theory, can be easily accommodated within the factor-analytic model by slightly increasing the dimensionality of the solution. It can be argued, therefore, that the limitation of metric factor analysis to ratio-scale data is more apparent than real.

Unfortunately, however, extracting biases and constants as extra factors has drawbacks when one tries to discover the preferred *rotation* for the factor axes; hence, an alternative approach is preferred. Authors such as Horst (1965), who applied two-way factor analysis directly to raw profile data, found a simple solution to the question of origin that bypassed full consideration of measurement issues: First, convert the data to deviation score form (usually as z-scores), either by rows, columns, or both, and then apply the factor analysis to these standardized scores. This approach is usually rationalized in an ad hoc fashion, as a method of concentrating on deviations from a group mean; the mean is taken to be an alternative reference point suitable for cases in which the true origin was unknown or undefined. As we shall see, however, this standardizing procedure can be justified in terms of a more theoretically based rationale.

With three-way arrays, the question of proper centering and scale adjustment becomes much more subtle; many alternative possibilities exist, and the "easy" or "natural" approaches are not always correct. It becomes important to develop an explicit rational basis for choosing among different procedures for preprocessing. We advocate here a set of criteria based primarily on an algebraic analysis of the effects of the preprocessing on the latent structure of the data. To do this, we must begin with a simple model of the raw data. This model includes terms representing the factors we wish to extract, along with terms representing the undesirable constants and biases that obscure the true origin and make the data interval scale (and row- or column-conditional) rather than ratio scale. We then study the effects of alternative preprocessing transformations on both the factors and on the unwanted components. "Good" preprocessing will remove the undesired components while leaving the latent factors relatively unchanged.

We first apply this theoretical analysis to the question of preprocessing two-way arrays and demonstrate that it confirms the appropriateness of the methods currently used in two-way factor analysis. We then extend it to the three-way case to discover which of the many possible three-way centering and standardizing techniques are in fact appropriate for three-way factor analysis.

A Model for Interval-Scale Data

Two-Way Case. Suppose we have a data matrix \mathbf{X} that fits the structural model of factor or principal component analysis, except that there are additional constants present that obscure the origin of the variables. There may be a constant term h offsetting the data as a whole from true ratio-scale properties. In addition, if the origin of the data is row- or column-conditional, then there would be additional constants h_i or h_j , which would describe how the i th row or j th column deviates from ratio scale (after adjustment by h). If we let x_{ij} be the element in the i th row and j th column of \mathbf{X} , our expression for the latent structure of a data point might be written as follows:

$$x_{ij} = \sum_r (a_{ir}b_{jr}) + h_i + h_j + h + e_{ij} . \quad (6-1)$$

Here, as earlier, the a_{ir} and b_{jr} terms represent the factor loadings for Modes A and B , respectively, and e_{ij} represents random error.

For ease of discussion, let us interpret (6-1) in terms of a concrete example. Suppose that the data matrix \mathbf{X} is a two-way array of stimulus ratings, with rows corresponding to stimuli (pictures of products) and columns corresponding to judges, and that two factors underlie the systematic ratio-scale part of the data. All judges are rating a common set of stimuli on a particular quality (such as attractiveness). Thus, the data entry described by (6-1) is the attractiveness score of stimulus i as rated by judge j . The a -loadings indicate how much each stim-

ulus has of each type of attractiveness on a ratio scale. For example, the two types of attractiveness might be the aesthetic and economic advantages of the products. The b -loadings (which in some cases would be called *factor scores* or *factor score estimates*) are proportionality coefficients that relate the size of a given subject's ratings to the amount of each type of attractiveness possessed by the stimulus. When considered in this way, it becomes apparent that two-way factor analysis is formulated in terms of ratio-scale quantities.

In addition to the $a_{ir}b_{jr}$ terms representing the factor contributions, we have included additional h_i , h_j , and h -terms in the model, representing the unknown constant offsets for each row and column, and an overall additive constant, which disturbs proportionality and thus cause the data to have conditional interval-scale rather than unconditional ratio-scale properties. Let us consider the interpretations that these extra terms might have in our hypothetical analysis of rating scales. In such a study, the true zero-point for attractiveness would be hard to define and would almost certainly differ from a zero data value. In fact, the responses would typically be made on a rating scale, with points on the scale ranging from 1 to 9 and the two ends of the scale labeled "unattractive" and "attractive." It is clear that a rating of "2" on this scale is unlikely to represent twice as much attractiveness as a rating of "1." Thus, an overall additive constant is necessary to shift the zero-point somewhere nearer the middle of the scale. This is provided by the h -term. But a single constant may not be enough. If a given rater views "unattractive" as a negative quality equal and opposite to "attractive," the neutral point may be at 4.5. But for other raters who view "unattractive" as simply a lack of attractiveness, it may be nearer the low end of the scale. Because of such subject-to-subject variation, an h_j term is also needed to specify the shifts in the baseline for the j th rater. Finally, there might also be shifts in the baseline from stimulus to stimulus, although it seems less likely in this particular example. In other cases, however, shifts from row to row might seem plausible; for example, if the data were test scores, with rows representing the particular tests and columns representing the subjects. To provide for the general case, our model includes both an h_i and h_j term.¹

Direct factor analysis of our hypothetical ratings data, without preprocessing, would be an example of fitting a ratio-scale model to row- and column-conditional interval-scale data. If we only fit r -factors, the obtained loadings would be distorted by the presence of the constants h_i , h_j , and h in the data. However, the fact that the ratings do not properly reflect the zero-point of the attributes being measured can be easily corrected within the context of the two-way factor model. One solution would be to extract two extra dimensions. If, as in our example above, there were two "real" dimensions present, then a third and fourth dimension would be needed. The dimensions could be defined so that

$$a_{i3} = h_i, \quad a_{i4} = 1 \quad (6-2)$$

and

$$b_{j3} = 1, \quad b_{j4} = h_j + h.$$

(Here we have absorbed the overall bias h into the fourth dimension. There is no reason why it might not have been divided between the two dimensions or absorbed into the third dimension instead; the choice is arbitrary.) With such additional factors, one can represent conditional origin interval-scale data within the framework of a ratio-scale factor model.

For this example, there should be a four-dimensional solution that would recover the factor loadings for the first two dimensions correctly while isolating the constant biases into a third and fourth dimension. However, because of error in the data, the four factors that explain the largest proportion of the variance will not exactly correspond to the two "true" factors plus the two factors representing the additive constants. If the error is not too large, it should nonetheless be possible to find a rotation in which these factors are closely approximated.

PARAFAC Three-Way Case. Both the PARAFAC model and Tucker's three-mode model focus on components that an analysis of variance would call *three-way interactions*. That is, they describe components whose pattern of change from one level of Mode A to the next depends simultaneously on the levels of Mode B and Mode C involved, as well as the levels of Mode A . The sizes of the factor contributions normally vary as a function of all three modes, because normally the loadings for a given factor vary in size across the levels of all three modes. Thus, in order to take all the unwanted offsets into account, we need to consider the consequences of not only the global constant h and "one-way" effects (the h_i , h_j , and now h_k terms) that are constant across two modes, but also the consequences of "two-way" effects, which arise out of the interaction of a particular pair of modes and are constant over a third mode.

For simplicity, let us return to our example involving rating-scale data. Suppose that each judge now rates each stimulus on several scales; in addition to "unattractive-attractive," he also rates the stimulus pictures on "practical-impractical," "fun-dull," "appealing-unappealing," "economical-uneconomical," and so on. This would generate a three-way array whose general entry would be the rating of stimulus i by judge j using scale k . As before, we might want to have an overall baseline term and terms for biases due to particular people, stimuli, and now scales. But we can also imagine that a particular person has an idiosyncratic interpretation of a particular scale and uses that scale as if it had a different neutral point. This would be a two-way interaction of person and scale that would be constant across stimuli. We might adjust for such effects by means of a doubly subscripted constant h_{jk} . Similarly, we should consider possible two-way interactions of particular people with particular stimuli and particular scales with particular stimuli. For example, if one of the stimulus pictures included a beautiful model, the scale "attractive-unattractive" might take on a special meaning with respect to this stimulus, which would uniformly alter the neutral point of ratings for this stimulus-scale combination across all subjects. Thus, within the PARAFAC framework, we might represent three-way

arrays of interval-scale data in terms of the following general form:

$$x_{ijk} = \sum_r (a_{ir}b_{jr}c_{kr}) + h_{ij} + h_{ik} + h_{jk} + h_i + h_j + h_k + h + e_{ijk} . \quad (6-3)$$

This representation could easily be adapted to incorporate the Tucker three-mode model or other metric three-way models by simply modifying the first term. In (6-3) the first term represents the PARAFAC-CANDECOMP factor model; by replacing it with the corresponding triple summation from Tucker's model and applying the algebraic transformations discussed below, we could prove the appropriateness or inappropriateness of different types of preprocessing for these models. It quickly becomes apparent, however, that the consequences of the transformations on the h -terms are the same regardless of the exact representation of the factors and that the consequences of the transformations on the factor component are similar for all current metric three-way models. Thus, all of the conclusions reached concerning representation of interval-scale data and the appropriateness of different centerings will apply equally to PARAFAC and Tucker's three-mode or other metric three-way models, with the exception of those considerations that involve the intrinsic axis property of the PARAFAC-CANDECOMP model.

Now, since the unsubscripted and singly subscripted h -terms can be absorbed into the doubly subscripted ones as the investigator sees fit, the general expression for a interval-scale three-way array might be written more simply:

$$x_{ijk} = \sum_r (a_{ir}b_{jr}c_{kr}) + h_{ij} + h_{ik} + h_{jk} + e_{ijk} . \quad (6-4)$$

Given this representation, let us consider how it can be incorporated within the ratio-scale framework of three-way factor analysis. The effects represented by the h_{ij} term, for example, cannot in general be represented as a single factor. In fact, this set of effects constitutes a two-way array (replicated identically across all levels of the third mode) which may have a complex structure, with a rank possibly equal to the number of rows or columns (whichever is less) in the h_{ij} table. Hopefully, however, the systematic part of the two-way biases will have a relatively low dimensionality; we might therefore approach the representation of these biases as a traditional two-way factor-analytic problem. We could decompose each of the two-way arrays of h -values into a set of factors in order to represent its effects in terms of our model.

Although decomposing a two-way table of, say, h_{ij} values would normally yield two-way factors, in order to incorporate these effects into our three-way factor model, we could express them in terms of three-way factors with constant loadings across a third mode. This would give us a three-way (such as PARA-

FAC) representation of the three-way array of bias components that is added to the data array by the h_{ij} term in equation (6-3). Obviously, the h_{ik} and h_{jk} terms could be represented in analogous fashion, and so we could rewrite our expression for the interval-scale data as:

$$x_{ijk} = \sum_r (a_{ir}b_{jr}c_{kr}) + \sum_s (a_{is}b_{js}c_{ks}) \quad (6-5) \\ + \sum_t (a_{it}b_{jt}c_{kt}) + \sum_u (a_{iu}b_{ju}c_{ku}) + e_{ijk} ,$$

with the constraints that:

$$\begin{aligned} \text{all } c_{ks} &= \text{a constant,} \\ \text{all } b_{jt} &= \text{a constant, and} \\ \text{all } a_{iu} &= \text{a constant.} \end{aligned} \quad (6-6)$$

Thus, two-way interactions can be decomposed into a set of two-way factors, which can be incorporated into the three-way model as factors that have constant loadings over a third mode. If desired, these two-way factors could further be decomposed into deviations plus *main effects* (that is, factors that are constant across two modes), to more closely parallel the longer form of (6-3). Such additional decomposition would be purely to facilitate some preferred scheme of interpretation, however, since no improvement in fit would be obtained by the extra terms and the particular manner of the decomposition would be somewhat arbitrary.

Disadvantages of Treating Constants as Extra Factors. Although (6-5) shows how the factor-analytic model can be considered appropriate for conditional-origin interval-scale data, the representation of bias and constant terms as extra factors can pose problems for the actual processes of factor estimation and rotation. As noted earlier, the presence of random error in the data normally prevents the extraction of factors with precisely constant loadings; factors often emerge with some of the constant variance mixed in with the variance of factors that are nonconstant across all modes. Even if it were possible to extract the constants unperturbed by error, existing methods for factor rotation do not automatically find the desired solution—one that has dimensions of the form specified in (6-6) or, in the two-way case, as described by the constraints of (6-2).

When fitting the PARAFAC model to three-way arrays, a further consideration is involved: We want to preserve the intrinsic axis property of the solution. As noted in chapter 5, however, it is generally necessary that factors show distinct patterns of variation across *all three modes* for their contributions to be unambiguously sorted out from other possible linear combinations with other factors (that is, in order for their axis orientation to be uniquely determined). It is apparent from (6-5) and (6-6) that the extra factors that we might incorporate to represent constants and two-way interactions do not show the necessary distinct patterns of variation across all three modes. Indeed, since they are defined by what are essentially two-way arrays

(replicated across a third mode), they show the classic rotational indeterminacy of two-way factor analysis. If the sum of squares due to these extra factors is substantial, then they tend to emerge even in solutions of low dimensionality, probably mixed in with the unique three-way factors that we are trying to identify. The indeterminacy of these two-way interaction factors causes the factors of interest to appear indeterminate, and the resulting solution in general is quite hard to interpret. Thus, the approach of treating unwanted constants as extra factors seems to have too many drawbacks to recommend it for normal three-way PARAFAC analysis. There is, however, an alternative way to deal with the unwanted constants: Remove them by appropriate preprocessing.

Transforming Interval-Scale into Ratio-Scale Data

Instead of including extra factors in our model to represent any unwanted constants or biases that may be present in the data, another approach is to remove such constants by special data transformations preceding the factor analysis, or in other words, by data preprocessing. First, we will show how this is done for two-way data and then extend it to the three-way case.

Two-Way Case. As mentioned earlier, the problem of arbitrary origin was traditionally dealt with in two-way factor analysis by means of *centering*, removing means from the rows and/or columns of a data matrix. This set the new origin at the group mean on each row or column. The procedure was justified as a way of focusing attention on the deviations from the mean. But is such a justification adequate? What information is lost or distorted by this procedure? In more precise terms: How is our interpretation of the underlying "true" factors of these centered scores different from the interpretation we would have obtained could we have factored the original uncentered data without the interfering constants?

If we algebraically examine the traditional centering process using our model of interval-scale data ([6-1]), we see that centering has very desirable properties. It removes the unwanted and unknown constants that obscure the origin of the data and at the same time performs only a minor transformation on the factors in which we are interested. Since this transformation does not usually interfere with our ability to interpret the solution, we are able to get essentially the same information that we would have obtained had the original data been ratio scale in the first place. This algebraic result provides a firm basis for concluding that centering is an appropriate method of preprocessing in two-way factor analysis.

Let us adopt the convention that a dot in place of a subscript means that the term has been averaged across the levels of that subscript. For example, if there are n levels to Mode A , then

$$x_{.j} = \frac{1}{n} \sum_{i=1}^n (x_{ij}) , \quad (6-7)$$

and similarly for other subscripted terms.

Let us return again to our example of the ratings matrix \mathbf{X} , where x_{ij} represents the rating of the i th stimulus by the j th judge. Suppose the original data differs from ratio scale because of an overall constant h and an offset h_j for each judge, but there are no differential offsets for each stimulus. Thus, our data point before centering has the underlying structure

$$x_{ij} = \sum_r (a_{ir}b_{jr}) + h_j + h + e_{ij} . \quad (6-8)$$

We wish to determine what happens to this structure after centering. We start with the description of the effects of centering defined in terms of the observed scores. If we let \dot{x}_{ij} represent the data point after centering across stimuli, then we can start with the expression:

$$\dot{x}_{ij} = x_{ij} - x_{.j} = x_{ij} - \frac{1}{n} \sum_r (x_{ij}) . \quad (6-9)$$

To determine the effects on the underlying structure, we substitute from (6-8) into (6-9) and simplify, obtaining:

$$\begin{aligned} \dot{x}_{ij} &= \sum_r (a_{ir}b_{jr}) + h_j + h + e_{ij} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \left(\sum_r (a_{ir}b_{jr}) + h_j + h + e_{ij} \right) . \end{aligned} \quad (6-10)$$

By distributing the summation over i and rearranging the terms of (6-10), we obtain:

$$\begin{aligned} \dot{x}_{ij} &= \sum_r (a_{ir}b_{jr}) + h_j + h + e_{ij} - \sum_r (b_{jr} \left(\frac{1}{n} \sum_{i=1}^n (a_{ir}) \right)) \\ &\quad - \frac{1}{n} \sum_{i=1}^n (h_j) - \frac{1}{n} \sum_{i=1}^n (h) - \frac{1}{n} \sum_{i=1}^n (e_{ij}) . \end{aligned}$$

Since the average of a constant is the same constant, the sixth and seventh terms are not affected by the summation. We can use the dot notation to simplify the remaining terms and obtain:

$$\begin{aligned} \dot{x}_{ij} &= \sum_r (a_{ir}b_{jr}) + h_j + h + e_{ij} - \sum_r (b_{jr}(a_{.r})) \\ &\quad - h_j - h - e_{.j} \end{aligned}$$

Collecting like terms, we obtain:

$$\dot{x}_{ij} = \sum_r ((a_{ir} - a_{.r})b_{jr}) + (e_{ij} - e_{.j}) .$$

If we let $\bar{a}_{ir} = a_{ir} - a_{.r}$ and $\bar{e}_{ij} = e_{ij} - e_{.j}$, we have the following simple expression for the underlying structure of the data matrix after centering:

$$\bar{x}_{ij} = \sum_r (\bar{a}_{ir} b_{jr}) + \bar{e}_{ij} . \quad (6-11)$$

The consequences of centering the data are simple: The factor loadings for one mode are centered, as are the error terms, and the constants h and h_j disappear. The centered data is now ratio scale and can be represented (aside from the random error) simply in terms of the underlying factors of attractiveness. We can recover these factors accurately after the centering operation, since the b -coefficients (in this example, person weights or factor score estimates) are unaffected by the centering, and the a -coefficients (or factor loadings) are the same except that the mean value of a for each factor has been subtracted out (that is, the loading matrix \mathbf{A} is column-centered). Thus, the shape of the profile of factor loadings is unaffected, and only the elevation is changed. Since most of the interpretation of a factor is based on the relative values of the loadings on the different stimuli, rather than the absolute size of the loadings, the centering operation does not normally interfere with our interpretation of the factor.

Similarly, if there had been bias or offset terms that differed across stimuli but were constant across persons (h_j terms), then centering across persons would have removed these terms. In this case the factor scores or b -terms would have been centered but not the loadings or a -terms. As noted earlier, the occurrence of h_j terms might be more plausible if our data consisted of ability test scores for a group of people; this is where centering across persons is most commonly employed.

Centering both rows and columns of \mathbf{X} can be accomplished by applying the second centering operation to the result obtained by the first centering. If we use \bar{x}_{ij} to represent an element of the double-centered data array $\bar{\mathbf{X}}$, then we could compute it as follows:

$$\bar{x}_{ij} = \bar{x}_{ij} - \bar{x}_{i.} .$$

Double-centering removes both h_i and h_j terms; it also column-centers both the Mode A and Mode B tables of factor loadings and double-centers the error terms. It otherwise leaves the latent structure unchanged. As has been indicated elsewhere (Horst 1965), the effects of double-centering are the same, regardless of the order of the centering operations performed; this becomes clear when one sees that double-centering can also be represented—more compactly but less transparently—as a single operation defined in terms of the original row and column means:

$$\bar{x}_{ij} = x_{ij} - x_{i.} - x_{.j} + x_{..} .$$

When the response scale is clearly offset from psychological zero, as in the rating scale example, the most important consequence of the centering operation is often the removal of the

overall constant h . This could be accomplished by centering in either direction. However, the elimination of the conditionality of origin can also be important, and this requires centering the correct mode.

