In this issue:

MULTIVARIATE MODELS FOR
THREE-DIMENSIONAL DATA

James W. Grice
Oklahoma State University

As early as the publication of Raymond Cattell's (1966) 'basic data relation matrix', or 'data box' researchers have been interested in data structures that extend beyond two dimensions. The development of statistical and mathematical models as well as accessible computer software for the analysis of three-dimensional data structures, particularly, has accelerated in recent years. A three-dimensional data structure might, for example, be comprised of 500 persons who rated themselves on 50 items from a personality questionnaire on 10 different occasions. The resulting $500 \times 50 \times 10$ data cube could be analyzed to explore the multivariate relationships between the three *facets*: persons, items, occasions. One general strategy would entail applying traditional factor analyses to different pairwise combinations of the facets by collapsing over the third facet. Gorsuch (1983, pp. 322-327) describes a number of variations on this approach. Another strategy, originating with Tucker (1964), would entail factoring all three facets simultaneously, yielding a 'core matrix' that describes their interrelations.

In this special issue of *Applied Multivariate Research* Kroonenberg, Harshman, and Murakami offer an excellent review of Tucker's original model and also compare it to Harshman's (Harshman & Lundy, 1984) popular Parafac model for analyzing three-dimensional data matrices. They furthermore provide clear guidance regarding practical decisions that must be made when employing either model, and they use a genuine parenting styles data set to exemplify the issues. While the models may appear complex at first glance, Kroonenberg and his colleagues show how the analyses parsimoniously uncover common styles of parenting while simultaneously revealing individual variability between families.

Leenen and Ceulemans also compare and contrast two different models for analyzing thee-dimensional data matrices. Their paper, however, is centered around the HICLAS method introduced by De Boeck and Rosenberg (1988) for modeling the hierarchical relations among binary variables. HICLAS essentially weds a form of binary factor analysis with set theory, and in this issue Leenen and Ceulemans compare and contrast two different models: INDCLAS and Tucker-3 HICLAS. Like Kroonenber's paper, their contribution reveals Tucker's legacy to analyzing three-dimensional data matrices. Leenen and Ceulemans analysis of two genuine data sets suggest that the Tucker-3 HICLAS model may yield more parsimonious results than the INDCLAS model. In an interesting example that is near and dear to my own heart, given my interests in person-centered statistics (see Grice, 2007), the authors also show how their technique can be applied to a single case. The intra-personal perceptions (or 'object relations') of a young

woman are explored, and some interesting findings regarding the stability of her emotional relations to her father and mother are revealed through the analysis.

As another strategy for analyzing three-dimensional binary data in this special issue, González, Tuerlinckx, and De Boeck discuss the 2sSEM model, which is essentially a log-linear approach with Bayesian estimation. They demonstrate how the 2sSEM model can be used to model the structure of a *sociomatrix*, which is a binary matrix that records the links between different people in a social network. After pointing out the limitations of aggregating across dimensions to analyze three-dimensional data and then presenting the formal model, they demonstrate their technique using data from secondary students. González et al.'s results reveal very interesting asymmetries in the students' perceptions of themselves regarding their popularity. The results furthermore confirm the loss of information that can occur when three-dimensional data structures are aggregated and analyzed with traditional two-dimensional approaches.

Finally, the similarities between Cattell's original data box and Generalizability Theory are obvious, including the use of the term 'facets' to refer to the dimensions that constitute the box. In their landmark book on Generalizability Theory, Cronbach, Gleser, Nanda, and Rajaratnam (1972) viewed analysis of variance (ANOVA) as a method for wedding data boxes to psychometric theory. Similarly, Gower (1975) introduced Generalized Procrustes Analysis (GPA) as a technique for exploring the structure of three-dimensional data matrices using ANOVA. In this special issue Grice and Assad explain the fundamental features of GPA, including the centrality of the 'consensus matrix' which is essentially the average of a series of rotated two-dimensional matrices. ANOVA is used to assess the variability of the different two-dimensional data matrices about the consensus matrix along the third facet in the data matrix. Grice and Assad describe the results of a study in which a small group of students were asked to rate themselves and each of their peers on Big Five trait scales. Results from the GPA revealed a moderate degree of similarity among the trait profiles even though the students had known each other for only seven days.

This special issue thus brings together a variety of different methods and statistical techniques for analyzing three-dimensional data matrices. They are common to the extent they permit researchers to extend beyond examining only two dimensions and to discover multivariate relationships that would otherwise be masked in aggregates or untapped by simpler analyses. In other words, these models open up new avenues of research. Of course with similarity, one must also find difference, and the methods in this special issue are not exempt from this dialectic. It could be argued that they differ primarily with regard to the statistical models underlying the data analysis and secondarily with regard to their technical features (e.g., examining classes with the HICLAS models). Perhaps this statement is an oversimplification, but a thorough and insightful comparison of the techniques is beyond the scope of this brief introduction. Such comparisons would certainly be beneficial, much like the comparisons that have been made between Multidimensional Scaling and Factor Analysis over the years. It is our hope that this special issue will spur such comparisons as well as novel investigations in psychology, marketing, education, and other domains of empirical thought where three-dimensional data structures are encountered.

INTRODUCTION

REFERENCES

Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand-McNally.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika, 53*, 361-381.

Gorsuch, R. L. (1983). *Factor analysis*. New Jersey: Erlbaum.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*, 33-51.

Grice, J. W. (2007). Person-centered structural analyses. In R. Robins, C. Fraley, and R. Krueger (Eds.) *Handbook of research methods in personality psychology* (pp. 557-572). New York: Guilford Press.

Harshman, R. A., & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122-215). New York: Praeger.

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart & Winston.

ANALYSING THREE-WAY PROFILE DATA USING THE PARAFAC AND TUCKER3 MODELS ILLUSTRATED WITH VIEWS ON PARENTING[1]

Pieter M. Kroonenberg
Leiden University, Leiden, The Netherlands

Richard A. Harshman[2]
University of Western Ontario, London, Ontario, Canada

Takashi Murakami
Chukyo University, Toyota, Japan

ABSTRACT

In this paper two major models for three-way profile data, i.e., the Parafac model and the Tucker3 model are discussed from the point of view of application. Topics treated are handling the data before analysis, model choice, choice of dimensionality, model fit, algorithmic hazards during the analyses, and interpretation and validation of the results. These issues are discussed in some detail so that prospective users can take guidance for analysing their own data.

The data provided by Japanese girls and their parents about the parenting style in their family are the major vehicle for demonstrating the issues touched upon. The general results from these data are that parental styles consisted of three groups of behaviours: Acceptance, Control and Rejection, and Discipline. Within families the parenting behaviours of fathers and mothers are seen as parallel rather than at cross purposes, both by the daughters and the parents themselves. Moreover, daughters and parents largely agree about the parenting style itself. Notwithstanding, there are also families in which daughters and parents disagree about the parenting style in particular about Acceptance and Control, but not about Discipline.

INTRODUCTION

The focus of this paper is the analysis of three-way profile data using the two most common models for such data, i.e., the Parafac model and Tucker3 model. Before we can discuss the models we have to look at three-way data themselves and how they have to be handled before analysing them with one the three-mode models[3]. After having considered these data preliminaries, we present the two analysis models in a rather conceptual fashion referring the reader for technical details elsewhere. How these models work in practice is the central theme of this

paper and this is illustrated with a detailed application of the Kojima data on judgements of parental behaviour towards their daughters (Kojima, 1975). Not only the working of the models is presented but also a comparison is made between them. Along the way we have formulated a number of recommendations which should allow applied researchers to start on a three-mode analysis of their own data and delve into the literature further.

Two-Way Profile Data: Setting the Scene

A common research goal in non-experimental psychology is to improve the understanding of a conceptual domain via studying a number of variables to measure aspects of that (multidimensional) domain. Often subjects are considered to be random samples from some population, but in reality they seldom are. Moreover, it is an empirical question whether in the sample subgroups exist which have different characteristics. Therefore, one should not only analyse the structures derived from the variable correlations, but also pay attention to individual differences or group differences during an analysis. A balanced interest in subjects and variables is called for.

Assume for the moment that we have drawn a random sample and that we want to assess the structure of the variables in a two-way (i.e., two-subscripted) data set consisting of individuals (rows) by variables (columns). Such data are an example of *two-way profile data* with the subjects having profiles across the variables. The variables or items could form a questionnaire constructed to contain a number different conceptual domains, and the aim of the investigation is to verify in a non-confirmatory way whether the items indeed group into their hypothesised domains. In such investigations, the prime interest is in the variables, and the subjects are exchangeable, one person is as acceptable as another as long as they come from the same population. This implies that individual differences are not of interest because the subjects are not treated as individuals but as members of a particular population.

In the case that the variables do not group into their domains, what to make of the variable grouping that one finds? If the researcher is primarily interested in refining the measurement device for the hypothesised domains, items are deleted that do not group as desired. Alternatively, wordings of items are modified in the hope this will put them within the intended domain, or new better items are constructed. With the thus amended measurement instrument, new data are gathered drawing again from the appropriate population, and so on until the instrument works as intended.

A different situation arises if the ultimate aim of the researcher is not the construction of a measurement instrument according to preconceived ideas, but the interest is in the psychological properties, processes, and structure in the domain from which the items are drawn. In other words, the data set at hand needs to be analysed in order to discover the underlying principles which guided the responses of the subjects. At the same time, because it is the data themselves which are of interest, there is a necessity to investigate the organisation of the subjects as well, especially if they cannot be guaranteed to be a truly random sample from an unstructured population. The purpose of such an analysis is not

merely descriptive and interpretive, but the aim of the researcher is to make an inductive, or theoretically principled, use of the analysis. In such a case, the investigator wants to learn more about the substantive domain, wants to investigate multi-item properties, such as "cognitive-affective dimensions" or other characteristics that span multiple items and account for their statistical correlations.

Multivariate methods in many textbooks are couched in terms of algebra and distributions of statistics, but there is also a geometric side to these methods (see, for instance, Wickens, 1994) especially in exploratory analyses, such as principal component analysis, multidimensional scaling, and correspondence analysis. An excellent book discussing multivariate methods along these lines is Legendre and Legendre (1998). In the extensions of component analysis discussed in this paper this geometric view is paramount.

The basic elements of the geometric view of multivariate statistics are that subjects can be seen as points defined in the space defined (or *spanned*) by the variables as is well-known in the case of scatterplots for two variables. In the same way, the variables can be seen as points (usually represented by arrows, vectors or axes) in the space defined by the subjects. This geometric symmetry is especially exploited in multiway analysis.

Suppose one wishes to reduce the multidimensional space spanned by many variables to a lower-dimensional one with the principal aim of investigating the relationships of the variables. In such a case one does not necessarily have to assume that important (psychological) principles underlie the data, but one simply wants to investigate the major relationships between variables and subjects in a lower-dimensional space without explicitly seeking to interpret the axes in the space. Necessarily, one wants the low-dimensional subspace to explain as much variance as possible to be sure to capture as much structural variance as possible in a few dimensions.

Principal component analysis and 'exploratory factor analysis' have been the major analysis vehicles for investigating low-dimensional subspaces to disentangle relationships between variables, subjects and their interrelationship. However, factor analysis, neither the technique nor the term, will feature in this paper. For expository purposes, the somewhat controversial position is taken here that there is a clear distinction between (exploratory) component analysis and (confirmatory) factor analysis, and the present paper deals exclusively with component techniques.

Three-Way Profile Data

Often data are more complex than individuals by variables, for instance when the individuals are measured under several conditions or at more than one time point. The resulting data have a three-way format of, for instance, subjects having scores on variables under several conditions (*three-way profile data*) or of situations which are judged on a number of scales by several individuals (*three-way rating data*). To tackle the investigation of the structure in such data, it seems natural to look for generalisations of the two-mode component analysis for their analysis. These generalisations or three-mode component models are the subject of this paper; in particular two of the most commonly used models for this pur-

pose, the Parafac model and the Tucker3 model. Several earlier didactically oriented papers exist which explain three-mode analysis in detail, among others Bro (1997); Harshman (1994b); Kroonenberg (1994); Van Mechelen and Kiers (1999); Kiers and Van Mechelen (2001), and many practical issues are treated in detail in Kroonenberg (2008). The present paper treats and illustrates the two major three-mode models in a single paper which we hope will provide insights into each method as well as an overall understanding of their relationship.

We will restrict ourselves here to three-way profile data to keep the discussion manageable; for a discussion of three-way rating data see Kroonenberg (2008, chap. 14). Subjects, variables, and conditions for the three modes will be the labels used, but obviously these labels are generic and in other sciences other types of three-way data occur. Even though many categorical variables are collected especially in the social and behavioural sciences, and three-mode methods exist to deal with them (see, for instance, Kroonenberg, 2008, chap. 17 and 18); in this paper we will limit ourselves to numeric variables. We will also not discuss data in which there exists a specific time structure or order relationship between the measurement conditions. The introduction of order information into an analysis introduces specific technical problems but also interpretational opportunities but treating them would take us too far afield.

Three-way profile data correspond to a multivariate repeated-measures design. The major difference between the standard multivariate analysis and our treatment with three-mode models is our focus on individual differences between the subjects as well as on the underlying correlational structure for the variables. If individuals are a random sample from one or more populations, i.e., the subject mode is a *stochastic mode*, only means and covariances are of interest. So that one may restrict oneself to their analysis without paying attention to the individuals themselves; see Bentler & Lee (1979); Oort, (1999), and their references for stochastic three-mode models. However, often the subjects are analysed as if they are a population rather than a (random) sample from a well-defined population. Moreover, there often is an interest in changes in the correlational structure between the variables across conditions. In such cases, three-mode analysis is particularly useful. The value of a three-mode analysis is enhanced if additional information is available about the subjects. If it is not, one can still fruitfully describe the differences found in terms of the data contained in the three-way array, but not in terms of other characteristics of the subjects.

If a data set is approached as a population or if one sees the three-mode analysis as a purely descriptive technique for the sample at hand, then sample size is not necessarily a serious concern. However, in small data sets, the description of the majority of the data can be easily distorted due to outlying data points, a specific level of a mode or both.

In three-way profile data, the subject mode often can be considered both as an individual differences mode and as a stochastic one, which allows for calling upon techniques which need stochastic modes for their application. For instance, given a stochastic mode with unknown distributional properties, the (statistical) stability of (parts of) a three-mode solution can nevertheless be assessed via a bootstrap analysis (see Kiers, 2004). Split-half procedures for assessing stability also assume a stochastic mode, as do imputation procedures for generating multiple data sets in the case of missing data (Kroonenberg, 2008, chap. 7).

Preprocessing Three-Way Profile Data

Because numeric variables often have different measurement scales and ranges, their values are not necessarily comparable and nearly always some form of preprocessing is required, i.e., means have to be removed (*centring*) and division by a scale factor has to be done as well (*normalisations*). The standard way to handle three-way profile data before the analysis proper is to centre per variable-condition combination (*jk*), usually per column across subjects, and to normalise per variable slice *j* so that $z_{ijk} = (x_{ijk} - \bar{x}_{\bullet jk})/\sum_{ik}\frac{1}{IK}(x_{ijk} - \bar{x}_{\bullet jk})^2)^{1/2}$, where x_{ijk} are the raw data and z_{ijk} the preprocessed data. A · refers to an index after summing over it.

This centring is recommended because it eliminates the average score on each variable for each occasion so that all scores are in deviation of the scores of the 'average person'. This is the person whose scores are exactly at the mean of each variable at each occasion. Thus all z_{ijk} are in deviation of the variable means and normalised per variable *j*. In longitudinal studies, this type of centring has the effect of removing the time trends over occasions, so that these trends have to be investigated independent of the variation around these trends. The suggested centring has the additional advantage that if there is a real meaningful, but unknown, zero point of the scale, it is automatically removed from the scores (see Harshman & Lundy, 1984b, for a detailed discussion of this point). A further effect of this centring is that the component coefficients of the subjects are in deviation of their means.

Normalising over all values of each variable means that the normalised deviation scores carry the same meaning for all occasions. Thus a standard deviation has the same size in the original scale for each occasion. A similar argument is used in multi-group structural equation modelling for using covariance rather than correlation matrices; Kroonenberg (2008, chap. 6) discusses the recommendation for normalisation in greater depth.

The effect of the recommended preprocessing is that the variable coefficients on the components can be adjusted in such a way that they can be interpreted as variable-component correlations as shown in Table 1. The recommended preprocessing is the only way to get this interpretation.

Objectives of Three-Mode Analysis

The major aim of applying three-mode analysis to three-way profile data is to unravel complex patterns of dependencies between the observations. All models have component matrices for each of the modes, thus one set of components for the subjects, one for the variables, and one for the conditions. Moreover, the models also contain weights which provide the size of the links between components of the different modes. In the Parafac model these weights are such that a component of one mode is linked only with one component of another mode, but in the Tucker3 model links between all components of all modes are allowed. This makes the latter model inherently more complex, but also more flexible in fitting it to data. Details of this are worked out when we discuss the models themselves.

Table 1. Kojima Data: Scale and Condition Coordinates

Scale	Order	CRPBI	Parafac Component-Scale Correlations			Tucker3 Rotated Components		
			1	3	2	1	2	3
Parental support or Acceptance								
Acceptance	1	AC	**.64**	.24	**.27**	**.45**	.01	.00
Acceptance individuation	13	AC	**.64**	.17	**.29**	**.44**	-.03	.05
Child centredness	2	AC	**.54**	.27	**.32**	**.41**	.05	.05
Positive involvement	7	AC	**.54**	**.30**	**.40**	**.42**	.08	.04
Behavioural and Psychological control								
Possessiveness	3	PC	.06	**.38**	**.47**	.17	**.24**	.14
Instilling persistent anxiety	15	PC	-.11	**.49**	**.41**	.07	**.34**	-.00
Intrusiveness	8	PC	.16	**.49**	**.37**	.21	**.28**	-.08
Hostile control	10	PC	-.17	**.61**	**.34**	.03	**.39**	-.08
Enforcement	6	FC	-.18	**.49**	**.34**	-.01	**.34**	-.08
Control	5	FC	.02	**.56**	**.32**	.13	**.34**	-.13
Rejection								
Control through guilt	9	PC	**-.34**	**.35**	**.36**	-.09	**.31**	.08
Withdrawal of relations	17	PC	**-.45**	.23	**.35**	-.16	**.27**	.18
Rejection	4	AC	**-.53**	.24	**.32**	-.22	**.27**	.18
Hostile detachment	16	AC	**-.58**	.16	.24	**-.27**	**.24**	.17
Lax discipline								
Inconsistent discipline	11	PC	**-.30**	.14	**.43**	-.08	.17	**.34**
Lax discipline	14	FC	**.37**	.04	-.00	.10	-.03	**.49**
Extreme autonomy	18	FC	**.32**	.02	-.20	.06	-.11	**.51**
Nonenforcement	12	FC	**.28**	.01	-.22	.04	-.12	**.47**
Standardised weights			.14	.12	.12	.20	.17	.08

<u>Conditions - Normalised Coordinates</u>

Daughter-Father		D-F	**.56**	**.70**	.23	**.69**	.01	
Daughter-Mother		D-M	**.59**	**.67**	.22	**.72**	-.01	
Father-Father		F-F	**.38**	.22	**.63**	-.01	**.72**	
Mother-Mother		M-M	**.44**	.12	**.72**	.01	**.70**	

The left-hand panel contains the component-scale correlations and the normalised condition coefficients from a three-component Parafac analysis. The right-hand panel contains the scales components after a Harris-Kaiser independent cluster rotation (see p. 33) and the varimax-rotated normalised condition components from a $4 \times 3 \times 2$-Tucker3 analysis. Both sets of components are orthogonal.

Cosines between the Parafac scale components: $\cos(\theta_{12}) = -.01$; $\cos(\theta_{31}) = -.02$; $\cos(\theta_{23}) = .77$. Cosines between the Parafac condition components: $\cos(\theta_{12}) = .81$; $\cos(\theta_{31}) = .92$; $\cos(\theta_{23}) = .53$.

Numbers of the scales is the sequence order in the CRPBI; they are used to label the scales in Fig. 4. *Official CRPBI scales*: PC = Psychological control; FC = Firm Control; AC = Acceptance.

Model and Dimensionality Selection

Even though we have referred to the three-mode models as the Parafac model and the Tucker3 model, it is better to think of classes of three-mode models, because each model can occur with different numbers of components. In practice we will therefore not only have to choose a type or class of model, but within those classes we will have to decide how many components we need for analysis, just as we have to choose the number of components in ordinary two-mode principal component analysis.

Interpretation

There are essentially four different ways in which the basic results from a three-mode analysis can be presented, i.e., (1) *Tables* of the coefficients or loadings for each mode—rotated or not; (2) Separate *pair-wise graphs* of the components per mode; (3) *All-components plots* showing all components of a single mode in a plot with the levels of the modes on the horizontal axis; (4) *Per-component plots* showing one component of each of the three modes in the same plot. As we shall see, the latter type of plots conforms most closely to the spirit of the Parafac model, because they allow for the simultaneous inspection of the coefficients of the three modes per component. The pair-wise component plots make more sense when the spatial characteristics of the solution are to be examined and this is more in line with the characteristics of the Tucker3 model. Further details will be discussed in later sections.

Validation

In validating three-mode results, one may look internally towards the analysis itself and investigate the statistical stability of the solution via bootstrap analyses; see Kroonenberg (2008, chap. 8). Furthermore, one may look at the residuals to see which parts of the data have not been fit well and to what extent there is much systematic information left. In Tucker models nearly every part of the model can be evaluated in terms of fit to the data. Another approach to validation is looking at information external to the data set available on the entities which make up the data set to see whether they can help to shed further light on the patterns found. For instance, design information for the variables, background information on the subjects, etc.

THREE-MODE MODELS

In this section we discuss the two most common three-mode models, the Parafac model and Tucker3 model. With these models the same type of data can be analysed but they differ in the assumed underlying structure in the data, the ease of fitting the models to data, and the number and type of choices one has make to find an appropriate model for the data at hand. We will take up each model in turn and discuss some of its most salient properties. For full details one should look in the reference mentioned earlier. However, before entering into the discussion of three-mode model, a brief recapitulation of principal component

analysis and the singular value decomposition will be given to set the scene for the three-mode discussion.

Singular Value Decomposition: Basis for PCA

Two-way profile data are usually collected in a two-way data matrix $\mathbf{X} = (x_{ij})$ of I subjects by J variables. The variables do not have necessarily the same scales and thus some type of normalisations will generally be required. We will refer to the entities within a mode as *levels*, thus the variables and subjects are the levels of their modes.

In standard (two-mode) component analysis, we attempt to fit a two-mode model to two-way profile data of subjects by variables as follows:

$$x_{ij} = \sum_{s=1}^{S} a_{is} b_{js} g_{ss} + e_{ij} = \sum_{s=1}^{S} a_{is}(b_{js} g_{ss}) + e_{ij} = \sum_{s=1}^{S} a_{is} \widetilde{b}_{js} + e_{ij}, \qquad (1)$$

where the a_{is} are the coefficients or scores for the subjects, and per component they can collected into a column vector $\mathbf{a}_s$. Such columns are also known as the 'subject' (or left) singular vectors. Similarly, the b_{js} are the coefficients for the variables, and per component they can be collected into the column vectors $\mathbf{b}_s$, which are the 'variable' (or right) singular vectors. The g_{ss} are the square roots of the eigenvalues (or *singular values*) which represent the standard deviations of the components given the standard centring and normalisations. The model,

$$x_{ij} = \sum_{s=1}^{S} a_{is} b_{js} g_{ss},$$

is known in the statistical literature as the *singular value decomposition*. As Figure 1 shows the g_{ss} constitute the diagonal of the matrix with singular values, $\mathbf{G}$, and in the light of the coming generalisation to three modes, this matrix is called the *two-mode core matrix*. If the g_{ss} are absorbed in the variable coefficients, the a_{is} are generally referred to as (standardised) component scores, while the $f_{js} = b_{js} g_{ss}$ are then referred to as loadings (see left-hand panel of Figure 1). The column vectors or components $\mathbf{a}_s$ together form the matrix $\mathbf{A}$ of component scores. The column vectors $\mathbf{f}_s$ form the matrix of component loadings $\mathbf{F}$.
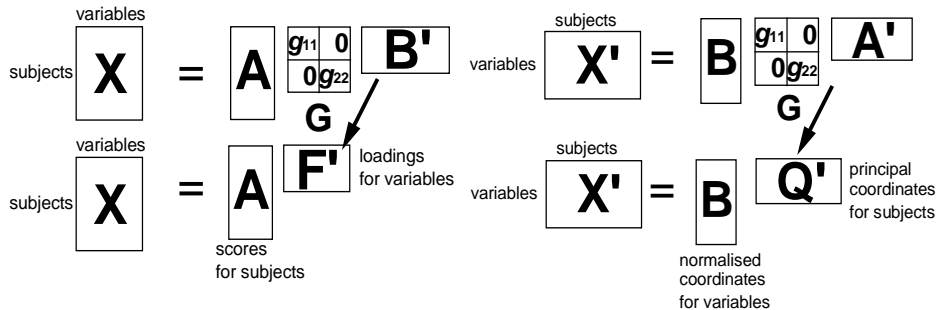

Figure 1. Singular value decomposition and principal component analysis.

If the g_{ss} are absorbed in the subject coefficients, the $\mathbf{q}_s$ are vectors of scores whose variance is equal to the eigenvalues (see right-hand panel of Figure 1). These two different ways of presenting the principal component model were

already present in the early days of component analysis (see Ten Berge & Kiers, 1997, for a complete discussion).

Parafac Models

Parallel Proportional Profiles

The fundamental idea underlying the Parafac model was first formulated by Cattell (1944) in the form of the principle of parallel proportional profiles which he explained in Cattell and Cattell (1955) as follows:

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

The parallel proportional profile principle has great intuitive appeal, exactly because it determines the orientation of component axes without recourse to assumptions about the desired pattern of loadings. In particular, there is no attempt to maximise simplicity or any other structural property *within* a given occasion. Instead, the proportional profiles criterion is based on a simple and plausible conception of how loadings of "true factors" would vary *across* occasions.

A further strength of the parallel proportional profile principle is that it is based on patterns in the data that could not have easily arisen by chance, or as an artefact of the way that the study was conducted. When it can be empirically confirmed that such proportional shifts are present and reliable, one has established an important fact about the data, one that is in accordance with the fact that there are underlying factors which have a particular axis orientation (see Harshman & Lundy, 1984a, pp. 147-152, 163-168, for a more detailed presentation of the argument).

Disadvantages of the Parafac model with respect to models or rotations with specific restrictions on the components is that no specific theories can be tested nor very simple, easily interpretable, components can be searched for. One way of attempting to get the best of both worlds is to use constrained Parafac models (see, for instance, Krijnen, 1993 and Bro, 1998, for several ways how this can be done). Furthermore, in practice very often data sets only support a limited number of components with parallel profiles, while other three-mode models can describe

additional patterns which do not conform to the parallel proportional profile principle. This is one of the important differences between the three-mode models discussed here.