Matrix Formulation of the Effects of Centering. In the preceding discussion, we used scalar notation to deduce the relationships between the form of a given factor before and after centering. We used this method because it easily generalizes to n -way arrays. However, with two-way arrays, there is a matrix representation of centering that allows one to demonstrate its effect on the latent structure most clearly. Horst (1965) defines a matrix that he calls a *centering matrix*. We will call this matrix L_n (for the n by n version of L needed for column-centering) or L_m (for the m by m version needed for row-centering). The matrix is defined as follows:

$$L_n = I - \frac{1}{n} (\mathbf{1}\mathbf{1}') , \quad (6-12)$$

where $\mathbf{1}$ is an n -element column vector, all of whose entries are 1, and I is an n by n identity matrix. The matrix L_n is symmetric, idempotent, and has rank $n - 1$. Its rows and columns sum to zero. Its diagonal elements are equal to $(n - 1)/n$, and off-diagonal elements are equal to $-1/n$. The matrix

$$\overset{*}{X} = L_n X \quad (6-13)$$

is the column-centered version of X ; that is, the sum of each column of $\overset{*}{X}$ is zero.

To determine the effects of centering on the latent structure of X , we may replace X by its representation in terms of factors. So, if

$$X = A B' + E , \quad (6-14)$$

where A is the n by r matrix of factor loadings and B is the m by r matrix of person loadings (factor scores or component scores), then by premultiplying both sides of (6-14) by L_n , we obtain

$$L_n X = L_n (A B' + E) \quad (6-15)$$

or

$$L_n X = (L_n A) B' + (L_n E) .$$

If we let $\overset{*}{A} = L_n A$ and $\overset{*}{E} = L_n E$, then

$$\overset{*}{X} = \overset{*}{A} B' + \overset{*}{E} . \quad (6-16)$$

That is, column-centering X simply column-centers the A matrix of factor loadings and the matrix of error terms, as we determined earlier in scalar notation. Similarly, we can row-center X by multiplying it on the right by an m by m matrix L_m , defined analogously to L_n ($L_m = I - (1/m)(\mathbf{1}\mathbf{1}')$ where $\mathbf{1}$ is a m -element

column vector all of whose entries are 1). The result is:

$$\mathbf{X} \mathbf{L}_m = (\mathbf{A} \mathbf{B}' + \mathbf{E}) \mathbf{L}_m ,$$

and so

$$\mathbf{X} \mathbf{L}_m = \mathbf{A} (\mathbf{B}' \mathbf{L}_m) + \mathbf{E} \mathbf{L}_m . \quad (6-17)$$

Thus, row-centering \mathbf{X} simply row-centers the matrix \mathbf{B}' and the error matrix \mathbf{E} . Since \mathbf{B}' is the transpose of the person-loading matrix \mathbf{B} , row-centering \mathbf{B}' is equivalent to column-centering \mathbf{B} .

Finally, as noted earlier, we can apply both row- and column-centering to a data matrix, in either order, with the same result: Both row sums and column sums will be zero. The effect of double-centering on the latent structure is:

$$\mathbf{L}_n \mathbf{X} \mathbf{L}_m = (\mathbf{L}_n \mathbf{A}) (\mathbf{B}' \mathbf{L}_m) + \mathbf{L}_n \mathbf{E} \mathbf{L}_m = \overset{*}{\mathbf{A}} \overset{*}{\mathbf{B}'} + \overset{\circ}{\mathbf{E}} . \quad (6-18)$$

Double-centering the data matrix simply column-centers both matrices of factor loadings and double-centers the error matrix.

Some Interpretations of the Effects of Centering. From these two-way results, we can state a rule that will enable us to generalize to three-way arrays: Centering across any mode of the data matrix simply centers the errors across that mode and column-centers the factor-loading matrix for that mode. For any given axis orientation, a column-centered factor matrix generally leads to the same interpretations as an uncentered factor matrix. Geometrically, after centering, the axes are parallel to the corresponding axes before centering; they have simply been translated as a unit from the original arbitrary origin to a new origin at the centroid of the configuration. Thus, centering by rows or columns can be said to *preserve the factor structure* underlying the data.²

The conclusion that centering the data across levels of a given mode centers the corresponding factor loadings also implies that centering removes unwanted components that are constant across that mode. We showed earlier how terms that are constant across levels of a given mode are equivalent to factors with constant loadings in that mode. Because centering the data across levels of a given mode column-centers the factor-loading matrix for that mode, those factors with constant loadings in the centered mode are transformed into factors with zero loadings in that mode and hence vanish.

The above discussion shows that the effects of different centerings are quite straightforward in the two-way case. Nonetheless, there has been considerable controversy on this point, with some authors claiming that different factors can be uncovered by different centerings. One reason that some investigators may have gotten this impression is that different centerings can give different relative *emphasis* to the several dimensions underlying a data set; the relative emphasis of dimensions before and after centering is determined by the relative sizes of the sums of squares of their original loadings compared to their column-centered loadings. This change in sums of squares will change the relative contributions of the dimensions to each unrotated factor

or principal component. If some nonnegligible components are discarded to obtain a lower-dimensional solution, then different parts of the factor space might be discarded in the centered versus uncentered solutions.

Perhaps the most serious obstacle to recognizing the sameness of factors before and after centering has been the rotational indeterminacy of the two-way model. Centering either mode will change the orientation of the principal components, and centering across levels of a given mode will also change the best simple structure orientation of the factors in that mode. These artificial differences have led authors to overlook the essential sameness of the centered factors.

For example, Horst (1965) examined the effects of centering on the underlying dimensions by comparing the basic structure (similar to the unrotated principal components or singular value decomposition) of a data matrix before centering to its basic structure after centering. But the orientation of axes in the basic structure solution is determined by the directions that successively maximize variance-accounted-for. These directions will usually be different after the translation of origin that results from centering, and so comparing the two basic structures amounts to comparing factors that have not only been translated but also rotated into different orientations. Naturally, this obscures the essential sameness of the dimensions. Although Horst presented a complex procedure for deriving the centered basic structure solution from the noncentered, and vice versa, he failed to mention the very simple relation between the two structures that applies *with suitable rotation to restore the equivalence of axis orientations in the factor space*.

The straightforward effects of centering may also be obscured if the pre- and postcentered versions are rotated to simple structure using Varimax, Oblimin, graphical rotation, or some other such criterion. Column-centering a factor-loading matrix changes the number and positions of near-zero entries in the matrix and thus affects the axis orientations obtained when simple structure or related analytic rotation techniques are applied to the dimensions of the centered mode. By centering across variables, for example, and then rotating the variable loadings to simple structure, one could easily get the idea that novel dimensions have been revealed by this ipsative analysis. Once again, this is an illusion caused by looking at the same configuration from a different angle and location. Furthermore, it could be argued that the empirical rationale behind rotation in search of hyperplanes does not directly apply to a mode that has been centered; any hyperplane that results from a number of variables being unaffected by a given factor should go through the true ratio-scale origin of the space and not the centroid origin that results from centering.

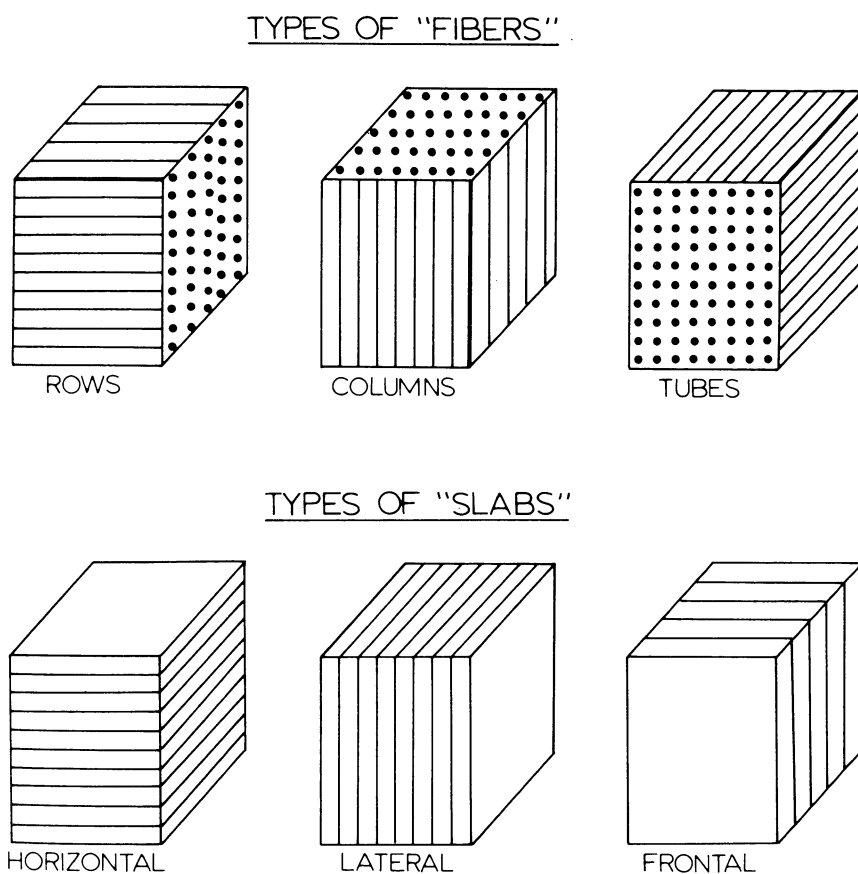
It must be acknowledged, however, that the error terms are also modestly affected by centering and hence the perturbations of the true factor loadings due to error may change slightly with different centerings. If the error is random with expected value of zero, the deviations from zero-mean error in the columns or rows of any particular matrix of observed errors will usually be small and so centering would not normally have much effect on these components.

Extension to Three-Way Centering. In three-way factor analysis of raw or profile data, the question of origin becomes considerably more complicated. Thus, we must start by defining some terms. There are three types of centering that might seem natural. We have called them *fiber-*, *slab-*, and *grand mean-* (*global-*) centering; Kruskal (chapter 2) calls them *one-way*, *two-way*, and *three-way centering*.³

Fiber- or one-way centering subtracts means computed over one-way subarrays. Since there are three kinds of fibers—rows, columns, and "tubes" (see Figure 6-1)—there are three corresponding varieties of fiber-centering. Slab- or two-way centering subtracts means computed over two-way subarrays. Since there are three kinds of slabs—lateral, horizontal, and frontal (see Figure 6-1)—there are also three types of slab-centering. Grand mean- or three-way centering subtracts the mean computed over the entire three-way array (that is, the *grand mean*). There is only one variety of three-way centering.

Fiber-centering removes means computed across all levels of a given mode, holding the level of the other two modes fixed. Thus, we say that fiber-centering is done across Mode A, across

Figure 6-1. Fibers and Slabs in a Three-Way Data Array



Mode *B*, or across Mode *C*. In slab-centering, means are computed across all levels of two modes, holding the third fixed. Thus, we say that slab-centering is done across Modes *A* and *B*, across Modes *B* and *C*, or across Modes *A* and *C*.

Examples of each type of centering are:

$$\bar{x}_{ijk}^* = x_{ijk} - x_{...} \quad \text{global- (three-way) centering;} \quad (6-19)$$

$$\bar{x}_{ijk}^* = x_{ijk} - x_{..k} \quad \text{slab- (two-way) centering} \quad (6-20)$$

across Modes *A* and *B*;

$$\bar{x}_{ijk}^* = x_{ijk} - x_{.jk} \quad \text{fiber- (one-way) centering} \quad (6-21)$$

across Mode *A*.

(Here, again, a dot in place of a subscript indicates that a mean was computed across all levels of the missing subscript.)

To determine which of these centering procedures would provide appropriate preprocessing for three-way factor analysis, we will algebraically determine their effects on a model of three-way interval-scale data. For example, when testing different centerings as possible preprocessing for PARAFAC, we will use a model of data that incorporates the PARAFAC model of the factor component, along with our standard terms for the two-way, one-way, and overall biases. Similarly, we can see which centerings provide appropriate preprocessing for Tucker's model by applying them algebraically to a model of interval-scale data in which the T2 or T3 representation of factors replaces the PARAFAC-CANDECOMP representation. It will become apparent, however, that the same conclusions about appropriate preprocessing hold for all current metric three-way factor models.

The model of three-way interval-scale data ([6-3]) is rather long and unwieldy. To simplify our discussion, we will break it into three parts. We can consider the expression for a data point to be constructed of error plus two different systematic parts: (a) a systematic trilinear part represented by the basic factor model (initially taken to be the PARAFAC-CANDECOMP model), which we will call t_{ijk} ; and (b) a systematic but troublesome part not represented by the basic PARAFAC-CANDECOMP model (except by adding factors with constant loadings in one or more modes), which we will call h_{ijk} . This latter part is a sum of the parts explicitly represented in (6-3) as two-way and one-way terms plus an overall constant. Thus,

$$x_{ijk} = t_{ijk} + h_{ijk} + e_{ijk} \quad , \quad (6-22)$$

where

$$t_{ijk} = \sum_r (a_{ir}b_{jr}c_{kr}) \quad (6-23)$$

and

$$h_{ijk} = h_{ij} + h_{ik} + h_{jk} + h_i + h_j + h_k + h \quad . \quad (6-24)$$

We are now able to test different types of centering to see if they eliminate the h_{ijk} component but leave the t_{ijk} component relatively unchanged (that is, changed only in a simple way that does not increase the dimensionality or obscure the interpretation). First, we note that we can investigate the effects on the two systematic components separately, since the linearity of the centering operation insures that the effects on each part are independent of the other part. For example, since

$$\begin{aligned} x_{.jk} &= \frac{1}{n} \sum_i (t_{ijk} + h_{ijk} + e_{ijk}) \\ &= \frac{1}{n} \sum_i (t_{ijk}) + \frac{1}{n} \sum_i (h_{ijk}) + \frac{1}{n} \sum_i (e_{ijk}) , \end{aligned}$$

we can rewrite this as:

$$x_{.jk} = t_{.jk} + h_{.jk} + e_{.jk} . \tag{6-25}$$

Substituting into our general expression for one-way centering and grouping like terms, we obtain:

$$\begin{aligned} x_{ijk} - x_{.jk} &= (t_{ijk} - t_{.jk}) + (h_{ijk} - h_{.jk}) \\ &\quad + (e_{ijk} - e_{.jk}) . \end{aligned} \tag{6-26}$$

In other words, centering the data by a given method can be thought of as independently centering each of the additive components of the data by that method.

Let us now ask what the effects of the different centerings are on the trilinear and bilinear parts of our model. We begin with fiber-centering, which happens to have desirable properties.

Centering of One-Way Arrays. Applying fiber- or one-way centering to the trilinear part of the model, we obtain:

$$\begin{aligned} t_{ijk} - t_{.jk} &= \sum_r (a_{ir}b_{jr}c_{kr}) \\ &\quad - \frac{1}{n} \sum_i (\sum_r (a_{ir}b_{jr}c_{kr})) \end{aligned} \tag{6-27}$$

$$\begin{aligned} t_{ijk} - t_{.jk} &= \sum_r (a_{ir}b_{jr}c_{kr}) \\ &\quad - \sum_r (b_{jr}c_{kr} (\frac{1}{n} \sum_i ((a_{ir})))) \end{aligned} \tag{6-28}$$

$$t_{ijk} - t_{.jk} = \sum_r (b_{jr}c_{kr} (a_{ir} - a_{.r})) . \tag{6-29}$$

If, as before, we place a "*" above a centered component, we obtain:

$$t_{ijk}^* = \sum_r (a_{ir}^* b_{jr}^* c_{kr}^*) . \quad (6-30)$$

Thus, fiber-centering has a very simple effect on the trilinear part of three-way arrays, strictly analogous to the effect of traditional centering on the bilinear part of the two-way arrays: It results in the same number of dimensions and the same loadings, except that the loadings for the mode across which the centering was done are themselves column-centered. In this case, since we centered across Mode A, the Mode B and C loadings were unaffected, and the Mode A loadings for each factor had the mean Mode A loading for that factor subtracted out.

Now, to complete our examination of fiber-centering, we must test its effects on the troublesome h_{ijk} part. We proceed in the same way as before:

$$h_{ijk} - h_{.jk} = (h_{ij} + h_{ik} + h_{jk} + h_i + h_j + h_k + h) \quad (6-31)$$

$$- (h_{.j} + h_{.k} + h_{jk} + h_{.i} + h_j + h_k + h)$$

$$h_{ijk} - h_{.jk} = (h_{ij} - h_{.j}) + (h_{ik} - h_{.k}) \quad (6-32)$$

$$+ (h_{jk} - h_{jk}) + (h_i - h_{.i})$$

$$+ (h_j - h_j) + (h_k - h_k) + (h - h)$$

$$h_{ijk}^* = h_{ij}^* + h_{ik}^* + 0 + h_j^* + 0 + 0 + 0 . \quad (6-33)$$

We find that fiber-centering across Mode A causes all the h -terms that do not have an i -subscript to vanish. In other words, fiber-centering across Mode A removes all the terms that are constant across the levels of Mode A. For the other components, which vary across levels of Mode A, the original terms are replaced by the corresponding i -centered versions.

The effect of fiber-centering on the error term is simply to center the errors across that mode; that is,

$$e_{ijk} - e_{.jk} = e_{ijk}^* . \quad (6-34)$$

Because our model of three-way interval-scale data treats all three modes symmetrically, it follows that the effects of centering across Mode B and Mode C are equivalent to the effects of Mode A centering, shown above, except that the affected modes are correspondingly changed. Thus, centering across Mode B causes the b_{jr} term of the trilinear part to be centered, the h -terms that do not have j -subscripts to vanish, and the error components to be centered across Mode B.

As with two-way arrays, a three-way array may be double-centered or triple-centered by centering across several modes in succession. The order of the operations does not matter, so long as the values input to each successive centering operation are the residuals from the prior stage of centering. Thus, if we first center over Mode A, then the doubly centered data point is computed as follows:

$$\hat{x}_{ijk} = \bar{x}_{ijk} - \frac{1}{m} \sum_j (\bar{x}_{ijk}) . \quad (6-35)$$

It does *not*, in general, equal

$$x_{ijk} - x_{.jk} - x_{i.k} .$$

With fiber-centering, it is possible to represent the effects of centering a three-way array by matrix notation. If we consider the three-way array to be composed of a set of p two-way arrays, each n by m , then we can represent double-centering across both Modes A and B by extending (6-18) to the three-way case as follows:

$$\{x_{ijk}\} = \{\mathbf{X}_k\} = \text{the three-way array} , \quad (6-36)$$

$$\mathbf{L}_n (\mathbf{X}_k) \mathbf{L}_m = \mathbf{L}_n (\mathbf{A} \mathbf{D}_k \mathbf{B}' + \mathbf{E}) \mathbf{L}_m , \text{ and} \quad (6-37)$$

$$\mathbf{L}_n (\mathbf{X}_k) \mathbf{L}_m = (\mathbf{L}_n \mathbf{A}) \mathbf{D}_k (\mathbf{L}_m \mathbf{B})' + \mathbf{L}_n \mathbf{E} \mathbf{L}_m . \quad (6-38)$$

If we represent the doubly centered data as $\hat{\mathbf{X}}$, then

$$\hat{\mathbf{X}}_k = \hat{\mathbf{A}} \mathbf{D}_k \hat{\mathbf{B}}' + \hat{\mathbf{E}} . \quad (6-39)$$

That is, double-centering the three-way array by means of two fiber-centerings is equivalent to double-centering each constituent \mathbf{X}_k matrix. The result is that the Mode A and Mode B factor-loading tables are column-centered, and the error component is double-centered.

If we represent our unwanted h -terms as extra factors that are constant across the levels of one or more particular modes, then we can see that double-centering eliminates all the factors that are constant in either Mode A or B , since both of their factor-loading tables are column-centered. The only h -terms that remain are those that vary across both Modes A and B but are constant across levels of Mode C .

The more stages of centering that are performed, the greater the number of h -terms that are removed. Single-centering removes the overall constant h and two of the three one-way effects, along with one of the three two-way effects. Double-centering removes the remaining one-way effect and another one of the two-way effects, leaving only one of the doubly subscripted h -terms remaining. Finally, it can easily be shown that triple-centering the data array removes all of the h -terms and leaves all three of the factor-loading tables for the trilinear part column-centered.

In theory, then, triple application of fiber-centering leaves the data in ideal condition, with all the troublesome constants and one- and two-way components removed. In practice, however, triple-centering can have an additional, less desirable effect. It sometimes causes too severe a reduction in the "signal-to-noise" ratio in the data, as will be discussed below. Since double-centering removes all the unwanted components except one, it can be very effective in cleaning up even difficult data. It is often

possible to choose the modes to be centered so that the two-way term remaining after double-centering is so small as to be insignificant. Thus, double-centering is often the optimum method of removing unwanted constants from one's data. Naturally, however, not all data require double-centering. Some may not require any centering at all, although this would be unusual.

Centering the data centers one or more tables of factor loadings, and this can sometimes affect one aspect of factor interpretation: It can become more difficult to distinguish unipolar factors (with high influence on a few variables and no effect on the rest) from certain kinds of bipolar factors (those with high influence on some variables and moderate opposite influence on many other variables). The problem does not arise with more clear-cut bipolar factors, with roughly equal numbers of loadings at each pole and a larger group of loadings closer to the origin. However, when the configuration of loadings does not make the bipolarity of a factor clear and the unipolar-bipolar distinction is important in a particular mode, then one might want to consider centering another mode instead.

Centering of Two-Way Subarrays. We can now compare the effects of fiber-centering versus slab-centering. Many investigators have considered slab-centering to be the most natural kind for three-way factor analysis. Yet, as we shall shortly see, it has undesirable effects. It does not remove all the h -terms and, most seriously, it distorts the trilinear part of the data, increasing its dimensionality.

In slab- or two-way centering, the mean is computed for all the points at a given level of a particular mode (namely, for all the points in that two-way subarray or slice of the data) and then subtracted from all the points at that level. For example, means might be obtained for each stimulus by averaging across persons and scales; the points are then centered as follows:

$$\bar{x}_{ijk} = x_{ijk} - x_{i..} = x_{ijk} - \frac{1}{mp} \sum_j \left(\sum_k (x_{ijk}) \right). \quad (6-40)$$

Now it is easily shown (by arguments parallel to those made above for fiber-centering) that slab-centering the data is equivalent to slab-centering each additive part in our representation of the data. Hence, we once again can simplify our discussion by considering the effects of centering on each of the three components of the data ([6-22]) separately. First, let us consider its effects on the trilinear part, which after slab-centering would have the form:

$$t_{ijk} - t_{i..} = \sum_r (a_{ir} b_{jr} c_{kr}) - \frac{1}{mp} \sum_r \left(\sum_j \left(\sum_k (a_{ir} b_{jr} c_{kr}) \right) \right). \quad (6-41)$$

Rearranging the terms, we obtain:

$$t_{ijk} - t_{i..} = \sum_r (a_{ir} \{ b_{jr} c_{kr} - \frac{1}{mp} (\sum_j [\sum_k (b_{jr} c_{kr})]) \}) \quad (6-42)$$

Unfortunately, no further simplification of (6-42) is possible. Obviously, the effect of slab-centering on the trilinear part of the data is not as simple as the effect of fiber-centering, since there is no direct representation in terms of changes in the Mode *B* or *C* loadings.

This example shows that slab-centering across Modes *B* and *C* removes the mean $b_{jr}c_{kr}$ product from each $b_{jr}c_{kr}$ part before it is multiplied by the a_{ir} coefficient. A more straightforward statement of the effects of slab-centering on the trilinear part is given by (6-41). Slab-centering across Modes *B* and *C* adds an artificial dimension, whose loadings in Mode *A* are -1.0 times the mean value in each Mode *A* slice, and whose loadings in Modes *B* and *C* are all equal to 1.0. Thus, the constant that is added to center the data becomes another constant factor, rather than simplify^{ing} the solution.

Let us now look at the effects of slab-centering on the h_{ijk} part of the data. We begin with:

$$h_{ijk} - h_{i..} = (h_{ij} + h_{jk} + h_{ik} + h_i + h_j + h_k + h) - (h_{i.} + h_{..} + h_{i.} + h_i + h_{.} + h_{.} + h) \quad (6-43)$$

$$h_{ijk} - h_{i..} = (h_{ij} - h_{i.}) + (h_{jk} - h_{.}) + (h_{ik} - h_{i.}) + (h_j - h_i) + (h_j - h_{.}) + (h_k - h_{.}) + (h - h) \quad (6-44)$$

The effect of centering across two-way subarrays is to eliminate any h -terms that are constant in those subarrays. In our example, we centered across the bc subarray for each level of i . This eliminated the h -terms that were constant across both b and c (that is, constant across all the cells in each bc slice). Thus, a single application of two-way centering will remove any overall constant h and a single one-way constant (for example, h_i). The preprocessing also alters the two-way h -terms. In our example, the h_{ij} and h_{ik} arrays were each single-centered, and the h_{jk} two-way array was globally centered (that is, the overall or grand mean for that array was removed).

The first important fact to notice about the effects of slab-centering on the h -terms is that far fewer terms vanish than with fiber-centering. The second point is that none of the two-way interaction terms, the doubly subscripted h -terms, vanish. Instead, these two-way arrays are partially centered. It is clear that a single application of slab-centering is less effective than a single application of fiber-centering in removing h -terms.

By applying the same algebraic analysis, it is easy to show that successively applying slab-centering across two or three

modes will still not remove all the unwanted h -terms. Triple slab-centering—for instance, by first performing bc -centering, then ac -centering, then ab -centering—will remove all the h -terms that are constant over two modes. Thus, it will remove all the one-way effects in the h -component of the data (reintroducing such effects from distortion of the t -component, however; see below). It will double-center all the two-way h -arrays but will not in general cause any of them to vanish.

Even though slab-centering removes a singly subscripted h -component, it reintroduces an unwanted constant of the same form because of its distortion of the t -component of the data. Thus, for example, any advantage to be gained from bc -centering because of the removal of h_j might be cancelled out by the addition of the extra dimension needed to fit the $t_{j..}$ term. This extra dimension has exactly the same form as h_j ; that is, it varies across Mode A and is constant across Modes B and C . If this extra dimension is smaller in mean-square contribution than the constant that is removed, the bc -centering might provide a net gain in interpretability of the solution. This could occur if the underlying factor structure of the data happened to be such that the products of the Mode B and C loadings had means of approximately zero. But, on the other hand, if the h_j bias in the data were small and the mean product of Mode B and C loadings were large for some factors, then the slab-centering could prove clearly detrimental.

Thus, we find that slab-centering is undesirable for several reasons: (a) it introduces unwanted constants because it transforms the trilinear or t -component of the data in a manner inconsistent with the factor model; (b) it does not remove two-way interaction terms; and (c) it removes only one of the one-way constants. If the basic requirements for good preprocessing are that it remove much of the unwanted part of the data and preserve the wanted part, then two-way centering is theoretically inadequate on both counts. The practical consequences of these inadequacies will depend on the particular structure of the data but will sometimes be quite substantial.

Three-Way Centering (removing the grand mean). The results for global-centering follow closely those for slab-centering and can be proven by very similar arguments. Thus, they will only be stated briefly without proof. Removal of the grand mean has little effect on the structure of the data. Its effect on the trilinear t -component of the data is again an inappropriate shift of origin, which introduces a spurious factor; in this case, the factor is constant across all modes, with contributions equal to $t_{...}$. All of the individual h -terms are globally centered; consequently, the only h -term to vanish is the overall constant h . The benefit versus the harm of global-centering thus depends on the relative size of $h_{...}$ compared to $t_{...}$, which in turn depends on the unknown values of the h -terms compared to the $a_{ir}b_{jr}c_{kr}$ loadings.

Appropriate Centering Independent of One's Model of the Biases. If one believed the data had the structure

$$x_{ijk} = \sum (a_{ir}b_{jr}c_{kr}) + h + e_{ijk} , \quad (6-45)$$

then one might be tempted to apply global-centering; that is, remove the grand mean, in order to get rid of the h -term. One might reason that a centering that treats all points uniformly would be appropriate since the conceptualized disturbance h affects all points uniformly. This would overlook the fact that global-centering also removes the mean from the trilinear part of the data, and this disturbs its ratio-scale properties. The proper constant to add to data with structure as in (6-45) is $-h$. However, since the size of h is unknown, then there is no obvious global adjustment that can be performed. Subtracting out the grand mean subtracts the quantity $(h + t_{...})$, which can distort the data a little or a lot, depending on the size of $t_{...}$.

A similar argument demonstrates why slab-centering is inappropriate, even when one is convinced that the only unwanted biases present are terms that are constant within two-way slices of the data (that is, when one is convinced that the only h -terms present are singly subscripted ones). Slab-centering will remove these h -terms but introduce equivalent singly subscripted offsets due to its effect on the trilinear part. For example, if bc -centering is used, the old h_i values will be eliminated, but a new set of h_i will be introduced which equal $-(t_{i..})$. And these spurious offsets will appear regardless of whether or not the original data had such offsets. Indeed one can start with synthetic data that is perfectly appropriate for the trilinear model and which is fit exactly in r dimensions and, by triple application of slab-centering, distort the data so that it takes $r + 3$ dimensions to fit it exactly. In Monte Carlo experiments that we have performed, the fit in r dimensions after slab-centering has been as low as two-thirds of the fit that would have been obtained with correct centering.

Summary. Surely, a minimum requirement of "appropriate" preprocessing should be that it preserves the appropriateness of any factor structure underlying the data. In particular, if we start out with error-free r -dimensional data appropriate to the model, the preprocessing should not make the data inappropriate to the model or create spurious factors and thus inflate the dimensionality. Fiber-centering preserves appropriateness, but slab- and global-centering do not.

Since the main objective of centering is to turn interval-scale data into ratio-scale data and eliminate conditionality of origin, appropriate centering methods should remove additive constants and one- and two-way effects that would otherwise interfere with the analysis. There is no single centering that removes all these effects, but application of one-way or fiber-centering to all three modes can do so. (However, application to two rather than three modes removes all but one of the possible biases and often provides preferable results.) Two-way or slab-centering does not generally remove two-way interaction components, and it replaces any preexisting one-way bias effects with new one-way biases. One must guess about the factor structure and the preexisting biases in order to decide whether the bias removed is likely to be more or less severe than the one introduced. Fortunately, however, one need not guess about the loading patterns of the latent factors in order to justify fiber-centering, since it does not introduce spurious components.

Multiplicative Adjustments:
 Reweighting and/or Equating Aspects of the Data

Principles Behind Reweighting

Not for Improving Appropriateness. The objectives of multiplicative adjustment are somewhat different from those of additive adjustment. PARAFAC does not use data rescaling to make data more consistent with the factor model. If there is size-conditionality in the data, the differences in scale usually vary across levels of a mode (that is, they are slab-conditional) and so are easily accommodated within the PARAFAC model. For example, if one subject has a more extreme response style than another and hence tends to use the endpoints of his rating scales, this is captured within the model by assigning the subject larger weights on all factors. Consequently, all entries in the row of the subject-weight matrix corresponding to that subject will be proportionally increased in order to reflect the larger size of his ratings. Similar effects occur in the variable-weight matrix if one variable is measured in different units than another. Thus, despite such slab-conditional variations in scale size, the dimensionality of the solution will not be affected nor will the structure of the factor-loading matrices (except as differences in the overall sizes of entries in different rows).

If however, there are *fiber*-conditional differences in size, the data will not be strictly appropriate for the PARAFAC model. This might happen, for example, if a subject arbitrarily uses extreme ratings when rating some stimuli and very moderate ratings when rating other stimuli, simply because of fluctuations in his response style that were independent of his perceptions of the stimuli. PARAFAC preprocessing does not have a satisfactory general method of making such fiber-conditional data appropriate for the PARAFAC model. However, the success of PARAFAC with real data suggests that such fiber-conditionality of scale size is usually either absent or insignificant and so does not seriously interfere with the data analysis process. The importance and effects of fiber-conditionality of size have yet to be thoroughly investigated.

While slab-conditionality of scale does not make the data inappropriate for the PARAFAC model, it does have other undesirable consequences that often make multiplicative adjustment desirable. It can complicate comparison of factor loadings between rows of a factor matrix and can cause some subjects or variables to have much more influence on the form of the final solution than others. To cope with such difficulties, PARAFAC incorporates size-rescaling options as a standard part of preprocessing. Here we discuss how size-rescaling can be used to change the relative influence of different parts of the data; this application is called *reweighting*. We then discuss how size-rescaling can be used to make factor loadings more directly comparable across slabs, as well as to facilitate the comparison of fit values and so forth across data sets; this second application is called *size-standardization*.

Need for Reweighting. With real data, the analysis model is never expected to fit perfectly; any solution is a compromise.

Hopefully, however, it is a compromise that most clearly resolves those aspects of the data of greatest interest or importance, even though it may lose some information concerning other less important aspects. To make such a beneficial compromise more likely, we will often need to influence how the compromise is to be resolved. Thus, with any data set, we should consider questions of relative sensitivity of the analysis to different aspects of the data. We may wish to modify the relative influence on the solution that can be exerted by different aspects—such as baseline versus deviations—or different segments of the data—such as one group of variables versus another. It may be appropriate to equalize the influence of different parts of the data (for instance, data from different subjects) or, conversely, to weight certain parts of the data more than others. In the following discussion, we consider how preprocessing can be used to accomplish such changes.

Definition of "Influence." What does it mean when we say that one part of the data has "more influence" on the form of the final solution than another? Basically, we mean that the structure underlying the influential part of the data has a stronger effect on the factor loadings. Suppose, for example, we compare the effects on the solution of randomly permuting or scrambling the elements within different subsets of the data. If one subset has more influence than another, then scrambling the elements in that subset will cause a larger change (on the average) in the factor loadings obtained by PARAFAC analysis than would be caused by scrambling the elements within the other subset. Alternatively, we could compare the effect on the loadings of split-half analyses when we exchange different subsets of the data between split-halves.

Reweighting to Modify "Influence." For least-squares fitting procedures, the influence of a given part of the data is determined by the amount that the error sum of squares can be reduced by fitting that part. Thus, to change the relative influence of particular data segments, the relative proportion of the total squared errors contributed by these parts must be changed. In some analysis procedures, this is accomplished by means of weights applied within the error computation algorithm; each error component is multiplied by the weight chosen for that part of the data before it is added to the error sum of squares. This provides a weighted least-squares solution, which allows the investigator to modify the influence of different parts of the data. However, the PARAFAC fitting procedure does not *directly* incorporate any method of assigning different weights to the errors contributed by various parts of the data; it simply minimizes the error sum of squares over the entire data set. Nonetheless, one can *indirectly* weight different sources of error—change the relative error sum of squares contributed by different parts of the data—by changing the actual size of different parts of the data itself.

Surprisingly, one useful type of multiplicative reweighting can be accomplished by additive adjustment of the data. A set of data points can be thought of as having two underlying aspects—baselines and deviations from baselines—and additive adjustment preprocessing alters their relative influence. In particular,

centering emphasizes deviations or change across levels of the centered mode by effectively multiplying baselines by zero and deviations by 1.0 in that mode.

In contrast, multiplicative adjustment alters the relative influence of one set of data points as compared to another. If all the data points within a particular subset of the data are increased in size, then the points in this subset become more influential in determining the form of the final solution. The opposite happens if they are decreased in size. There are two reasons why one might want to change the relative importance of different subsets of the data. The first is to deliberately make the influence of certain segments of the data unequal; one might want to reduce the relative importance or impact of data from certain segments of the three-way array (such as certain variables, subjects, stimuli, and so on) because these are considered unreliable or unimportant. The second reason is the converse of the first, that is, to equalize the weight given to different aspects of the data. For example, certain segments of the data (such as certain variables) may happen to contain much larger numbers than other segments because of arbitrary differences in scale of measurement, yet we may not want them to influence the form of the final solution more than any other variables. Reweighting can equate the influence of the segments.

Weighted Least-Squares

Weighted Least-Squares via Preprocessing. By multiplying all the data points for a particular variable by a constant k , we accomplish exactly the same effect as if we weighted all the errors involved in fitting that variable by the constant k (and also multiplied the factor loadings for that variable by k). Thus, we can obtain a weighted least-squares solution by ordinary least-squares fitting, if it is preceded by appropriate preprocessing and followed by compensatory rescaling of the resulting factor loadings.

For example, suppose we wished to include a particular variable in our analysis in order to see how the factors would load on it, but we suspected that the data for that variable might not have the same structure as the other data in the analysis or might otherwise be inappropriate for the PARAFAC model determined by the rest of the data. Thus, we might want the data for that variable to have very little effect on the final solution. We could accomplish this by dividing all the values for that variable by some constant, such as 10^3 , so that they would be very much smaller than all the other data in the array. Two things would happen as a result. First, the factor loadings for that variable would become very small— 10^{-3} times the size they would otherwise have had. Second, both the fitted parts and the *residuals* for the data points involving that variable would be reduced by a factor of 10^{-3} . As a consequence of this second fact, there would be no appreciable contribution to the error sum of squares from that variable. The least-squares fitting criterion would determine the factor loadings for the rest of the data without being influenced by any need to reduce the errors contributed by that variable, since they would already be so small. When the

final solution is obtained, we could recover the appropriately scaled factor loadings for the variable in question by multiplying its factor loadings by 10^{+3} .