The Basic Parafac Model

Three-way profile data can be collected in an *I* by *J* by *K* three-way data array $\mathcal{X} = (x_{ijk})$ of *I* subjects by *J* variables by *K* conditions. The variables have not necessarily the same scales and thus some type of normalisations will generally be required. Again, we will refer to the entities within a mode as levels, thus the variables, subjects, and conditions are the *levels* of their modes.

The basic Parafac model for a three-way array $\mathcal{X}$ with elements x_{ijk} consisting of the scores of *I* subjects on *J* variables under *K* conditions is a direct extension of standard two-mode PCA (Equation 1), and has the form,

$$x_{ijk} = \sum_{s=1}^{S} a_{is} b_{js} c_{ks} g_{sss} + e_{ijk}, \tag{2}$$

where the subject scores a_{is}, the variable loadings b_{js}, and the condition weights c_{ks} are the elements of the component matrices **A**, **B**, and **C**, respectively. Note that the model shows great similarity to the two-mode case (Equation 1), but it has an additional set of coefficients for the conditions. In the Parafac model, all modes take identical roles so that there is no a priori designation of variable coefficients as loadings in the sense of variable-component correlations and subject coefficients as standardised scores as is common in two-mode analysis. However, we often use these terms in this paper for ease of presentation without assuming the specific interpretations attached to them in two-mode analysis.

The weight of the *s*th component, g_{sss}, indicates its importance. These weights can be interpreted in the same way as their two-mode counterparts, the singular values g_{ss}, be it that they are not necessarily standard deviations. Moreover their squared values g_{sss}^2 in general do not add to the total variance explained unless the components are uncorrelated. The squared values do, however, indicate the variability accounted for by their components and indicate their importance in reconstructing the data. In standard two-mode component analysis the g_{ss} fit onto the diagonal of a square matrix (see Fig. 1), and analogously the g_{sss} can be collected on the diagonal of a cube (see Figure 2; top). Such a diagonal is called a superdiagonal and the cube $\mathcal{G}$ (i.e., $g_{pqr} = 0$ if $p \neq q \neq r$) is referred to as a core cube.
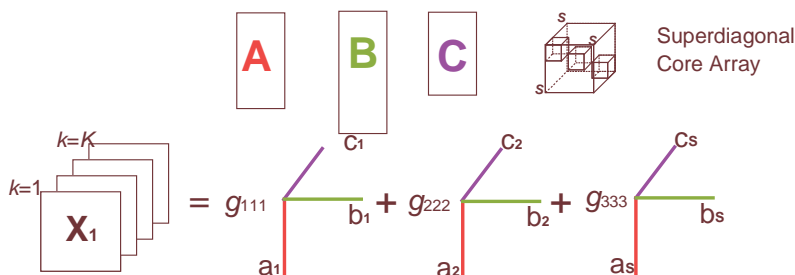

Figure 2. The Parafac model: Sum of combinations of components; one from each mode for each component.

THREE-MODE MODELLING OF THREE-WAY PROFILE DATA

The papers by Harshman and Lundy (Harshman & Lundy, 1984a, 1984b; Harshman, 1984) together are the most complete treatment of the model and many related issues to date. More recent treatments are to be found in the following books: Krijnen (1983), Bro (1997), Smilde, Geladi, and Bro (2004), and Kroonenberg (2008).

Uniqueness and explicit models

Compared to the Tucker3 model to be discussed, the Parafac model has, given the number of components, a unique solution apart from trivial rescalings and reorderings of the components, given there is ''adequate'' variation across all three modes; to use the term originally used in Harshman (1970). The easiest set of conditions which define adequacy is that at least in one of the modes the components are nonproportional, i.e., $\mathbf{c}_s \neq k\,\mathbf{c}_s$ for any constant k. Adequate variation in all three modes is central to the uniqueness of the model and the presence of systematic variation in the data; full details can be found in Kruskal (1984), Carroll and Chang (1970), and Harshman (1972b).

The uniqueness makes the Parafac model very attractive if it is known or hypothesised that it gives an adequate representation of the structural patterns in the data. Harshman uses this uniqueness property (or 'intrinsic axis property', as he also calls it) to search for ''real'' psychological factors basing himself on Cattell's parallel proportional profiles as was indicated in the quote in the opening paragraph of this section. The parameters in the model depend only on one of the indices i, j, k, thus they may be seen as proportionality constants for entire components. Thus, the conditions weight each $a_{is}b_{js}$ with the same weight, c_{ks}, irrespective the variable j or the subject i. For example, the component $\mathbf{a}_s$ is proportionally enlarged or decreased by an amount c_{ks} for a level k of the third mode. Thus across conditions the components $\mathbf{a}_s$ are parallel; hence the name of the model.

Tucker Models

The Tucker models form the other class of models which are generalisations of two-mode principal component analysis, and they are named after their proposer Ledyard R Tucker. Tucker (1966) was the first to propose workable three-mode models, to present algorithms for fitting them, and to show a number of real-data examples. There are several variants of the basic model proposed by Tucker, such as the Tucker3 and Tucker2 models. In this paper we will only discuss the Tucker3 model which features components for all three modes. The central difference between the Tucker3 model and the Parafac model is its property that each mode has its own components rather than that there is one set of components to which all modes refer. In two-mode analysis, the components in the subject and variable spaces are uniquely linked to each other and can therefore be considered to be the same (all $g_{ss'}$ are zero). However, this is not so in the Tucker3 model where any components of one mode can be linked to any component of the other modes. On the other hand, the exclusive link between components of different modes is retained in the Parafac models.

The Tucker3 model can be used to model a data array into all its components as is true for PCA in the two-mode case, so that one can always find a complete solution for any three-mode data set. The model applied to a three-way array X with elements x_{ijk} has the form:

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} , \qquad (3)$$

where the scores a_{ip}, the loadings b_{jq}, and occasion coefficients c_{kr} are the elements of the component matrices **A**, **B**, and **C**, respectively. The g_{pqr} indicate the weight of the link between the pth component of the first mode, qth component of the second mode, and the rth component of the third mode. The g_{pqr} can be collected in a rectangular 'box' called the *core array* $\mathcal{G}$, and the e_{ijk} are the errors of approximation (see Fig. 3). In this presentation we will not go into the question, whether the e_{ijk} are really random disturbance terms or that they represent information which may be decomposed further. The symmetry with respect to the modes (see Equation 3) makes that all modes take identical roles so that there is no a priori need to designate variable coefficients as loadings and subject coefficients as scores. However, also for the Tucker3 model we often use these terms for ease of presentation without assuming the specific interpretations attached to them as in two-mode analysis. The Tucker3 model can also be written in matrix form, but we refrain from this in this paper; for the technical details see Kroonenberg (2008, chap. 4).
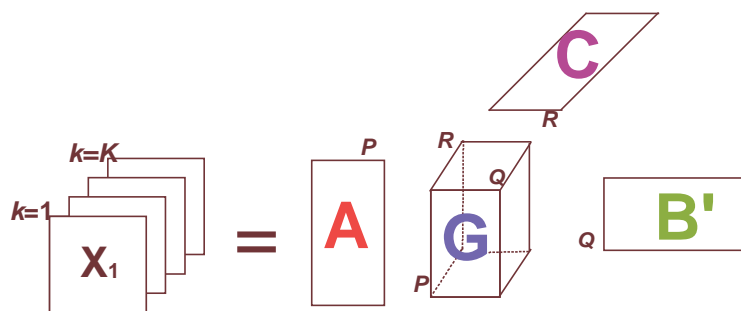

Figure 3. The Tucker3 model

THE KOJIMA GIRLS' DATA

Background, Instrumentation, and Sample

The perception of parental behaviour by parents and their children were the central concern of the study from which the illustrative data in this paper were drawn (Kojima, 1975). Kojima wanted to validate the component (or factorial) structure of the questionnaires used to measure parental behaviour. He argued that this could be done by comparing the reactions of parents and their children on the same (type of) instrument, and in separate analyses he found great similarity. This formed the basis for presuming that the perceptions or perceptual dimensions of parents and children are sufficiently the same to be included in a single analysis such as presented here.

THREE-MODE MODELLING OF THREE-WAY PROFILE DATA

The instrument used in his research was the Japanese version of the Child's Report of Parent Behaviour Inventory—CRPBI (Schaefer, 1965). In order to have parents judge their own behaviour, Kojima (1975) developed a strictly parallel parental version of this inventory (PR-PBI). The CRPBI is a three-point Likert-type questionnaire designed to assess children's perceptions with respect to parental acceptance, permitted psychological autonomy, and level of parental control. In English nearly identical forms are used for indicating behaviours of mothers and fathers, but owing to the structure of the Japanese language, it was possible to make a single version suitable for both parents. The substantive questions to be addressed in our analysis are (1) to what extent the structures of the questionnaire subscales are independent of who judges the parent behaviour and (2) whether individual differences exist between judges, and how they can be modelled and presented.

The data in this section are ratings expressing the judgements of parents with respect to their own behaviour towards their daughters and that of their daughters with respect to their parents. Thus, there are four conditions: both parents assessing their own behaviour with respect to their daughters—Father-Own behaviour (F-F), Mother-Own behaviour (M-M); the daughters' judgement of their parents' behaviour—Daughter-Father (D-F), Daughter-Mother (D-M). Collectively we will refer to conditions or to judges depending on the context. The judgements were made of parents of 150 middle-class Japanese eighth-grade girls on the 18 subscales of the inventory (see Table 1). Thus the three-way profile data consist of a 150 (girls) × 18 (scales) × 4 (judgement combinations) data array. Rather than only speaking of girls, we will at times refer to 'families' as well.

Objectives of the Analysis

Kojima performed separate component analyses for each of the judgement conditions, and evaluated the similarities of the components using congruence coefficients. He also used Tucker's interbattery procedure (Tucker, 1958). With the Parafac model we will search for a single set of parallel proportional components for the scales valid in all conditions simultaneously.

The three-component Parafac solution presented in the next section should be seen as an example of the kind of answers that can be obtained with the model. We do not claim that this model solution is necessarily the best or most detailed Parafac analysis that can be obtained from the data. In later sections, we will discuss such issues in more detail.

Dimensionality Selection

Before the Parafac analysis itself, the data were preprocessed in one of the ways recommended for three-way profile data. The condition means per scale $\bar{x}_{\bullet jk}$ were removed and the scales were normalised. That is, per scale the centred data, $z_{ijk} = x_{ijk} - \bar{x}_{\bullet jk}$, were divided by the square root of the average sum of squares

$$s_j = \left(\frac{1}{IK}\sum_{ij} z_{ijk}^2\right)^{1/2} = \left(\frac{1}{IK}\sum_{ij}(x_{ijk} - \bar{x}_{\bullet jk})^2\right)^{1/2}.$$

The first objective of a Parafac analysis is to establish how many reliable parallel proportional components can be sustained by the data. This issue is dealt

with in detail in the section on the Parafac model, where it is investigated to what extent there exist inappropriate or degenerate solutions, and to what extent the components cross-validate in split-half samples.

As shown in detail in the section on the Kojima girls' data, the fitted sums of squares for the one- to four-component solutions were 16%, 30%, 39%, 46% respectively, but the reliability of the three- and four-component solutions are questionable. Requiring the first (subject) mode to be orthogonal improves the solutions at all dimensionalities. Their proportional fitted sums of squares are only proportions of one percent less than the original solutions (for more details see below). For this reason and comparability with the solution of Kojima's boys' data set reported in Kroonenberg (2008, chap. 13), we chose to present the orthogonal three-component solution with a fit of 38%. The amount of fit is probably reasonable for this type of three-mode data. It should, however, be pointed out that, as in two-mode analysis, it would be a mistake to look for a specific amount of fit. The purpose of component analyses is to find an adequate representation of the (major) structural patterns in data which may contain considerable amounts of noise. Whether these structures explain a large part of the total variability is also related to other aspects of the data such as their reliability, the homogeneity of the population from which the sample was drawn. From an inspection of the fit of each subject, scale, and condition, it could be seen that there were no levels seriously fitting much better than other levels, and there were no large groups of levels which did not fit at all. This gave sufficient confidence to proceed with this solution.

Interpretation

As the Parafac model specifies that per component s the same source of variation is underlying the variability in each mode, a single interpretation for a component is appropriate as it is in two-mode component analysis. To understand this interpretation, two aspects are involved in each set of components of the Parafac model: (1) the components themselves ($\mathbf{a}_s = (a_{is})$, $\mathbf{b}_s = (b_{js})$, $\mathbf{c}_s = (c_{ks})$) and (2) how they combine to generate the estimated observed score (or structural image of the observed score): $\sum_s g_{sss} a_{is} b_{js} c_{ks}$. Each of the components s corresponds to a source of variation. In particular, this source is in the system generating the measurements, but varies in impact across the levels of each mode. If parental love is the source in question, then items measuring parental love should have high scores on the variable component, children who are judged to receive much parental love should have high scores on the child component, and conditions in which parental love is particularly evident should have high scores on the condition component. The effect of the source is then expressed equally by the variation across levels of every mode, i.e., parental love increases the score for all children in the same manner and also for all conditions in the same manner. The interpretation of the source is made by considering the patterns of variation in its relative impact across levels of each of the modes. Basically, one interprets all modes because the component is an entity that is the same source of variation being moderated both by levels of one mode and by the levels of another mode. The interpretation of the Parafac model for three-way profile data usually starts with the variables (scales in the present example), but the interpretation of the

component or the source of the variation in the data should be in line with the observed variability in the other modes.

The top left-hand part of Table 1 gives the component-scale correlations, i.e., the structure matrix; the right-hand part will be discussed in the section on the Tucker3 model. The scales have been arranged to supply as coherent a picture as possible, and the components 2 and 3 have been interchanged to correspond as well as possible with the Tucker3-solution to be discussed later. The commonly found dimensions underlying the CRPBI, are Psychological Control, Firm Control, and Acceptance (see Table 1 for their constituent items). Largely the items of each dimension group on the Parafac components, however the Parafac components themselves do not concur with the "official" dimensions. For instance, there is no separate Firm Control dimension. In the bottom part of the table we find the normalised coordinates for the conditions. Noteworthy is that all values are positive. The first component shows the consensus between the daughters' judgements and their parents be it that the girls has slightly higher coefficients than their parents indicating that their judgements are more extreme, the second component is primarily valid for the parents, and the third for the daughters. Note, that it is not a question of contrasts as all coefficients are positive. In other words the second and third components carry comparatively much less weight for the daughters and the parents, respectively.

Figure 4 shows the per-component plot for Components 1 and similar plots can be made for the other two components. As Kojima's interest centred on the similarity in structure of the scale space, we take this as our starting point for the interpretation.

The first scale component is largely the Acceptance - Rejection dimension often found for the CRPBI with Lax discipline siding with Acceptance, but it also contains three scales which are supposed to belong to the Psychology dimension (Inconsistent discipline, Control through guilt and Withdrawal of relations). All judging conditions are positive indicating that all judges had similar views on parental behaviour. Thus overall, parents and their daughters gave similar judgements. Whether the parents were characterised by the judges as Accepting or Rejecting varied enormously per family.

The term of the Parafac model displayed in Figure 4 is the product of all first components, $g_{sss}a_{i1}b_{j1}c_{k1}$. We observe that the judges' coefficients c_{k1} are all positive. The signs of a_{i1} and b_{j1} in the product $a_{i1}b_{j1}$ determine whether the parents of girl i are judged as rejecting or accepting. High positive scores for a_{i1}, such as those of daughters 29, 60, and 81, combined with the positive coefficients of the accepting scales b_{j1} indicate that the parents are judged particularly accepting and not rejecting, and vice versa for the girls at the other end of the component, i.e., all judges see the parents of girls 57, 13, and 138 not as accepting, but as rejecting parents. If the small differences between judges are to be believed, girls judge their parents to be slightly more extreme than the parents judge themselves.

To illustrate the above argument numerically, suppose that Acceptance (scale j) has a value of $b_{j1} = .40$ on the first component. In addition, the judging condition k, Daughter judges parent has a value of $c_{k1} = .43$. Furthermore, the weight of the first component $g_{sss} = 2.1$, then the combined value of $g_{sss}b_{j1}c_{k1} = 2.1 \times .40 \times .43 = .36$. For Girl 29 who has a component score of $a_{29,1} = .20$, the first component's contribution to her Acceptance score in case of the daughter-judges-
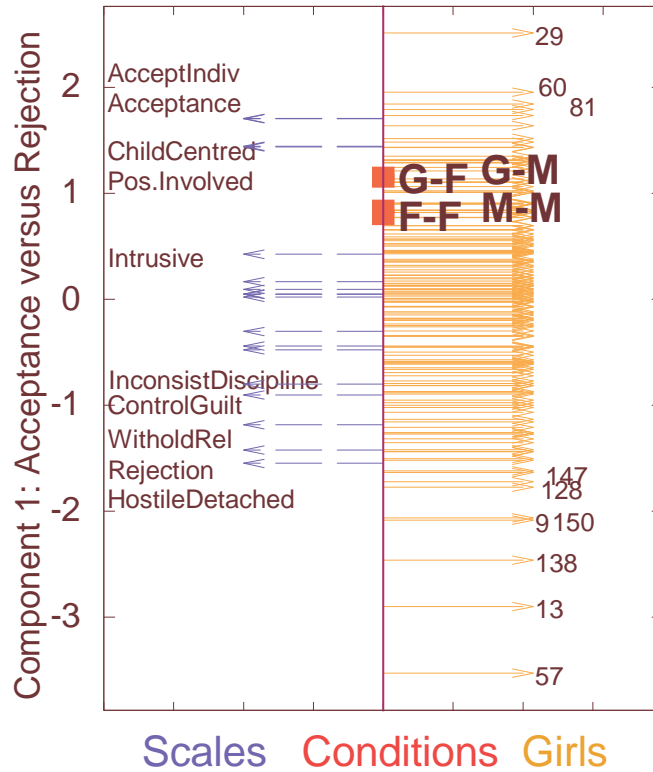
Figure 4. Kojima Data: First components of the Girls' mode, the Scale mode, and the Condition mode. The solid arrows on the left-hand side refer to the scales, the dashed arrows on the right-hand side to the daughters, and the four bold marks on the vertical line to the conditions, i.e., D-F: Daughter judges her father's behaviour towards her; D-M: same with respect to her mother; F-F: the father judges his own behaviour towards his daughter; M-M: same for the mother.

parent condition is $.36 \times .20 = .07$. However, for Girl 138 with a component score of -.20, the contribution is -.07. Thus on the basis of this component Girl 29 judged her parents as more accepting (rather than more rejecting), while Girl 138 judged her parent as more rejecting (rather than more accepting). As the zero point represents the mean due to the centring, Girl 29 scores above and Girl 138 below average on Accepting in the daughter-judges-parent condition.

All scales have positive and comparable coefficients on the second component. In addition, the coefficients of the judges are all positive, but those of the parents as judges are much higher than those of their daughters. Given that the positive coefficients of the parents, this component seems to represent something like a response style. Some judges especially parents give more extreme answers on all scales than others. For parents whose daughters are on the positive side of the girls' axis the response style serves to make the parents' judgements more positive, while for the girls on the negative side the parents' judgements are more negative on all scales.

The third scale component is (positively) dominated by Behavioural and Psychological control, be it that there is a bit of a similar response-style flavour as with the parents in the second component. Only Lax discipline is exempt. From the condition components we see that this component reflects primarily the daughters' judgements of their parents. And the interpretation proceeds along the lines as with the parents on the second component.

Summary of the Analysis

There is substantial agreement between parents and their daughters about the parents' behaviour, but some differences between them could be observed. The differences between families were, however, much larger than the within family judgements. This evident from the considerable differences between the girls (between-family differences), and the comparatively small differences in the judging conditions (within-family differences). Parallel proportional differences were primarily evident in the contrast between Acceptance versus Rejection. It is not easy to compare these results with those of most component analyses of similar data, due to the use of a different model. When discussing analyses using the Tucker3 model, we will come back to comparisons of these results.

An interesting question arises from the analysis using the Parafac model. If we follow Cattell and Harshman that components derived with Parafac have more "real" meaning than (rotated) maximum-variance components, one may wonder what the meanings are of the Parafac components in the present study. The analysis gives the impression that the scales exhibit one real contrast (Acceptance - Rejection) and that the other components capture primarily response styles. Of course, these findings need to be confirmed by analysing other similar data sets in the same way, preferably some of the defining data sets of Schaefer. It is clear that the variable-component correlations (see Table 1) only partially follow the "official" grouping of the originators of the instrument.

THREE-MODE ANALYSIS: PRACTICAL ISSUES

In this section we will discuss how to carry out three-mode analyses with the major three-mode models. We will discuss such issues as choosing an analysis within each class, comparing solutions, plotting, properties of the solutions, and what to do in case things do not go according to plan. Various points are illustrated with the Kojima girls' data.

Parafac Model

The primary aim of the analyses in this section is to discuss the practical issues in connection with carrying out a Parafac analysis on three-way profile data. For this purpose, we will lean heavily on the *magnum opus* of Parafac's godfather, Richard Harshman (Harshman & Lundy, 1984a, 1984b; Harshman & DeSarbo, 1984).

Objectives

The aim of most analyses with the Parafac model is to uncover the existence of components which show parallel proportional profiles, and if possible to identify these components as ''real'' ones which carry true substantive meaning. In many chemical applications this is not a problem as the components often correspond to physical or chemical properties of the substances under investigation. Harshman, Bro and co-workers have shown several analyses where the uncovered components could be given such a status; e.g. Harshman (1994a), Harshman, Ladefoged, & Goldstein (1977), Bro (1998), Smilde et al. (2004a). However, in the example of the previous section, the correspondence between the components and the theory of parental behaviour is far more difficult to establish, especially because no non-orthogonal solutions could be found. If stable solutions are present, this will give valuable information to the substantive researchers about possible underlying parallel proportional profiles. Thus, given stable Parafac components have emerged, it should spur researchers on to make sense of these components using parallel proportional profiles.

Data and Design: Types of variability.

Because the Parafac model is based on parallel proportional profiles (PPP) and therefore sensitive to violation of this principle in the data, it is important to pay attention to the possibility that such profiles might not be present. However, often the only way to check this is via an analysis with the model itself. The proportional profile requirement is also known as trilinearity, which refers to the property that the model is linear in each mode given the other two modes (see Equation 2), where the components of one mode are linear given the values of the other two modes.

In Harshman & Lundy (1984a, p. 130) the PPP principle is explained in terms of models or components underlying the data. In the *system variation model*, the components ''reside in the system under study and through the system affect the particular objects; the [component] influences exhibited by particular objects would thus vary in a synchronous manner across a third mode.'' In ''the *object variation model*, separate instances of the [component] can be found in each of the objects, and these within-object [components] would not be expected to show synchronous variation across levels of a third mode.''[4]

Typically, object variation is not in accordance with the Parafac model. Harshman & Lundy (1984a) suggest that if the object variation is in one mode only, the fully-crossed data can be converted to cross-product matrices such that the object-variation mode 'disappears', in the same way as individuals are no longer visible in a correlation matrix but only the variables. Then these cross-product matrices can be investigated for parallel proportional profiles in the other modes with the standard Parafac model.

Model and Dimensionality Selection

The choice of the 'best' or most appropriate solution of the Parafac model is not an easy one and the procedure to arrive at such a solution requires consider-

able attention to details. Because the model is identified or unique given the data and the number of components, it may not fit due to the lack of parallel profiles in the data. Often several analyses at each plausible number of components are necessary to find an appropriate solution if it exists at all. Again the Kojima girls' data, but not the boys' data, are a case in point.

Uniqueness.

Harshman (1970) provided the first discussion of the uniqueness of the Parafac model. Krijnen (1993, chap. 2) provided a further discussion of the concept of uniqueness. The major thrust of his argument was that one should distinguish between weak and strong uniqueness. In the former case, the unique solution is one in which the obtained solution is surrounded by a number of non-unique solutions which fit almost as well as the unique one, and thus Harshman's claim that the unique solution should be given preference over all other solution is not quite as strong. He argues that in such a case it might be advantageous to take a nearly as good solution if it is easier to interpret, for example because via rotations the components can have a simple structure. A solution is strongly unique if there are no non-unique models which have almost as good a fit as the Parafac model itself.

Harshman (1972a) (see also Krijnen, 1993, p. 28, 29) showed that a necessary condition for the uniqueness of a solution is that no component matrix has proportional columns. Krijnen suggested checking for weak uniqueness by comparing the fit of a regular solution of the Parafac model with that of a model with two proportional columns in any of the component matrices. If the difference is small the solution is considered weakly unique; for an example, see Kroonenberg (2008, pp. 323ff.). A similar approach may be taken by comparing the fit of the regular unique Parafac solution with that of any other Parafac model, say with orthogonality constraints on one of the component matrices.

The accepted way to compare two components, say $\mathbf{x}$ and $\mathbf{y}$ is via Tucker's congruence coefficient,

$$\phi_{xy} = \frac{\sum_i x_i y_j}{(\sum_i x_i^2)^{1/2} \, (\sum_j y_j^2)^{1/2}} \, , \tag{4}$$

which is similar to a correlation coefficient except that the components are not in deviation of their means (Tucker, 1951; Ten Berge, 1986). If the components are already in deviation of their means as is generally the case for the subject mode in profile data, the congruence coefficient is equal to the correlation coefficient. For the Kojima girls' data, Table 2 shows the congruence coefficients for the one-through three-component analyses with orthogonal subject components. The reason for the orthogonality is explained below.

For the Kojima girls' data, the orthogonal subject component spaces seem to be nearly unrelated components as most congruence coefficients are nowhere near one (Table 2). However, regression analyses show that both the one-dimensional and the two-dimensional solutions are contained in the three-dimensional solution, and the three-dimensional solution itself is again imbedded in the four-dimensional one. This shows that for the subjects the lower-dimensional spaces are nested in the higher-dimensional ones, but that the orientation of the axes in each of the component spaces is different from that of the lower-dimensional one.

Table 2. Congruence Coefficients of Several Parafac Solutions to Evaluate Type of Uniqueness of the Standard Solution.

Analysis	Numbers of components	PF1	PF2		PF3			Fit
		1	1	2	1	2	3	
PF1	1	1.00	.78	.62	-.18	.70	.68	.160
PF2	1	.78	1.00	.00	.47	.60	.64	.301
PF2	2	.62	.00	1.00	-.88	.33	.33	
PF3	1	-.18	.47	-.88	1.00	.00	.00	.383
PF3	2	.70	.60	.33	.00	1.00	.00	
PF3	3	.68	.64	.33	.00	.00	1.00	

Therefore, it seems that the parallel proportional profiles in these data are not very strong and that one calls their existence into question.