We can algebraically develop the general rationale for this preprocessing approach to weighted least-squares. In an unweighted PARAFAC analysis of a three-way array, we seek to minimize the quantity

$$\sum_i \left(\sum_j \left(\sum_k ([e_{ijk}]^2) \right) \right) . \quad (6-46)$$

Since e_{ijk} is equal to the difference between the observed x_{ijk} and the PARAFAC estimate of x_{ijk} , (6-46) is equivalent to:

$$\sum_i \left(\sum_j \left(\sum_k ([x_{ijk} - \sum_r (a_{ir} b_{jr} c_{kr})]^2) \right) \right) .$$

With the standard PARAFAC fitting procedure, all squared errors of a given size count equally toward the total error sum that is to be minimized. Now suppose, for example, that we wanted to weight the errors for some variables more than others. If variables corresponded to levels of Mode A, we could accomplish this weighting by multiplying each error from level i of Mode A by a weight factor w_i . We now seek to minimize

$$\sum_i \left(\sum_j \left(\sum_k ([w_i e_{ijk}]^2) \right) \right) , \quad (6-47)$$

or

$$\sum_i \left(\sum_j \left(\sum_k ([w_i \{x_{ijk} - \sum_r (a_{ir} b_{jr} c_{kr})\}]^2) \right) \right) ,$$

which is equal to:

$$\sum_i \left(\sum_j \left(\sum_k ([w_i x_{ijk} - \sum_r (w_i a_{ir} b_{jr} c_{kr})]^2) \right) \right) . \quad (6-48)$$

Now, suppose we consider a set of weighted raw data points, \check{x}_{ijk} , such that $\check{x}_{ijk} = w_i x_{ijk}$, and likewise consider a row-weighted Mode A factor-loading matrix $\check{\mathbf{A}}$, such that $\check{a}_{ir} = w_i a_{ir}$. Then, we could rewrite (6-48) as:

$$\sum_i \left(\sum_j \left(\sum_k ([\check{x}_{ijk} - \sum_r (\check{a}_{ir} b_{jr} c_{kr})]^2) \right) \right) . \quad (6-49)$$

To find the \check{a} , b , and c loadings that will minimize (6-49), we merely need to perform an ordinary least-squares PARAFAC analysis of the weighted data matrix $\check{\mathbf{X}}$. But minimizing (6-49) is

in fact equivalent to minimizing (6-47). Thus, we see that an ordinary least-squares analysis of the weighted data yields the required weighted least-squares solution. However, since the \hat{a}_{ir}^* values are equal to a $w_i a_{ir}$, they must be multiplied by the inverse of the weights applied to the data to give the a_{ir} loadings that are used to predict unweighted \mathbf{X} .

We can readily see how weighting the variables will result in much closer fit to the data values that have large weights than to those that have small ones. We need only represent this weighting as premultiplication of \mathbf{X}_k by a diagonal matrix \mathbf{W} whose i th diagonal element is the weight for the i th variable. Now, if

$$\hat{\mathbf{X}}_k^* = \mathbf{W} \mathbf{X}_k = \hat{\mathbf{A}} \mathbf{D}_k \mathbf{B}' + \hat{\mathbf{E}}_k^*,$$

then

$$\mathbf{X}_k = \mathbf{W}^{-1} \hat{\mathbf{X}}_k^* = \mathbf{W}^{-1} \hat{\mathbf{A}} \mathbf{D}_k \mathbf{B}' + \mathbf{W}^{-1} \hat{\mathbf{E}}_k^*. \quad (6-50)$$

In this representation of \mathbf{X}_k , the rows of the factor loading matrix $\hat{\mathbf{A}}$ are multiplied by the inverse of the weights applied to the data, and the rows of the error matrix are multiplied by the inverse weights, as well. Now $\hat{\mathbf{E}}_k^*$ was an ordinary least-squares error matrix for the estimation of $\hat{\mathbf{X}}_k^*$, and so most of the rows and columns of this matrix will usually have errors of comparable size. However, the rows of $\mathbf{W}^{-1} \hat{\mathbf{E}}_k^*$ in (6-50) will usually be very unequal in size. Where the diagonal weight matrix \mathbf{W} had a large diagonal entry, the corresponding row of the error matrix $\mathbf{W}^{-1} \hat{\mathbf{E}}_k^*$ will have small entries, and vice versa. Therefore, those rows of \mathbf{X} that were made large by \mathbf{W} are associated with small errors of fit in (6-50), while those rows that were made small are associated with large errors of fit. On the other hand, if we thought that there were big imbalances in the sizes of the errors in \mathbf{X} before reweighting, then reweighting could be used to make the errors more equal.

Two Bases for Reweighting. One approach to justifying the reweighting of variables in this way is to assume that we are making the variance of the error terms uniform throughout the data array. This then justifies a statistical model with fixed error variance, as mentioned in Kruskal (1981; see also chapter 2) and discussed later in this chapter. Another approach, however, is based on the realization that in addition to *random* error, there is likely to be *specification* error for each component of the data. That is, there is likely to be systematic error because the model is only an approximation to a more complex systematic structure. From such a perspective, we would often like to reweight data to adjust the trade-offs between different sources of systematic error that arise as part of this approximation process. In essence, we may want to equate the sensitivity of the analysis to the specification error in the different parts of the data or to deliberately make the sensitivity unequal for theoretical reasons. Such alteration of the sensitivity of the model to different parts of the data is also sometimes done to obtain "robust" statistical procedures.

Appropriateness of Different Rescaling Methods

When dealing with three-way data, we see that there are several different ways in which rescaling can be performed. We could apply multiplicative adjustments to fibers or one-way subarrays, to slabs or two-way subarrays, or to the entire three-way array. To determine the appropriateness of different rescaling schemes, we proceed in a manner similar to our investigation of centering. We start with a model of the structure underlying the data and then consider the effects of the rescaling on this structure. If the structure is preserved or only modified in a simple way that does not adversely affect interpretation, then the rescaling method is said to be *appropriate*. If, at the same time, the objectives of rescaling are achieved, then the method is said to be *effective*.

For investigation of the effects of rescaling, we can consider a simplified trilinear model of the structure underlying the data. We could assume that one- and two-way components have been removed by prior centering operations. Even if we do not want to make this assumption, the simple trilinear model is appropriate since, as we have shown earlier, the one- and two-way components can be embedded in it by simply considering them to be extra factors.

One-Way or Fiber-Rescaling. In fiber- or one-way rescaling, multiplicative adjustments are made to one-way subarrays—to rows, columns, or "tubes." It does not matter what determines the size of these adjustments, but for simplicity we can assume that they adjust the fibers so that they each have mean-square of 1.0. If these scale factors are designated by the letter s followed by subscripts, we have three different directions of rescaling that are possible: rescaling of each row of each slab by the coefficient s_{ik} , rescaling of each column of each slab by the coefficient s_{jk} , and rescaling of each tube by the coefficient s_{ij} . Let us consider, as an example, rescaling of columns. Our representation of the rescaled data is:

$$x_{ijk}^* = s_{jk} x_{ijk} , \quad (6-51)$$

or

$$x_{ijk}^* = s_{jk} \sum_r (a_{ir} b_{jr} c_{kr}) + s_{jk} e_{ijk} , \quad (6-52)$$

or

$$x_{ijk}^* = \sum_r (a_{ir} (s_{jk} b_{jr} c_{kr})) + s_{jk} e_{ijk} . \quad (6-53)$$

The expression (6-53) cannot be represented by any simple modification of the three-way structure that existed before standardization. The factor contributions to each column (each jk pair) are subject to a potentially different rescaling, but no such column-specific rescaling is feasible within the PARAFAC or Tucker model. Hence, the number of PARAFAC factors required

to fit the data after such preprocessing will in general be higher than the number of factors required before such preprocessing. Empirical tests with error-free synthetic data show that fiber-rescaling of just one mode will add more than three extra dimensions to the data structure and decrease the fit value at the original true dimensionality by as much as 20–40%.

We conclude that fiber-rescaling of the data points is not appropriate for the PARAFAC model. Similarly, if we carry through the same reasoning with the Tucker model, we obtain similar results at each step and end up with:

$$\dot{x}_{ijk} = \sum_r (a_{ir} \sum_s (\sum_t (s_{jk} b_{js} c_{kt} g_{rst}))) + s_{jk} e_{ijk} . \quad (6-54)$$

Thus, it is apparent that fiber-rescaling is also inappropriate for the Tucker model.

Two-Way or Slab-Rescaling. When performing slab-rescaling, we multiply all the elements in a given two-way subarray of the three-way array by the same constant (for size-standardization, we use the constant that gives a mean-squared data value of 1.0 in the rescaled slab). For example, if we rescale Mode *B*, we multiply the entries in each level *j* of Mode *B* by some multiplier s_j . The algebraic representation of the effects of this transformation is:

$$\dot{x}_{ijk} = s_j \sum_r (a_{ir} b_{jr} c_{kr}) + s_j e_{ijk} , \quad (6-55)$$

which equals

$$\dot{x}_{ijk} = \sum_r (a_{ir} (s_j b_{jr}) c_{kr}) + s_j e_{ijk} . \quad (6-56)$$

Thus, slab-rescaling has a simple representation in terms of the PARAFAC or Tucker model: It just multiplies each row of the Mode *B* loading matrix by the corresponding constant s_j . Similar results hold for slab-rescaling of Mode *A* or *C*. In all cases, the dimensionality of the data remains the same, and the loadings for two of the three modes are unaffected.

Slab-rescaling can also be represented in matrix terms as multiplication of the rescaled mode by a diagonal matrix with diagonal entries equal to the rescaling coefficients. For example, rescaling of Mode *B* would be accomplished by means of an m by m diagonal scaling matrix \mathbf{S}_b , and its effects on the underlying structure would be written as follows:

$$\begin{aligned} \dot{\mathbf{X}}_k &= \mathbf{X}_k \mathbf{S}_b = \mathbf{A} \mathbf{D}_k \mathbf{B}' \mathbf{S}_b + \mathbf{E}_k \mathbf{S}_b \\ &= \mathbf{A} \mathbf{D}_k \dot{\mathbf{B}}' + \dot{\mathbf{E}}_k , \end{aligned} \quad (6-57)$$

where $\dot{\mathbf{B}}' = \mathbf{S}_b \mathbf{B}'$ and $\dot{\mathbf{E}}_k = \mathbf{E}_k \mathbf{S}_b$. Since \mathbf{S}_b is diagonal, each row j of $\dot{\mathbf{B}}'$ is equal to the corresponding row of \mathbf{B}' multiplied by a constant s_j .

In the affected mode, the transformation can be visualized geometrically as moving each point along its radius from the origin, so that its distance from the origin is s_j times its prior distance. When the s_j are selected so that all the levels of the rescaled mode have equal mean-square, then the points are projected out onto the surface of a hypersphere of unit radius. However, if the model does not fit the data perfectly, then the points will fall short of the surface of the hypersphere.

Now that we have seen the simple effects of rescaling a single mode, it is easy to show that multiple rescalings—such as rescaling both the levels of Mode A and Mode B —can be accomplished without changing the dimensionality of the solution and without changing the patterns within rows of the factor loading matrices. In general, if

$$\hat{x}_{ijk} = s_i s_j s_k \left(\sum_r (a_{ir} b_{jr} c_{kr}) + e_{ijk} \right), \quad (6-58)$$

then

$$\hat{x}_{ijk} = \sum_r ((s_i a_{ir}) (s_j b_{jr}) (s_k c_{kr})) + s_i s_j s_k e_{ijk}. \quad (6-59)$$

We conclude that two-way rescaling is appropriate for PARAFAC; similar arguments show that it is also appropriate for Tucker's three-mode factor analysis.

Slab-Rescaling of Several Modes. While the effect of multiple rescalings on the factor loadings is straightforward, the effect on the slab mean-squares is somewhat more complicated. If rescaling of Mode A is followed by rescaling of Mode B , the effects of the second rescaling will modify the results of the first one. For example, if we size-standardize within the levels of Mode A so that each level has a mean-square of 1.0 and then size standardize within the levels of Mode B , the result will generally be that the levels of Mode A no longer have a mean-square of 1.0. However, if one wants to size-standardize both Mode A and B simultaneously, this can be accomplished by iteratively applying standardization to first one mode, and then the other, and then the first again, and so on until the results converge on a solution in which both modes have a mean-square arbitrarily close to 1.0 for all levels. Likewise, such iterative procedures can be used to size-standardize all three modes simultaneously. Iterative rescaling is an option of the PARAFAC program, and experience with this procedure demonstrates that it usually converges in a small number of iterations (usually 3–6, seldom more than 20).

The purpose of iterative rescaling is to discover the multiplications that will jointly standardize several modes simultaneously. By representing the process of iterative rescaling in matrix terms, we can see that the final result could be accomplished by a single rescaling matrix $\tilde{\mathbf{S}}$ applied to each rescaled mode. If we use \mathbf{S}_{a1} to represent the rescaling matrix applied to Mode A on iteration 1, and similarly \mathbf{S}_{b1} for rescaling Mode B on iteration 1,

and so forth, we can represent the process of iterative rescaling Modes A and B as follows:

$$\begin{aligned} \dot{X}_k &= \dots S_{a_4} S_{a_3} S_{a_2} S_{a_1} X_k S_{b_1} S_{b_2} S_{b_3} S_{b_4} \dots & (6-60) \\ &= \tilde{S}_a X_k \tilde{S}_b \end{aligned}$$

Iterative rescaling of several modes to obtain a mean-square of 1.0 in each level of each rescaled mode produces the same effect on the data regardless of the order in which the modes are rescaled. For example, iteratively standardizing the modes in the order $ABABA \dots$ produces the same converged result as is obtained by starting the sequence with Mode B (that is, using the order $BABA \dots$). This mode-order independence is a highly desirable property, which obtains with PARAFAC preprocessing but not with other plausible methods. Rescaling based on variance rather than mean-square does not have this property (unless the two criteria are equivalent, due to prior centering that causes slab means in each standardized mode to be zero), nor does rescaling in which each iteration involves centering and rescaling a given mode before the next mode is preprocessed.

Note that we have been consistently referring to standardization of the *mean-square* (or, equivalently, *root-mean-square*) of the levels of a mode, rather than standardization of the variances. There are several reasons for this. We feel that this is a more general criterion, one that applies equally well both to cases in which we also center the data and thus remove baselines (hence, in which a mean-square of 1.0 is equivalent to a variance of 1.0) and to cases in which we do not center because we consider the baselines of interest (hence, in which equating mean-squares does not generally equate variances). In the latter case, it would seem illogical to standardize the data on the basis of mean-squared deviations (variance) when the quantities being minimized—and which determine the influence of the various data points—are not deviations but rather the mean-squared data values themselves (or, more precisely, their residuals).

The most conclusive reason, however, for size-standardizing the mean-square (or root-mean-square) rather than the variance is that use of the mean-square makes standardization of multiple modes possible. It is generally not possible to standardize several modes to unit variance simultaneously. For example, when we iteratively rescaled the slabs of an uncentered three-way array such that on each step unit variances were produced within the levels of a rescaled mode, the procedure converged on a solution that only had unit variances in the slabs of the last mode rescaled. The slab variances in the other rescaled mode or modes were uniform but were not equal to 1.0. In one case, for example, the variances were close to 1.2 in one mode and 1.1 in another. After a few cycles, level variances in all modes remained unchanged by successive iterations.

The PARAFAC iterative rescaling procedure is closely related to the iterative proportional fitting procedure used to equate the marginal frequencies in contingency tables. The convergence and other properties of iterative proportional fitting are discussed by Bishop, Fienberg, and Holland (1975), among others.

Three-Way Rescaling. For completeness, we mention the case in which all entries in a three-way array are multiplied by the same constant s . If the three-way array constitutes the entire data set under analysis, then three-way rescaling will have no effect on the solution other than to multiply all the coefficients by s in the mode selected to represent the scale of the data. For example, if Modes *A* and *B* were standardized and Mode *C* represented the scale of the data, then all the Mode *C* loadings would be multiplied by s . Those fit measures that are scale-dependent (such as mean-square error) will be multiplied by s^2 or s , depending on the measure.

There are occasions when one might want to multiply only a part of the total three-way array by some constant. As noted earlier, this technique can be used to perform a weighted least-squares analysis. For example, one might wish to fit a model determined primarily by variables 1–15 and have the solution only modestly affected by the characteristics of variables 16–30. A simple approach would be to scale variables 1–15 so that their mean-squares were 1.0 but scale variables 16–30 so that their mean-squares equaled some smaller constant, such as .2. The resulting solution will, in some sense, be five times more sensitive to the characteristics of variables 1–15 than to variables 16–30. In terms of squared errors, the solution will be 25 times more sensitive to variables 1–15. In addition, the two groups of variables will have different sized loadings; all loadings for variables 16–30 will have been effectively multiplied by .2, and so these loadings should be multiplied by 5 to restore them to equivalent scale before they are compared to the loadings of variables 1–15.

Multiplicative Adjustment for Data Standardization

In the preceding section, we discussed multiplicative adjustments used to alter the relative *influence* of different parts of the data. In this section, we consider multiplicative adjustments for a different purpose: To facilitate comparisons of estimated parameters either within or between solutions. One may want to standardize the size of different parts of a data set—such as different variables, subjects, and so forth—to facilitate comparison of the loadings across these parts or to standardize the size of the data as a whole to permit comparison of loadings, mean-square error values, and likewise across several data sets. Finally, size-standardization is used in conjunction with centering so that useful added interpretations of the factor loadings (as correlations) becomes possible.

Standardization of Levels within a Data Set

It often happens that different variables included in a three-way data set might have different and incomparable units of measurement. It would be meaningless to compare, for instance, the size of tremor in millimeters with reaction time in milliseconds. In such cases, the differences in size of measurements for these variables should generally be removed by size-standardization;

otherwise, they will act as a confounding source of size differences in factor loadings. They are typically removed in two-way factor analysis.

Sometimes, differences in overall size between levels arise from the same sources as differences within levels. In the Harshman, Ladefoged, and Goldstein (1977) analysis of variations in tongue shapes, certain locations on the vocal tract showed small variations in tongue position, whereas others showed much larger variations. Yet it was not considered appropriate to use standardization to remove these overall size differences, because they were interpreted as being a part of the same overall linguistic pattern that generated the within-location variations in position. The differences in mean-squares between levels of that mode were interpreted as a natural consequence of the differences in effects of the underlying tongue factors.

There are times, however, when one should consider size-standardization within levels of a mode, even when all the levels of that mode are measured in the same nonarbitrary units. This can happen when there are large, naturally occurring differences between the size of the values corresponding to different levels of a given mode. For example, in one study, each of 24 persons was videotaped while performing 7 tasks involving manipulation of blocks to solve verbal or spatial puzzles (Hampson forthcoming). The videotapes were then scored for the frequencies of 38 different behaviors made during performance of each task. In the resulting $24 \times 7 \times 38$ three-way array, some classes of behaviors had much smaller overall frequencies than others. Consequently, the factor loadings for these behaviors were much smaller than the loadings for other behaviors, even when the low-frequency behaviors were in some sense crucially related to the given factor. This interfered with the interpretation of the factor-loading matrix for the behavior mode. Therefore, in some analyses, Hampson size-standardized the levels of the mode corresponding to behavior type. This made the sizes of loadings for the different behaviors directly comparable. In general, after such size-standardization, a small loading means that there is little relation between the factor and that level of the mode (for example, that class of behavior); a large loading means a strong relationship.

However, there is sometimes a disadvantage to standardizing the size of the different behaviors. This problem arises in the Hampson example. Some cells in the data array correspond to behaviors with very small overall frequencies and are thus based on such small samples of that type of behavior that they are unreliable estimates of the relative population frequencies for those behaviors. When slabs containing many such cells are increased in size to make the size of their factor loadings directly comparable to those of different behaviors, there is a substantial increase in the proportion of error in the overall data set. The most unreliable low-frequency data values are increased in influence, and the most reliable high-frequency data values are decreased in influence. Consequently, the solution changes. If the reweighting is severe, there could be a deterioration of the stability and thus the interpretability of the solution. In general, the investigator employing reweighting should be alert to

the danger of overinflation of error.

An alternative method of making the data values directly comparable is to standardize the loadings but not rescale the data. For example, one could divide each row of the factor-loading matrix for a given mode by the standard deviation of the data at that level of that mode. This provides standardized loadings without altering the relative weighting of the data values during the analysis, an approach that is discussed in more detail later. When deemed useful, this approach could be combined with judicious data reweighting.