Multiple Solutions.

Given that the Parafac model is a restrictive one, i.e., not all data sets have an acceptable Parafac solution, one has to run analyses with different numbers of components to determine whether the data contain parallel proportional profiles. In other words, when solving for the Parafac model we are really concerned with model fitting and not so much with data approximation. This also means that both using too few and using too many components can lead to unsatisfactory solutions. In general, it turns out that specifying too many components is more problematic than too few, because violation of the parallel proportional profile principle is more likely for later components. When two components are close together in terms of their contributions to the fit of a too small model, different analyses may pick-up different components, and only when the number of components is increased can stable solutions be found. Whether this occurs is very much data dependent; see, Murakami and Kroonenberg (2003) for an example.

Split-Half Strategy.

One way to get insight in the stability of the components is to split the data in half and perform a separate analysis on both parts. If there is a true underlying solution, it should show up in both analyses. Important restrictions to this procedure are that there must be a stochastic mode for which splitting makes sense. However, such a mode almost always exists in three-way profile data, in particular the subject mode. When there is no stochastic framework, splitting might not be possible. For instance, in some typical experiments in analytical chemistry splitting the data set in halves does not make sense. The other caveat is that there must be sufficient 'subjects' in the mode that is being split. In other words, both splits must be large enough to minimise the influence of the idiosyncrasies of specific individuals. How much is large enough is difficult to say in general, because much depends on the noise level of the data and the clarity of the underlying structure. Harshman and DeSarbo (1984) discusses the split-half procedure

in great detail with illustrative examples (see also Kiers & Van Mechelen, 2001); see Kroonenberg (2008, pp. 326ff) for split-half analyses of the Kojima boys' data.

Degeneracy.

Harshman (1970) was the first to note the problem of non-converging solutions in which two components tend to become perfectly negatively correlated, and in which the g_{sss} increased without bound. Such solutions from a Parafac analyses are called *degenerate*, and this phenomenon has been a topic of intensive research ever since; for a recent paper, see, Stegeman (2006). The basic problem causing the degeneracy is that algorithms to compute the model parameters cannot cope with data that do not conform to the Parafac model, and therefore produce degenerate, uninterpretable solutions.

Krijnen and Kroonenberg (2000) (see also Krijnen, 1993, p. 13ff.) discussed a number of measures to assess whether an algorithm is tending towards such a degeneration solution, and they suggest a number of heuristic values for these measures. Their approach was inspired by and an improvement of earlier work of Harshman and Lundy (1984a, p. 272). To assess degeneracy, we need the cosines between two components in a single mode, which is the same as calculating the *congruence coefficient* between the components. As an example, the cosine between the sth and s'th component of the first mode is $\cos(a_{s,s'}) = (\mathbf{a}_s{}' \mathbf{a}_{s'})$. If we define f_s as the $I \times J \times K$ (column) vector of consisting of the terms $a_{is} b_{js} c_{ks}$ (see Equation 2) then the cosine $\theta_{s,s'}$ between f_s and $f_{s'}$ is the triple cosine product:

$$\cos(\theta_{s,s'}) = \cos(\alpha_{s,s'}) \cos(\beta_{s,s'}) \cos(\gamma_{s,s'}) \tag{5}$$

If $\cos(\theta_{s,s'})$ is approaching -1, there is almost certain a degenerate solution. It signifies that the two components f_s and $f_{s'}$ have become proportional, and this is explicitly 'forbidden' in the Parafac model. The conclusion can be further supported by creating an $S \times S$ matrix of the $\cos(\theta_{s,s'})$ and inspecting its smallest eigenvalue. If it gets, say, below .50, degeneracy might be present. In addition, one should assess the condition number of the triple-cosine matrix (i.e., the largest eigenvalue divided by the smallest one). If the condition number is getting large, say somewhat arbitrarily, larger than 5, a degeneracy is likely. Both the smallest eigenvalue and the condition number are indicators whether the triple-cosine matrix is of full rank as it should be for a Parafac solution. The best way to confirm degeneracy is to run the same analysis again but with an increased number of iterations and possibly with a more stringent criterion. If the indicators get worse, degeneracy is getting more and more likely.

For the Kojima girls in an unconstrained three-component solution the triple-cosine product between the first and second component was -.95, the smallest eigenvalue of the triple-cosine matrix was .06, and the condition number of that matrix was 23.4. Finally, the standardised component weights, i.e., the g_{sss}^2/Total sum or squares, were 3.9, 3.6, and .3, while they should be smaller than 1.0. Clearly, the three-component solution is a degenerate one. In fact, also the two- and the four-component solutions are degenerate. When degeneracy is suspected, it is absolutely necessary to use several starting positions for the algorithm, because ending up with one degenerate solution does not necessarily mean that there is no properly convergent solution; see especially Paatero (2000) for an enlightening discussion of this point. To circumvent degeneracy, one may pose

constraints on the solution such as requiring one of the modes to be orthogonal or imposing nonnegativity of the components. Not uncommonly abandoning the Parafac model and exploring the data with a Tucker3 model might be a good option and this was the approach taken with the Kojima girls' data.

Searching for Convergent Solutions.

When searching for a satisfactory number of components, analyses with different numbers of components have to be run and, if necessary, restrictions have to be placed on one or more of the component matrices. Parafac analyses for Kojima's girls' data were run with 1 through 4 components both with and without orthogonality restrictions on the subjects' components and, for purposes of assessing stability, the data set was randomly split in halves (S1 and S2) and analysed with orthogonal solutions.

The first objective in getting acceptable solutions is to have convergent and non-degenerate solutions. None of the non-restricted analyses came up with an acceptable solution as all of them were degenerate. Because of this, only results of solutions with orthogonality restrictions are discussed. Note that because there are only four levels in the third mode, going beyond four components does not make much practical sense. It should be noted that having convergent solutions does not necessarily mean that the solutions are the same, a matter taken up in the next section.

Table 3. Fit Measures for Parafac Solution with 1 through 4 Components

Model	Data set	Proportional Fit	No. of Iterations	Core Consistency	Standardised Component Weights			
PF1	All	.160	56	1.00	.160			
	S1	.175	55	1.00	.173			
	S2	.161	62	1.00	.161			
PF2	All	.301	225	.99	.152	.148		
	S1	.315	81	.93	.160	.155		
	S2	.312	66	1.00	.161	.151		
PF3	All	.383	266	.20	.141	.122	.121	
	S1	.405	135	.09	.141	.141	.122	
	S2	.400	254	.67	.144	.134	.122	
PF4	All	.459	112	.01	.128	.127	.122	.081
	S1	.470	221	.18	.135	.128	.121	.086
	S2	.443	103	.01	.147	.128	.123	.076

Notes: S1, S2 = Split-half data sets

The proportional fitted sums of squares for the four solutions of the full data set were .16, .30, .38, and .46 showing that at least 46% of the variability in the data could be modelled with a parallel proportional profiles model. The two split-half samples show similar explained variability. Also on the other measures the

split-half samples show comparable results to the complete sample. Note that the standardised component weights of most solutions show at least two nearly equal values which might explain why the components of the successive solutions did not align as was evident in Table 2.

Examining Convergent Solutions.

Detailed analyses of convergent solutions via component comparisons using the congruence coefficients are necessary to establish their stability. These comparisons may also be used as an additional way to assess the uniqueness of the solution, because assuming that there is a unique solution in the population, the same components should recur in the overall solution and in the split-half solutions, barring sampling errors and possible instability due to insufficient subjects. It is an empirical question whether 76 subjects in the split-half samples are enough to establish stability. When comparing solutions we could take our lead from Lorenzo-Seva and Ten Berge (2006), who found in an empirical study that subjects judged components to have a fair similarity when the congruence coefficient was in the range .85-.94, while two components with a congruence coefficient higher than .95 were considered to be identical for interpretation.

Detailed analyses showed that there is no stability across the overall orthogonal analyses and the split-half ones. This suggests that there either is little stability or that there are serious violations of the underlying model. This is contrast with the boys' data where we could conclude that the same components were present in the overall solutions and the split-halves, be it that there were considerable differences between the split-halves in that analysis.

Assessing the Parafac Core Array.

Bro (1998) (see also Bro & Kiers, 2003) suggested using the *core consistency* as a measure for evaluating Parafac models. This measure is based on assessing how far away the core array derived from the Parafac components is from a superdiagonal core array, i.e., the cube of size $S \times S \times S$ in which only the g_{111}, g_{222}, ..., g_{SSS} have sizeable values, and all other core elements are near zero. Bro (1998, p. 113-122) also proposed to construct a *core consistency plot* which has the core values on the vertical axis and the core elements on the horizontal one with the superdiagonal elements plotted first. For the Kojima girls' data, core consistency plots for the three-dimensional orthogonal solution and for the second split-half set (S2) are presented in Figure 5.
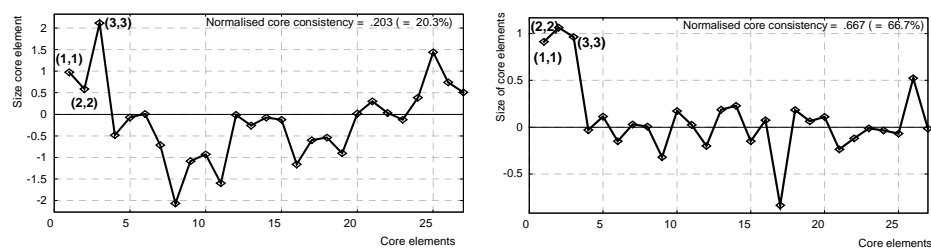

Figure 5. Core consistency plots for the Kojima girls' data. Left: Three-component Parafac solution for all data; Right: Three-component solution for subset S2.

All two-component solutions had a core consistency of just about 100%, indicating a perfect superdiagonal core array (see Table 3). The three-component orthogonal solution (Figure 7) with a core consistency of 20% shows that there are quite a few larger-than-zero elements in the core array, and the three super-diagonal elements are not even the largest ones. However, the three-component solution (Figure 7) of split-half sample S2 showed are far better core consistency (67%) in contrast with the other split-half sample (9%) (Table 3).

Interpretation

Above we have illustrated the presentation of the results via a table with the scale coefficients (Table 1) also via the per-component plot of the first components of each of the three modes (Figure 4). If one wants to make pair-wise component plots, it is not the components themselves which should form the coordinate axes as they are correlated, but a set of orthogonal axes which span the same space. This is necessary to obtain the correct interpretation of distances in such plots (Kiers, 2000). Given the problematic solution we will not show such a plot in this paper, but examples can be found in Kroonenberg (2008).

Parafac with Constraints.
In a previous section we introduced the idea of constraining the components of a Parafac solution to evaluate uniqueness, but constraints have a far wider use. For instance, in intelligence tests all subtests have to correlate positively with each other, therefore a nonnegative first component is obligatory and even two non-negative components are desirable (Krijnen and Ten Berge, 1992). Thus substantive issues may require restrictions on the components.

Parafac Core Arrays.
As mentioned above, one of the major reasons for degeneracies is that the data contain so-called *Tucker structure* (Harshman & Lundy, 1984b). Under the Parafac model, each component of each mode is exclusively linked to one single component of each of the other modes. Thus in the Parafac model only terms such as $a_{is}b_{js}c_{ks}$ exist in which the s is the same for all modes. However, in some data also the term $a_{is'}b_{js}c_{ks}$ with $s' = s$ explains a substantial part of the variability in the data. In other words, there are two components $\mathbf{a}_s$ and $\mathbf{a}_{s'}$, which have links the components $\mathbf{b}_s$ and $\mathbf{c}_s$. This is the standard situation in the Tucker3 model, but prohibited in the Parafac model, and therefore solutions including such terms are said to have *Tucker structure*. The core consistency plot (Figure 5) showed that there were indeed many non-zero elements in the core array computed from the three-component Parafac model for the Kojima girls' data.

Conclusion from the Parafac Analysis

The scale and condition spaces of the Kojima girl data showed clear interpretability, but a closer examination of the various solutions brought a considerable number of problems to light. That this is particular to this data was clear from the analyses of the parallel data of the boys, which were presented in Kroonenberg (2008, chap. 13). By looking in detail at the quality of the solutions via fit meas-

ures, comparing the original solution with split-half samples, inspecting congruence coefficients between components, and examining core consistency, it became clear that the Parafac model is not the most appropriate, or easiest to interpret, model for these data. In the next section we will examine the same data but with a Tucker3 model to establish whether using this model a better insight can be gained into the judgements of parental behaviour.

Tucker3 Model

Objectives

The Tucker3 model is primarily used to find a limited set of components with which the most important part of the variability in the data can be described. The model is thus especially useful for data reduction and for exploration of variability. The main reason that the Tucker model is not as directly useful as the Parafac model for the search of developing or identifying general principles or rules that underlie patterns in the variables is its rotational freedom. The basic result of applying the Tucker3 model to three-way data are component spaces in which any orientation of the axes is as good as any other one in terms of fit of the solution to the data. However, if descriptions of the patterns are desired in terms of the original variables, subjects, and conditions, rather than as latent entities, the rotational freedom can be extremely helpful in unravelling these patterns.

In three-way profile data the variability analysed takes mostly the form of squared normalised deviations from the variable means at each condition (see section on preprocessing). Thus, a large part of the variability in the original data, in particular the variability between the means per condition, is not contained in the three-mode analysis itself. When interpreting results, this should always be borne in mind. However, what facilitates the interpretation is that the removed means represent the scores of the average subject, i.e., the person who has an average score on each variable under each condition. In many applications, the variable×condition matrix of means contains important information in its own right and should be carefully analysed, for instance via a biplots (Gabriel, 1971). An important caveat is that such a matrix of means is most effectively interpreted if the sample over which the means are calculated is in some way representative of an identifiable population. In the present example this means that we should satisfy ourselves that the girls represent some meaningful sample of a population of Japanese girls.

Types of Variability.
In contrast with the Parafac model, the Tucker3 model can handle both system and object variation, and therefore can be used to fit data in which the correlations between the components change over time. It is especially the simultaneous handling of these types variation which make the Tucker3 model extremely useful to investigate the patterns within and between modes in situations where little is known a priori about them. Alternatively, the model can be used when one does not want to prejudice the analysis or cannot find a satisfactory Parafac analysis.

Model and Dimensionality Selection

When confronted with three-way data the choice of the most appropriate Tucker3 model is not always an easy choice because the numbers of components can vary independently for each mode. In this section we will try to find a reasonably fitting Tucker3 model for the Kojima girls' data by first evaluating a series of models with varying numbers of components in the three modes.
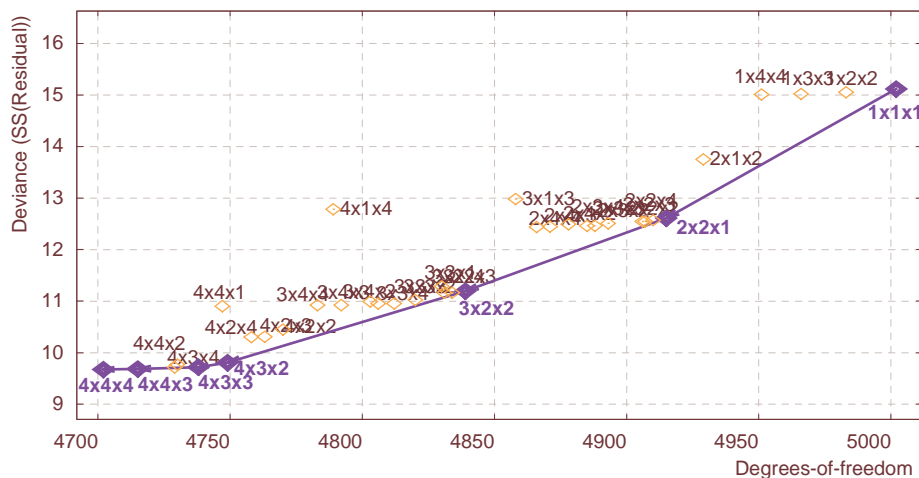

Figure 6. Deviance plot for all Tucker3 models $P \leq 4$ and $Q \leq 4$ and $R \leq 4$. The solid line connects optimal choices based on the deviance convex hull.

There are two types of plots useful for model selection: (1) the *deviance plot* (Figure 6) displaying the deviance or residual sums of squares against the degrees-of-freedom, and the *three-mode scree plot* (Figure 7) showing the deviance against the total number of components, i.e., N = P + Q + R (Timmerman & Kiers, 2000; Kiers & Der Kinderen, 2003). The models with the best SS(Residual)/*df* ratios and those with the best SS(Residual)/*N* ratios lie on what is called the *convex hull*, which is the curve drawn in each of these figures. To assist the choosing a model, one may use the Ceulemans-Kiers (2006) *st*-criterion; see also Kroonenberg (2008, p. 182ff.). This criterion looks for the sharpest angle in the convex hull in the deviance plot. For the present data this occurs for the 4×3×2-Tucker3 model, but it should be noted that the convex hull does not show much curvature. The preferred model indicated by Cattell's "elbow" in the three-mode scree plot is the same 4×3×2-Tucker3 model. This model contains a large number of components for the girls which may cause difficulties in interpretation. An interesting alternative could be the more parsimonious 3×3×2-model, as it lies on the convex hull in the deviance plot and only just misses it in the three-mode scree plot. However, it has the disadvantage of only two scale components which may be simplifying the scale space too much.

Unlike the Parafac model, the Tucker3 model has no serious problems with multiple solutions, non-convergence, and similar algorithmic difficulties. Timmerman and Kiers (2000) conducted an extensive simulation study and they only found problems when the numbers of components deviate much from the real
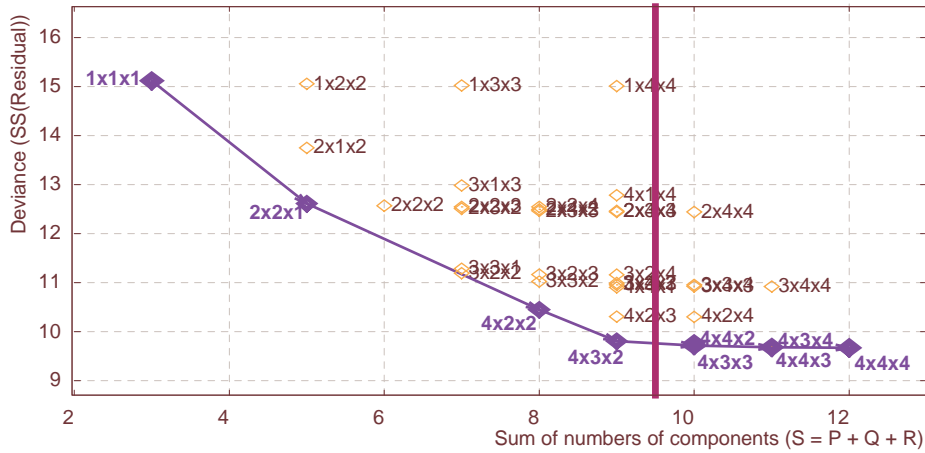
Figure 7. Kojima data: Three-mode scree plot for all Tucker3 models with *P* ≤ 4 and *Q* ≤ 4 and *R* ≤ 4. The horizontal line indicates that the models at the right of the line should probably not be considered.

number of components analysed, and even then primarily when random starts was chosen, rather than a rational ones.

It is instructive to compare the fit of the Tucker3 model with those of comparable orthogonal Parafac ones. For the two-component models the proportional fitted sums of squares are: .3151 (Parafac), .3016 (Tucker3 - 2×2×2), and for the three-component models the proportional fitted sums of squares are: .4046 (Parafac), .3914 (Tucker3 - 3×3×3). Clearly on the basis of fit alone there is not much to choose between comparable models.

Evaluating Fit.

The fit of a model can be partitioned in several ways to assess how well different parts of the data are fitted. In particular, one can look at the fit of the components of each mode, at the fit of combinations of components from different modes, at the levels within each mode, and at the fit of each data point. In this way it is possible to make a detailed evaluation of the quality of the solution, and search for ways to improve the analysis by expanding or restricting the model.

Table 4. Proportional Fit of the Components of the Three Modes for $4 \times 3 \times 2$-Tucker3 Model

				Components		
	Mode	Sum	1	2	3	4
1	Girls	.45	.16	.14	.09	.07
2	Scales	.45	.24	.15	.07	
3	Conditions	.45	.37	.09		

31

Components. The overall fit of the 4×3×2-Tucker3 model is 41% which is a fairly common value, considering the size of the data and the fact that the fit refers to individual data points rather than to covariances.

Girls. Due to the three-way orthogonality of the Tucker3 model, it is possible to assess the fit or lack of it for each level of each mode separately (Ten Berge, De Leeuw, & Kroonenberg, 1987). *The sums-of-squares plot* (Figure 8) shows that there are several girls who have large total sums of squares coupled with a good fit (e.g., G57 and G138), some girls have large total sums of squares coupled with a low fit (e.g., G104). Girls located near the origin have both small residual and fitted sums of squares (e.g., G44). Because the scales are centred this means that the latter have scores close to the means of the variables. The large dot on the dotted line indicates the point of average fit and average residual sum of squares.
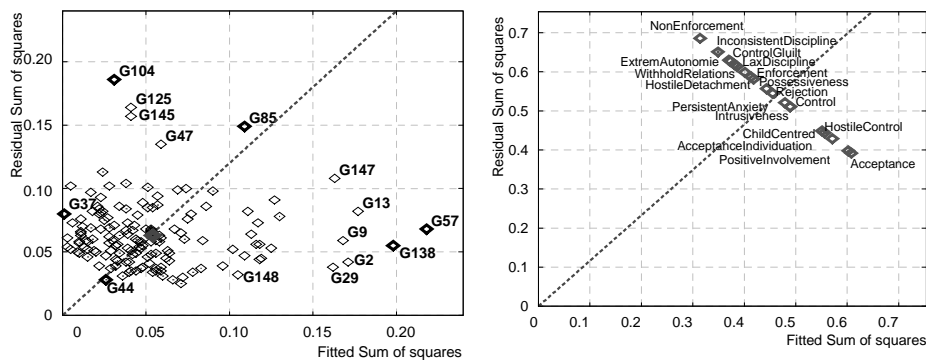

Figure 8. Kojima girls' data: Sums-of-squares plots showing how well the model fits the data of each girl (left) and each scale (right). The dashed line starting from the origin indicates the locus of all girls (scales) with the same average residual/fit ratio. Girls (Scales) above the average line have a worse residual/fit ratio than average while the girls (scales) below the line have a better residual/fit ratio. The scales lie on a line of equal total sum of squares because they were normalised.

Scales. Figure 4 shows that Nonenforcement is the worst fitting scale in the 4×3×2-solution, while Acceptance, Positive involvement, and Acceptance of individuation fit best. Whether a scale fits better than another one, is not related to the type of disciplining involved, but has something to do with the consistency and variability of the answers. The normalisations ensured that all variables had equal total sum of squares so that the scales are aligned on the slanted line of equal total sum of squares in the graph.

Judgement conditions. As there are only four conditions, there is no need to construct a sums-of-squares plot. The Girl-Father and Girl-Mother conditions have a proportional fit of .49 and .56, respectively. The Father-Father and Mother-Mother conditions have a better fit, i.e., .62 and .73, respectively. The (modest) difference is probably due to more inconsistency in the girls' judgements over the scales.

Fit of combinations of components. Apart from determining an appropriate number of components and their contributions to the overall fit, one may also investigate the contributions of the component combinations to the fit of the model to the data. As long as the components are orthogonal within the modes, the core array can be partitioned into the explained variabilities by the combinations of components by squaring the core elements and dividing them by the total sum of squares after preprocessing (see on preprocessing). Thus, the proportion explained variability of combination (p,q,r) is $g_{pqr}^2/\text{SS(Total)}$, where $\text{SS(Total)} = \sum_i \sum_j \sum_k z_{ijk}^2$ with z_{ijk} the preprocessed data. For details on the Kojima girls' data see below.

Interpretation

In the Parafac analysis, we interpreted the results from the scales in a straight-forward manner, i.e., we described the patterns of the components of the scales and condition modes. On the whole, these patterns were relatively clear without additional manipulation of the output itself. Moreover, Parafac components have the proportional profile property which gives them a special status, and it is primarily the patterns per component which need to be interpreted rather than the patterns across components. However, from a 'simple structure' point of view looking at patterns across components, many scales in the Parafac solution have sizeable coefficients on more than one component, so that the components are all but simple. Because of the rotational freedom, the results from a Tucker analysis can be investigated for simple structure and that is what we will proceed to do in this section. In particular, we will examine plots of components per mode, rotations, both of component spaces and the core array, joint representations of the modes, and using constraints.

Displaying Components and Core Array.

As mentioned in above, there are essentially four different ways in which the components may be presented: (1) *Tables* of the coefficients or loadings for each mode—rotated or not (Table 1); (2) Separate *pair-wise graphs* of the components per mode (not shown); (3) *All-components plots* showing all components of a single mode in a plot with the levels of the modes on the horizontal axis (not shown); (4) *Per-component plots* showing one component of each of the three modes in the same plot (Figure 4). The third option is especially useful when there is a natural order for the levels of a mode, but this was not the case here (for an example see Figure 11.10 in Kroonenberg (2008, p. 268).

Scale components. The solution for the scales after rotation was already presented in Table 1. The rotational procedure, called the *Harris-Kaiser independent cluster rotation* is described and evaluated in detail in Kiers and Ten Berge (1994). In essence it is an oblique rotation realised via a varimax rotation on an orthonormalised component matrix and its aim is to find as clearly separated clusters of variables as possible. After rotation the normalised scale components show indeed a much clearer cluster structure than in the Parafac case. The components can now be designated as Acceptance, Rejection and Control, and (Lax) Discipline, which again does not completely concur with the official grouping of the scales. However, the components no longer have the parallel proportional profile property.

Table 5. Unrotated and rotated condition components and their fit

Condition	Unrotated		Rotated		Proportional
	1	2	1	2	SS(Res)
Girl judging Father behaviour (D-F)	**.53**	**-.44**	**.73**	.03	.49
Girl judging Mother behaviour (D-M)	**.54**	**-.48**	**.68**	-.03	.56
Father judging his behaviour (F-F)	**.46**	**.55**	.01	**.79**	.62
Mother judging her behaviour (M-M)	**.46**	**.52**	-.01	**.62**	.73
Proportional fit of components	.37	.09	.25	.21	

Note: Unit length components; Varimax rotation on the orthonormal components.

Judgement components. The mode of the judgement conditions is more conveniently presented in a table (Table 5), because there are so few levels. From this table, it follows that the girls judge their parents' behaviour as being rather similar, and also the parents give similar judgements about their own behaviour. The model selection graphs showed that adding condition components will not really increase the fit, so that no further important distinctions between the judgements can be found. In other words, it makes sense to talk about parental behaviour rather than father or mother behaviour, be it that the daughters mostly but not always agree with their parents about its nature.