Standardization of Overall Size across Data Sets

As noted earlier, when PARAFAC is used to perform direct fitting of profile data, the output loadings are usually standardized in such a way that factor loadings for one of the three modes *reflect the scale of the data*—namely, they are in the same units as the data and of the appropriate size to predict the observed data values. Similarly, when cross-product or covariance data is analyzed, the output loadings are usually standardized such that the two identical modes jointly reflect the scale of the data. To facilitate comparison of such loadings between data sets, it is often useful to standardize the overall size of the data values. In two-way factor analysis, this is usually accomplished by conversion of the data on each variable to *z*-scores. With PARAFAC, similar standardization options are available to set the mean-square for each level of one or more modes to 1.0 before the analysis and thus equate the overall size of the data. When no slab-standardization was desired, one could simply divide all of the data points in the array by the root-mean-square value of the data.

Standardization to give Added Meaning to Factor Loadings

It was noted in chapter 5 (and in more detail in appendix 5-1) that particular combinations of centering and size-standardization would permit factor loadings to take on additional meaning. For example, if the data were size-standardized within levels of Mode *A* (so that the mean-square for each variable is 1.0) and centered across the levels of Mode *B* and/or *C* (so that the mean for each variable is 0), then the data consists of *z*-scores for each variable. If the output is also standardized so that Mode *A* reflects the scale of the data, then Mode *B* and *C* loadings can be interpreted as estimates of factor scores with Mode *A* as factor loadings of the traditional kind (factor pattern coefficients in the oblique case, and correlations between factors and variables in the orthogonal case). However, the joint application of both centering and size rescaling requires more complex iterative preprocessing, as will be discussed below.

Combined Additive and Multiplicative Adjustments

Often, an investigator would like to accomplish several of the above listed objectives of preprocessing in the same data set. Centering one or more modes is almost always advisable in order

to remove unwanted constant components and two-way interactions, as well as to emphasize interesting variations rather than baselines. At the same time, equalization of the influence of different variables and subjects is also usually desirable, along with the standardization of the data that facilitates interpretation of the resulting factor loadings. However, the joint use of centering and size-standardization raises new issues and complications, because the two types of preprocessing interact.

Interaction among Preprocessing Operations

The joint application of both size-standardization and centering requires more complex iterative preprocessing than either alone, since some multiplicative adjustments disturb previous additive adjustments, and vice versa. For example, if the data are first fiber-centered on Mode B and then slab size-standardized on Mode B , the standardization disturbs the effects of the centering; consequently, the data is standardized but is no longer centered. If we recenter Mode B , then the standardization of Mode B is disturbed, although the mode is closer to equal mean-squares within levels than it was initially. To jointly accomplish both the centering and the size-standardization of a given mode requires repeated iteration of the preprocessing operations.

In general, the interactions of the multiplicative and additive preprocessing steps can be described as follows: Size-standardization of any mode disturbs prior centering on that mode but not on the other two modes; centering on a given mode not only disturbs prior standardization of that mode but of the other two modes as well. Thus, there are only a restricted number of combinations of additive and multiplicative preprocessing that can be performed without iteration between the additive and multiplicative stages. Pure centering requires no iteration whatsoever. Pure standardization of more than one mode requires iteration within the standardization step (as described in [6-60]). Iteration between centering and standardization steps is only required if both centering and size-standardization are being performed on the same mode. (Here we always refer to fiber-centering and slab-standardization, since we established earlier that these are the appropriate forms of these operations.)

Matrix Representation

The results of iterative application of joint additive and multiplicative preprocessing can be represented in compact matrix terms. Suppose that both centering and size-standardization are requested for Modes A and B of a three-way array. The same centering and standardization properties would thus be required across all levels of Mode C . If \mathbf{X}_k represents the slice of that array at the k th level of Mode C , we can represent the resulting operations as follows:

$$\mathbf{X}_k = \dots (\check{\mathbf{S}}_{a2} (\mathbf{L}_n (\check{\mathbf{S}}_{a1} (\mathbf{L}_n \mathbf{X}_k \mathbf{L}_m) \check{\mathbf{S}}_{b1}) \mathbf{L}_m) \check{\mathbf{S}}_{b2}) \dots \quad (6-61)$$

$$= \mathbf{M}_a \mathbf{X}_k \mathbf{M}'_b \quad .$$

That is, all the operations performed on the left-hand side of \mathbf{X}_k can be represented by a single matrix \mathbf{M}_a ; similarly, the operations on the right-hand side can be represented by \mathbf{M}_b . Note that each size-standardization step in (6-61) is the result of an iterative standardization, since, in this example, two different modes were being standardized, and the standardization of the first would disturb the standardization of the second. Thus, in (6-61), the $\hat{\mathbf{S}}$ matrices each represent the compressed product of a series of individual iterations, so that, for example, $\hat{\mathbf{S}}_{a1}$ in (6-61) is equivalent to the whole string of diagonal rescalings summarized in (6-60) as $\hat{\mathbf{S}}_a$. Note also that the rescaling and centering matrices are not specific to the k th slice but rather are applied uniformly across all slices of the three-way array.

The properties of the summary standardization matrices \mathbf{M}_a and \mathbf{M}_b are still under study. It can be shown, for example, that they are necessarily column-centered. (Note, however, that the right-hand transformation in [6-61] is written \mathbf{M}_b and hence is row-centered as it occurs there.) When the initial step in the iterative sequence is centering, they are necessarily double-centered. More importantly, it is known that when \mathbf{M}_a and \mathbf{M}_b are determined by iterative preprocessing, their form is not unique for a given \mathbf{X} and a particular set of centering and/or standardizing conditions but depends also on the sequence of operations used, and, in particular, on whether the initial operation is centering or size-standardization. As a consequence, the preprocessed data that results are also not unique. Fortunately, however, with real data, the different solutions obtained by starting the iterative preprocessing sequence with centering versus standardization seem to lead to similar results, at least in the cases we have tried to date.

To simplify the relationship between the raw and the preprocessed data, we now sometimes employ only part of the iterative sequence, equivalent to the initial centering followed by the standardization operations shown with subscript 1 in (6-61). This *one cycle* preprocessing gives results similar to those obtained when iterative preprocessing is carried to convergence but simplifies the extended PARAFAC model associated with the analysis. (The relationship between preprocessing issues and the extended PARAFAC model will be discussed briefly later in this chapter.)

General Linear Preprocessing

More General Preprocessing Transformations

An Example Involving First-Differences. Suppose that we are studying economic time-series data and that each successive level of Mode C represents the same variables, measured for the same industries, but in successive years. For some purposes, we might want to focus on the rate of changes in such data and ignore the overall trends. This could be done by taking first-difference scores, which would be the discrete analog of the first derivative of the variables across time. To prepare a data set for factor analysis, we could construct a preprocessed data array

in which each Mode C slab represents the difference between two successive levels of the raw data array. But would such a preprocessed array be appropriate as input for a PARAFAC analysis? If the initial structure of the raw data were appropriate for the model, would the transformed data also be appropriate?

This example prompts us to inquire into a wider range of possible data preprocessing operations and to ask how we can decide whether these are "appropriate" in the sense used earlier. For example, does the taking of first-differences of successive levels of a mode constitute appropriate preprocessing? If we start with a model of our data and carry through the algebraic representation of this preprocessing, we find that it does. With error-free data, the solution remains unchanged except that the Mode C loadings for the preprocessed data are equal to the first-differences of the adjacent levels of the original Mode C factor-loading matrix. Thus,

$$\begin{aligned} \overset{*}{x}_{ijk} &= x_{ijk} - x_{ij(k-1)} \\ &= \sum_r (a_{ir} b_{jr} (c_{kr} - c_{(k-1)r})) + (e_{ijk} - e_{ij(k-1)}) . \end{aligned}$$

This result shows that the taking of first-differences has the same desirable properties that we noted above for fiber-centering and slab-preprocessing. That is, it does not inflate the dimensionality of the array, and it has straightforward consequences on the factor-loading matrices. As before, the effect of the preprocessing on the factor loadings mirrors its effect on the data.

Matrix Formulation. These parallels are not coincidental but rather can be seen as part of a more general pattern. To understand this pattern, it is useful to formulate the preprocessing in matrix terms. For our example of taking differences of successive years, we can represent the preprocessing transformation in matrix terms by slicing our three-way array into lateral slabs, which are variables by occasions, rather than frontal slabs, which are variables by industries. We then represent the arbitrary n by p slab for the j th level of Mode B (for the j th industry) as follows:

$$\overset{*}{X}_j = X_j T_{\text{diff}} \quad ,$$

where T_{diff} is a p by $(p - 1)$ matrix that has the form:

$$\begin{array}{cccccccc} -1 & 0 & 0 & 0 & 0 & \dots & 0 & \\ 1 & -1 & 0 & 0 & 0 & \dots & 0 & \\ 0 & 1 & -1 & 0 & 0 & \dots & 0 & \\ 0 & 0 & 1 & -1 & 0 & \dots & 0 & \\ 0 & 0 & 0 & 1 & -1 & \dots & 0 & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & \end{array}$$

The resulting matrix $\overset{*}{\mathbf{X}}_j$ is an n by $(p - 1)$ matrix whose rows represent variables and whose columns represent first-differences of the value of those variables over successive pairs of years.

Linear Transformations Provide "Appropriate" Preprocessing. More generally, a very broad family of preprocessing operations can be written as linear transformations or recombinations of the levels of a mode. For example, we commented earlier that the result of combined fiber-centering and slab-standardization applied to several modes requires iteration to determine the correct transformations but once determined can be written in terms of single linear transformation matrices for each mode (namely, \mathbf{M}_a , \mathbf{M}_b , and \mathbf{M}_c), as defined in (6-61). It can easily be shown, in fact, that any kind of preprocessing that can be expressed as a linear recombination of the levels of a given mode (and thus in matrix terms as multiplication of slices of \mathbf{X} by some common matrix \mathbf{T}) will be appropriate preprocessing in the sense meant in this article.

Consider any preprocessing representable by multiplying all the slices of the matrix \mathbf{X} by the transformation \mathbf{T} . We obtain the following simple expression for the effect of this preprocessing on the latent structure of the preprocessed data $\overset{*}{\mathbf{X}}$:

$$\begin{aligned}\overset{*}{\mathbf{X}}_k &= \mathbf{X}_k \mathbf{T} = \mathbf{A} \mathbf{D}_k (\mathbf{B}' \mathbf{T}) + (\mathbf{E}_k \mathbf{T}) \\ &= \mathbf{A} \mathbf{D}_k \overset{*}{\mathbf{B}}' + \overset{*}{\mathbf{E}}_k ,\end{aligned}\tag{6-62}$$

or, for Tucker's model T3 we would write:

$$\begin{aligned}\overset{*}{\mathbf{X}}_k &= \mathbf{X}_k \mathbf{T} = \mathbf{A} \mathbf{H}_k (\mathbf{B}' \mathbf{T}) + (\mathbf{E}_k \mathbf{T}) \\ &= \mathbf{A} \mathbf{H}_k \overset{*}{\mathbf{B}}' + \overset{*}{\mathbf{E}}_k ,\end{aligned}\tag{6-63}$$

where the constituent matrices are as defined in chapter 5. The effects on the latent structure can be interpreted as simply an application of the linear transformation \mathbf{T} to the factor-loading matrix for the transformed mode and to the corresponding mode of the three-way error matrix. Hence, the preprocessing is appropriate since (a) the dimensionality of the data is not inflated, and (b) the post-preprocessing structure bears a straightforward relationship to the original structure. As in all the other cases considered earlier, the transformation of the affected factor loadings is of the same kind as the transformation applied to the preprocessed mode and should thus be easily intelligible to the investigator. The preprocessing is effective because when it is applied to data that is inappropriate (in ways anticipated by a plausible model of our raw data), many of the inappropriate characteristics of the raw data disappear and new spurious components are not introduced. In the case of first differences, we might have in mind a model of the raw data in which there are large but irrelevant components that change in unpredictable ways but only slowly so that they are very similar from one occasion to the next. First-difference preprocessing would greatly reduce these irrelevant components and bring out the changes of interest. It would also remove two of the three singly subscripted and one of the three doubly subscripted h -terms in our model of conditional-origin interval-scale data.

In this particular example, we apply the preprocessing to Mode *B*. However, since all three modes have equivalent status, it can be shown that the same properties hold when the preprocessing is applied to any of the three modes. Furthermore, since such a transformation of one mode does not interfere with the factor structure of any other mode, the effects of double and triple preprocessing can be obtained by simply applying the several preprocessing transformation matrices to their respective factor-loading tables.

The underlying unity of all these appropriate transformations can be seen most clearly by treating them all as instances of a generalized kind of matrix multiplication.

Appropriate Preprocessing as Array-Matrix Multiplication

Multiplying a Three-Way Array by a Two-Way Matrix. Mathematicians have generalized the operation of matrix multiplication to include the multiplication of a three-way array by a two-way matrix (for instance, see Kruskal 1977). The logic of this generalization is straightforward. Consider the matrix product

$$Y = X T .$$

Suppose we initially take X to be an n by m two-way array and postmultiply it by an m by u matrix T to obtain an n by u product matrix Y . The columns of Y are simply linear recombinations of the columns of the original X . Here, the weights contained in the j th column of T are applied to the *columns* of X , which are then summed to produce the j th column of Y .

Now suppose that X is not an n by m two-way matrix but an n by m by p three-way array. Instead of thinking of X as composed of m vertical *columns*, we now consider it to be made up of m vertical *slabs*. Each slab is n by p . The product Y is now an n by u by p three-way array, made up of u vertical slabs. In this case, the weights contained in the j th column of T are applied to the *vertical slabs* of X to produce the j th vertical slab of Y . An alternative way of looking at the multiplication of a three-way array by a matrix is to consider it as a set of parallel two-way matrix multiplications. If the n by m by p array X is postmultiplied by an m by u matrix T , the result obtained is the same as if each n by m slice of the three-way array is postmultiplied by the m by u matrix T .

In a similar fashion, for premultiplication of the three-way array X by a u by n matrix \tilde{T} , the n elements in the i th row of \tilde{T} are the weights applied to the horizontal slabs of X so that their sum will equal the i th horizontal slab of Y . Finally, if a p by u matrix T is applied to the third mode, the p elements in the k th column of T are the weights applied to the frontal slabs of X to obtain the k th frontal slab of Y . (For a picture of the three kinds of slabs, see Figure 6-1.)

Appropriate Preprocessing is Array-Matrix Multiplication. From the definition of array-matrix multiplication just given, it should be apparent that all examples of appropriate preprocessing that we have discussed can be considered instances of multiplication of a three-way array by one or more matrices. In fact, all

examples of appropriate preprocessing of which we are presently aware fall into this class. Such array-matrix multiplication generates the class of operations that can be represented as linear transformations of a mode, as in (6-62), and thus have the desirable properties that make them appropriate as preprocessing for three-way factor analysis.

This unified characterization of appropriate preprocessing methods resolves the apparent anomaly that emerged earlier in which the appropriate additive adjustment seemed to require a different kind of operation (fiber transformation) than was required by the appropriate multiplicative adjustment (slab transformation). We can now interpret both kinds of preprocessing, along with their various iterative combinations, as special cases of a single kind of operation: multiplication of a three-way array by a matrix. At the same time, we find that the inappropriate preprocessing methods (slab-centering and fiber-standardization) cannot be interpreted in this way.

Of course, not all matrices constitute effective preprocessing transformations, even when correctly applied to the data via array-matrix multiplication. In order to determine which family of preprocessing transformations is most appropriate for a given kind of data set, it is still essential to consider a mathematical model of the structure of the data and examine the effect of proposed preprocessing transformations on that model, as was done above. In particular, the terms representing the inappropriate part of the data should vanish, and those representing the appropriate part should retain their simple form. And in any given analysis, one must further select the optimal member or members of the family of appropriate preprocessing transformations, based on the particular characteristics of the data set at hand.

Some Practical Guidelines for Application of Preprocessing

To conclude this section, we should like to give the reader some brief summary guidelines for selection of preprocessing procedures, based partly on general principles and partly on experience:

1. It is almost always desirable to center across at least one mode. This removes unwanted constants and provides an approximation to ratio-scale data. Centering two modes is often optimal but rarely is centering three.
2. If there are meaningless differences in origin across the levels of a given mode—for example, due to different arbitrary or incomparable measurement scales—it is usually advisable to center to remove these differences.
3. In selecting which modes to center, consider where there are likely to be effects that do not vary across levels of a mode. In other words, identify likely singly subscripted and doubly subscripted h -terms, and center those modes in which the

unwanted effects are constant. For example, each subject's overall biases, which are constant across the ratings of all stimuli, can be removed by centering across stimuli.

4. For reasons not yet well understood (but probably related to the complex individual differences in relationships that lead to degenerate solutions, as described later), it is usually undesirable to center across the person mode. This result is surprising, since centering across persons is a traditional method of preprocessing in two-way factor analysis. However, in three-way intrinsic axis methods, such centering sometimes seems to emphasize those individual differences that are inconsistent with the model and that cause problems for the determination of axis orientations. Centering across persons may not be problematic if used as preprocessing for Tucker's T2 or T3 model, in which axis orientation is determined by reference to outside criteria.
5. If there are meaningless arbitrary differences in scale size across the levels of a given mode—perhaps due to differences in the unit of measurement—it is generally advisable to remove these differences by standardizing within the levels of that mode.
6. One may sometimes want to remove overall differences in mean-square across levels of a mode, even when these differences are meaningful and result from genuine differences in the process being measured (such as different frequencies of movements, in the example given in the section on Multiplicative Adjustment for Data Standardization), to equalize the influence of different levels on the solution.
7. Reweighting other than size-standardizing (such as for weighted least-squares analysis) may sometimes seem advisable, as described in the section on Reweighting and/or Equating Aspects of the Data.
8. With three-way rating-scale data—for example, stimuli by scales by persons or judges—good results have been repeatedly obtained when the scales and the stimuli are centered and the scales and the persons are size-standardized.
9. More generally, it has often been found useful to standardize two modes and center two modes, but not the same two modes, so that one mode is simply centered, one mode is simply standardized, and one mode is both centered and standardized.
10. It is often necessary to try several different preprocessing methods and compare their effects on the form of the factor-analytic solutions that result. Select the preprocessing that produces the most interpretable solution.
11. Technically inappropriate centerings can sometimes work out fairly well in practice if the distortions they introduce are

smaller than the distortions they remove. Nonetheless, there seems no reason to use an inappropriate method when appropriate ones exist that work at least as well and do not risk introducing serious distortions into the solution. Many examples of both appropriate and inappropriate preprocessing are given by Kroonenberg (1983).⁴

THE EXTENDED PARAFAC MODELS

PARAFAC Preprocessing as a Basis for More General Models

Extension to Conditional-Origin Interval-Scale Data

Two Perspectives. In our earlier treatment of the topic, we interpreted centering as a method of making a wide class of data appropriate for a more restricted model. However, the opposite perspective is also possible. We can also consider centering to be a way of making the PARAFAC model less restricted and hence appropriate for more general types of data.⁵ In any PARAFAC analysis in which preprocessing is employed, the combined preprocessing-factor extraction procedure can be thought of as fitting an extended PARAFAC model. There is a family of such models, and the particular model that is fit in any particular analysis is determined by the specific preprocessing options selected (for instance, which modes are centered for that analysis). Similar extensions of Tucker's model can also be formulated.