Core array. The core array contains the weights for the links between the components from different modes (Table 6). In the unrotated case, we are always dealing with principal components, so that in almost all cases the first element of the core array, here $g_{111} = 1.57$, is the largest one.

The second largest element in the unrotated core array, $g_{221} = 1.46$ is a combination of the second girl component, the second scale component and the first condition component, which is an impossible combination for a Parafac model. Note, that the adjusted core array after rotation of the condition and scale components has many more mid-sized elements than the original core array. This general effect, i.e., distribution of variability, is the price to pay for having a simpler component structure. Kiers (1998b) has devised procedures to simultaneously simplify the components and the core array.

A detailed examination of the rotated core array proceeds via a comparison of the sizes of the links between the components. The rotated core array contains a fair number of medium-sized elements, which complicates its interpretation. However, due to the clear meaning of the rotated Tucker3 components (see Table 1) it is easier to investigate the relationships between the components from the three modes. The substantive interpretation of the rotated core array makes use of the fact that after rotation we have separate components for the daughters and the parents (see Table 1). This translates itself into separate panels for the daughters (left-hand panel) and the parents (right-hand panel). Looking at the first two rows G1 and G2 corresponding to the first two girl or rather family components, we see that the patterns are similar for the judges and that on the whole the core values are very similar as well. This indicates that, for families having highly positive or

Table 6. Tucker3 core array

	Raw Core Array						%Explained Variability					
	C1			C2			C1			C2		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
						Unrotated Core Array						
G1	**1.57**	-.62	.05	.13	-.14	.14	**13.8**	.2	.0	.1	.1	.1
G2	-.60	**1.46**	-.01	.14	-.08	.11	2.0	**11.9**	.0	.1	.0	.1
G3	-.03	-.09	**.81**	**.91**	.12	.05	.0	.0	**3.7**	**4.6**	.1	.0
G4	.04	.09	**.79**	**-.73**	.25	.11	.0	.0	**3.3**	**4.0**	.3	.1

						Rotated Core Array						
	Daughters			Parents			Daughters			Parents		
	Accept	Control	Lax D	Accept	Control	Lax D	Accept	Control	Lax D	Accept	Control	Lax D
G1	**1.21**	-.07	.27	**1.05**	-.32	.40	**8.2**	.0	.4	**6.1**	.6	.9
G2	.09	**1.28**	.10	.14	**.90**	.23	.0	**9.1**	.1	.1	**4.5**	.3
G3	**-.76**	.12	.40	.42	-.37	**.68**	**3.2**	.1	.9	1.0	.8	**2.5**
G4	.23	-.39	.58	-.51	.41	.52	.3	.9	1.8	1.4	1.0	1.5

The numbers in **bold** indicate the larger elements of the core array. C stands for condition components; S for scale components; G for girls components. The rotated core is the result from a counter-rotation with respect to the varimax rotation of the condition components (see right-hand panel of Table 5). Accept = Acceptance; Control = Control & Rejection; Lax D = Lax discipline

negative values on these components, the daughters and their parents have a similar view on parenting style. Families with highly positive values on the first family component show an accepting style of parenting $g_{111} = 1.21$ and $g_{112} = 1.05$ and a tendency towards lax discipline $g_{131} = .27$ and $g_{132} = .40$; for families with highly negative values this is the reverse. Families with highly positive values on the second family component show primarily a controlling and reject-ing style of parenting $g_{221} = 1.28$ and $g_{222} = .90$ with small values for the other core elements. The third family component (G3) shows a contrast in views between the daughters and parents. Daughters: a negative value for Acceptance, near zero for Control and a moderately positive value for Lax discipline (-.76, .12, .40); Parents: a moderately positive value for Acceptance, a moderately negative value for Control and a positive value for Lax discipline (.40, -.37, and .68). The fourth family component signifies a contrast between daughters and parents in judging parental style especially in the control and rejection scales, but the judges agree in their view of the discipline exerted by the parents.

Obviously, actual families can have scores on several components and thus are combinations of the patterns described above. It would have been insightful to have additional information on the families to make a further in depth analyses of parenting style of Japanese families. Unfortunately such information is not avail-able as Kojima's primary interest centred on the usability of the CRPBI in Japan, rather than on a full description of the perceptions on parental behaviour by Japanese families.

Joint Biplots.
A large part of the analysis of three-way data concentrates on the links bet-ween the components of the three modes, the size of which is contained in the core array. The interpretation of these links can be hard when no clear meaning can be attached to the (principal) components themselves. The principal compon-ents represents after all directions in the component space, which are not neces-sarily the directions of maximal interpretability. Not being able to assign substan-tive meaning to components puts restrictions on the possibility of the interpreta-tion of combinations of such components. This should be evident from the above interpretation of the Kojima core array that leaned very heavily on the meaning of the components in the scales and condition modes. To get around this, one may construct a *joint biplot* of the components of two modes (the *display modes*) given a component of the third (the *reference mode*). Detailed expositions of these plots can be found in Kroonenberg (2008) which also contains a number of examples and a joint biplot for the boys' data (Kroonenberg, 2008, p. 341). In most analy-ses, they form the most powerful representation of the results of a Tucker3 analy-ses.

Rotations of Components and/or Core Arrays.
The basic solution of the Tucker3 model is based on the orthonormality of the components in each mode. This is not a restriction due to the model, but one that has to be made in order to obtain a solution. However, once a solution is obtained it may be transformed without changing the fit to the data. This is in strong con-trast with the Parafac model which due to its parallel proportional profiles puts strong restrictions on the solution and rotating the component matrices automati-

cally leads to a loss of fit. Each component matrix may be rotated both with orthogonal transformations, such as varimax, and with oblique rotations or transformations, such as oblimin and the Harris-Kaiser independent cluster rotation (see Kiers & Ten Berge, 1994). The transformational freedom is a nuisance in the sense that there exists no unique solution for the model, at the same time it may be a blessing as it may be used to find more easily interpretable directions, which turned out to be case in the present example. Most proposed criteria for transformations in the Tucker3 model are orthogonal as is evident in the work by Kiers, summarised for instance in Kiers (1998b). This paper also treats the case of searching for overall simplicity in the model by simultaneous transforming the components and the core array. Applications where the combined procedure has been used are, for instance, Van Mechelen & Kiers (1999), Murakami and Kroonenberg (2003), and Kroonenberg (2008, p. 249ff.).

Tucker3 Model with Constraints.

Given that the Tucker3 model is primarily used in an exploratory sense, most applications use models without constraints on the parameters. However, theoretical developments have been extensive in the field of examining core arrays with as many zero elements as possible; see Kiers (1998a) for an overview. With respect to putting constraints on the components several authors have made proposal in this area and shown examples of their use, such as Klapper (1998), Bro and Sidiropoulos (1998), Bro (1998), Timmerman (2001), and Timmerman and Kiers (2002).

Internal validation of the model

Earlier we looked at the relative fit of the components and the combinations of the components via the core array, but we have not yet investigated the residuals from a model at the level of the individual data points. Similarly to other least squares procedures we may construct a residual plot of the standardised residuals verses the (standardised) predicted values. From such a plot (not shown) for the three-component Parafac, for instance, one can see that there are some larger standardised residuals, but given that we have 10,800 residuals that is only to be expected. The general conclusion from that residual plot is that the residuals are well-behaved. A further view at the adequacy of the solution can be had by looking at the distribution of the standardised residuals, which show as expected a well-behaved normal distribution. There are only a limited number (13) outlying observations, and there is no reason to be seriously worried about them.

CONCLUSION

In this paper an extensive overview is given of several practical issues in connection with applying three-models to real empirical data. It is hoped that this paper will make it easier for relatively uninitiated researchers to apply these techniques to unravelling complex patterns which exist in some three-way data sets. By placing the two major three-mode models side-by-side, an overview has been given of the strengths and weaknesses of these models. No real comprehen-

sive comparison, however, was made as this would require a lengthier treatment than was possible in the present context.

What has to be emphasised, however, is that the models may give very comparable results. In such cases, the nature of the data, the requirements and preferences of the researcher as well as ease of interpretation, play important roles in model selection. Finally, it should be remarked that it is often advantageous to use both types of model to get a good understanding of what the data have to tell. Eclecticism in model choice often leads to more insight than dogmatism.

REFERENCES

Bentler, P. M., & Lee, S.-Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology, 32*, 87-104.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems, 38*, 149-171.

Bro, R. (1998). *Multi-way analysis in the food industry. Models, algorithms, and applications*. Unpublished doctoral dissertation, University of Amsterdam, Amsterdam, The Netherlands. (http://www.models.kvl.dk/users/rasmus/thesis/thesis.html (Accessed April 2009.)

Bro, R., & Kiers, H. A. L. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics, 17*, 273-286.

Bro, R., & Sidiropoulos, N. D. (1998). Least squares algorithms under unimodality and non-negativity constraints. *Journal of Chemometrics, 12*, 223-247.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of ''Eckart-Young'' decomposition. *Psychometrika, 35*, 283-319.

Cattell, R. B. (1944). ''Parallel proportional profiles'' and other principles for determining the choice of factors by rotation. *Psychometrika, 9*, 267-283.

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

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology, 59*, 133-150.

Gabriel, K. R. (1971). The biplot graphic display with application to principal component analysis. *Biometrika, 58*, 453-467.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an ''explanatory'' multi-modal factor analysis. *UCLA Working Papers in Phonetics, 16*, 1-84.

Harshman, R. A. (1972a). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics, 22*, 111-117.

Harshman, R. A. (1972b). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics, 22*, 30-44.

Harshman, R. A. (1984). "How can I know if it's 'real'?" A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling. In H. G. Law, C.W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 566-591). New York: Praeger.

Harshman, R. A. (1994a). Substituting statistical for physical decomposition: Are there applications for parallel factor analysis (PARAFAC) in non-destructive evaluation? In P. V. Malague (Ed.), *Advances in signal processing for nondestructive evaluation of materials* (pp. 469-483). Dordrecht: Kluwer.

Harshman, R. A. (1994b, 24-26 June). *Three- and four-way factor analysis of brain scans, satellite photos, and other image data using both standard and novel multi-way models*. Paper presented at the Meeting of the Psychometric Society, Urbana IL.

Harshman, R. A., & DeSarbo, W. S. (1984). An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 602-642). New York: Praeger.

Harshman, R. A., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America, 62*, 693-707.

Harshman, R. A., & Lundy, M. E. (1984a). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis (pp.* 122-215). New York: Praeger.

Harshman, R. A., & Lundy, M. E. (1984b). Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216-284). New York: Praeger.

Kiers, H. A. L. (1998a). Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Computational Statistics and Data Analysis, 28*, 307-324.

Kiers, H. A. L. (1998b). Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *Journal of Classification, 15*, 245-263.

Kiers, H. A. L. (2000). Some procedures for displaying results from three-way methods. *Journal of Chemometrics, 14*, 151-170.

Kiers, H. A. L. (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics, 18*, 22-36.

Kiers, H. A. L., & Der Kinderen, A. (2003). A fast method for choosing the numbers of components in Tucker3 analysis. *British Journal of Mathematical and Statistical Psychology, 56*, 119-125.

Kiers, H. A. L., & Ten Berge, J. M. F. (1994). The Harris-Kaiser independent cluster rotation as a method for rotation to simple component weights. *Psychometrika, 59*, 81-90.

Kiers, H. A. L., & Van Mechelen, I. (2001). Three-way component analysis: Principles and illustrative application. *Psychological Methods, 6*, 84-110.

Klapper, D. (1998). *Die Analyse von Wettbewerbsbeziehungen mit Scannerdaten*. Heidelberg, Germany: Physica-Verlag.

Kojima, H. (1975). Inter-battery factor analysis of parents' and children's reports of parental behavior. *Japanese Psychological Bulletin, 17*, 33-48.

Krijnen, W. P. (1993). *The analysis of three-way arrays by constrained PARAFAC methods*. Leiden: DSWO Press.

Krijnen, W. P., & Kroonenberg, P. M. (2000). *Degeneracy and Parafac* (Technical report). Groningen, The Netherlands: Department of Psychology, University of Groningen.

Krijnen, W. P., & Ten Berge, J. M. F. (1992). A constrained PARAFAC method for positive manifold data. *Applied Psychological Measurement, 16*, 295-305.

Kroonenberg, P. M. (1994). The TUCKALS line: A suite of programs for three-way data analysis. *Computational Statistics and Data Analysis, 18*, 73-96.

Kroonenberg, P. M. (2008). *Applied multiway data analysis*. Hoboken NJ: Wiley.

Kruskal, J. B. (1984). Multilinear methods. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 36-62). New York: Praeger.

Legendre, P., & Legendre, L. (1998). *Numerical ecology* (2nd English ed.). Amsterdam: Elsevier.

Lorenzo-Seva, U., & Ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*, 57-64.

Murakami, T., & Kroonenberg, P. M. (2003). Individual differences in semantic differential studies and their analysis by three-mode models. *Multivariate Behavioral Research, 38*, 87-96.

Oort, F. J. (1999). Stochastic three-mode models for mean and covariance structures. *British Journal of Mathematical and Statistical Psychology, 52*, 243-272.

Paatero, P. (2000). Construction and analysis of degenerate PARAFAC models. *Journal of Chemometrics, 14*, 285-299.

Schaefer, E. S. (1965). Childrens reports of parental behavior: An inventory. *Child Development, 36*, 413-424.

Smilde, A. K., Bro, R., & Geladi, P. (2004). *Multi-way analysis in chemistry*. Chichester, UK: Wiley.

Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for p×p×2 arrays of rank p + 1 or higher. *Psychometrika, 71*, 483-501.

Ten Berge, J. M. F. (1986). Some relationships between descriptive comparisons of components from different studies. *Multivariate Behavioral Research, 21*, 29-40.

Ten Berge, J. M. F., De Leeuw, J., & Kroonenberg, P. M. (1987). Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika, 52*, 183-191.

Ten Berge, J. M. F., & Kiers, H. A. L. (1997). Are all varieties of PCA the same? A reply to Cadima and Joliffe. *British Journal of Mathematical and Statistical Psychology, 50*, 367-368.

Timmerman, M. E. (2001). *Component analysis of multisubject multivariate longitudinal data*. Unpublished doctoral dissertation, Department of Psychology, University of Groningen, Groningen, The Netherlands. (http://irs. ub.rug.nl/ppn/217649602. Accessed May 2007.)

Timmerman, M. E., & Kiers, H. A. L. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology, 53*, 1-16.

Timmerman, M. E., & Kiers, H. A. L. (2002). Three-way component analysis with smoothness constraints. *Computational Statistics & Data Analysis, 40*, 447-470.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personell Research Section Report No. 984). Washington, D. C.: Department of the Army.

Tucker, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika, 23*, 111-136.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279-311.

Van Mechelen, I., & Kiers, H. A. L. (1999). Individual differences in anxiety responses to stressful situations: A three-mode component analysis model. *European Journal of Personality, 13*, 409-428.

Wickens, T. (1994). *The geometry of multivariate statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Footnotes

1. Parts of this paper are based on Chapter 16 in Kroonenberg (2008); 2008©John Wiley & Sons Limited; reproduced and adapted with kind permission.
2. Due to Richard Harhman's untimely death on 10 January 2008, in finishing this paper we were unable to benefit from his deep insights into three-mode analysis.
3. In this paper the word *way* will primarily be used to designate three-way *data*, while the word *mode* will be used in connection with three-mode *methods* and *models*.

THREE-WAY HIERARCHICAL CLASSES: A COMPARISON OF THE INDCLAS AND TUCKER3-HICLAS MODELS

Iwin Leenen
Universidad Complutense de Madrid

Eva Ceulemans
Katholieke Universiteit Leuven

ABSTRACT

Hierarchical classes (HICLAS) models for multi-way multi-mode data constitute a unique family of classification models in that (a) they simultaneously induce a hierarchical classification (of the elements) of each mode and (b) they link the hierarchical classifications together by an association relation that yields a predicted (or reconstructed) value for each cell in the data array. For the case of three-way three-mode binary data, the most prominent HICLAS models include INDCLAS and Tucker3-HICLAS. In this paper, we compare the latter two models, introducing the underlying theory of both in substantive terms and showing how a Tucker3-HICLAS analysis may result in a simpler model than that yielded by INDCLAS, although the former is mathematically more complex than the latter (which it includes as a special case). We illustrate by two applications: a study on anger responses in frustrating situations and a case-study on emotions in interpersonal relations.

INTRODUCTION

Since De Boeck and Rosenberg's (1988) seminal paper on hierarchical classes, the HICLAS approach has expanded to a distinct family of classification models within the literature. The original model is a deterministic model for so-called two-way two-mode binary data. In this terminology (due to Tucker, 1964), the number of ways refer to the dimensionality of the data array, while the number of modes are the number of distinct sets these dimensions refer to. A well-known example of two-way one-mode data is a proximity (e.g., correlation) matrix, where rows and columns (two ways) of the matrix refer to the same set of entities/variables (one mode). For examples of two-way two-mode binary data, one may think of a child by item 0/1 matrix indicating success/failure of each child on each item, or object by attribute presence/absence data, person by choice object select/nonselect data, etc. Typically, deterministic models for two-way two-mode binary data (see, e.g., the models introduced by Coombs, 1964) include

an equation that assigns a predicted (or reconstructed model) 0/1 value to each cell of the matrix. That is to say, these models account for the *association relation* between the two modes. In the same line, De Boeck and Rosenberg's model accounts for the association relation in a two-way two-mode binary array, but, as a unique feature, their model represents two other types of relations as well, namely the *equivalence and hierarchy relations* defined on the two modes of this array, which yield the typical two-sided hierarchical classification. Each of the three types of relations may be of substantive interest: For example, accounting for the association relation in person by object select/nonselect data may reveal the underlying choice mechanisms (Van Mechelen & Van Damme, 1994); by representing the equivalence relations in patient by symptom presence/absence data, HICLAS may shed light on the diagnostic classification system of psychiatrists (Van Mechelen & De Boeck, 1989); and the hierarchical relations are of key relevance in person perception (Gara, 1990; Gara & Rosenberg, 1979), differential emotion psychology (Kuppens, Van Mechelen, Smits, De Boeck, & Ceulemans, 2007), and knowledge space theory (Doignon & Falmagne, 1999; Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990).

De Boeck and Rosenberg's (1988) original model has been extended in various directions: Variants have been proposed that differ in the way the association, equivalence, and hierarchy relations are represented (see, e.g., Van Mechelen, De Boeck, & Rosenberg, 1995); generalizations have been developed that allow for a direct modeling of nonbinary data (Schepers & Van Mechelen, 2006; Van Mechelen, Lombardi, & Ceulemans, 2007); and other models replaced the deterministic framework by a probabilistic one (Leenen, Van Mechelen, Gelman, & De Knop, 2008). Another extension of the original model comprises HICLAS models for N-way N-mode data (with $N > 2$, in its general form described by Ceulemans and Van Mechelen, 2003), that is, data that can be represented in an N-dimensional array with the N dimensions referring to N different sets of elements. Substantively interesting HICLAS models for multi-way multi-mode data include the INDCLAS (Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999) and Tucker3-HICLAS model (Ceulemans, Van Mechelen, and Leenen, 2003; and variants of the latter, see Ceulemans and Van Mechelen, 2004, 2005), which are models for three-way three-mode binary data.

The present paper has a twofold aim: First, we want to provide an easy introduction to the INDCLAS and Tucker3-HICLAS models, focusing on substantive aspects. Therefore, we will recapitulate the theory behind these models giving special attention to the substantive meaning of the various aspects of the models. Furthermore, a number of tools that have been proposed to deal with the model selection problem—that is, the problem of selecting which INDCLAS or Tucker3-HICLAS model yields the best description of the data (best referring to an optimal trade off between simplicity/parsimony and goodness of fit)—are discussed and illustrated, as this topic is particularly relevant for practical users. Second, we want to elaborate on the comparison of the INDCLAS and Tucker3-HICLAS models, focusing on the results that can be expected when applying both models to the same data set. This will be illustrated by two applications. In the first application, it will be shown that a Tucker3-HICLAS model, which implies a generalization of the association rule in INDCLAS, may yield simpler hierarchical classifications. In the second application, we show how a Tucker3-HICLAS

model may yield a considerably better fit to the data as compared to an INDCLAS model of about the same complexity.

The paper is further organized as follows: In the next section, we give a precise definition of the association, equivalence, and hierarchy relations in three-way three-mode binary data and show how these relations are represented by the INDCLAS and Tucker3-HICLAS model, respectively. Subsequently, we discuss some of the more practical aspects (including model selection) when either model is fitted to a given data set. Thereafter, we discuss the two applications where we compare INDCLAS and Tucker3-HICLAS results. In the final section, we present some recent developments in HICLAS analysis and discuss their usefulness when applying the models discussed in this paper.

The INDCLAS and Tucker3-HICLAS Model

Consider an $I \times J \times K$ data array $\underline{\mathbf{X}} = (x_{ijk})$ with entries $x_{ijk} \in \{0,1\}$.[1] So as not to make the exposition unnecessarily abstract, we will illustrate the theoretical concepts introduced in this section by means of a hypothetical example. In particular, we will assume in this guiding example that four staff members in a health service—a clinical psychologist, a trainee, a general practitioner, and a psychiatrist—each interviewed (the same) six patients at intake—a drug addict, a paranoiac, a demented man, a compulsive cleaner, an anorexic girl, and a man suffering delirium—and, after the interview, judged the presence/absence of seven psychiatric symptoms—hallucinations, delusions, anxiety, depression, obsessive behavior, social isolation, and disorientation—in each of these patients. The data, which can be organized in a 4 (judges) $\times$ 6 (patients) $\times$ 7 (symptoms) data array $\underline{\mathbf{X}}$, with $x_{ijk} = 1$ if judge i considers symptom k present in patient j and $x_{ijk} = 0$ otherwise, are shown in Table 1.

We now define three types of relations on such a three-way three-mode data array. Thereafter, we will show how the INDCLAS and Tucker3-HICLAS models account for these relations.

Equivalence, Hierarchy, and Association in a Binary Three-Way Array

Equivalence. Equivalence relations are defined on the set of elements of a given mode. As a result, for three-way three-mode data, three equivalence relations are considered. For the data in Table 1, an equivalence relation is defined on (a) the set of four judges, (b) the set of six patients, and (c) the set of seven symptoms. A pair of elements of a given mode is equivalent if and only if the submatrices associated with each of them are identical. In order to compare the submatrices associated with the respective elements, a representation of the data such as in Figure 1 may be helpful. The figure represents the data in three different ways, depending on which mode is used to segregate submatrices. To make the comparison easier, a value of one in a cell is represented by a cross, while cells with a zero value are left blank. Figure 1(a) shows the patient by symptom submatrix of each judge, from which it is easily checked that the clinical psychologist and the psychiatrist have identical submatrices and, hence, are equivalent for these data. In the same way, from Figure 1(b), which shows the patient by judge submatrix of each symptom, we can derive, for example, that hallucinations,

Table 1. Hypothetical Three-Way Three-Mode Binary Data Array

Clinical psychologist

	Hallucinations	Delusions	Anxiety	Depression	Obsessive behavior	Social isolation	Disorientation
Drug addict	1	1	1	1	1	1	1
Paranoiac	1	1	1	0	0	1	1
Demented man	1	1	1	1	1	1	1
Compulsive cleaner	0	0	1	1	1	1	0
Anorexic girl	0	0	1	1	1	1	0
Man with delirium	1	1	1	1	1	1	1

Trainee

	Hallucinations	Delusions	Anxiety	Depression	Obsessive behavior	Social isolation	Disorientation
Drug addict	1	1	1	0	0	1	1
Paranoiac	1	1	1	0	0	1	1
Demented man	1	1	1	1	1	1	1
Compulsive cleaner	0	0	1	1	1	1	0
Anorexic girl	0	0	1	1	1	1	0
Man with delirium	1	1	1	0	0	1	1

General practitioner

	Hallucinations	Delusions	Anxiety	Depression	Obsessive behavior	Social isolation	Disorientation
Drug addict	1	1	1	1	1	1	1
Paranoiac	1	1	1	0	0	1	1
Demented man	1	1	1	0	0	1	1
Compulsive cleaner	0	0	0	0	0	0	0
Anorexic girl	0	0	0	0	0	0	0
Man with delirium	1	1	1	1	1	1	1

Psychiatrist

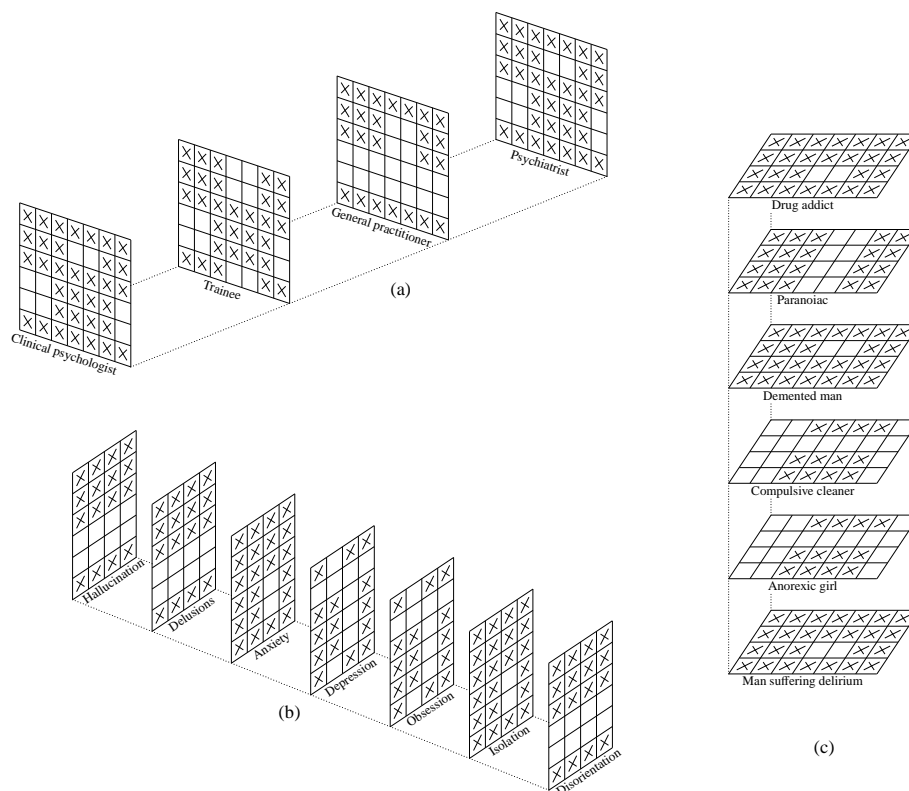|	Hallucinations	Delusions	Anxiety	Depression	Obsessive behavior	Social isolation	Disorientation
Drug addict	1	1	1	1	1	1	1
Paranoiac	1	1	1	0	0	1	1
Demented man	1	1	1	1	1	1	1
Compulsive cleaner	0	0	1	1	1	1	0
Anorexic girl	0	0	1	1	1	1	0
Man with delirium	1	1	1	1	1	1	1

Figure 1. Three alternative representations of the data in Table 1: segregated (a) by judges, (b) by patients, and (c) by symptoms. Cells with a data value of 1 are filled with a cross; cells with a data value of 0 are left empty.

delusions, and disorientation are pairwise equivalent. Likewise, from Figure 1(c), which allows for a comparison of the judge by symptom submatrices of different patients, we learn, for example, that the compulsive cleaner and the anorexic girl are equivalent. The substantive interpretation of the equivalence relations follows from their definition: Equivalent judges assign the same symptoms when considering a particular patient; equivalent patients are assigned the same symptoms by each judge; and, for any judge, equivalent symptoms either co-occur or they are both absent in any patient. Note further that the equivalence relations induce a classification of the elements in the respective modes. For example, the equivalence relation on the symptoms induces the following three classes of equivalent symptoms: {hallucinations, delusions, disorientation}, {anxiety, social isolation}, {depression, obsessive behavior}.