Extension via Centering. Since PARAFAC centering is based on an explicit model of conditional-origin interval-scale data, this model ([6-3] or [6-4]) can be considered part of an extended PARAFAC model. From this perspective, the means that are removed by centering are considered additional terms in the PARAFAC representation.

Suppose, for example, that we center across all three modes before performing a PARAFAC analysis. This centering process can be thought of as estimation and then removal of the first three components of the following model:

$$x_{ijk} = h_{ij} + h_{jk} + h_{ik} + \sum_r (a_{ir}b_{jr}c_{kr}) + e_{ijk} . \quad (6-64)$$

In addition to the ratio-scale components that were part of the basic PARAFAC model (5-3), this extended model directly incorporates the offsets that make the data interval scale and fiber-conditional in origin.

While the form of (6-64) is fairly compact, we should keep in mind that it implicitly incorporates the additional constant terms given in (6-3). The doubly subscripted h -terms of (6-64) can be considered to have absorbed and can thus also represent the effects of the unsubscripted additive constant h , the singly subscripted main effects h_i , h_j , and h_k , as well as the two-way interactions. If desired, the *grand mean* and *main effects* (to use

analysis-of-variance terminology) can be extracted from the two-way h_{ij} , h_{jk} , and h_{ik} arrays by conventional methods. The resulting model would be closely related to those proposed by Gollob (1968) for the two-way case and by Gower (1977) for the three-way case.

As noted earlier, one could also subject these h -arrays to two-way factor analysis to bring out further structure. For example, we saw how h_{jk} two-way interactions could be due to one or more factors that are constant across Mode A . Two-way factor analysis of this array could uncover these dimensions. The factors would, of course, be rotationally indeterminate and would therefore need to be rotated to simple structure or to congruence with factors extracted by PARAFAC from the three-way interactions (see below). These additional factors could then be incorporated into the analysis model by giving them constant loadings in the third mode, much as in (6-5) and (6-6).

To provide the most complete example, the extended model that results from triple-centering was represented in (6-64). As noted earlier, however, triple-centering often reduces the size of the systematic part of the data too much without appreciably reducing the size of the random error; thus, it impairs the signal-to-noise ratio of the data. On occasion, we have recovered meaningful factors from triple-centered data, but we usually obtain better results with double-centering. The mode left uncentered, if carefully chosen, can be one that has minimal constant components not already removed by the other centerings.

The extended model that would result from double-centering would have two of the three h -terms in (6-64). For example, if Mode C were not centered, the h_{ij} term would not be part of the extended model being fit in that particular analysis.

One-Stage and Two-Stage Least-Squares. From the extended model perspective, PARAFAC can be considered to be a two-stage least-squares fitting procedure. Since the mean is a least-squares estimate, computation of means as part of centering is the first least-squares stage of the procedure. Least-squares fitting of the factor model to the residuals is the second stage.

Unpublished theorems by Kruskal (1977) show that when the centering is across *fibers*, this two-stage least-squares procedure is equivalent to a one-stage least-squares fit of an extended model (such as [6-64]). That is, the procedure of first estimating and extracting fiber means and then performing PARAFAC analysis of the residuals produces exactly the same model parameters, fitted values, and residuals as would be produced by using a more general least-squares procedure (such as a hill-climbing procedure) to directly solve for the least-squares fit of all the means and factor loadings at the same time.

Kruskal's theorems also show that this equivalence of two-stage and one-stage least-squares procedures does *not* hold if slab-centering or global- (grand mean-) centering is used (Kruskal 1977). These other kinds of centering do not as neatly allow one to treat the combined centering-factoring procedure as the fitting of an extended model. Thus, although Kruskal's results were not known at the time PARAFAC centering was being developed, they provide an additional reason for preferring fiber-centering over other kinds of centering for three-way factor analysis.

Further Extensions: Additive Factor Change

Additive Change in One Mode. The system variation model ([5-3]) permits only multiplicative changes in factors across levels of a mode. That is, the factor weights (for instance, factor or component scores) can change from one occasion to the next only by being multiplied by some coefficient, as shown in (5-6). However, we might want to consider a more general expression for the way that component scores are related across occasions, one that incorporates additive shifts in the factor baseline, as well as multiplicative shifts in the factor size or variance.

We might expect additive shifts in the factor score baseline values in, for example, a study of animal stress and adaptation in novel environments.⁶ Suppose we measure animals' behavior—including number of vocalizations, exploratory movements, attempts to escape, and so on—during several successive exposures to the novel environment. We could then apply PARAFAC to these data in an attempt to identify a few factors—such as general activity, stress, and so forth—that would explain the behavioral differences among animals and across occasions.

Besides proportional changes in these factor scores across occasions, we might also anticipate additive shifts in their baseline values due to the animals' adaptation to the new environment. Hence, we would want a more general model than (5-3) to incorporate these baseline shifts. With the desired extended model, the expression for f_{jkr} (the factor score of the j th animal on the k th occasion for factor r) would not simply be $b_{jr}c_{kr}$ but instead would have the general linear form

$$f_{jkr} = (b_{jr}c_{kr} + d_{kr}(b)) , \quad (6-65)$$

where the b - and c -terms are defined as usual, and the $d_{kr}(b)$ term represents the additive shift in factor r on occasion k that is constant across levels of Mode B (namely, animals). This model is more general than that for system variation ([5-6]) but less general than the one for object variation ([5-7]); here, the additive shifts are the same for all animals.

Expression (6-65) can be viewed as a linear regression equation that predicts f_{jkr} , the factor score of the j th animal on occasion k from the animal's basic level of factor r . The prediction equation multiplies the base level of the factor, b_{jr} , by a slope term for occasion k , c_{kr} , and adds an intercept term for occasion k , d_{kr} . Thus, the full PARAFAC representation becomes:

$$x_{ijk} = \sum_{r=1}^q (a_{ir}(b_{jr}c_{kr} + d_{kr}(b))) + e_{ijk} , \quad (6-66)$$

which can be rewritten:

$$x_{ijk} = \sum_{r=1}^q (a_{ir}b_{jr}c_{kr}) + \sum_{r=1}^q (a_{ir}d_{kr}(b)) + e_{ijk} . \quad (6-67)$$

Thus, we obtain a PARAFAC model with up to q extra factors. There is one extra factor for each dimension that shows nonzero additive shifts across occasions.

The extra factors in (6-67) have two distinguishing characteristics: (a) they are constant in Mode B ; and (b) factor r in the second set has the same Mode A loadings as factor r in the first set. These two characteristics make it possible to fit models like (6-67) as another extension of the original PARAFAC analysis procedure.

The extra factors generated by the baseline shifts in the r original factors appear as two-way interactions constant over Mode B and would thus contribute to the h_{jk} terms of the extended PARAFAC model. By centering on Mode B , these extra components are extracted during the first stage of fitting and so are eliminated from the second stage. Thus, the presence of factor change of this more general kind does not interfere with extraction of the correct form of the Mode A factor loadings, nor does it interfere with the estimation of each Mode B base level for each factor (the b_{jr} terms).

To identify the extra factors contributed by the additive shifts, we could rotate factors extracted by two-way analysis of the h_{jk} terms into a position where the Mode A loadings resemble as closely as possible the Mode A loadings obtained from the PARAFAC analysis of the centered data. Alternatively, one could simply use regression to estimate weights for these factors. Starting with the assumption that there are dimensions in the h_{jk} table that have the same Mode A factor loadings as the dimensions subsequently extracted from the centered data, one could simply estimate their weights on occasion k by using regression to predict the raw data from the Mode A loadings.

Additive Change in Several Modes. In a completely general PARAFAC model that treats all three modes even-handedly, one would want to allow for additive shifts of factor baselines in any of the three modes. For a given factor r , there are six possible two-way additive contribution factors: two with the same Mode A loadings (one of these two is constant in Mode B and the other is constant in Mode C), two with the same Mode B loadings (constant in Mode A or C), and two with the same Mode C loadings (constant in Mode A or B). The full model would have the form:

$$x_{ijk} = \sum_r (a_{ir}b_{jr}c_{kr} + a_{ir}d_{kr(b)} + a_{ir}d_{jr(c)} + b_{jr}d_{kr(a)} + b_{jr}d_{ir(c)} + c_{kr}d_{jr(a)} + c_{kr}d_{ir(b)}) . \quad (6-68)$$

The simultaneous interpretation of all the shifts in such a model is not straightforward; indeed, it is not clear that all these shift terms would ever be needed simultaneously. Nonetheless, (6-68) displays all the varieties that might be needed in one or another situation. (We will not pursue this model further in this chapter.)

Perhaps the most important point to be made about additive factor shifts is that they need not interfere with the identification of the basic q three-way factors when proper centering is employed. Furthermore, the shift scores can in fact be estimated

after the central PARAFAC analysis by factor score estimation methods.

Extensions via Reweighting and Size-Standardization

The PARAFAC model can also be extended by incorporation of the effects of multiplicative preprocessing. Suppose, for example, that we had rescaled data values within levels of all three modes before performing a PARAFAC analysis, perhaps to size-standardize within the levels of all three modes. As a result, we analyze a data array whose entries were equal to $s_i s_j s_k x_{ijk}$, where s_i is the constant multiplier for the i th level of Mode A (the i th diagonal element of the rescaling matrix \hat{S}_a , as defined earlier in this chapter), s_j is the multiplier for the j th level of Mode B, and s_k is the multiplier for the k th level of Mode C. We might then consider our solution to represent the fitting of the extended model

$$x_{ijk} = \frac{1}{s_i} \frac{1}{s_j} \frac{1}{s_k} \left(\sum_r (a_{ir} b_{jr} c_{kr}) + e_{ijk} \right), \quad (6-69)$$

or equivalently,

$$x_{ijk} = \sum_r (\overset{*}{a}_{ir} \overset{*}{b}_{jr} \overset{*}{c}_{kr}) + \frac{1}{s_i} \frac{1}{s_j} \frac{1}{s_k} (e_{ijk}), \quad (6-70)$$

where $\overset{*}{a}_{ir} = a_{ir}/s_i$, $\overset{*}{b}_{jr} = b_{jr}/s_j$, and $\overset{*}{c}_{kr} = c_{kr}/s_k$. Our interpretation of (6-69) and (6-70) depends on our motivation for applying the rescalings. We will mention two basic motivations: (a) an attempt to achieve uniform error standard deviations, and (b) a desire to differentially weight particular subsets of the errors.

Making Error Standard Deviations Uniform. We could adopt the perspective advocated by Kruskal (chapter 2) and assert that the rescaling was applied because we believed that the original \mathbf{X} array had entries with greater error in some levels of each mode than others and that our goal in rescaling was to make \mathbf{X} appropriate for a model in which all the e_{ijk} in (6-69) are drawn from distributions with the same standard deviation. The scaled errors in (6-70) then represent our estimates of how the size of the error in \mathbf{X} actually varies, and so the terms s_i and so forth are an extension of the model to explicitly represent error standard deviations.

Of course, as Kruskal points out (see chapter 2), we cannot actually observe the error, and so the rescalings are determined on the basis of our beliefs about the errors. Usually, its size is assumed to be approximately proportional to the size of the data itself. There are several justifications for such an assumption. For example, there are many psychological situations in which this proportionality is known to hold, such as when subjects are asked to make judgments of the properties of stimuli (for instance, the weight of objects). More generally, it is often thought that the variables in a given data set are roughly equal in reliability so that the random component will be approximately the same propor-

tion of the total variance across all variables. Of course, this will not be exactly true, but in the absence of other knowledge, it can often be considered a best approximation; it would be unusual to have very precise and reliable observations in the same data set with very imprecise and unreliable ones. Furthermore, if we knew that we had such a combination of reliable and unreliable observations, then we would presumably not rescale the data uniformly, or at least we would not adopt this uniform-error justification for our rescaling.

If a substantial part of the error is thought to be *specification* error—systematic error introduced because the model is not strictly appropriate for the data (for example, when the model is too simple to represent all the complexities of the data)—then it is often reasonable to consider this specification error roughly proportional to the size of the data. The assumption is that the model has roughly the same degree of inappropriateness for all the different observations (for all variables, subjects, and so on). Again, in the absence of other knowledge, this is often a good working assumption. It implies that those levels of a mode in which the mean-square is larger probably have error components that are also larger by roughly the same proportion.

However, if we expected that certain parts of the data would have greater specification error than others, we might choose rescaling coefficients that were not based on size-standardization of the data itself but instead were based on our external estimates of the error components in different parts of the data. In this case, the preprocessing would probably be a noniterative multiplication of the levels of particular modes by particular values.

Weighting Errors Differentially. Alternatively, we could view (6-70) as a weighted least-squares model. In this case, the s_j , s_j , and s_k values might be determined by other means than data-standardization. When the weighting of different parts of the data is based on our estimate of the reliability of those parts, or on our beliefs about the relative inappropriateness of the PARAFAC model for different parts of the data, this weighted least-squares approach does not differ much from the approach that aims to equate error terms across the data. However, when our weights are based on assessment of theoretical importance of different parts of the data or on the relative "cost" associated with not fitting different parts of the data (in some practical application), then the weighted least-squares approach introduces a novel rationale for (6-70).

Rescaling Loadings but Not Data. Recall from our earlier treatment of multiplicative adjustment that the second reason for rescaling aspects of the data was to facilitate comparison of loadings through size-standardization within levels of a particular mode. This method of achieving comparability of levels of a mode is appropriate if one also wants to equalize the influence of the levels of that mode on the solution, since such influence equalization is a necessary consequence of size-standardization.

However, suppose one does *not* think that the levels of a particular mode should have an equal influence, perhaps because those levels with small values are based on fewer observations and hence are less reliable (as in the example given in the sec-

tion on Multiplicative Adjustments for Data Standardization). Then, a different method of obtaining comparable loadings is called for. In an earlier section, we briefly mentioned an alternative approach that involves rescaling the loadings but not the data. Here, we explain the extended model that this implies.

The size of a factor loading is influenced by two things: (a) the strength of the relationship between factor and variable—such as how much the factor influences a particular variable, subject, occasion, and so on; and (b) the overall size of the data values that the factor loading is supposed to predict—such as whether the variable is measured in large units or small ones. For example, the larger the data values at the i th level of Mode A (in comparison to other levels of that mode), the larger the corresponding row of factor loadings will have to be (in comparison to other rows of the Mode A factor-loading matrix). When these variations in overall size are deemed extraneous, we need to disentangle their effects from loading differences that are due to variations in strength of factor-variable relationship so that the two influences can be interpreted separately.

We can construct an extended version of the PARAFAC model in which these two different influences on factor-loading size are represented by separate parameters in the model. As a first approximation, we could divide each row of the factor-loading matrices A , B , and C by the root-mean-square of the corresponding level of the data, obtaining row-standardized loading matrices \check{A} , \check{B} , and \check{C} . To maintain accurate representation of the data, we would compensate for these changes in the loadings by incorporating new \dot{s} -coefficients that explicitly represent the differences in overall size or scale for the levels of each mode. This extended model would be written as follows:

$$x_{ijk} = \dot{s}_i \dot{s}_j \dot{s}_k \sum_r (\check{a}_{ir} \check{b}_{jr} \check{c}_{kr}) + (e_{ijk}) , \quad (6-71)$$

where \dot{s}_i is the scale coefficient (in this case taken to be the root-mean-square of the i th level of Mode A), and \check{a}_{ir} is the size-standardized factor loading defined as a_{ir}/\dot{s}_i . This model is equivalent to (5-3) except that the expression for the systematic part has been expanded slightly; the effects of differences in data size across levels are now represented by separate terms, leaving the factor-loading matrices to express unambiguously the strength of relationship between a factor and a variable, person, and so forth. Note particularly that in both (5-3) and (6-71), the error term is unweighted. The size-standardized factor loadings of (6-71) can be obtained by an unweighted analysis. Hence, the potential problems of the weighted model in (6-70), such as inflation of error in levels with small mean-squares, do not arise.

Issues Concerning Estimation of Size-Coefficients. In the foregoing discussion, we suggested estimating the \dot{s} -terms in (6-71) by simply taking the root-mean-square of the corresponding levels of the data. In some situations, however, this might not represent the effects of data size as completely or correctly as desired. That is, a data matrix multiplied by the inverse of

these \hat{s} -values would not generally have all the effects of data scale removed. This is because the effects of the several standardizations on level sizes would interact; the resulting data array would thus have modest variations from a mean-square of 1.0 across the different levels of two modes. (The last mode standardized, of course, would not have this problem.) For some purposes, it would be more desirable to use \hat{s} -values determined by iterative size-standardization of a "spare copy" of the data (that is, define the \hat{s} to be the diagonal elements of the \hat{S} matrices defined above).

The issue of which \hat{s} -coefficients to use in (6-71) would seem to hinge on the following question: What is the most appropriate weighting of the elements within a given slab when computing the relative size of that slab compared to other slabs in the same mode? One approach is to use the "raw" slab root-mean-squares. This implies that each data point in the slice is taken as it is, independent of what position it has in other modes, and with no consideration of how the weighting coefficients of other modes might reduce or enlarge it to accomplish removal of scale effects from the other modes. Another approach mentioned above is to use the \hat{s} -values that would be determined by an iterative size-standardization of the data. By this method, the investigator equates the relative size of the slabs in a given mode *after removing* size differences due to variations across levels of other modes. This implies that each data point in a given slice is weighted by the \hat{s} -values determined for the other two modes before the \hat{s} -value for that mode is computed. We are not experienced enough yet to know when each perspective is more appropriate; indeed, the choice may depend on the particular data application and perspective of the investigator. We speculate, however, that the latter approach might seem more appropriate when two ways of the data correspond to the same entities; an example of this as part of a two-way DEDICOM analysis of frequencies of different automobile trade-ins can be found in Harshman, Green, Wind, and Lundy (1982), along with a more detailed discussion of rationale and comparison of results with those obtained by reweighting the data before analysis.

Even when one or two (but not all three) modes of the data have been rescaled to reweight the influence of different parts, it still might be deemed desirable to incorporate data size-coefficients into the final model. One would then have a mixed model, which would combine an attempt to separate factor importance from data size effects, and which would also have a weighted error term. We have not yet had any experience with application of these more complex adjustments for data size differences and thus do not know whether they have important advantages over the simpler models.

*Complications Arising
from Combined Additive
and Multiplicative Preprocessing*

As noted earlier, combined additive and multiplicative preprocessing is accomplished by iterative procedures. This approach was adopted on pragmatic grounds: (a) it is a method for discovering

the transformations that accomplish several different standardization objectives simultaneously (for instance, set fiber means across two modes to zero and simultaneously set slab levels to mean-square of 1.0 in two modes); and (b) data preprocessed in this way often reveals meaningful structure that was not accessible before preprocessing. Often, when meaningless degenerate solutions are obtained with a given data set, the right combination of iterative preprocessing options will allow the investigator to recover meaningful, interpretable solutions.

However, when we consider the effects of iterative combined centering and size-standardization in the light of possible extended models of the data, we realize that such procedures introduce unanticipated subtleties and complications of interpretation. More than one way of preprocessing will accomplish a given set of standardization objectives, but the different methods result in different standardized data sets, which can sometimes yield dissimilar solutions. One is then faced with the question of how to determine if a particular approach is "optimal." Even more perplexing is how to relate the solution for iteratively preprocessed data to an extended model of the original data. We are still investigating these issues of data preprocessing and cannot yet present a fully satisfactory solution; however, we mention them here to introduce the reader to our current thinking on them.