Hierarchy. Apart from the equivalence relation, a hierarchy relation is defined on the set of elements of each mode. An element i is hierarchically below an element i' if and only if the submatrix of i is a proper subset of the submatrix of i'. In tracing hierarchy relations, Figure 1 may again be helpful. From panel (a), we conclude, for example, that the trainee is hierarchically below the psychiatrist: If the trainee's matrix contains a cross, then the psychiatrist's matrix contains a

cross as well. This example shows that the hierarchy relation can alternatively be interpreted as an if-then type of relation (see also, Van Mechelen, Rosenberg, & De Boeck, 1997): *If* the trainee judges a symptom to be present in a patient, *then* also the psychiatrist judges that symptom to be present in that patient. Likewise, from Figure 1(b), we can read that hallucinations are hierarchically below anxiety. That is, if a patient suffers from hallucinations (according to some judge) then (s)he suffers from anxiety as well (according to the same judge). Finally, Figure 1(c) shows that the paranoiac is hierarchically below the drug addict: If a symptom is believed to be present in the paranoiac, then it is believed to be present in the drug addict as well.

The hierarchy relation defined on the elements of a given mode directly produces a hierarchical organization of the classes induced by the equivalence relation. Such hierarchically organized classifications can be conveniently represented by a Hasse diagram (which is a graphical representation of a partially ordered set as a directed graph, the direction implicitly being upward, see Davey & Priestley, 2002). Figure 2 shows a Hasse diagram representing the hierarchical classification of the patients. Similar Hasse diagrams can be constructed for the judges and the symptoms as well. In such a representation, the boxes represent classes and the straight lines connecting the boxes indicate that the corresponding classes are hierarchically related. Equivalence and hierarchy relations are often alternatively called *set-theoretical relations* (as they are defined by comparing the sets of positive responses in the submatrices of different elements).
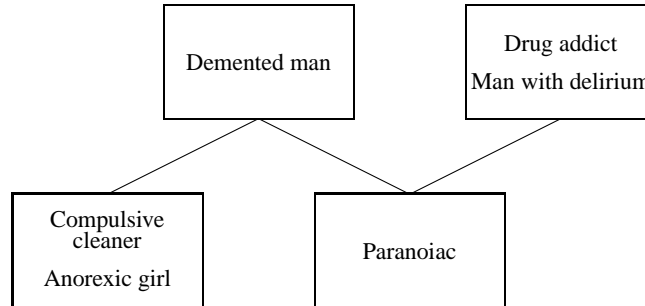

Figure 2. Hasse diagram representing the hierarchical classification of the patients as implied by the data in Table 1.

Association. The association relation links the three modes together. Loosely speaking, the association relation is "the" relation showed in the data array; more precisely, a judge i, a patient j, and a symptom k are associated if and only if the corresponding cell in the data array has a value of one. Whereas the set-theoretical relations are defined on each mode separately, giving rise to three equivalence and three hierarchy relations in a three-way three-mode array, there is only a single association relation, linking the three modes together. Analogously to the way that the hierarchy relation on the elements implied a hierarchical organization of the classes, the association relation on the elements implies an association of the classes. Hence, the association relation can be said to link the three hierarchical classifications together. For example, from the data in Table 1, we read that the clinical psychologist, the drug addict, and hallucinations are

associated; as a result, the corresponding classes {clinical psychologist, psychiatrist}, {hallucinations, delusions, disorientation}, and {drug addict, man suffering delirium} are associated as well.

Representation of Equivalence, Hierarchy, and Association in INDCLAS

Algebraically, an INDCLAS model for an $I \times J \times K$ binary array $\underline{\mathbf{X}}$ is a triple of binary matrices ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) that satisfy a set of restrictions so as to represent the relations on $\underline{\mathbf{X}}$ that were introduced in the previous section. The dimensions of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are $I \times R$, $J \times R$, and $K \times R$, respectively, where the integer R is called the *rank* of the model. The R columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ correspond to R *bundles* and these matrices themselves are called *bundle matrices*. The bundles can be conceived of as latent features that are related in some way to the elements in the three modes. In some cases, the bundles can be given a substantive interpretation. As an example, consider Table 2, which shows an INDCLAS model of rank $R = 3$ for the hypothetical data of Table 1.[2] For this example, we will interpret the bundles in terms of latent syndromes (for applications on similar, genuine data, see Leenen, Van Mechelen, De Boeck & Rosenberg, 1999; Van Mechelen & De Boeck, 1989), namely affective disorder, psychotic disorder, and substance abuse disorder, where the bundle matrix $\mathbf{A}$ indicates for each of the ($I = 4$) judges which of the ($R = 3$) latent bundles/syndromes (s)he takes into account, $\mathbf{B}$ indicates for each of the ($J = 6$) patients from which of the ($R = 3$) syndromes (s)he suffers, and $\mathbf{C}$ indicates for each of the ($K = 7$) symptoms of which of the ($R = 3$) syndromes it is characteristic. We see, for example, that the trainee takes into account the affective and psychotic disorders, but not the substance abuse disorder (possibly due to a lack of experience with the latter). Similarly, the paranoiac suffers from a psychotic disorder but not from affective and substance abuse disorders; and hallucinations are characteristic of psychotic and substance abuse disorders, but not of affective disorders.

One may note parallels between the models for hierarchical classes analysis and principal component analysis (PCA) and, in particular, the conceptual similarity between a bundle matrix in a HICLAS model on the one hand and a loading matrix or (component) score matrix in PCA on the other hand. Both a bundle and a loading/score matrix indicate for each of a number of *observed* entities (e.g., judges, patients, symptoms, persons, tests, items, etc.) the loading on or connection with a number of *latent* variables (with the obvious difference that loading and score matrices in PCA commonly have real-valued entries, while bundle matrices in HICLAS are binary).

We now discuss the restrictions that are imposed on the three INDCLAS bundle matrices so as to make them represent the abovementioned three types of relations.

Equivalence. If two elements of a given mode are equivalent (or, in other words, belong to the same class), then in the bundle matrix for that mode, the rows corresponding to these elements must be identical. For example, the clinical psychologist and the psychiatrist, which were found to be equivalent in the data of Table 1 (see, Figure 1(a)), have identical rows in the judge bundle matrix $\mathbf{A}$ (both are related to all latent bundles/syndromes). Likewise, the symptoms in the

Table 2. INDCLAS Model for the Hypothetical Data of Table 1.

	Syndromes		
Judge bundle matrix A	Affective Disorder	Psychotic Disorder	Substance Abuse Disorder
Clinical psychologist	1	1	1
Trainee	1	1	0
General practitioner	0	1	1
Psychiatrist	1	1	1

Patient bundle matrix B			
Drug addict	0	1	1
Paranoiac	0	1	0
Demented man	1	1	0
Compulsive cleaner	1	0	0
Anorexic girl	1	0	0
Man with delirium	0	1	1

Symptom bundle matrix C			
Hallucinations	0	1	1
Delusions	0	1	1
Anxiety	1	1	1
Depression	1	0	1
Obsessive behavior	1	0	1
Social isolation	1	1	1
Disorientation	0	1	1

class {hallucinations, delusions, and disorientation} have identical rows in the symptom bundle matrix **C** (the three being characteristic of the psychotic as well as the substance abuse disorder but not of the affective disorder).

Hierarchy. Two elements of a given mode that are hierarchically related must have rows in the corresponding bundle matrix that are in a subset/superset relation. For example, we have already pointed out that, in the hierarchy relation defined on the judges, the trainee is below the psychiatrist; therefore, in the judge bundle matrix **A**, the trainee's row is a (proper) subset of the psychiatrist's row: The psychiatrist takes into account all the syndromes that the trainee takes into account (and more). Similarly, in the patient bundle matrix **B**, as the paranoiac is below the drug addict in the patient hierarchy, the row of the former is a subset of the row of the latter: The paranoiac suffers from a subset of the syndromes that the drug addict suffers from.

Association. Saying that a hierarchical classes model represents the association relation means that the model includes a rule which allows to derive the binary value of a given data cell from the bundle matrices. For the INDCLAS model, this rule reads as follows: Three elements (from different modes) are associated if and only if there exist a latent bundle to which each of them is related. In the context of our guiding example: A judge i indicates that patient j suffers from symptom k if and only if there exists a syndrome that (a) judge i takes into account, (b) patient j suffers from, and (c) symptom k is characteristic of. Mathematically, this implies that the following equation holds for all i, j, and k:

$$x_{ijk} = \bigoplus_{r=1}^{R} a_{ir} b_{jr} c_{kr}, \tag{1}$$

where $\oplus$ denotes the Boolean sum (i.e., with $1 \oplus 1 = 1$). One may check that, with model (**A**, **B**, **C**) in Table 2, Equation (1) holds for all cells of the data in Table 1. For example, the trainee judges the anorexic girl to suffer from depression (i.e., $x_{254} = 1$) "because" there exists a syndrome (viz., affective disorder as represented by the first bundle) (a) that is taken into account by the trainee (i.e., $a_{21} = 1$), (b) from which the patient suffers (i.e., $b_{51} = 1$), and (c) of which depression is characteristic (i.e., $c_{41} = 1$).

Leenen, Van Mechelen, De Boeck, and Rosenberg (1999) proposed a comprehensive graphical representation of the INDCLAS model from which (most of) the relations accounted for can be easily read. The graphical representation of the INDCLAS model in Table 2 is displayed in Figure 3. The upper part of the figure (above the circles in the center) shows the Hasse diagram with the hierarchical organization of the patients (in the same way as in Figure 2), while the bottom part (under the circles) shows, upside down, the Hasse diagram with the hierarchical organization of the symptoms. Each hierarchy has at the bottom $R = 3$ *bundle-specific classes* (of which one or more may be empty). The elements in a bundle-specific class are exclusively related to one bundle (or syndrome) and, as a result, have exactly one value which equals one in their row of the bundle matrix. Whereas in the symptom hierarchy, all bundle-specific classes are empty (note that neither of the rows in bundle matrix **C** of Table 2 has exactly one value
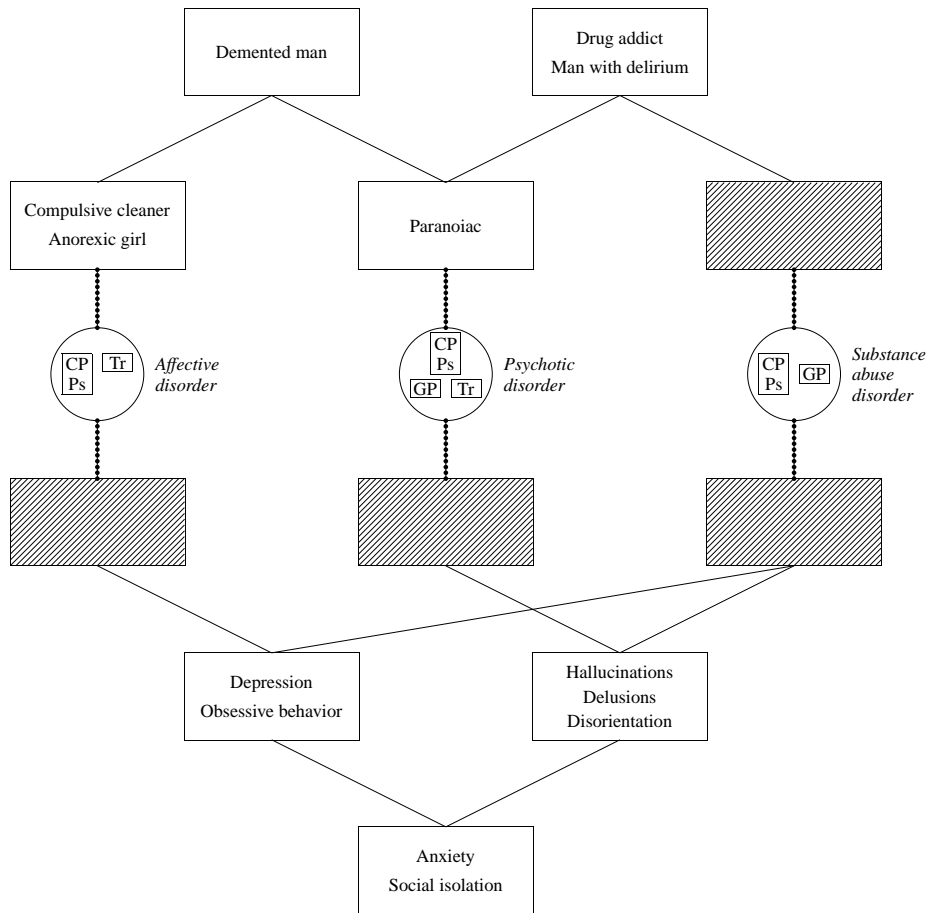
Figure 3. Graphical representation of the INDCLAS model of Table 2. CP, Tr, GP, and Ps stand for Clinical psychologist, Trainee, General practitioner, and Psychiatrist, respectively.

of one), the bundle-specific classes in the patient hierarchy are (i) {compulsive cleaner, anorexic girl}, (ii) {paranoiac}, and (iii) an empty class. The compulsive cleaner and the anorexic girl are exclusively related to the affective disorder, while the paranoiac suffers only from the psychotic disorder. In order to represent the association relation, the bundle-specific classes of the patient and symptom hierarchy are pairwise connected by circles, where each circle corresponds to a particular bundle/syndrome and contains the (classes of) judges who are related with that bundle (i.e., who take that syndrome into account). For example, the leftmost circle, corresponding to the affective disorder, contains the judge classes {Clinical psychologist, Psychiatrist} and {Trainee}, that is, the judges that have a one in the first position of their row in **A**. As a consequence, the association relation is read as follows: If (and only if) there exists a (downward) path from patient j to symptom k that passes through a circle that contains judge i, then judge i, patient j, and symptom k are associated. Thus, we can derive, for example, that the trainee indicates that the demented man suffers from depression, but that the general practitioner does not.

The graph does not show the hierarchical classification for the elements that are represented within the circles in the centre (in this case, the judges).[3] Optionally, the hierarchical classification of these elements can be represented separately by means of a Hasse diagram, such as in Figure 4. Note that, like for the symptoms, the bundle-specific classes for the judges are all empty. These classes are nevertheless included so as to clarify with which syndromes the judges are related.
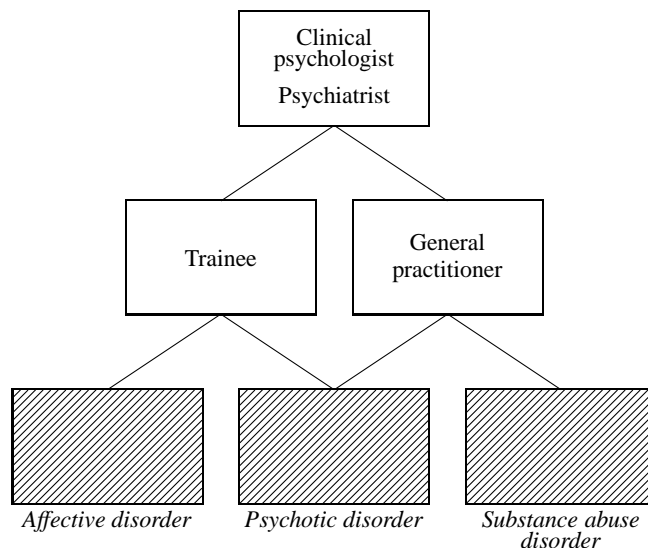

Figure 4. Graphical representation of the hierarchical classification of the judges in the INDCLAS model of Table 2.

The reader may wonder whether imposing the restriction in Equation (1) does not by itself imply that the bundle matrices correctly represent the set-theoretical relations. Or, in other words, is it possible that matrices **A**, **B**, and **C** correctly combine to $\underline{\mathbf{X}}$, while these matrices violate the restrictions with respect to equivalence and/or hierarchy? An example showing that this is possible indeed is given in Table 3. On the one hand, the data array $\underline{\mathbf{X}}$ can be perfectly reconstructed by Equation (1) from the bundle matrix **C*** in this table and the bundle matrices **A** and **B** in Table 2. The set-theoretical relations, on the other hand, are not represented correctly. For example, the equivalent symptoms hallucinations and delusions do not have identical rows in **C***; and disorientation is hierarchically below social isolation in $\underline{\mathbf{X}}$, while in **C*** the row of the former is not a subset of the row of the latter. This triple (**A**, **B**, **C***) is characterized as a decomposition of X that is not set-theoretically consistent.

Representation of Equivalence, Hierarchy, and Association in Tucker3-HICLAS

In an INDCLAS model, the three modes share the same *R* bundles; that is to say, the matrices **A**, **B**, and **C** have the same number of columns and the columns of these bundle matrices are in a one-to-one relationship (which is nicely illustrated in Figure 3 where the bundle-specific classes of patients and symptoms are

Table 3. INDCLAS Symptom Bundle Matrix, which Together with Matrices A and B in Table 2 Correctly Represents the Association Relation, but Fails to Give a Correct Account of the Set-Theoretical Relations on the Symptoms as Implied by the Data in Table 1.

	Syndromes		
Symptom bundle matrix C*	Affective Disorder	Psychotic Disorder	Substance Abuse Disorder
Hallucinations	0	1	0
Delusions	0	1	1
Anxiety	1	1	0
Depression	1	0	1
Obsessive behavior	1	0	1
Social isolation	1	1	0
Disorientation	0	1	1

connected through the bundle they have in common). The Hasse diagrams for judges and symptoms in Figures 3 and 4, however, show that, if only the hierarchical organizations of these modes are considered, a smaller number of bundles may be sufficient to represent them. This motivated the development of Tucker3-HICLAS, which, like INDCLAS, implies a decomposition of a binary three-way three-mode array $\underline{\mathbf{X}}$ into bundle matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, but unlike INDCLAS, allows the bundle matrices to differ with respect to the number of columns. We write $I \times P$, $J \times Q$, and $K \times R$ for the dimensions of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively. It follows that the rank is no longer characterized by a single number; rather, we will denote the rank of a Tucker3-HICLAS model as (P,Q,R). Obviously, with the bundle matrices having a different number of columns and without further specification, it would be unclear how the bundles from different bundle matrices are connected (note that a one-to-one relationship like in INDCLAS is not generally possible anymore). Therefore, Tucker3-HICLAS adds a three-dimensional binary *core array* $\underline{\mathbf{G}}$, of dimension $P \times Q \times R$, which specifies $g_{pqr} = 1$ if bundle p of $\mathbf{A}$, bundle q of $\mathbf{B}$, and bundle r of $\mathbf{C}$ are connected, and $g_{pqr} = 0$ otherwise.

By means of the Tucker3-HICLAS model of rank (2,3,2) in Table 4, we now show how the model represents the relations defined on the data array $\underline{\mathbf{X}}$ in Table 1.

Equivalence and hierarchy. Tucker3-HICLAS imposes identical restrictions on **A**, **B**, and **C** as INDCLAS with respect to the representation of the set-theoretical relations. In the context of our guiding example, equivalent judges (respectively patients, symptoms) have identical rows in **A** (resp. **B**, **C**) and two judges (resp. patients, symptoms) that are hierarchically related have rows in **A** (resp. **B**, **C**) that are in a subset-superset relation. Note that, as to the set-theoretical relations, no restrictions are (directly) imposed on the core array **G**.

Association. In Tucker3-HICLAS, **A**, **B**, **C**, and **G** combine to **X** by the following association rule:

$$x_{ijk} = \bigoplus_{p=1}^{P} \bigoplus_{q=1}^{Q} \bigoplus_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr}.$$

More substantively, within the context of the model in Table 4, this reads: A judge i indicates that patient j suffers from symptom k if and only if (a) a judge bundle, say p, exists with which judge i is related, (b) a patient bundle, say q, exists with which patient j is related, (c) a symptom bundle, say r, exists with which symptom k is related, and (d) judge bundle p, patient bundle q, and symptom bundle r are connected in **G**. For example, the general practitioner judges that the drug addict suffers from depression (i.e., $x_{314} = 1$) as (a) the general practitioner is related to judge bundle A-II (i.e., $a_{32} = 1$), (b) the drug addict is related to patient bundle B-III (i.e., $b_{13} = 1$), (c) depression is related to symptom bundle C-I (i.e., $c_{41} = 1$), and (d) judge bundle A-II, patient bundle B-III, and symptom bundle C-I are connected in **G** (i.e., $g_{231} = 1$).

It is easily seen that the Tucker3-HICLAS model includes the INDCLAS model as a special case: A rank R^* INDCLAS model (**A**, **B**, **C**) is identical with a rank (P,Q,R) Tucker3-HICLAS model (**A**, **B**, **C**, **G**) with $P = Q = R = R^*$, and $g_{pqr} = 1$ if $p = q = r$, and $g_{pqr} = 0$ otherwise (i.e., **G** is a unit superdiagonal array, Kiers, 2000).

A graphical representation of the Tucker3-HICLAS model in Table 4 is given in Figure 5. The Hasse diagram for the judges implied by the model is shown in Figure 6. The relations of equivalence, hierarchy, and association are read from these graphs analoguously as from the INDCLAS graphs in Figures 3 and 4. Note that, on the one hand, the empty judge and symptom classes have disappeared in Tucker3-HICLAS, leaving simpler hierarchical classifications for these modes, while, on the other hand, the associations among the judges, patients, and symptoms (represented by the connections in the center of Figure 5) are slightly more complex than in the INDCLAS graph.

Fitting a Three-Way HICLAS Model to Observed Data: Practical Considerations

In going through the previous section, the reader might have raised the objection that it is particularly unlikely to find in a genuine three-way data array pairs of elements that are perfectly equivalent or that are hierarchically related without any exception. Indeed, in practical applications, the data almost never can be perfectly described by a simple HICLAS model, where with ''simple'' we mean ''of low rank''. (In this respect, remind that the rank poses an upper limit on the

Table 4. Tucker3-HICLAS Model for the Hypothetical Data of Table 1.

Judge bundle matrix A

	A-1	A-II
Clinical psychologist	1	1
Trainee	1	0
General practitioner	0	1
Psychiatrist	1	1

Patient bundle matrix B

	B-I	B-II	B-III
Drug addict	0	1	1
Paranoiac	0	1	0
Demented man	1	1	0
Compulsive cleaner	1	0	0
Anorexic girl	1	0	0
Man with delirium	0	1	1

Symptom bundle matrix C

	C-I	C-II
Hallucinations	0	1
Delusions	0	1
Anxiety	1	1
Depression	1	0
Obsessive behavior	1	0
Social isolation	1	1
Disorientation	0	1

Core array $\underline{\mathbf{G}}$

A-I

	C-I	C-II
B-I	1	0
B-II	0	1
B-III	0	0

A-II

	C-I	C-II
B-I	0	0
B-II	0	0
B-III	1	0

number of classes in a mode, namely 2^R if the bundle matrix for that mode has R columns, as such, forcing a partition of its elements in 2^R hierarchically related classes). Although theoretically, for any data array, the rank can be increased up to a point that the HICLAS model represents the data perfectly (see, for a general treatment of existence in decomposition models for binary data, Leenen, Van Mechelen, & De Boeck, 1999), in practical applications one usually allows for nonperfect HICLAS models but that are simple and easy to interpret. In the context of three-way HICLAS models, this means that the data array $\underline{\mathbf{X}}$ is approximated by an array $\underline{\hat{\mathbf{X}}}$, which can be perfectly reconstructed from a simple model, while an array $\underline{\mathbf{E}}$ accounts for the discrepancy between model and data:

$$\underline{\mathbf{X}} = \underline{\hat{\mathbf{X}}} + \underline{\mathbf{E}},$$

where $\varepsilon_{ijk} \in \{-1,0,1\}$. The proposed INDCLAS and Tucker3-HICLAS algorithms (described in Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999, and Ceulemans et al., 2003, respectively) search, given a data array $\underline{\mathbf{X}}$ and a rank specified by the user, for a model [i.e., a triple ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) and a quadruple ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\underline{\mathbf{G}}$), respectively] that correctly represents the association, equivalence, and hierarchy relations in an array $\hat{\underline{\mathbf{X}}}$, while minimizing the errors in $\underline{\mathbf{E}} = \underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}$, that is, while minimizing the loss function:

$$L\left(\hat{\underline{\mathbf{X}}}\right) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{ijk} - \hat{x}_{ijk})^2. \tag{2}$$

Since the values x_{ijk} and $\hat{x}_{ijk}$ are binary, the least squares loss function in (2) is equivalent to a least absolute deviations loss function and returns the number of cells with different values for $\underline{\mathbf{X}}$ and $\hat{\underline{\mathbf{X}}}$. Such cells are called discrepancies. HICLAS algorithms aim at minimizing the number of discrepancies between a particular HICLAS model and observed data. Often, the value on loss function (2) is divided by the number of cells ($I \cdot J \cdot K$), which yields the proportion of discrepancies. Note that the original minimization algorithms do not include restrictions on the bundle matrices and/or the reconstructed data array, which implies that the analysis is exploratory. Ceulemans, Van Mechelen, and Kuppens (2004) recently introduced some tools for fitting restricted HICLAS models, as such allowing for a confirmatory HICLAS analysis; this new development will be briefly discussed in the last section of this paper.
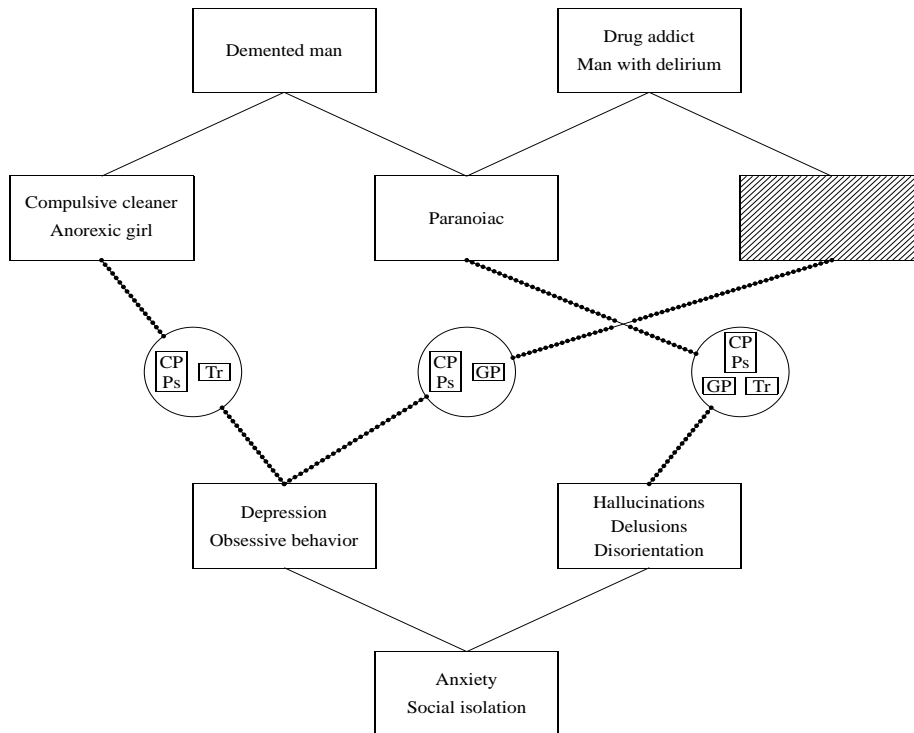

Figure 5. Graphical representation of the Tucker3-HICLAS model of Table 4. CP, Tr, GP, and Ps stand for Clinical psychologist, Trainee, General practitioner, and Psychiatrist, respectively.
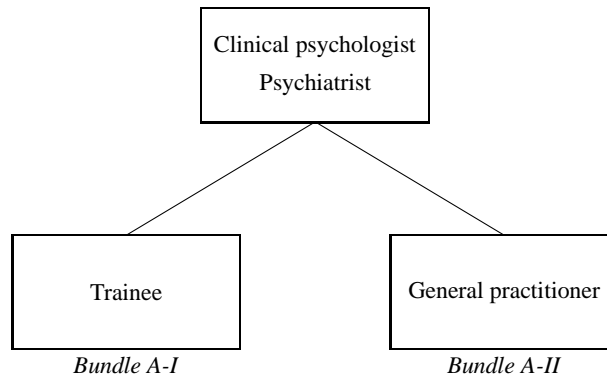
Figure 6. Graphical representation of the hierarchical classification of the judges in the Tucker3-HICLAS model of Table 4.

Although in some cases the rank of a HICLAS model follows from a theory about the data generation process, in most applications the basis for prespecifying the rank of the model (required as input to the fitting algorithms) is unclear. A common strategy in that case is fitting models with different ranks in successive runs of the algorithm and selecting a model that gives an optimal trade-off between simplicity and fit to the data (e.g., proportion of discrepancies). This idea has been concretized in various heuristics for rank selection, including ''visual methods'' applying an ''elbow criterion'' on the rank by goodness of fit plot (i.e., the scree test and its extensions such as the convex-hull approach, see Cattell, 1966; Ceulemans & Van Mechelen, 2005) and ''pseudo-statistical methods'' (such as the pseudo-binomial test, Leenen & Van Mechelen, 2001, and the pseudo-AIC strategy, Ceulemans & Van Mechelen, 2005). In the application sections, we will illustrate some of these heuristics and briefly discuss their rationale (for the details, we refer to the original papers).

As a final note, we mention that an INDCLAS or Tucker3-HICLAS model may suffer from an indeterminacy, in some sense similar to the indeterminacy in principal component analysis. The problem in HICLAS is definitely much smaller than in PCA, because in most applications the HICLAS model can be shown to be unique. In contrast with PCA, however, the problem is much harder to delineate in the HICLAS case (where alternative solutions cannot be obtained through a ''rotation'' of the solution). Ceulemans & Van Mechelen (2003) identified some sufficient conditions for uniqueness, one of them being that an INDCLAS model is unique if the bundle-specific classes are nonempty. A similar, slightly more complex proposition (including a condition for the core array $\mathbf{G}$ as well) applies to the Tucker3-HICLAS model.

Application 1: Behavioral Signatures in Situation-Response Profiles

We now present the results of an INDCLAS and Tucker3-HICLAS analysis of questionnaire data previously reported by Vansteelandt and Van Mechelen (1998). These authors presented 54 first-year psychology students with 23 frustrating situations (see Table 5), and asked them to indicate for each situation which behaviors from a set of 15 anger responses (mentioned in the bottom half
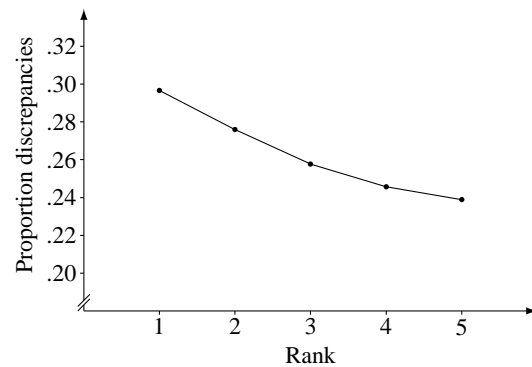
Figure 7. Plot of proportion of discrepancies by rank of five INDCLAS models for the situation-response data of Application 1.

of Figure 8) they would display, yielding a $54 \times 23 \times 15$ binary array $\underline{\mathbf{X}}$, with x_{ijk} = 1 if person i displays behavior k in situation j, and $x_{ijk} = 0$ otherwise. The aim of the study was to construct from these data a triple typology model, that is, a model with (a) a typology (i.e., partition in classes) of persons, situations, and responses, (b) where the types in a typology are hierarchically organized, and (c) where the model would provide a characterization of the person types in terms of if(situation)—then(response) rules (called ''behavioral signatures'', Shoda, Mischel, & Wright, 1994), as such linking the three typologies together. To this end, Vansteelandt and Van Mechelen analyzed the data array using the INDCLAS model. Here, we reanalyze the data with both the INDCLAS and the Tucker3-HICLAS algorithm. As will become clear in the discussion below, the solutions obtained for INDCLAS and Tucker3-HICLAS models of comparable complexity (i.e., of about the same rank) fit the data almost equally well. Therefore, we will concentrate on a comparison of the structural aspects of the INDCLAS and Tucker3-HICLAS models for these data.

INDCLAS Analysis

We successively applied the INDCLAS algorithm so as to find best fitting INDCLAS models in each of the ranks 1 to 5 and obtained models with proportions of discrepancies equal to .297, .276, .258, .246, and .239, respectively. In order to select one of them for further discussion, we applied both visual and pseudo-statistical rank selection heuristics. Visual heuristics are based on an analysis of a fit by rank plot (see Figure 7) and, generally speaking, search for an ''elbow'' in such a plot (i.e., for a sharpest angle, Cattell, 1966). Pure visual inspection of the plot in Figure 7 would lead to a model in either rank 3 or 4. A formalization of the ''elbow'' idea is Ceulemans et al.'s (2003; see also Leenen & Van Mechelen, 2001) scree test, which calculates, for each rank R, the difference between the decrease in discrepancies when going from rank R - 1 to R and the decrease in discrepancies when going from rank R to R + 1, and selects the rank that maximizes this difference. A variant (see Ceulemans & Van Mechelen, 2005) consists of calculating the ratio rather than the difference and selecting the rank that maximizes the ratio.[4] If applied to the proportion of discrepancies for the five

Table 5. List of Frustrating Situations Used in the Situation-Response Questionaire of Application 1 With Key Words for Later Reference in Italic

1. You are talking to someone and he (she) does *not answer* you

2. You accidentally *bang your shins* against a park bench.

3. The *grocery* store closes just as you are about to enter.

4. Your instructor *unfairly accuses* you of cheating on an examination.

5. Someone has splashed *mud* over your new clothing.

6. Someone has *lost* an important *book* of yours.

7. Someone persistently *contradicts* you when you know you are right.

8. You have just found out that someone has told *lies* about you.

9. You are driving to a party and suddenly your car has a *flat tire*.

10. Someone makes an error and *blames* it on you.

11. You arrange to meet someone and he (she) does *not show* up.

12. You are reading a mystery and find that the last *page* of the book is *missing*.

13. You are trying to study and there is *incessant noise*.

14. You miss your *train* because the clerk has given you faulty information.

15. You are waiting at the *bus stop* and the bus fails to stop for you.

16. You are typing a term paper and your *typewriter* breaks.

17. You are in a *restaurant* and have been waiting a long time to be served.

18. You use your last 10c to call a friend and the operator *disconnects* you.

19. Someone has opened your personal *mail*.

20. You *wake up early* to get to a special 8 a.m. class and the instructor doesn't show up.

21. You are carrying a cup of *coffee* to the table and someone bumps into you.

22. You are very tired and have just gone to sleep, when you are *awakened* by the arrival of some friends.

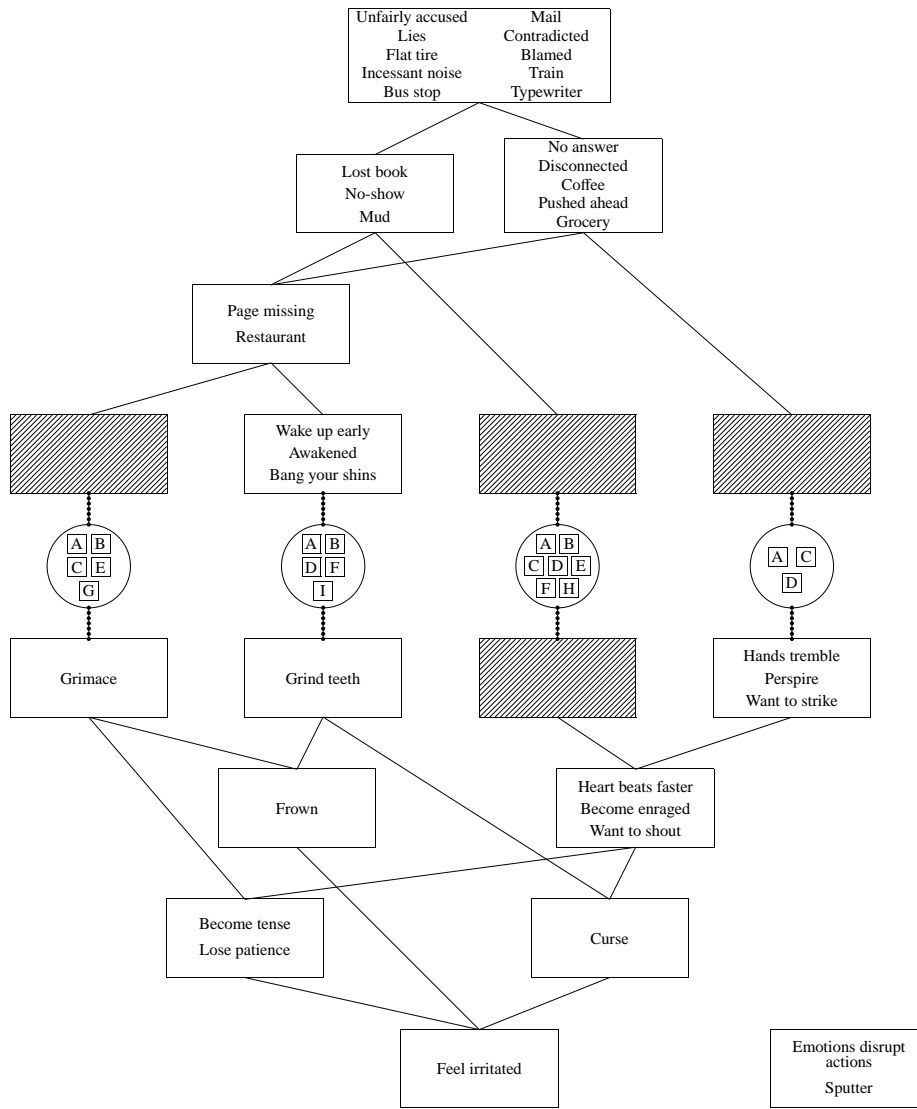23. Someone *pushes ahead* of you in a theaterticket line.

Figure 8. Graphical representation of the rank 4 INDCLAS model for the situation-response data of Application 1. Person classes in the circles are represented by capital letters. Person classes with a single element are omitted.

ranks, the difference-based scree test selects a rank 3 model, while the ratio-based variant selects a rank 4 model. In the proposed pseudo-statistical methods, the deterministic HICLAS model is considered within a probabilistic framework, assuming a stochastic process by which the cells of an underlying true binary array $\mathbf{T}$ may switch their values (either from 0 to 1 or vice versa) to give rise to the observed data $\mathbf{X}$. The rationale of Leenen & Van Mechelen's (2001) pseudo-binomial heuristic is then to select the lowest rank R for which the number of discrepancies in rank $R + 1$ is not significantly lower than the number of discrep-

ancies in rank R. Finally, the pseudo-AIC test proposed by Ceulemans & Van Mechelen (2005) is based on Akaike's (1973; see also Bozdogan, 1987) information criterion (AIC) and selects the model in rank R, that minimizes

$$-2\log_e L(\hat{\theta}) + 2T, \tag{3}$$

where $L(\hat{\theta})$ is the likelihood of the model given the data within the presumed probabilistic framework and T is the number of model parameters (i.e., the number of elements in the bundle matrices and, if applicable, the core array). When applying the pseudo-binomial and the pseudo-AIC test to the five fitted models, both point to the model in the highest rank (so, even a rank 6 model or higher may be considered most appropriate by these heuristics).

Considering, on the one hand, that most of the applied heuristics point to an INDCLAS model of rank 4 or higher, and, on the other hand, the difficulty in interpreting and representing higher-rank models (remind that in a rank 5 model, the Hasse diagram of a single mode may include up to 32 classes), we will retain the rank 4 model for further discussion. It is not unusual that the different rank selection heuristics yield inconsistent results. Anyhow, a decision on the rank of the model is preferably based not only on the outcome of one or more numerical procedures, but also on substantive considerations (interpretability). As an aside, we mention that Vansteelandt and Van Mechelen (1998) selected a rank 3 model based on an elbow ''discovered'' by visual inspection and substantive interpretability of the resulting solution.
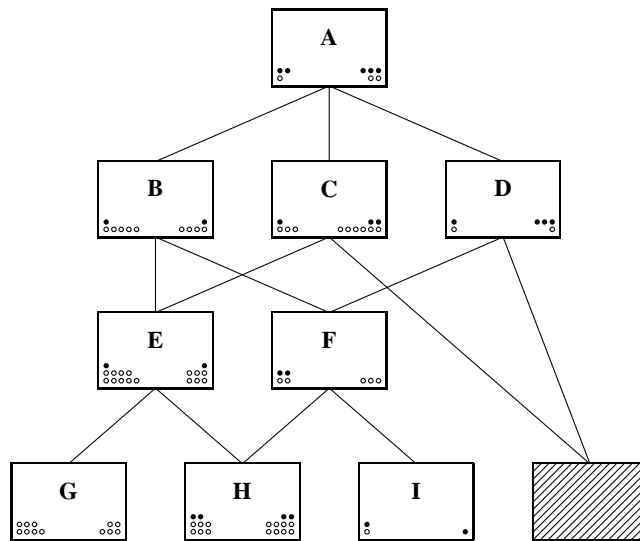

Figure 9. Graphical representation of the person typology in the INDCLAS and Tucker3-HICLAS models of Application 1. The number of white and black dots in the bottom left [right] corner of each box indicates how many women and men, respectively, belongs to that class in the INDCLAS [Tucker3-HICLAS] model. For simplicity's sake, the null class (with four females for each model) and single-person classes (four in INDCLAS and two in Tucker3-HICLAS) are omitted.

The rank 4 model is shown in Figures 8 and 9. From these figures, we see how the INDCLAS model meets the objectives aimed for by the original authors. A classification of the elements in each of the three modes is obtained, yielding a triple typology of persons, situations, and responses, where each of the typologies has an internal hierarchical structure. From Figure 8, we see that the 23 situations group into five types, where the type (class) at the top of the hierarchy include the situations that provoke the widest range of anger reactions. Indeed, from the definition of hierarchy it follows that if some situation provokes a particular anger reaction in a person, any situation that is hierarchically higher provokes that reaction in this person as well. The five situation classes almost form a complete order, with only at the third level two classes that are not hierarchically related. A complete order is indicative of a quantitative dimension, while classes that are not hierarchically related differ qualitatively (Gati & Tversky, 1982). The qualitative difference between the classes at level three might relate to the one at the right side (with situations ''no answer'', ''disconnected'', etc.) covering situations implying incidental harm, while the one at the left side (with the situations ''lost book'', ''no-show'', etc.) seems to include situations where the harm is a consequence of neglect with the cause not being present in the situation.

The 15 responses are spread across 9 classes, with, as a consequence, several classes containing only a single response. Nevertheless, a substantive meaning can be given based on the bundles shared by the same classes. That way, classes which load on the fourth bundle (i.e., classes at the right side of the hierarchy in Figure 8) are more related to fysiological and more automatic responses, while bundles I and II seem indicative of more voluntary responses. As an aside, note the response class at the right bottom of Figure 8, which is not related to any other class. This class is called the *null class* and contains the elements that are not related to any bundle (i.e., the element's row in the bundle matrix only contains zero's, hence its name); this implies that (discrepancies aside) the responses in this class are never displayed by any person in any situation. In a graphical representation of a HICLAS model, the null class is put aside or (most often) not displayed at all. When going up in the hierarchy of the responses, the classes include more general responses. For example, feeling irritated is common to all situations (in the sense that it is the ''first'' response displayed in any situation, i.e., if a situation provokes anger responses from a person, these responses necessarily include feeling irritated), while perspiring, trembling, and wanting to strike are specific to the most frustrating situations only.

The person typology has persons distributed across 14 classes (including the null class, which contains persons who, according to the model, never display any anger response). So as to avoid cluttering up the graphical representations, we omitted in Figures 8 and 9 four single-person classes. The person types higher up in the hierarchy generally display more anger responses. For an interpretation of the person typology, one may look for a relation between characteristics of the person and the types they belong to. As a simple example, we will examine the distribution of both sexes across the classes. In Figure 9, we have indicated the number of women (white dots) and men (black dots) at the left bottom of each box. At simple inspection, it seems that men are relatively higher up in the hierarchy than women, which would mean that men are more prone to displaying anger responses than women. This impression is confirmed by a significant

polychoric correlation between a person's level in the hierarchy and his/her sex (r = .39, with 95%-confidence interval [.06,.73]). More advanced (and far more interesting) comparisons between different types in the person typology may look for relations with a person's subjective appraisals, evaluative affective judgements, and/or his action tendencies (see Mischel & Shoda, 1995; Vansteelandt & Van Mechelen, 1998).

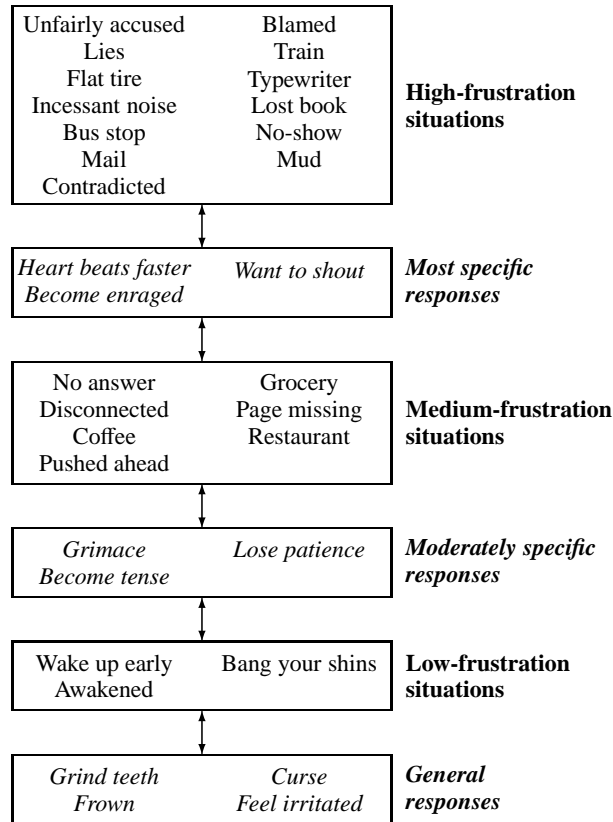|		
Unfairly accused Lies Flat tire Incessant noise Bus stop Mail Contradicted	Blamed Train Typewriter Lost book No-show Mud	**High-frustration situations**
Heart beats faster *Become enraged*	*Want to shout*	***Most specific responses***
No answer Disconnected Coffee Pushed ahead	Grocery Page missing Restaurant	**Medium-frustration situations**
Grimace *Become tense*	*Lose patience*	***Moderately specific responses***
Wake up early Awakened	Bang your shins	**Low-frustration situations**
Grind teeth *Frown*	*Curse* *Feel irritated*	***General responses***

Figure 10. The total order of situation types (roman font) and response types (italic font) for Person type B in the INDCLAS model for the data in Application 1.

The INDCLAS association rule connects the typologies of situations, responses, and persons together. This allows for deriving, for each person type, a set of if-then rules that characterize them in terms of the responses they display in different situations. As an example, consider person type B. From Figure 8, it follows that these persons differentiate between (i.e., display distinct behaviours in) three groups of situations, which may be labeled high-, medium-, and low-frustration. Indeed, as persons of person type B are not related to the fourth bundle, they do not differentiate between situations in the top class (with ''unfairly accused'', ''lies'', etc.) and in the level-three class with ''lost book'', ''no-show'', and ''mud'', and, as a result, all these situations are regrouped into a single class

when considering person type B separately. For the same reason, also the class with, amongst others, ''no answer'', ''disconnected'' and the class with ''page missing'' and ''restaurant'' are joined together for person type B. Along the same lines, the behaviors for these persons are regrouped in three response classes, from most general to most specific. Figure 10 shows how these situation and response types are related: For a given situation type, persons of person type B display all responses below the situation type. Hence, if such a person is in a low-frustration situation, then (s)he will display the general anger responses such as grind teeth, curse, and so on; if (s)he is in a medium-frustration situation, then (s)he will additionally grimace, become tense, and lose patience; finally, if (s)he finds him-/herself in a high-frustration situation, then (s)he will display the previous responses and the specific, fysiological/automatic responses. Such a set of if(situation)-then(response) rules for a given person (type) has been called his/her ''behavioral signature'' (Shoda et al., 1994), as it describes his/her particular way of behaving across situations of a given domain.

Tucker3-HICLAS Analysis and Comparison With INDCLAS

We applied the Tucker3-HICLAS algorithm to the same data to obtain models for all combinations of ranks (P, Q, R) with $P, Q, R \in \{1,...,5\}$. Obviously, here the rank selection problem is considerably larger than in INDCLAS although, as pointed out by Ceulemans et al. (2003), some models are overparametrized and need not be evaluated (with $P, Q, R \in \{1,...,5\}$, only 74 of the $5^3 = 125$ models need to be evaluated). We will apply the convex-hull based approach proposed by Ceulemans and Van Mechelen (2005) for model selection, which is an extension of the scree test and is based on a plot of the proportion of discrepancies by the sum of the number of bundles $P + Q + R$. In a first step, this method retains all models that are on the lower boundary of the convex hull for the set of points. As shown in Figure 11, this procedure applied to the 74 models for the current data set retains seven models. (Remind that the convex hull of a set S of points is the smallest convex set containing S; in this case, the set of points included in the polygon in Figure 11). A further selection of the retained models can then be
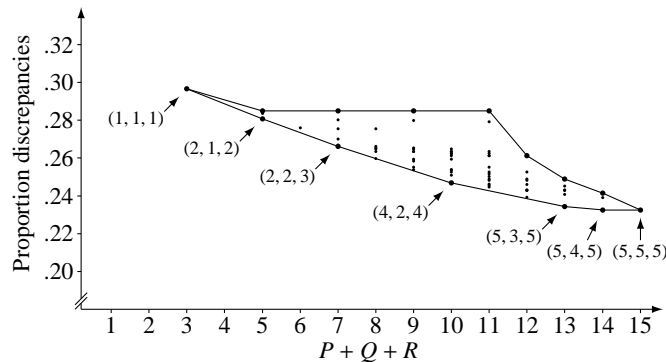

Figure 11. Plot of proportion of discrepancies by sum of the number of bundles (P + Q + R) for the Tucker3-HICLAS models fitted to the situation-response data of Application 1, with the ranks of the models on the lower boundary of the convex hull.

obtained by applying a scree test similar to the one we applied for model selection in INDCLAS. With this approach, model (5,3,5) and (4,2,4) are almost indifferent (with a slight preference for the more complex model). For simplicity's sake, we will retain the (4,2,4) model for further discussion. Note that the fit of the retained model (with a proportion of discrepancies equal to .247) is about the same as that of the INDCLAS model discussed in the previous section.
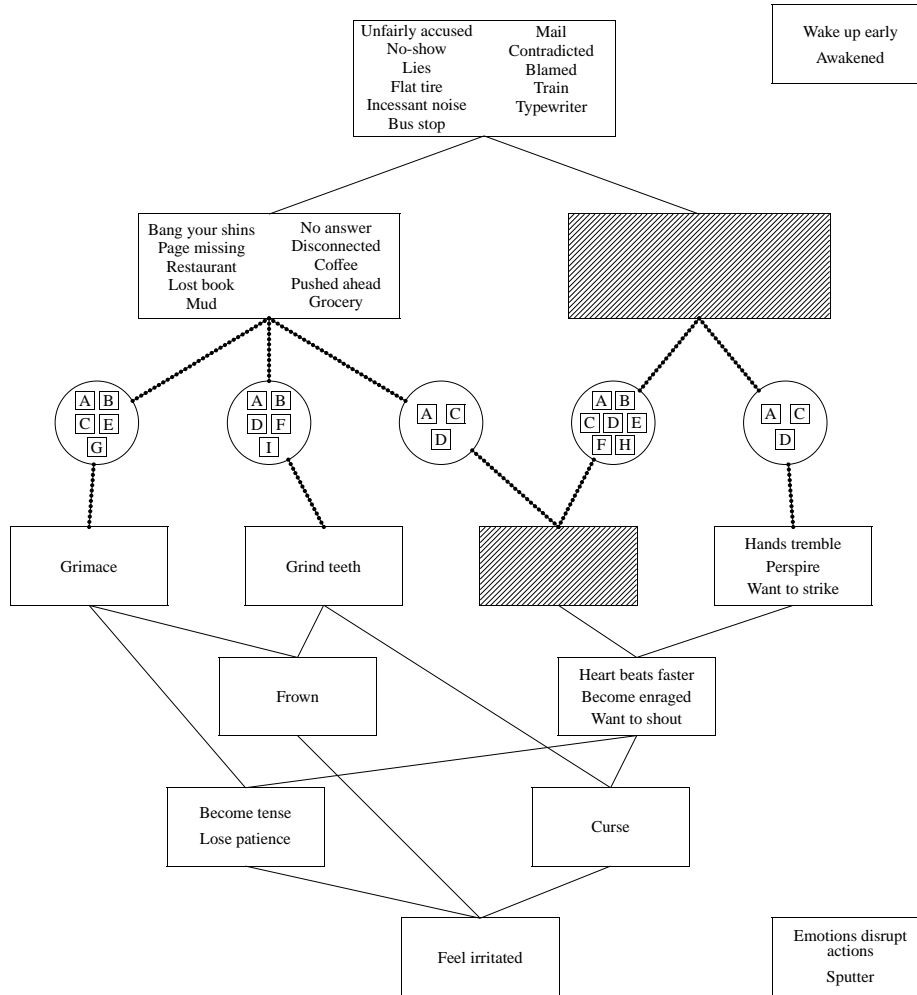

Figure 12. Graphical representation of the rank (4, 2, 4) Tucker3-HICLAS model for the situation-response data of Application 1. Person classes in the circles are represented by capital letters.