Dependence on Order of Operations. While the results obtained from iterative combinations of additive and multiplicative preprocessing depend on the order in which the preprocessing operations are performed, things could have been much worse. The nature of this order dependence is less serious with PARAFAC preprocessing than with any other method of iteratively combining multiplicative and additive preprocessing of which we are aware.

It is useful to distinguish two kinds of order dependence that iterative preprocessing schemes might be subject to: (a) *mode-order* dependence; and (b) *operation-order* dependence. Mode-order dependence occurs when the result of iterative preprocessing depends on the order in which the modes are preprocessed. In contrast, operation-order dependence occurs when both centering and standardization are involved in iterative preprocessing; the converged result depends on whether the process begins with centering or with standardization.

Let us first consider an example of mode-order dependence. Suppose we want to center and size-standardize both Mode *A* and Mode *B* of a two-way array. If we do this by first centering and standardizing Mode *A*, then centering and standardizing Mode *B*, and then repeating *A*, *B*, *A*, and so on until the process converges, we end up with a different final result than if we start by centering and size-standardizing Mode *B*, then *A*, then *B*, *A*, and so forth until convergence. Both converged matrices are doubly standardized; that is, they have both variances of 1.0 and row and column means of zero. But their particular data entries—and more importantly their factor structures—will differ quite substantially. Thus, this procedure has the undesirable property that the result obtained by preprocessing the transpose of a matrix is not the transpose of the result obtained by preprocessing the original matrix itself.

This type of order dependence was first noted by Cattell (1966), who experimented with the method of iteratively centering *and* size-standardizing one mode before proceeding to the other, as described above. Cattell's discouraging findings caused some investigators (such as Kroonenberg 1981a) to take a relatively gloomy view of attempts to iteratively standardize several modes. However, as we pointed out earlier, pure size-standardization of several modes, if iterated and carried to convergence, is independent of the order in which the modes are standardized. Thus, PARAFAC size-standardization (like PARAFAC centering) is not subject to mode-order dependence.

When performing combinations of additive and multiplicative preprocessing that require iteration, PARAFAC uses a different sequence of operation from that employed for two-way arrays by Cattell. It first centers all requested modes, then iteratively standardizes all requested modes, then returns to update the centering, and so on. Because in the most recent versions of PARAFAC full iterative size-standardization is employed between each centering step, the combined procedure is a sequence of individual parts, each of which is free of mode-order dependence; therefore, the total preprocessing is also free of mode-order dependence. With PARAFAC, the transpose of a preprocessed matrix or three-way array is thus equal to the result that would be obtained by preprocessing the transpose of the original matrix or array, provided the requested preprocessing objectives are transposed appropriately.

While PARAFAC preprocessing procedures are free from mode-order dependence, they are subject to operation-order dependence. With some synthetic data sets, when we compare results obtained by starting with centering versus starting with size-standardization, we find that we have two different outcomes, both of which fulfill the preprocessing objectives but have quite different factor structures in the preprocessed modes (although the same number of dimensions as the original data). How, then, is one to decide which operation to use first? And how does one argue for the optimality of a given solution when a different result could have been obtained by a different order of preprocessing operations? We currently employ the following tentative principles for preprocessing.

When, as is usually the case, we suspect that major components of the raw data are constant in one or more modes, and we think that removing them would make the data more appropriate for the model, we consider it preferable to center as the first stage of iterative preprocessing. This eliminates these unwanted components and lets successive preprocessing and analysis proceed on the parts of the data of greatest interest. When the raw data contains such constants, standardization as a first stage of preprocessing would disturb the constants and presumably make their extraction more difficult.

If, on the other hand, we suspect that the major constant biases and two-way interactions have been obscured by different scales of measurement, subject response styles, or other influences on the sizes of entries in different levels of particular modes, then it might be necessary to standardize these modes to make these artifactual components appropriately constant before

centering. When there is a serious question as to whether it is more appropriate to size-standardize first or instead to center—or whether to do only a partial size-standardization, then center, then a fuller iterative size-standardization and centering—it is necessary to try out the different procedures and compare the interpretability of the resulting solutions.

To date, when we have used real data to compare the results of initial centering versus initial standardization, we found very little difference between the PARAFAC solutions obtained with the two kinds of iterative preprocessing. For example, both preprocessing methods were used for analysis of the semantic-differential ratings of automobiles and celebrities described by Harshman and De Sarbo (appendix C). To our surprise, the correlations and factor congruence measures for the corresponding factors in the two solutions were usually .99; none was below .98. Similar results have been obtained with other data sets of the same general kind. But at the same time, parallel comparisons with certain synthetic data sets have resulted in factor correlations more in the range of .7. We do not yet understand what properties of the real data sets make them so relatively immune to preprocessing order-dependence. It may be that the problem of order-dependence is of greater theoretical than practical concern, but it might instead be that we have not checked out the degree of order-dependence on a sufficiently wide range of data types.

The Relation between Original and Preprocessed Data. Earlier in this chapter, we demonstrated how fiber-centering preserves a very simple relationship between the structure of the original data and that of the preprocessed data. In the error-free case, the factor-loading tables of the centered modes are themselves column-centered. Similarly, we determined that slab-rescaling also preserves a simple relation between the original and preprocessed solutions; namely, the rows of the factor-loading table for the rescaled mode or modes are multiplied by the same rescaling coefficients as were applied to the corresponding levels of the data (again, in the error-free case). This was found to be true even when several modes were iteratively rescaled. But when we considered the combined iterative application of centering and size-standardization, we were not able to describe the effect on the loadings in any simple straightforward way. The transformation of the factor loadings is described by the M_a (or M_b or M_c) matrix, whose general properties are not well understood.

The possible complexity of the transformation imposed by iteratively centering and standardizing a given mode becomes clearer when we consider the effects geometrically. The centering operation shifts the origin of the axes to the centroid of the points. Then the size-standardization operation shifts these points along their radii from this origin out toward the surface of the unit hypersphere. If we consider the points in the full dimensionality that provides perfect fit, then the points are in fact moved to the surface of a hypersphere of unit radius in this high dimensional space. This shift redistributes the points, and so their centroid is no longer at the origin. The successive centering operation shifts the origin to the new centroid and the process continues until the points are distributed in such a way

across the surface of the hypersphere that their centroid is at its origin.

Algebraically, we can describe the consequences of the successive operations applied to the data in terms of the equivalent operations applied to the factor-loading matrix. That is, the factor-loading matrix is sequentially column-centered and then row-rescaled until the process converges.

One theoretical consequence of this apparent complexity is that we are not able to write a simple extended PARAFAC model. The best we can do—at least until we understand the properties of the M matrices better—is to write an opaque expression, such as:

$$X_k = (M_a^{-1} A) D_k (M_b^{-1} B)' + M_a^{-1} E_k M_b'^{-1}. \quad (6-72)$$

This example represents the model resulting from iterative preprocessing of Modes A and B but not C . Similar models would result from other combinations of preprocessing.

Perhaps when using fully iterative preprocessing we would do better to talk of the extended PARAFAC *procedure*, rather than an extended model for the original data. The objective of such an extended procedure is to reveal certain structure that can be found in the data after other unwanted effects are removed.

"One-Cycle" Preprocessing. One method of simplifying the relationship between the original data structure and the preprocessed data structure is to truncate the iterative process after only one full iteration. Thus, preprocessing would consist of centering all requested modes, followed by iterative size-standardization of all requested modes (or vice-versa, depending on whether initial centering or initial size-standardization was used). This "one-cycle" preprocessing allows us to write a simple extended PARAFAC model for the data; the preprocessing matrix M has the simple form $\tilde{S}L$ or $L\tilde{S}$. The scalar form of the extended PARAFAC model can be written as:

$$x_{ijk} = \frac{1}{s_i} \frac{1}{s_j} \frac{1}{s_k} [h_{ij} + h_{jk} + h_{ik} + \sum (a_{ir}b_{jr}c_{kr}) + e_{ijk}], \quad (6-73)$$

if initial centering is used, and in a slightly different form—with the h -terms outside of the parentheses and the s -terms inside—if initial size-standardization is used.

Because of the usual rapid convergence of the iterative preprocessing procedure, one-cycle preprocessing has produced results quite similar to full iterative preprocessing. The interpretation of the factors has been virtually identical. However, all the preprocessing objectives are not fully met at the end of one cycle. If, for example, initial centering is used, then the subsequent standardization usually disturbs the zero-means in the preprocessed modes. Thus, the h -terms of the extended model

do not represent the entire two-way mean components in the data. Also, the trilinear part of the model generates fitted values whose means are not exactly zero. While this is different from the familiar roles played by such terms in two-way models, we see nothing wrong with such a representation.

The *h*-terms of the three-way model are defined by their distinctive functional roles (having to do with offsets arising from conditional origins, factors constant across one mode, plus part of the two-way effects in the trilinear component), rather than by the fact that they express all the two-way components of the fitted part of the data.

Other Approaches. Kettenring (1983) discusses ways of extending the information obtained in analysis by combining PARAFAC and analysis of variance. Models that combine the two perspectives have also been proposed by Gollob (1968), Gower (1977), and Lohmoller (1979, as noted in Kroonenberg [1983]). (The reader is referred to Kroonenberg [1983, 135-41] for further discussion of these models.)

An Orthogonally Constrained Model: Degenerate Solutions and Their Remedies

We have seen several extensions of PARAFAC's applicability that result from implicit *generalizations* of the algebraic model. We now discuss an important extension of applicability that results, surprisingly, from placing a *restriction* on the model. The restriction consists of orthogonality or zero correlation constraints on the factor-loading patterns in one or more modes of the solution. For some three-way data sets it appears that the restricted model has a "robustness" that allows it to provide more meaningful solutions, despite its formal inappropriateness.

While the usefulness of an orthogonality constraint has been demonstrated repeatedly in the last few years, we are only now beginning to understand why it works. The following account should be considered simply as a brief progress report of ongoing research.

The Classical Patterns of Degeneracy

During the initial period after development of PARAFAC, its primary applications were indirect fitting to accomplish multidimensional scaling (see Gandour and Harshman 1978; Terbeek 1977; Terbeek and Harshman 1972) and direct fitting of physiological/acoustic profile data (see Harshman, Ladefoged, and Goldstein 1977; Harshman and Papcun 1976; Lindau, Harshman, and Ladefoged 1971). It generally worked well in these applications. Eventually, however, PARAFAC began to be applied to a wider range of data types, particularly profile data of more psychological origin, such as semantic differential ratings. With these data, we would frequently observe certain characteristic but uninterpretable patterns of factor loadings that we have come to call *degenerate solutions*. We saw that the occurrence of these degenerate solutions was one of the most serious problems interfering with wide application of PARAFAC to profile data.

Degenerate solutions are characterized by factors with very high correlations (.80-.98) in Modes *A* and *B*, and usually Mode *C*, as well. However, Mode *C* is typically used for person weights; thus, correlations are sometimes lower (.50-.80) in this mode, even in degenerate cases, perhaps because of the lower reliability of person weights that are based on only one subject's data. Mode *C* is also typically the one mode left uncentered, so it shows high cross-products even when it does not show such high correlations.

The loading patterns on these factors are usually quite difficult to interpret, and in any case, because of the high correlations, there would be only one interpretation for the several dimensions in the solution. In the classic pattern of "hard-core" degeneracy, the correlations between a pair of factors will have a triple product across modes that is negative; that is, their Mode *A* correlation times their Mode *B* correlation times their Mode *C* correlation will be large and negative. Therefore, a high negative correlation between factors will be found in either one or in all three modes.

Circumstances Yielding High Factor Correlations. We now distinguish three or four conditions in which highly correlated factors can occur. The first is not considered evidence of degeneracy, whereas the other three represent successively more extreme types of degenerate solutions.

The first condition in which high correlations might be observed occurs when more factors are extracted from a data set than are actually present in the data set or can be supported by the data set. It can be demonstrated with synthetic data that highly correlated factors will sometimes occur in this situation, but several features distinguish this case from one in which a true degeneracy is present. The highly correlated factors emerge after the extraction of several meaningful nondegenerate factors. The triple product of factor correlations for these extra factors might sometimes be high and negative but at other times might be high and positive and would more often be lower in magnitude. The high correlations, when found, presumably are due to chance collinearity as a result of the random axis orientation in the subspace spanned by the redundant dimensions. Thus, such correlated dimensions were not necessarily interpreted as a symptom of degeneracy but rather as an indication that we had exceeded the correct dimensionality for a given data set.

With error-free synthetic data, the highly correlated dimensions resulting from extracting too many factors will vary across different starting positions. With real data containing error, the solution may or may not be unique, since the fitting of error might cause one orientation to be strongly preferred. However, the arbitrary nature of the axis orientation for these extra dimensions should be revealed by comparing solutions obtained in two split-halves of the data set, in which the random error would not be the same.

The second occasion on which high correlations are observed occurs when the data have not been adequately preprocessed. If the dimensionality is low and yet we observe the "classical" degenerate pattern of very high factor correlations (with one or three of them negative), we may suspect a degenerate solution,

particularly if consideration of other information (such as fit values, number of dimensions extracted, replicability of the degenerate solution across split-halves, theoretical expectations, and so on) suggests that there should be more systematic dimensions present. If, as is often the case, better preprocessing allows additional meaningful dimensions to be uncovered, then we know that we were dealing with a "soft" or easily overcome degeneracy.

The third situation is degeneracy that persists even after several attempts at carefully chosen preprocessing. Then we might be confronted with hard-core degeneracy. This would be most strongly confirmed if the pattern of degenerate dimensions was replicable across split-halves of a data set and if imposition of orthogonality constraints caused a noticeable drop in the fit value, yet revealed much more meaningful dimensions (which, sample size permitting, should be replicable across split-halves).

There is a fourth situation that could develop, although we have not yet observed it except with synthetic data. A hard-core degenerate solution might occur that could not be overcome, even by imposition of orthogonality constraints. In such a case, imposition of orthogonality constraints would presumably produce solutions that were not interpretable and perhaps not replicable across split-halves.

"Soft" (easily overcome) Degeneracies. Sometimes highly correlated and uninterpretable dimensions even occur in a two-dimensional analysis of a given data set. Yet examination of the fit-versus-dimensionality curve for these data sets often suggests that more than one dimension is present. Our Monte Carlo studies have shown that highly correlated dimensions can sometimes be obtained as one consequence of analyzing data that does not have the independent patterns of variation for each factor needed to determine a unique solution (for instance, when several factors are constant in a given mode) or when artificial additive constants (h -terms) have been added to the data. We originally thought that all degenerate solutions were caused by such extraneous components or conditionality in the data, and PARAFAC preprocessing methods were developed primarily so that meaningful solutions could be recovered with data of this kind. For many data sets, the preprocessing does just that.

As an example, consider the analysis discussed by Harshman and De Sarbo (appendix C). The data consisted of ratings of 25 stimuli (celebrities and automobiles) on 39 rating scales by 34 individuals. (Appendix C, Table C-1 gives the loadings obtained before data preprocessing.) Note that the first factor is constant in two of the three modes; its Mode A loadings are close to 1.0 and its Mode C loadings are close to 3.5, which is the center of the 7-point rating scale. Its Mode B loadings show a pattern of variation across scales that represents a strong positive evaluation, indicating that subjects took the zero-point for these stimuli to be of very high positive evaluation. We might suppose that this dimension is adjusting for the arbitrary displacement of the origin from true zero, that is, the h -term in (6-3). However, it shows variations in the baseline across stimuli and so would more accurately be described as representing $h + h_j$ in (6-3). The explanation is not as simple as this, however. The second and

third dimensions do not represent independent patterns of influence freed from the baseline that the first factor represented. Instead, they are very highly intercorrelated and also resemble the first dimension in Mode *B* but show more complex patterns in Modes *A* and *C*. The correlations among dimensions (appendix C, Table C-4) reveal the classic high negative triple product across modes. However, there is a dramatic difference between this solution and the one obtained after iterative preprocessing: The highly correlated dimensions have disappeared, and the solution is much more interpretable (see appendix C, Tables C-5 and C-6, and Figures C-1 through C-3).

Difficult or "Hard-Core" Degeneracies. With the development of PARAFAC preprocessing procedures, we were frequently able to prevent or overcome degenerate solutions. However, there remained some solutions that were not improved, even after the most careful and extensive preprocessing efforts. Even when attempts were made to encourage noncorrelated dimensions (for instance, by using an orthogonal approximation to the solution as a starting position for further iterations), the solution would quickly return to the degenerate form. Furthermore, this degenerate form was consistent across different random starting positions and replicable across split-halves of the data. And so it would seem that the highly correlated dimensions were strongly determined by the data itself. Indeed, any less correlated solutions obtained by subsequent application of orthogonality constraints usually had noticeably poorer fit. This indicated that despite the high correlations, successive dimensions in these degenerate solutions did not represent redundant information.

The pattern of factor correlations obtained in these cases always showed a characteristic "signature": The triple product of the Mode *A* times Mode *B* times Mode *C* correlation between a pair of factors not only was quite high, but invariably was negative in sign. Yet both theoretical arguments and empirical tests with synthetic data indicated that if the problem was simply one of too many dimensions or inadequate determination of axis orientation by the data, then only some high correlations would be expected, and for these, the triple product of correlations in the three modes would be positive at least as often as it was negative. It seemed that something additional was causing the high negative correlations, that some unknown phenomenon was forcing a special form on the solution. We soon found a practical method of overcoming the problem, but only now are we beginning to understand theoretically what causes it.

Application of Orthogonality Constraints

In an attempt to block the classical pattern of degeneracy, a special option was incorporated into the PARAFAC program that allowed the user to constrain the factors in one or more modes to be orthogonal (or uncorrelated, which would be more appropriate if the mode contained all positive loadings). We reasoned that if highly correlated factors were prohibited by means of such a constraint, then interpretable dimensions might emerge.

The constraint worked surprisingly well. In most cases, it was sufficient to apply the constraint to only one of the three

modes. This prevented the degenerate dimensions from emerging in all modes and provided dimensions with quite distinct interpretations. Once again the systematic nature of the degenerate solutions was emphasized. Apparently high correlations in *all three* modes (with either one or three negative correlations) were necessary for the fit value to benefit from the degeneracy. If high correlations were prevented in one mode, then they did not form in the other two. In the case of three-way rating-scale data, the constraint was generally applied to the scales mode, although sometimes the stimulus mode was used with success (Dawson 1982; also see below). Interestingly, the constraint did not act as effectively when applied to the subject mode.

Not only were the dimensions obtained after application of the constraint almost always interpretable, but we have begun to accumulate evidence that they are in some sense empirically valid. One example of such evidence is Harshman and De Sarbo (appendix C). In this study, preprocessing fixed the initial "soft" degeneracy, and the classic pattern of "hard core" degeneracy did not emerge until four dimensions were extracted, at which point only two dimensions participated in the degeneracy. When split-half analyses of the data set were performed, however, highly correlated dimensions emerged in the three-dimensional solutions (due perhaps to the small sample size [17] in each half). But by applying orthogonality constraints to the scales mode, we were able to obtain interpretable three-dimensional split-half solutions. In this case, we knew what the valid three-dimensional solution should look like, based on the interpretable full-sample solution. Thus, it is noteworthy that the solutions that emerged from the constrained split-half analyses very closely resembled the interpretable three-dimensional unconstrained solution obtained with the total sample.

In another recent PARAFAC application, Dawson (1982) performed an analysis of judges' rating-scale reactions to a set of metaphors. His unconstrained two-dimensional analysis yielded highly degenerate solutions. By application of the orthogonality constraint in the stimulus mode, however, he obtained a two-dimensional solution that was not only highly interpretable but congruent with theoretical predictions.