The graphical representation of the (4,2,4) model is shown in Figure 12 (with the hierarchical classification of the persons in Figure 9). Interestingly, the typology for the responses is identical to that in the INDCLAS model (Figure 8) and also the person typology, although not exactly identical, strongly resembles the typology in INDCLAS. This is not necessarily the case, even though both the

response and the person typology are based on four underlying bundles in the Tucker3-HICLAS as well as the INDCLAS model. With respect to the situations, the Tucker3-HICLAS model presents a strong simplification as compared to INDCLAS, with three situation classes that are totally ordered for all persons. As the model indicates that the two situations ''wake up early'' and ''awakened'' in the null class do not provoke any frustration in any of the persons and, therefore, can be left aside when interpreting the model, the situations that possibly provoke any anger reaction can be classified in a low-frustration type (''bang your shins'', ''no answer'', etc.) and a high-frustration type (''unfairly accused'', ''mail'', etc.).

The simplification in the situation typology is compensated for by a slightly more complex association rule, which no longer implies a one-to-one relation among the bundle-specific classes of the three typologies like in the INDCLAS model. However, the behavioral signatures of the respective person types are as easy to read from the graphical representation as in INDCLAS. For example, one might examine the if(situation)—then(response) rules of Person type B and find that situation and response classes can be totally ordered, similar to Figure 10. Thanks to the simplification in the situation typology, the retained Tucker3-HICLAS model then guarantees that situation and response types are totally ordered for *each* person type. With the response and person typology as well as the goodness of fit being practically identical for the INDCLAS and Tucker3-HICLAS model for these data, the simplification in the situation typology (and the resulting total order of situation and response types for each person) can be considered a net gain of the latter model.

Application 2: A Single-Case Study on Interpersonal Emotions

In our second application, we apply the INDCLAS and Tucker3-HICLAS models to three-way three-mode data obtained from a single (21-year old, female) individual (Geluykens, 2000), who was asked every 15 days over a five-month period (i.e., at 10 occasions) which of 40 emotions she experienced at that moment towards six target persons (mother, father, partner, and three relations that she characterized as unstable). This yielded a $10 \times 6 \times 40$ (occasion by target person by emotion) data array $\underline{\mathbf{X}}$, with $x_{ijk} = 1$ if, at occasion i, this young woman indicated that she experienced emotion k towards target person j, and $x_{ijk} = 0$ otherwise.

We fitted INDCLAS models with rank R ranging from 1 to 5 and Tucker3-HICLAS models with rank (P, Q, R) satisfying $P, Q, R \in \{1,...,5\}$ to these data. The dots in Figure 13 represent the proportions of discrepancies associated with the fitted Tucker3-HICLAS models in function of the sum of the number of bundles $P + Q + R$. For the sake of a first (rather crude) comparison, we have also represented, by means of squares, the proportions of discrepancies of the five fitted INDCLAS models, using $R + R + R$ as the corresponding complexity measure on the abscissa. It should however be noted that comparing the complexity of INDCLAS and Tucker3-HICLAS this way implies a disadvantage for the former model; as explained previously, a rank R INDCLAS model is a restricted rank (R, R, R) Tucker3-HICLAS model, yet the complexity measure used on the abscissa in Figure 13 returns equal values for both these models. Therefore, we

first derive the convex hull for the set of Tucker3-HICLAS models and then we compare the Tucker3-HICLAS models that are located on the lower boundary of the convex hull with the INDCLAS models in rank 1 to 5 by their pseudo-AIC value, as defined in Equation (3). The pseudo-AIC test, as it accounts for the additional number of parameters in the core array of the Tucker3-HICLAS model, implies a fair comparison of the latter model with the more restricted INDCLAS model.
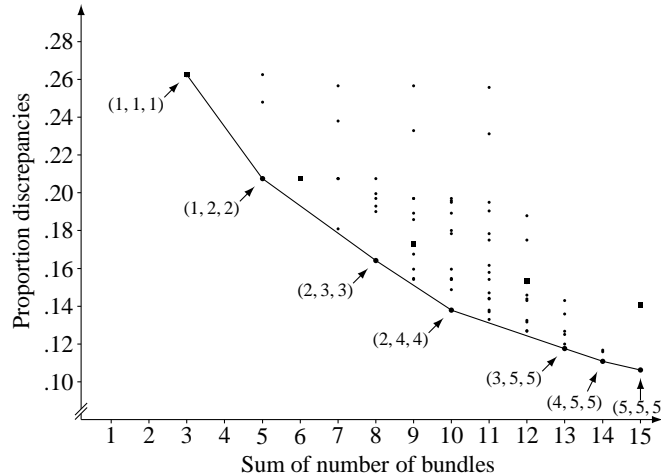

Figure 13. Plot of the proportion of discrepancies by the sum of the number of bundles for the INDCLAS (squares) and Tucker3-HICLAS (dots) models fitted to the interpersonal emotion data of Application 2, with the ranks of the (Tucker3-HICLAS) models on the lower boundary of the convex hull.

As shown in Figure 13, the lower bound of the convex hull contains seven Tucker3-HICLAS models. As an aside, one may note that applying a scree test (similar to the procedure applied in the previous application) would select the rank (2,4,4) model from this set. Including the five INDCLAS models to the set of seven Tucker3-HICLAS models on the convex hull yields the twelve models shown in Table 6. From this table, we further read that the rank (2,4,4) Tucker3-HICLAS model has the lowest pseudo-IAC value. Moreover, we see that the INDCLAS models have considerably worse fit than their Tucker3-HICLAS counterparts: The selected rank (2,4,4) Tucker3-HICLAS model decreases the number of discrepancies in the INDCLAS rank 4 model with about 10% (and fits even slightly better than the INDCLAS rank 5 model). Here, with the rank (2,4,4) Tucker3-HICLAS and the rank 4 INDCLAS model having about the same degree of parsimony, the significant improvement in goodness of fit can be considered a net gain of the former model. Given that Tucker3-HICLAS models for these data clearly fit better than the INDCLAS models, we limit further (substantive) discussion to the retained (2,4,4) Tucker3-HICLAS model.

Figure 14 displays the graphical representation of the selected model. We first discuss the hierarchical classification obtained for each of the three modes. The circles in the center of the figure show how the ten time points are grouped in three classes. Interestingly, each class contains a set of consecutive time points, as

Table 6. Discrepancies, Number of Parameters, and Pseudo-AIC Values for the INDCLAS models and a Selected Set of Tucker3-HICLAS Models Fitted to the Interpersonal Emotion Data of Application 2.

Model type	Rank	Number of discrepancies	Proportion discrepancies	Number of parameters	Pseudo-IAC value
INDCLAS	1	630	.263	56	2875.15
Tucker3-HICLAS	(1,1,1)	630	.263	57	2877.15
Tucker3-HICLAS	(1,2,2)	498	.208	106	2663.00
INDCLAS	2	498	.208	112	2675.00
INDCLAS	3	415	.173	168	2546.31
Tucker3-HICLAS	(2,3,3)	394	.164	176	2495.27
INDCLAS	4	368	.153	224	2504.55
Tucker3-HICLAS	(2,4,4)	331	.138	236	2397.58
INDCLAS	5	337	.140	280	2507.45
Tucker3-HICLAS	(3,5,5)	282	.118	335	2407.19
Tucker3-HICLAS	(4,5,5)	266	.111	370	2411.62
Tucker3-HICLAS	(5,5,5)	255	.106	405	2435.29

such subdividing the complete five-month period into three phases (where the three phases comprise occasions 1 to 6 [T1-T6], occasions 7 to 9 [T7-T9], and the single last occasion [T10], respectively). Note that the grouping in consecutive time points does not follow from a restriction in the model, but is important from a substantive point of view. The hierarchical relations among the three phases, which are not directly shown in Figure 14, are of less substantive interest and will not be discussed further.

With a grouping into five classes, we can hardly speak of a classification of the six target persons. We observe that the partner and one of the unstable relations classify together, which implies that the participant experienced the same emotions towards these two persons across the five-month period. Furthermore, we observe that the emotions this woman experienced towards her father are a combination of the emotions towards her partner (and Unstable relation A) on the one hand and her mother on the other hand. Surely, it is too far-fetched to adhere psycho-analytical interpretations about the father figure of this woman on this result only, but if such results were obtained from an analysis of a patient's data in a clinical setting, they may yield hypotheses to a psychiatrist, which then can be further investigated in one or more therapy sessions.

In the hierarchical classification of the 40 emotions, we see that all positive emotions are related to the first bundle (at the left side in Figure 14), with higher-up in the hierarchy more general positive emotions (e.g., friendly) and in the
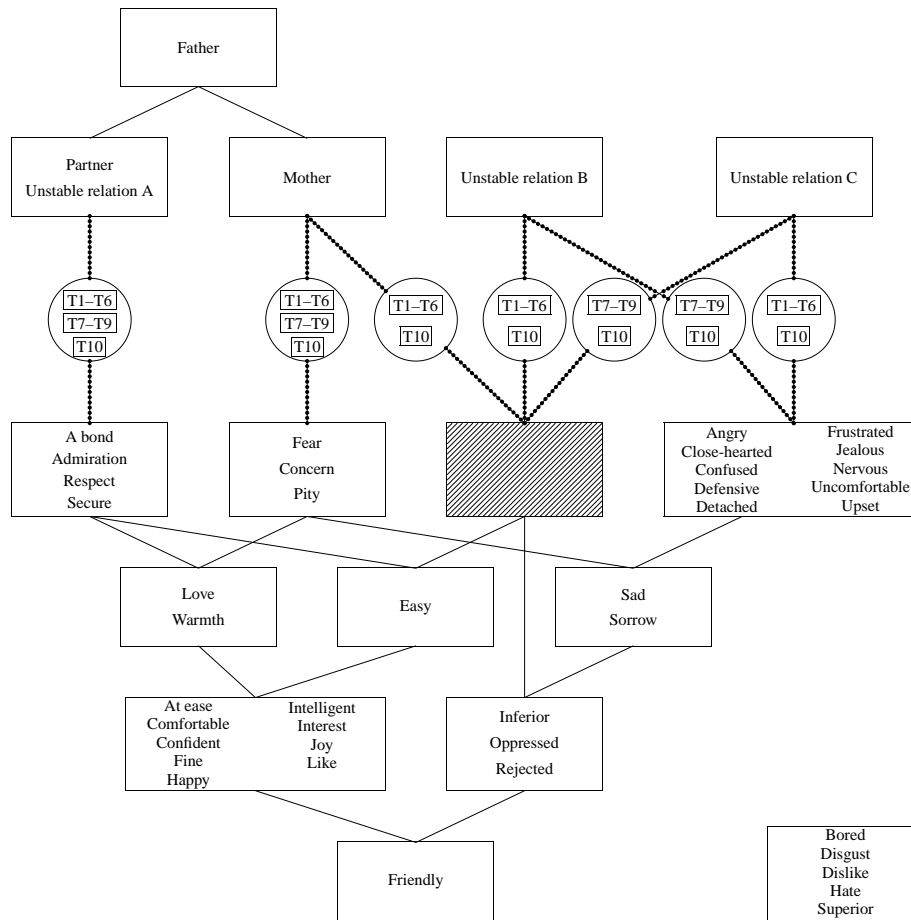
Figure 14. Graphical representation of the rank (2, 4, 4) Tucker3-HICLAS model for the interpersonal emotion data of Application 2.

bottom class strong positive feelings (e.g., admiration). Emotions that are not related to the first bundle are either negative (particularly those related to the fourth bundle at the right side of the graph) or indicate ambivalence (such as concern and pity, which are exclusively related to the second bundle). Note further the five extremely negative emotions in the null class, which are not experienced towards any of the target person at any moment.

We now proceed with a discussion on how the three modes link together, that is, on the emotional evolution towards the respective target persons across time. Because five of the six target persons are in different classes that are almost not hierarchically related to each other, we will discuss this evolution for each of them separately, drawing parallels between them whenever needed.

We can read from the graph that the partner and Unstable relation A are completely stable relations (at least in the period under study) as the emotions experienced towards them do not vary at all across occasions. (This means that the characterization of unstability of Relation A is due either to past experiences before the onset of the study or to a prospect of possible unstability in the future.)

Furthermore, we can derive that the participant of the study experiences these two relations as entirely positive, as they elicit the full range of positive feelings over the complete five-month period.

With respect to her parents, we derive that also the relation with her father is entirely stable, although in addition to the full set of positive emotions, he elicits some (mildly) negative and ambivalent feelings as well. The analysis further shows that she has a slightly more negative relation with her mother as compared to her father given that her mother elicits the same emotions as her father, except the positive emotions of feeling a bond, admiration, respect, and feeling secure (during the entire period) and feeling easy (during the second phase [T7-T9]). The latter emotion is otherwise the only one that changes across time (feeling easy towards her mother at the beginning and the end, but not intermediate), so that the relation with her mother can be characterized as ''almost stable''.

The Unstable relations B and C are truly unstable, indeed. Note first that neither of these relations elicit strong positive feelings (such as feeling a bond, admiration, love, warmth, etc.) at any occasion. But they do elicit mild positive feelings (feeling at ease, comfortable, easy, etc.), although, for Relation B, this is only the case at the beginning (or first phase, T1-T6) and the end (or third phase, occasion T10) of the studied period, and, for Relation C, from the second phase onwards (phases two and three, occasions T7-T10). Both relations also elicit rather strong negative emotions (angry, close-hearted, etc.): For Relation B, this occurs at the end (during phases two and three) and for Relation C at the start and at the last occasion (during phases one and three). From these results, we may conclude that Relation B turned from positive to negative, while Relation C went from negative to positive and at the very last occasion back to negative.

It would definitely be interesting to confirm or disconfirm the conclusions drawn from this analysis in a discussion of the results with the participant. Unfortunately, no details are available about the target persons. Nevertheless, this study shows that a Tucker3-HICLAS analysis of longitudinal data from a single person may provide insightful results and interesting hypotheses about evolutions in this person.

Concluding Remarks

In this paper, we have discussed the INDCLAS and Tucker3-HICLAS models, which both are three-way extensions of De Boeck and Rosenberg's (1988) original hierarchical classes model. We have shown the similarities and differences between both models from a substantive rather than from a pure mathematical point of view. We further illustrated how a Tucker3-HICLAS analysis may improve on the results from an INDCLAS analysis. This improvement may manifest itself in two possible ways, which, to be sure, are aspects of the same underlying principle. First, a Tucker3-HICLAS model possibly implies a simplification of the hierarchical classifications of one or more modes as compared to an INDCLAS model, without the simplification (substantially) affecting the goodness of fit. This was exemplified in the hypothetical example in the first section of this paper as well as in the analysis of the situation-response data in Application 1. Second, a Tucker3-HICLAS model of about the same complexity as an INDCLAS model (in terms of the number of parameters) may imply a sig-

nificant improvement in goodness of fit, which was illustrated in the application with interpersonal emotion data.

In the remainder of this paper, we want to draw the attention to some recent developments within the HICLAS family. In particular, we will discuss three that are especially interesting for practical users of INDCLAS and Tucker3-HICLAS as they may help users in obtaining robust models (i.e., models that can be expected to replicate across studies with a similar set up) and in providing sound substantive interpretations of these models. A first development relates to imposing restrictions on the model based on some a priori theory about the structure of the data, which is nowadays a common practice in factor analysis and is known as a confirmatory (as opposed to an exploratory) analysis. In the context of hierarchical classes analysis, Ceulemans et al. (2004) introduced various possible constraints and organized them in a taxonomy. They accordingly discussed adaptations to the existing algorithms so as to fit constrained models to observed data and provided tools so as to compare the goodness of fit of a constrained model to the corresponding unconstrained model. Such constraints are interesting both because of substantive reasons (a priori knowledge can be incorporated in the model) and for reasons of robustness of the fitted model (considering that in constrained models a smaller number of parameters has to be estimated). As an example, one may think of a confirmatory analysis using data similar to the hypothetical data in the first section of this paper, where (some of) the values of the symptom bundle matrix are fixed a priori based on a theory about which symptoms are characteristic of the hypothesized syndromes. Another example was already hinted at in the second application (on interpersonal emotions), in which one of the modes involved was a time mode and where we mentioned that an a priori constraint may be imposed so as to assure that the time classes are sets of consecutive moments.

When we discussed the hierarchical classification of the persons in the first application, we mentioned the possibility of relating external information to the obtained model and showed how traditional statistics (in this case, a polychoric correlation, but other statistics, e.g., based on analysis of variance) may be used to relate differences on a single covariate (in this case, sex) to differences in class membership. Sometimes, however, more than one covariate is available for the elements of a given mode, such that the external information may be organized in a two-way two-mode array (which then, as a consequence, has one mode in common with the data at hand). Suppose as an example that in the first application, in addition to the person by situation by response data, we have also acquired self-descriptions on a number of personality traits from each of the participants in the study; these external data can then be organized as a person by trait matrix (e.g., indicating whether or not the person assigns him-/herself the trait). For the case that this additional two-way two-mode array is binary, Wilderjans, Ceulemans, and Van Mechelen (2008) recently proposed a general framework, where the INDCLAS or Tucker3-HICLAS model for the person by situation by response data is estimated simultaneously with a two-way HICLAS model (such as De Boeck & Rosenberg's, 1988, model) for the person by personality trait data, with the restriction that the person bundle matrix in the three-way HICLAS model is identical to the person bundle matrix in the two-way HICLAS model. The advantages of such a simultaneous approach are twofold: (a) more robust estima-

tion of the common bundle matrix (in this case, the person bundle matrix) is obtained and (b) the two submodels are integrated in an overall model from which the connection among the different modes are readily understood. In the example at hand, for instance, the overall model sheds light on how the structure of individual differences in behavioral signatures relates to individual differences in personality traits.

Finally, a third recent development comprises Leenen et al.'s (2008) Bayesian extension of hierarchical classes analysis, which, if considered for the models discussed in this paper, leads to Bayesian INDCLAS or Tucker3-HICLAS models. Such models would explicitly include a parameter for the probability that some random error process changes a predicted value of one to an observed value of zero or vice versa, and, as such, account for the discrepancies between predictions ($\underline{\mathbf{X}}$) and data ($\underline{\mathbf{X}}$). Estimation techniques that are common in Bayesian analysis, such as Markov Chain Monte Carlo algorithms, are then employed to fit the model to the data. The most important advantages of the Bayesian approach include: (a) Rather than pinning the user down to just a single best-fitting model solution for a given data array (as the traditional INDCLAS and Tucker3-HICLAS algorithms do), the Bayesian algorithm aims at returning *all* good model solutions (with ''good'' refering to goodness of fit to the data). Subsequently, common patterns across this set of good models may be looked for, as such separating robust from incidental results. (b) Statistically sound tools for model selection (particularly, rank selection) and model testing become available; these tools allow for the testing of very specific model assumptions with tests driven by substantive considerations.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267-281). Budapest, Hungary: Academiai Kiado.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.

Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 174-243). Chicago: Rand McNally.

Ceulemans, E., & Van Mechelen, I. (2003). Uniqueness of *N*-way *N*-mode hierarchical classes models. *Journal of Mathematical Psychology,* 47, 259-264.

Ceulemans, E., & Van Mechelen, I. (2004). Tucker2 hierarchical classes analysis. *Psychometrika, 69*, 375-399.

Ceulemans, E., & Van Mechelen, I. (2005). Hierarchical classes models for three-way three-mode binary data: Interrelations and model selection. *Psychometrika, 70*, 461-480.

Ceulemans, E., Van Mechelen, I., & Kuppens, P. (2004). Adapting the formal to the substantive: Constrained Tucker3-HICLAS. *Journal of Classification, 21*, 19-50.

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2003). Tucker3 hierarchical classes analysis. *Psychometrika, 68*, 413-433.

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Davey, B. A., & Priestley, H. A. (2002). *Introduction to lattices and order* (Second ed.). Cambridge, UK: University Press.

De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika, 53*, 361-381.

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin, Germany: Springer.

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review, 97*, 201-224.

Gara, M. A. (1990). A set-theoretic model of person perception. *Multivariate Behavioral Research, 25*, 275-293.

Gara, M. A., & Rosenberg, S. (1979). The identification of persons as supersets and subsets in free-response personality descriptions. *Journal of Personality and Social Psychology, 37*, 2161-2170.

Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 325-340.

Geluykens, K. (2000). *Scripts over emoties binnen relaties* [Emotion scripts within relations]. Unpublished master's thesis, Katholieke Universiteit Leuven.

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics, 14*, 105-122.

Kuppens, P., Van Mechelen, I., Smits, D. J. M., De Boeck, P., & Ceulemans, E. (2007). Individual differences in patterns of appraisal and anger experience. *Cognition & Emotion, 21*, 689-713.

Leenen, I., & Van Mechelen, I. (2001). An evaluation of two algorithms for hierarchical classes analysis. *Journal of Classification, 18*, 57-80.

Leenen, I., Van Mechelen, I., & De Boeck, P. (1999). A generic disjunctive/ conjunctive decomposition model for *n*-ary relations. *Journal of Mathematical Psychology, 43*, 102-122.

Leenen, I., Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1999). INDCLAS: A three-way hierarchical classes model. *Psychometrika, 64*, 9-24.

Leenen, I., Van Mechelen, I., Gelman, A., & De Knop, S. (2008). Bayesian hierarchical classes analysis. *Psychometrika, 73*, 39-64.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*, 246-268.

Schepers, J., & Van Mechelen, I. (2006). *The real-valued model of hierarchical classes.* Manuscript submitted for publication.

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology, 67*, 674-687.

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 109-127). New York: Holt, Rinehart, & Winston.

Van Mechelen, I., & De Boeck, P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. *Journal of Social and Clinical Psychology, 8*, 276-287.

Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika, 60*, 505-521.

Van Mechelen, I., Lombardi, L., & Ceulemans, E. (2007). Hierarchical classes modeling of rating data. *Psychometrika, 72*, 475-488.

Van Mechelen, I., Rosenberg, S., & De Boeck, P. (1997). On hierarchies and hierarchical classes models. In B. Mirkin, F. R. McMorris, F. S. Roberts, & A. Rzhetsky (Eds.), *Mathematical hierarchies and biology* (pp. 291-298). Providence, RI: American Mathematical Society.

Van Mechelen, I., & Van Damme, G. (1994). A latent criteria model for choice data. *Acta Psychologica, 87*, 85-94.

Vansteelandt, K., & Van Mechelen, I. (1998). Individual differences in situation-behavior profiles: A triple typology model. *Journal of Personality and Social Psychology, 75*, 751-765.

Wilderjans, T., Ceulemans, E., & Van Mechelen, I. (2008). The CHIC model: A global model for coupled binary data. *Psychometrika, 73*, 729-751.

Author Note

Iwin Leenen is now at the Instituto Mexicano de Investigación de Familia y Población, Mexico; Eva Ceulemans is at the Department of Educational Sciences, K.U.Leuven.

Correspondence concerning this article should be addressed to Eva Ceulemans, Department of Educational Sciences, K.U.Leuven, Andreas Vesaliusstraat 2 - PO Box 3762, B-3000 Leuven, Belgium. E-mail: Eva.Ceulemans@ped.kuleuven.be.

Footnotes

1. We adhere to the notation proposed by Kiers (2000) in representing three-dimensional arrays by underlined bold uppercase symbols (while for matrices, the standard, nonunderlined bold uppercase symbol is used).

2. As we will explain in the next section, a model may be obtained by applying the INDCLAS or Tucker3-HICLAS algorithm to the three-way three-mode data. The Tucker3-HICLAS algorithm (which include the INDCLAS algorithm as a special case) is freely available from http://ppw.kuleuven.be/okp/t3h (upon filling out a form, you will be sent a link to a ZIP file with the program and some example data sets). The reader who is interested in the intricacies of the algorithm is referred to the original papers (Ceulemans et al., 2003; Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999).

3. Note that the mode of which the elements/classes appear in the circles is somehow selected arbitrarily. In most cases, the user will opt for the mode for which the hierarchical organization of its elements is of least interest.

4. The reader may wonder why we do not simply calculate the respective angles by standard geometry formulae. In this respect, note that the order among the angles is not invariant under linear transformations of the badness-of-fit statistic used (e.g., different results may be obtained from a plot with number of discrepancies instead of proportion of discrepancies on the ordinate). The scree tests proposed by Ceulemans and colleagues overcome this drawback.

ANALYZING STRUCTURAL RELATIONS IN MULTI-VARIATE DYADIC BINARY DATA

Jorge González
Francis Tuerlinckx
Paul De Boeck
K.U.Leuven, Belgium

ABSTRACT

In social network studies, most often only a single relation (or link) between the actors is investigated. When more than one link has been recorded, the two-way sociomatrix becomes a three-way array with the set of links being the third way. In this paper, we present a model which simultaneously accounts for the three ways in the data. Random effects are used to model the between-actor variability, both on senders and receivers side. In addition, structural relations between the linking variables are investigated. The model is applied to a study of popularity and strength in a class of students. It is shown that popularity can be seen as a linear function of strength on the receivers' side, but not on the senders' side.

INTRODUCTION

Dyadic data arise from the relationships of a pair of objects, called a *dyad*. Examples of objects are people, organizations, or countries, and examples of relationships are friendship, competition, or collaboration. In this paper we will use the term *actors* to refer to the objects, and *links* to refer to relationships. In the field of social networks (e.g., Wasserman & Faust, 1994), the set of actors and the links between them are called a network. Social network analysis aims at the study of the possible structures underlying the links between actors. The data are stored in a rectangular array in which the same group of actors is considered as *senders* in the rows and *receivers* in the columns. The entries of the matrix represent the links between actors. If a group of n actors is considered, then the link between sender i and receiver j is recorded in the variable X_{ij}. In the binary case, we have that $X_{ij} = 1$ if i has a link with j, and $X_{ij} = 0$ otherwise. The resulting $n \times n$ matrix is commonly known as *sociomatrix*. Figure 1 shows what the structure of the sociomatrix looks like.