In certain special cases, it is useful to apply orthogonality constraints to two or all three modes to obtain a solution whose axes are oriented in the directions more closely analogous to unrotated principal components. In Snyder, Walsh, and Pamment (1983), this approach was justified by special theoretical considerations and provided a satisfying solution. In general, however, it is desirable to apply the orthogonality constraint to only one mode. This allows the orientation of axes to be established by the proportional profiles criterion without undue influence from the orthogonality constraints. Constraints applied to only one mode would allow the intrinsic axis property to determine the choice among the infinite set of solutions that are orthogonal in one mode; however, double or triple constraints would force the solution to be some compromise between the unique orientation determined by proportional profiles and that determined by principal-components-type criteria.

As we gain experience with these constraints, we are becoming

increasingly convinced that the distortions that they impose on a solution generally have relatively minimal effects on the interpretation, as can be demonstrated by applying the constraints to synthetic data with oblique axes in all three modes. In exchange for these minor distortions, the constrained solution is robust in the sense that it usually has a much more meaningful interpretation than the unconstrained degenerate solution does. Sometimes it is useful to apply these constraints even when actual degeneracies are not present in the unconstrained solution. Moderately high correlations and dimensions that are difficult to interpret might also justify use of the constraint.

Searching for the Cause of "Hard-Core" Degeneracies

As we mentioned earlier, we initially attributed the problem of degenerate solutions to data peculiarities, such as extra constants, two-way interactions, and so on. This explanation was at least partly correct, since the problems with many data sets disappeared after application of adequate preprocessing. However, there were degeneracies that were not corrected in this fashion. They consistently showed a high negative product of factor correlations for the three modes. Also, we found that orthogonally constrained solutions sometimes had substantially poorer fit values than unconstrained solutions in the same dimensionality, even though the unconstrained solutions appeared much more redundant than the constrained ones. (However, careful examination of the fitted data produced by a two-dimensional degenerate solution suggested that the two highly correlated dimensions were not as redundant as they might seem. Because of the high negative correlation between the dimensions, their common part was largely cancelled out when they were added together to fit data, and the subtle differences between them became magnified to produce different and more interpretable dimensions.) Thus, we had to seek some other special characteristic of the data that could account for these properties of the degenerate solutions.

Hypothesis Testing by Data Synthesis. Since the most striking characteristic of the degenerate solutions is the fact that successive dimensions look like only slightly modified versions of previous ones, we began with the premise that the degenerate solutions were an attempt to fit individual variations in the patterns of the loadings. If each subject had points that were idiosyncratically displaced relative to those of other subjects, then the PARAFAC program might extract two or more similar dimensions to more closely approximate the subjects' alternative versions of a given dimension.

Therefore we generated "perturbed" subject spaces; that is, we synthesized profile-type data from two underlying "true" factors that had been additively "perturbed," with a different random perturbation used for each subject. We then performed a two-dimensional PARAFAC analysis of the data to see if the solution would be degenerate. The perturbations themselves were uniform random numbers selected from the range (-1,1), a different one for each loading that was to be perturbed. Before being added to the loadings, they were rescaled so that the mean-

squared perturbation value for a factor in a given mode was equal to some proportion of the mean-squared true loading value for that factor; for example, sometimes we used perturbations that had a mean-squared value equal to half of the mean-squared loading value. Since we took Mode C to represent person weights, we never distorted the true loadings in Mode C but selectively perturbed the factors in one or both of the other two modes.

In addition to varying the size of the perturbations, we also varied the amount of additional error (using 0%, 5%, and 20%) and the relative size of the underlying factors in the data. Regardless of the combination of perturbation and error in the data, however, we were not able to produce fully degenerate solutions. We frequently obtained factors that were highly correlated (that is, 0.77 or above) in one mode; occasionally we saw correlations above 0.7 in two modes, but never in all three. The solution that was most like a degenerate one had factor correlations of 0.68, -0.99, and 0.73 in Modes A, B, and C, respectively; it was obtained by analyzing error-free $18 \times 18 \times 18$ data with one underlying factor that was almost three times the size of the other and with the larger factor equally perturbed in both Modes A and B. On the basis of these tests, therefore, we concluded that perturbations of the underlying factors were not an adequate explanation of the degenerate solutions that we had seen. However, idiosyncratic variations in the location of points in each subject's space might contribute to some instances of highly correlated factors in one or perhaps two modes.

Hypothesis Testing by "Filtering" the Data. Finally, to test the assumption that the problem was due to some pattern of individual perturbations in the loadings of each dimension, we decided to remove all such perturbations (while retaining stretches and other linear transformations) and see if the degeneracies disappeared. For this test, we used real data rather than synthetic data. We applied a "filter" to the data for each person, making sure that each row of that person's data fit into the common row space for all subjects and similarly that each column fit into the common column space.

To accomplish this "filtering," we first determined an arbitrary set of basis vectors for the common space by averaging the data across persons and then performing a component analysis or singular value decomposition of the matrix of means. Suppose we represent the average subject matrix as \bar{X} ; then we can write this decomposition as follows:

$$\frac{1}{p} \sum_k^p X_k = \bar{X} = P D Q' .$$

To test the cause of degeneracies in a u -dimensional PARAFAC solution, we used u -dimensional filters; that is, we used the ~~orthonormal~~ matrices P_u and Q_u , consisting of the u columns of P and Q , respectively, to construct orthogonal projection matrices that projected each person's data into the u -dimensional subspace of the common space. first

The filtered matrix for subject k is called F_k and has the property that each column of F_k is some linear combination of the columns of P_u and each row is some linear combination of the columns of Q_u . This is accomplished with the orthogonal projection matrices $P_u P_u'$ and $Q_u Q_u'$, as follows:

$$F_k = P_u P_u' X_k Q_u Q_u' . \quad (6-74)$$

We then submitted this filtered version of the data to PARAFAC analysis. The results were clear cut. Although the fit value increased dramatically—from R -squared of .44 to .94 in one case and from .28 to .89 in another—the loadings obtained in Modes A and B were almost identical to those obtained by analysis of raw data. The solution was still degenerate, although Mode C loadings were sometimes much less correlated. This demonstrated to us that it was not some idiosyncratic fluctuations or perturbations of individual loadings that were the cause of the degenerate solutions, since these had all been removed. Instead, the source of the degeneracies had to lie in the linear transformations that related each subject's space to the common space. Indeed, differences in the subjects' linear transformations of the common space constituted the only possible differences left after the filtering operation, yet the degeneracies still appeared.

Such individual differences, where each person's space was a different linear transformation of a common space, had to go beyond individual differences in the amount of stretch of the factor axes, since such differences could readily be represented by PARAFAC without degeneracies. There must also have been differences in either the obliqueness of dimensions or the directions of stretch (that is, individual rotations of axes before stretching or contracting them), or perhaps both. These are just the kind of more general transformations permitted by the Tucker three-mode model. In other words, it appeared that in the hard-core degenerate cases, the problems may have been caused by the fact that the data showed the kind of variations that could only be compactly represented by the Tucker T3 or T2 models; it appeared that the patterns in the data were too complex for the PARAFAC model.

Our hypotheses thus became the following: The hard-core degenerate solutions were attempts by PARAFAC to approximate the more general patterns of variation consistent with Tucker's model, using a bizarre combination of the more restricted PARAFAC factors; when we applied the orthogonality constraint, we somehow prevented this approximation procedure and forced PARAFAC to find a solution that had lower overall fit but presented the underlying dimensions in a more straightforward, interpretable manner. In other words, we hypothesized that constrained PARAFAC provided a meaningful fit to a *subset* of the Tucker structure that had a simple PARAFAC representation.

Successful Synthesis of Data Causing Degeneracies. As a first test of this hypothesis, we generated data using the Tucker T3

model and analyzed it with PARAFAC to see if we could find T3 structures that produced the characteristic signature of degenerate solutions found with real data. We were successful. Furthermore, we began to see which features of the Tucker structures would give rise to the degeneracy. For the same pattern of off-diagonal cells, we synthesized data sets in which we varied the size of the superdiagonal elements of the Tucker core matrix (the elements for which $r = s = t$). When these were large compared to the other elements of the core (such as 2 or 4 versus -.5 to .9) then the resulting data had a structure dominated by aspects for which the PARAFAC model was appropriate, and thus the solution was not degenerate. However, when we synthesized data with superdiagonals that were only moderately larger than the other entries (such as 1.0 to 1.5 versus -.5 to .9), we obtained degenerate solutions. The sign pattern for the off-diagonal cells of the core matrix also seemed to affect the likelihood of obtaining degenerate solutions. Some patterns (for instance, negative entries in certain locations) increased the likelihood of obtaining degenerate solutions, but other patterns of core matrix entries did not produce degeneracies at all. However, since these investigations are still in the preliminary stages, further generalizations would be premature.

When the degenerate PARAFAC dimensions obtained upon analysis of the synthetic data were correlated with the "true" Tucker dimensions used to generate it we found a very interesting pattern. A particular dimension (for instance, dimension 1) would be well recovered in one mode, but another (for instance, dimension 2) would be well recovered in the other mode. This surprising pattern was similar to a phenomenon that was observed in the Harshman and De Sarbo (appendix C) analysis. When the "true" dimensions of the three-dimensional solution—that is, those obtained with the total sample—were correlated with the degenerate dimensions of the three-dimensional solutions (found when analyzing the split-half samples), we sometimes found that the "redundant" dimension of the degenerate solution resembled the second true dimension in Mode A but the third true dimension in Mode B. This similarity in the behavior of our synthetic and real data analyses further indicated that we were "on the right track" with our hypothesis about the cause of degenerate solutions.

Testing Recovery via the Orthogonality Constraint. As a next step, we tried applying the orthogonality constraint to the synthetic T3 data to see if it aided in recovering the "true" underlying dimensions. (By "true" dimensions, we now mean the ones associated with large superdiagonal cells in the T3 core matrix used to generate the data.) The success depended, once again, on the size of the superdiagonal cells versus the off-diagonal cells. For all but the most severe cases, however, the orthogonality constraint worked quite well, providing solutions in which the PARAFAC dimensions correlated 0.84 to 0.99 with the "true" Tucker dimensions. Furthermore, the intrinsic axis property of the PARAFAC solution allowed the "true" T3 dimensions associated with large superdiagonals to be recovered and interpreted *without rotation*.

Current State of this Research

Currently, we are collaborating with J. B. Kruskal of Bell Laboratories on a series of mathematical and empirical investigations of how and why the T3 model can give rise to such patterns of degeneracy. We are also exploring psychological models of patterns of individual differences (and thus particular core matrix structures) that might result in degeneracies and are beginning to examine real data that causes degeneracies to determine the core matrix that would be estimated by suitable Tucker analysis of such data.

As part of this work, Kruskal has been able to confirm mathematically that T3 data in which the elements of the core matrix have certain sizes will lead to degeneracy when fit by a PARAFAC model. For example, he can define relations among the core matrix elements that determine whether a $2 \times 2 \times 2$ core matrix has trilinear rank of three or two. When the core is rank 2 and far above the boundary to the rank 3 region, the PARAFAC solution will have nondegenerate form and should recover the "true" dimensions well. When the core is rank 3, however, unusual things can happen, including degenerate two-dimensional PARAFAC solutions. Some preliminary results were reported in two papers presented at the 1983 Psychometric Society meetings (Harshman, Kruskal, and Lundy 1983; Kruskal, Harshman, and Lundy 1983).

To summarize, it seems likely that the latent structure of many three-way profile data sets is well approximated by the PARAFAC-CANDECOMP model (after preprocessing); these do not yield degenerate solutions. And for others that may not conform quite as closely to the model, preprocessing may often make them suitable for PARAFAC analysis by changing the size of different elements in the core matrix so that the preprocessed data will yield acceptable solutions even without orthogonality constraints; this might account for many cases of "soft" degeneracies. For still other data sets, however, the T3 or T2 model may be the only sufficiently general representation, even after preprocessing; these data would give rise to "hard core" degeneracies.

Nonetheless, by use of orthogonality constraints, many of these latter cases might be usefully subjected to PARAFAC analysis. The orthogonality constraint apparently forces PARAFAC to fit a *subset* of the Tucker variations, those that can be simply and meaningfully expressed in terms of orthogonal PARAFAC dimensions.

To understand the full complexity of such data, of course, one needs to apply the Tucker model(s), perhaps by means of Kroonenberg and de Leeuw's (1980) TUCKALS programs (see also Kroonenberg 1981b, 1981c). However, it might be useful to use them in conjunction with the orthogonally constrained extended PARAFAC model, since it has the intrinsic axis property and so might help to determine an empirically meaningful rotation of the Tucker solution. In the same vein, it might be useful to add some Tucker-related options to PARAFAC, such as a method of estimating the core matrix associated with a given set of PARAFAC axes. There would seem to be a convergence of the two perspectives, and methods of combining the Tucker and PARAFAC

FAC-CANDECOMP approaches would seem to provide a valuable direction for future research.

More Basic Generalizations

In these chapters, we have considered PARAFAC1 and its extensions. We have only noted in passing that there are also fundamental generalizations of the model, such as PARAFAC2 (mentioned in chapter 5). These would allow the proportional profiles and intrinsic axis methods to be applied to data that currently cannot be properly fit, even by the extended versions of the PARAFAC1 model. For example, PARAFAC2 (Harshman 1972) allows oblique axis analysis of cross-product matrices, and PARAFAC3 allows even more general structures to be investigated. DEDICOM (Harshman 1978; Harshman, Green, Wind, and Lundy 1982) allows a type of factor analysis or multidimensional scaling of a matrix of relationships that are not necessarily symmetric (for instance, the number of telephone calls between different towns, alphabetic confusions, international trade balances, and so on). This model has both two- and three-way versions. There is also a "linked-mode" version of PARAFAC1, in which two or more different three-way data sets that share a common mode (for example, if all were based on the same stimuli) can be analyzed with the constraint that the dimensions for the common or linked mode be the same. Discussion of generalizations such as these is beyond the scope of this chapter. We merely mention that work on such models is being undertaken and that they may further extend the domain of data to which three-way factor-analytic and multidimensional scaling methodologies can be usefully applied.

NOTES

1. We should note in passing that less subjective measures can also have the problem of an uncertain origin. Interval-scale data can arise even when the measures have a superficial true zero (for example, counts of the number of observations that fulfilled certain criteria, or the number of people in a certain location at a certain time) if there is a threshold before nonzero counts begin to be observed. Even though such counts represent a ratio-scale measure of the directly observed quantities (for example, the number of observations fulfilling a particular criterion), they may represent only an interval- or ordinal-scale estimate of the underlying *influence* that one is trying to get at by means of those quantities. Horst (1965) gives the example of items designed to measure intelligence. A count of the number of test items passed successfully might seem like a ratio-scale value with a true zero, but if the items were all of sufficient difficulty, then younger children or less-bright adults might fail all the items and yet still have considerable intelligence.

2. Some might object that the effect of centering cannot always be this simple, since when X is full column rank (rank m for an n by m matrix), row-centering will reduce it to rank $m - 1$. But even this is just the effect of a translation of axes.

For example, consider a 3×3 square matrix representing the position of 3 points in a three-dimensional space. Centering this matrix shifts the points so that the origin of the space is at the centroid of the points. This makes all three points lie on a plane through the origin and hence reduces the dimensionality of their space from three to two, simply by translation of axes.

3. Others have informally called them "spaghetti," "lasagna," and "meatloaf" centering.

4. We do not agree with some of Kroonenberg's (1983) preprocessing recommendations, for reasons that should be apparent after reading the preceding section of this chapter. His discussion of preprocessing and many examples are most instructive, however, and his book is a valuable source of information about three-mode factor analysis in general. Unfortunately, we did not receive it in time to discuss it elsewhere in this chapter.

5. This perspective was first suggested to us by Kruskal (personal communication, 1980) and is explicitly mentioned in Kruskal (1983), reprinted in this volume. We have also been recently informed by Kroonenberg (personal communication, 1983) that Lohmoller (1979) has considered three-way models with some additional additive terms. We have not yet seen this work, however, and so cannot comment on it.

6. This example is based on a study being done at the University of Western Ontario by P. Ossenkopp, L. Sorensen, and D. Mazmanian.

REFERENCES

- Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass.: The MIT Press.
- Cattell, R. B. 1966. The data box: Its ordering of total resources in terms of possible relational systems. In *Handbook of multivariate experimental psychology*, ed. R. B. Cattell. Chicago: Rand McNally.
- Dawson, M. 1982. Multidimensional responses to metaphor. Master's thesis, University of Western Ontario.
- Gandour, J. T., and R. A. Harshman. 1978. Cross-language differences in tone perception: A multidimensional scaling investigation. *Language and Speech* 21:1-33.
- Gollob, H. F. 1968. A statistical model which combines features of factor analysis and analysis of variance techniques. *Psychometrika* 33:73-116.
- Gower, J. C. 1977. The analysis of three-way grids. In *The measurement of intrapersonal space: Dimensions of intrapersonal space*, vol. 2, ed. P. Slater. London: John Wiley and Sons.
- Hampson, E. Forthcoming. Three-way factor analysis of hand movement patterns during verbal and nonverbal tasks. University of Western Ontario.
- Harshman, R. A. 1972. PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics* 22:30-44 (University Microfilms International, Ann Arbor, Michigan, Order No. 10,085).

- . 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.
- 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., 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., J. B. Kruskal, and M. E. Lundy. 1983. Some theoretical and practical relationships between Tucker's three-mode factor analysis and PARAFAC-CANDECOMP analysis. Paper presented at the annual meeting of the Psychometric Society, Los Angeles, June.
- Harshman, R. A., P. Ladefoged, and L. Goldstein. 1977. Factor analysis of tongue shapes. *Journal of the Acoustical Society of America* 62:693-707.
- Horst, P. 1965. *Factor analysis of data matrices*. New York: Holt, Rinehart and Winston.
- 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. Scaling of input data for three-mode principal component analysis. Rijksuniversiteit Leiden.
- . 1981b. *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.
- . 1981c. *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.
- . 1983. *Three-mode principal component analysis: Theory and applications*. Leiden, The Netherlands: DSWO Press.
- 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. 1977. Some least-squares theorems for matrices and N-way arrays. Manuscript, Bell Laboratories, Murray Hill, New Jersey.
- . 1983. Multilinear methods. In *Statistical data analysis*, ed. R. Gnanadesikan. Providence, R.I.: American Mathematical Society.
- 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.
- Lindau, M., R. A. Harshman, and P. Ladefoged. 1971. Factor analysis of formant frequencies of vowels. *UCLA Working Papers in Phonetics* 19:17-25 (University Microfilms International, Ann Arbor, Michigan, Order No. 10,085).
- Lohmoller, J. B. 1979. Die tridmodale faktorenanalyse von Tucker: Skalierungen, rotationen, andere modelle. *Archiv fur Psychologie* 131:137-66.

- 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.
- 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.
- Stevens, S. S. 1946. On the theory of scales of measurement. *Science* 103:677-80.
- Terbeek, D. 1977. A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics* 37:1-271 (University Microfilms International, Ann Arbor, Michigan, Order No. 10,085).
- Terbeek, D., and R. A. Harshman. 1972. Is vowel perception non-Euclidean? *UCLA Working Papers in Phonetics* 22:13-29.
- Weeks, D. G., and P. M. Bentler. 1979. A comparison of linear and monotone multidimensional scaling models. *Psychological Bulletin* 86:349-54.