Note that if $X_{ij} = 1$, this does not necessarily mean that $X_{ji} = 1$. For example, i may like j without j liking i. Note also that the main diagonal of the matrix is commonly not available (NA) because self-ties do not exist or are undefined.

Several characteristics of the data in the sociomatrix can be summarized using simple descriptive statistics called graph-theoretic statistics. For example, one
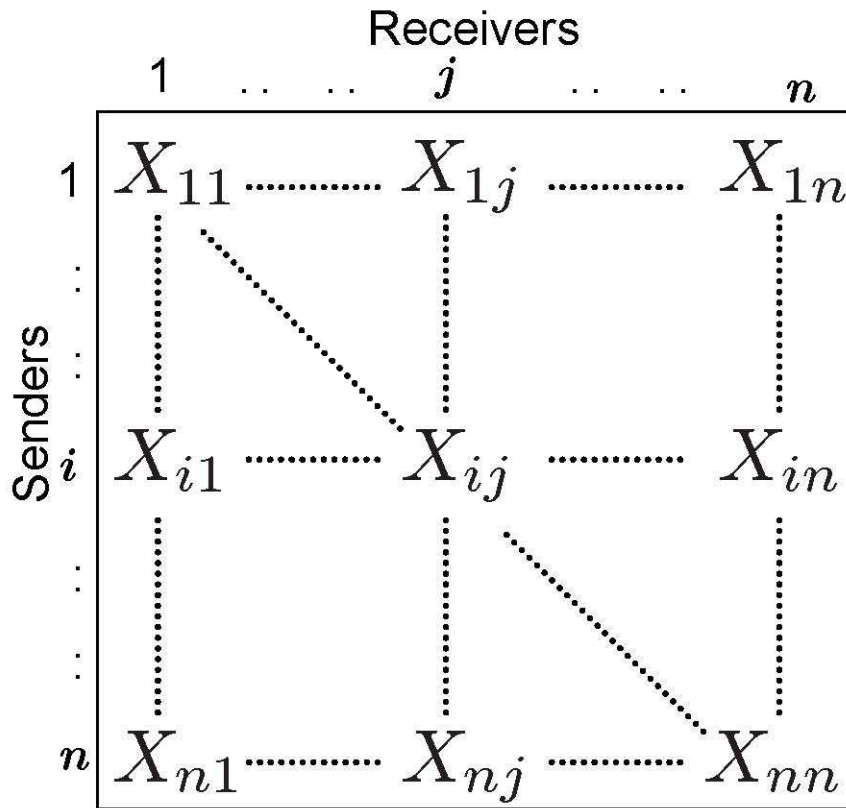
Figure 1. An example of a sociomatrix. The observation X_{ij} represents the link between sender i and receiver j.

may compute the total number of links $X_{++} = \sum_{i,j} X_{ij}$, or the number of mutual dyads (or reciprocated links) $M = \sum_{i<j} X_{ij} X_{ji}$. Another informative statistic is the number of relational links that actor i originates, called the outdegree $X_{i+} = \sum_{j=1}^{n} X_{ij}$ or vice versa, the indegree or degree of prestige (receptivity) $X_{+j} = \sum_{i=1}^{n} X_{ij}$, etc. For a complete description of these statistics and their interpretation see Wasserman and Faust (1994).

Besides these simple descriptive measures, more complex statistical models have been developed for the analysis of sociomatrices. Holland and Leinhardt (1981) made use of log-linear models and introduced the so-called p_1 family of models. These models, and all the subsequent extensions of it (e.g., Fienberg & Wasserman, 1981; Wasserman & Galaskiewicz, 1984; Wang & Wong, 1987), assume that the probability distribution of the random variable X_{ij} is a function of both the graph-theoretic statistics and parameters associated to them. However, these models were criticized because they assumed independence between actors' relationships. A more flexible class of models which take into account network dependence is the $p*$ class of models (e.g., Wasserman & Pattison, 1996; Anderson, Wasserman, & Crouch, 1999). A different approach is the use of random effects models (Van Duijn, Snijders, & Zijlstra, 2004; Hoff, 2003). The latter approach is also taken in this paper.

In the literature, most attention has been paid to univariate networks in which only a single link between actors is studied. Fienberg, Meyer, and Wasserman (1985); Pattison and Wasserman (1999); Gill and Swartz (2004); Zijlstra, Van Duijn, and Snijders (2006), are exceptions because they propose models for extensions to multivariate networks. However, it is very often the case that more than one link within a network is recorded. When more than one link is recorded, the data are stored in a three-dimensional array of dimension $n \times n \times r$ with r being the number of links recorded. An example is shown in Figure 2. A model accounting for the three-dimensions of the data array is then desirable.

A simple solution to handle the three-way data array would be the aggregation over the senders or the receivers. The resulting two-way data could be fitted with standard models of social networks (see references above). However, aggregation has severe disadvantages. After aggregation over the senders, one can only answer questions with respect to differences and correlations across receivers. Of course, to investigate differences and the correlational structure on the senders side, one can always aggregate over the receivers, but the net result is two unrelated models. For example, if strength and popularity are the two links under study, when aggregating over senders, the questions asked concern whether persons who are judged to be strong are also considered to be popular. When aggregating over receivers, one might ask whether persons who rate others as strong tend to rate them also as popular. The relations between the links (strength and popularity) in the two models can be similar or even identical but they don't need to be, neither on statistical grounds, nor on psychological grounds.

In this paper we follow the approach of random effects models to simultaneously take into account the variability that exist among senders and receivers for each link. Moreover, we would like to study the structure of the relations between the links, both on the sender and on the receiver side. This will lead to a so-called "double structure" model.

The double structure nature of the model is an interesting feature because the structures at the senders' and receivers' side do not necessarily have to be identical. In an example that will be elaborated in detail in the application section, senders rate receivers on two features: strength and popularity. It will be assumed that senders and receivers take separate (but possibly correlated) positions on a latent strength dimension and on a latent popularity dimension. In a next step, the latent popularity score is made a function of the latent strength score. Psychologically, this means that both the sender and the receiver have a contribution in the existence of the link. In other words stated, some senders tend to have more links than others based on their psychological characteristics and social position, and some receivers attract more links than others based on the same or other psychological characteristics and characteristics of their social position. It is hypothesized that among pupils, popularity is related to strength. Pupils who are considered strong would be considered also popular, whereas those who are considered weak are seen as not popular. The correlation between strength and popularity is a natural one over receivers because they are the persons who are rated. A correlation over senders would mean that raters who consider people in general as strong, would consider people in general also as popular. Such a correlation cannot reflect a reality but only a perception or opinion about others.
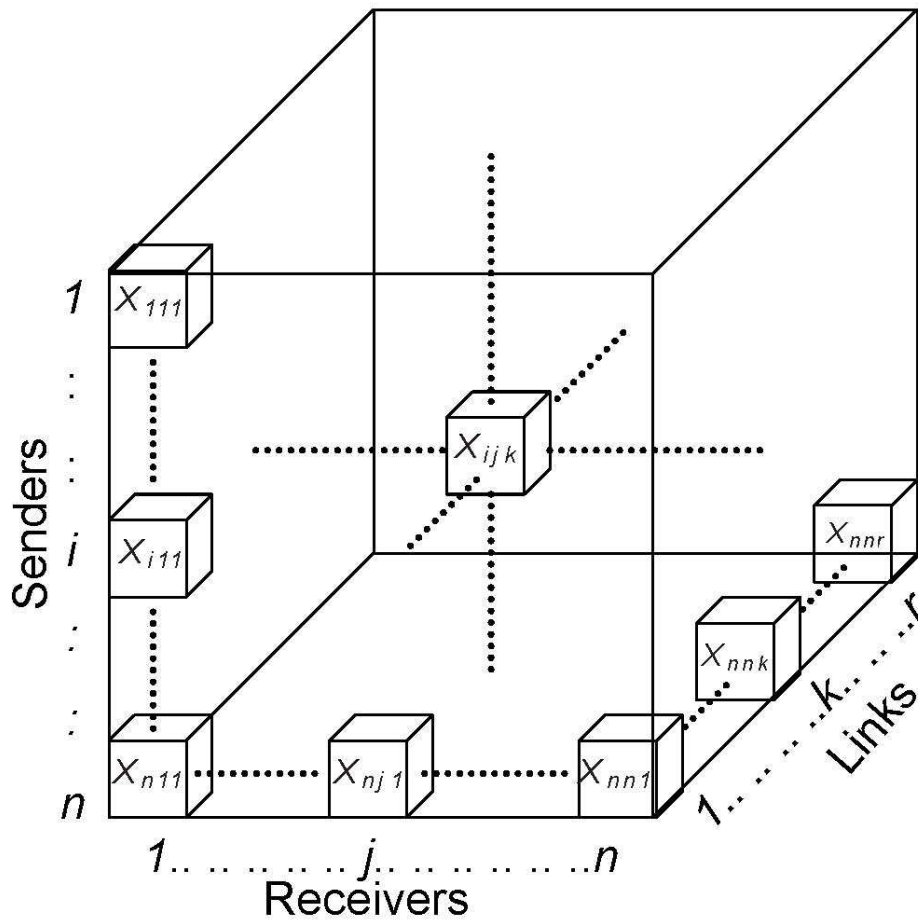
Figure 2. Three-way array of senders by receivers by links. Each slide of the cube is a sociomatrix. The observation X_{ijk} represents the tie between sender i and receiver j regarding link k.

For example, when the correlation is positive over receivers, it means that the receivers who are considered stronger are also considered more popular. One may assume that it reflects an external reality, possibly a collectively constructed reality in the social group under consideration, but more than just a purely internal conception of the world.

When the correlation is derived over the senders, then it reflects the relation between attributing strength and attributing popularity, independent of the receivers, because the correlated effects are sender effects that contribute independently of the receivers. Any such correlation is therefore probably based on mental conceptions or social attitudes from the part of the senders.

The models presented in this paper can disentangle the two kinds of contributions, separate the two correlations and compare them. If the correlations agree, then the correlation based on the receiver effects are suspicious, in that it may have been induced in an indirect or direct way by an implicit theory and not by an

objective reality. However, when the two do not agree, this is much likely, and the correlation of the receiver effects may then not be seen as rooted in people's implicit theory.

In the following, our model to analyze three-way dyadic data is introduced, and the estimation of the model following a Bayesian approach is explained. A study of popularity and strength in a class of students is presented. We finalize the paper with some final remarks and discussion.

The Model

The basic model we will apply here to analyze multivariate social network data is a special case of the double-structure structural equation model (2sSEM) as described in González, De Boeck, and Tuerlinckx (2008). González et al. (2008) focused on the analysis of a three-dimensional data array with binary data on persons, situations, and emotional responses, such that a simultaneous study of the individual differences and situational differences structure of the three-mode data set was possible. We will use the model here for the analysis of multivariate network data in which a sample of persons has been asked to rate each other on a set of personality items. The items can be thought of measuring latent variables. In addition, we consider structural relations between the links under study. The structure of individual differences between the raters in terms of the personality items is not necessarily equal to the structure of the individual differences of the rated persons in terms of the same personality items. As explained earlier, one could even argue that the equality of both structures is suspicious since then the former might be a mere reflection of beliefs from the part of the raters. Here we briefly introduce the model and how it can be estimated. For a more complete description the reader is referred to González et al. (2008).

The 2sSEM model is a generalized linear mixed model (GLMM) (e.g., McCullogh & Searle, 2001), with a linear predictor, a logit link function, a distribution from the exponential family and population distributions for the random effects. The probability of sender i having a link with receiver j is modeled as a function of sender's and receiver's effects and a general effect of the link under study. We account for differences in senders and receivers considering their effects as random effects.

Model for one link

The observed random variable X_{ij} is used to denote whether sender i has a link with receiver j. We have that $X_{ij} = 1$ if i has a link with j, and $X_{ij} = 0$ otherwise. Let π_{ij} be the probability of sender i having a link with receiver j, i.e., $\pi_{ij} = P(X_{ij} = 1)$. Then,

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \text{logit}(\pi_{ij}) = \eta_{ij} = \theta_i + \beta_j + \tau, \tag{1}$$

where θ_i and β_j represent an effect for sender i and for receiver j ($i,j = 1,....,n$), respectively, and τ is a general fixed effect of the link under study. It is assumed that $(\theta_i, \beta_i)^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$, a multivariate normal distribution with vector mean $\mathbf{0}$ and covariance matrix $\Sigma_{\theta\beta}$. Here, the random variable X_{ij} follows a Bernoulli distribution with parameter π_{ij}, η_{ij} is the linear predictor, and the logit is the link function.

Model for multiple links

When more than one link is considered, we add an additional index k ($k = 1,....,r$), in Equation (1). The model reads now as

$$\log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \text{logit}(\pi_{ijk}) = \eta_{ijk} = \theta_{ik} + \beta_{jk} + \tau_k, \tag{2}$$

where θ_{ik} refers to the score of sender i on the k-th link (latent trait), and, likewise, β_{jk} refers to the score of receiver j on the k-th link.

Suppose that in addition we are interested in the study of possible relations between the links under study. For example, one could hypothesize that link 2 is linearly related with link 1 and one would like to know whether this structural relation is the same for both, senders and receivers. The model equations would read as shown in (3)

$$\eta_{ij1} = \theta_{i1} + \beta_{j1} + \tau_1,$$
$$\eta_{ij2} = \theta_{i2} + \beta_{j2} + \tau_2, \tag{3}$$

where $\theta_{i2} = \omega\theta_{i1} + \theta_{i2}^*$ and $\beta_{j2} = \lambda\beta_{j1} + \beta_{j2}^*$. The parameters ω and λ are regression coefficients and θ_{i2}^*, and β_{j2}^* are error terms with distribution $(\theta_{i2}^*, \beta_{i2}^*)^T \sim N(\mathbf{0}, \Sigma_{\theta^*\beta^*})$. It is also assumed that $(\theta_i, \beta_i)^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$.

Estimation

We follow a Bayesian approach (e.g., Gelman, Carlin, Stern, & Rubin, 2003) to estimate the model. In the Bayesian approach, the uncertainty about a parameter vector ϑ is expressed in the prior distribution $p(\vartheta)$. Thus, prior distributions are to be assigned for all the parameters of interest in ϑ. Denoting the likelihood of the data $\mathbf{X}$, given ϑ, by $p(\mathbf{X} \mid \vartheta)$, the inference is then based on the joint posterior distribution $p(\vartheta \mid \mathbf{X})$, which after applying conditional distribution rules can be written as
$$p(\vartheta \mid \mathbf{X}) \propto p(\mathbf{X} \mid \vartheta)p(\vartheta). \tag{4}$$

It follows that the posterior distribution is proportional to the likelihood times the prior. Note that

$\vartheta = (\theta, \beta, \tau, \sigma_\theta^2, \sigma_\beta^2, \rho_{\theta\beta})$, for the model in Equation 1, and

$\vartheta = (\theta, \beta, \tau, \omega, \lambda, \sigma_\theta^2, \sigma_\beta^2, \rho_{\theta\beta}, \sigma_{\theta^*}^2, \sigma_{\beta^*}^2, \rho_{\theta^*\beta^*})$ for the model in Equation 3, and $\mathbf{X}$ are the data. The last elements of ϑ are the components of the matrices $\Sigma_{\theta\beta}$ and $\Sigma_{\theta^*\beta^*}$, respectively.

As prior distributions for τ, ω, and λ, we use diffuse normal distributions with zero mean and variance 10000. These prior distributions are diffuse in the sense of having a large variance that does not contain any prior information about the location of the parameter. For the covariance matrices, inverse Wishart priors with an identity scale matrix and 2 degrees of freedom were used. These prior distributions were chosen to represent little prior information.

After drawing a sample from the posterior distribution, the posterior sample mean and standard deviations can be used to summarize the distribution of each parameter of interest. Samples from the posterior distribution can be simulated using Markov Chain Monte Carlo (MCMC) (Gelman et al., 2003) techniques. We use WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003), a software program

for Bayesian analysis of complex statistical models using MCMC techniques. The WinBUGS code used to fit the models in this paper is shown in Appendix A.

Model checking

To assess the goodness of fit of the models, we use posterior predictive checks (PPC) (Gelman, Meng, & Stern, 1996). Using this technique, lack of fit can be assessed by the tail-area probability or *p*-value of a test quantity, such as a global fit statistic. A reasonable range for the *p*-value is considered to be the interval 0.05 and 0.95 (Gelman et al., 2003). In this context, a *p*-value near to 0.5 would indicate a good fit, whereas more extreme values are an indication of a poor fit.

The definition of a global goodness-of-fit test quantity for the model, and the description for computing Bayesian *p*-values are shown in Appendix B.

Application

Data

The data are from two classes of 13-14-year-old pupils in a secondary school in Belgium. The first class has 22 students whereas the second one has 21 students. Students were presented a questionnaire with 7 questions concerning each of their classmates. The questionnaire sheet consisted of a grid with different letters to identify the classmates in the rows, and with 7 questions in columns. We will focus our study on the first class and the following two questions: 1) How strong do you feel with regard to this classmate?, and 2) How popular do you think is this classmate? For the first question, the following 5-point scale was used: 0=very weak, 1=weak, 2=neutral, 3=strong, and 4=very strong. For the second question, the following 5-point scale was used: 0=not popular at all, 1=a little popular, 2=considerably popular, 3=very popular, and 4=extremely popular. In order to illustrate the binary data approach which is quite common for social networks, the original data were dichotomized recoding the values 0, 1, and 2 as 0, and 3 and 4 as 1 in question 1). In question 2) the values 1, 2, 3, and 4 were recoded as 1, and 0 remains 0 in order to obtain an optimal classification. The dichotomized data are shown in Appendix A. The basic question we ask with respect to these data is whether one's perceived popularity can be explained by one's perceived strength. We wonder whether also on the sender's side, the given ratings of popularity can be considered as a function of the given strength ratings.

For the strength link, the total number of ties (or ones in this case) is 97. The outdegree, i.e., the number of ties sent by each student varies considerably, ranging from 0 to 21 with a mean of 4.41 and standard deviation 6.25. The indegree, i.e., the number of ties received varies between 1 and 12 with a standard deviation 2.46.

For the popularity link, the total number of ties is 312, with outdegree varying between 6 and 21 with mean 14.18 and standard deviation 5.46. The indegree varies between 7 and 19 with standard deviation 3.91.

Models

We fit three different models to the data denoted as Model 1, Model 2, and Model 3. In Model 1 the logit of the probability of sender *i* having a strength tie with receiver *j* is modeled as a function of a sender parameter $\theta_{i,str}$ plus a receiver

parameter $\beta_{j,str}$ plus a parameter representing the effect of the strength link τ_{str}. The model equation reads as

$$\eta_{ij,str} = \theta_{i,str} + \beta_{j,str} + \tau_{str}, \tag{5}$$

where $(\theta_{i,str}, \beta_{i,str})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$.

Model 2 is exactly the same model as Model 1 but considers the popularity link. The model equation is written as

$$\eta_{ij,pop} = \theta_{i,pop} + \beta_{j,pop} + \tau_{pop} \tag{6}$$

where $(\theta_{i,pop}, \beta_{i,pop})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$. Note that Models 1 and 2 consider each only one link so that they are two separate models, one for strength and one for popularity.

In Model 3, on the other hand, both the strength and popularity links are modeled together and, moreover, popularity is made a linear function of strength. The model equation can be written as

$$\eta_{ij,str} = \theta_{i,str} + \beta_{j,str} + \tau_{str}$$
$$\eta_{ij,pop} = \underbrace{\omega\theta_{i,str} + \theta_{i,pop}^*}_{\theta_{i,pop}} + \underbrace{\lambda\beta_{j,str} + \beta_{j,pop}^*}_{\beta_{j,pop}} + \tau_{pop} \tag{7}$$

where θ, β, and τ are interpreted as explained before, ω and λ are regression sender and receiver coefficients, respectively, and θ^* and β^* are error terms. It is assumed that $(\theta_{i,str}, \beta_{i,str})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$, and $(\theta_{i,pop}^*, \beta_{i,pop}^*)^T \sim N(\mathbf{0}, \Sigma_{\theta^*\beta^*})$. Model 3 considers both links at the same time assuming that popularity is linearly related with strength at both senders and receivers side.

Models 1 and 2 are conceptually inferior to Model 3 in that analyzing multiple links simultaneously provides more general results compared to two analyzes of only one single link at a time. When analyzing multiple links at the same time, it is possible to investigate the relation between the links, and to compare the structure in question between senders and receivers. In despite of this, and only for explorative purposes, we fit a fourth model (Model 4) in which both links are allowed to correlate at sender's and receiver's side but there is no specific hypothesis about linear dependency between the links. Because the latent values can also correlate between senders and receivers this model actually assumes a four-variate normal distribution for the latent values. The model equation read as

$$\eta_{ij,str} = \theta_{i,str} + \beta_{j,str} + \tau_{str}, \tag{8}$$
$$\eta_{ij,pop} = \theta_{i,pop} + \beta_{j,pop} + \tau_{pop},$$

with $(\theta_{i,str}, \theta_{i,pop}, \beta_{i,str}, \beta_{i,pop})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$. The covariance matrix $\Sigma_{\theta\beta}$ is unstructured and all (co)variances should be estimated (in the results section we will report the values for the correlations). The variances are denoted as $\sigma_{\theta_{name}}^2$ or $\sigma_{\beta_{name}}^2$ where *name* can be equal to *str* or *pop*. Notation for correlations is very similar (e.g., the correlation between sender's strength and receiver's popularity is denoted as $\rho_{\theta_{str}\beta_{pop}}$). Model 4 can be considered an exploratory model that can be used to understand the data better.

Results

The posterior *p*-values for the four fitted models are 0.40, 0.55, 0.44, and 0.42, respectively. The *p*-value stems from the posterior predictive check (PPC) as

explained earlier (see also Appendix B), and refers to the absolute goodness of fit. The parameters estimates are presented in Table 1.

Table 1. Parameter estimates for the four fitted models. Posterior standard deviations appear between parentheses.

	Model 1	Model 2	Model 3	Model 4
$\sigma^2_{\theta_{str}}$	8.41 (4.63)		7.73* (3.69)	7.44* (3.45)
$\sigma^2_{\theta_{pop}}$		4.84* (2.45)		4.42* (2.11)
$\sigma^2_{\beta_{str}}$	1.50 (0.77)		1.26 (0.68)	1.38 (0.71)
$\sigma^2_{\beta_{pop}}$		1.51* (0.66)		1.50* (0.65)
$\rho_{\theta_{str},\theta_{pop}}$				0.04 (0.24)
$\rho_{\theta_{str},\beta_{str}}$	-0.37 (0.27)		-0.41 (0.25)	-0.41 (0.27)
$\rho_{\theta_{str},\beta_{pop}}$				0.41 (0.23)
$\rho_{\theta_{pop},\beta_{str}}$				0.03 (0.28)
$\rho_{\theta_{pop},\beta_{pop}}$		-0.44* (0.22)		-0.41 (0.23)
$\rho_{\beta_{str},\beta_{pop}}$				-0.66* (0.18)
τ_{str}	-2.57* (0.69)		-2.49* (0.61)	-2.48* (0.60)
τ_{pop}		1.45* (0.45)	1.56* (0.55)	1.42* (0.46)
ω			0.06 (0.22)	
λ			-0.92* (0.36)	
$\sigma^2_{\theta*}$			5.60 (2.94)	
$\sigma^2_{\beta*}$			0.84 (0.46)	
$\rho_{\theta*_{pop},\beta*_{pop}}$			-0.58* (0.22)	

Note. $\sigma^2_{\theta_l}$ = senders' variance for the l-th link, (l = str, pop); $\sigma^2_{\beta_l}$ = receivers' variance for the l-th link, (l = str, pop); $\rho_{\theta_{str},\theta_{pop}}$ = correlation between senders' strength and popularity link; $\rho_{\theta_l,\beta_{l'}}$ = correlation between senders' l-th link and receivers l'-th link, (l, l' = str, pop); $\rho_{\beta_{str},\beta_{pop}}$ = correlation between receivers' strength and popularity link; τ_l = effect of link l, (l = str, pop); ω = senders' regression coefficient; λ = receivers' regression coefficient; $\sigma^2_{\theta*}$ = residual senders' variance; $\sigma^2_{\beta*}$ = residual receivers' variance. Note. Model 1 = separate model for strength link; Model 2 = separate model for popularity link; Model 3 = joint model for strength and popularity links, with causal relations; Model 4 = joint model for strength and popularity links, without causal links but free correlations instead. Note. * = significant using the rule (estimate/posterior sd) > |2|.

From these results several conclusions can be drawn. First, the single link models each have a good fit to the data. The variability at the sender side is larger than the variability at the receiver side, both for strength ($\sigma^2_{\theta_{str}}=8.41 > \sigma^2_{\beta_{str}}=1.50$) and popularity ($\sigma^2_{\theta_{pop}}=4.84 > \sigma^2_{\beta_{pop}}=1.51$). For Model 1 the value τ_{str}=-2.57 can be interpreted as an indication of the probability of having a strength-tie when all the other effects are set to zero (on the logistic scale). In other words, for an average sender and an average receiver, the probability of having a strength-tie is exp(-2.57)/(1 + exp(-2.57))=0.07. The low value of this probability reflects the small number of ties presented in this link (97 out of 462) and means that most pupils don't feel stronger than their classmates. For Model 2, the probability of having a popularity-tie is exp(1.45)/(1 + exp(1.45))=0.81 and this higher value reflects the larger number of ties exhibited by this link (312 out of 462) meaning that most pupils find their classmates to be popular. Second, the multiple link models have both a good fit to the data, as may be derived from the PPC p-value. Note that the estimates of the parameters Model 3 has in common with Model 4 are similar for the two models; the results of common part are fairly stable regardless of the specific structure of the model. The most important parameters in Model 3 are the regression coefficients ω and λ. It is clear that the relation between strength and popularity is different for receivers and senders. The high and negative regression weight λ=-0.92 corresponds with a correlation $\rho_{\beta_{str},\beta_{pop}}$=-0.66 between $\beta_{j,str}$ and $\beta_{j,pop}$ (Model 4), and may be considered a reflection of a social reality, although the causal process is unclear. The regression weight for the senders is much smaller ω=0.06 and not significant, and the corresponding correlation is $\rho_{\theta_{str},\theta_{pop}}$=0.04 (Model 4). This is a clear example of how the structure of two modes of a three-mode data array can be quite different. In this particular case, these two modes are the same persons, in their roles of sender and receiver. This makes the finding all the more interesting. A similar correlation for the two would mean that there are reasons to believe that the receiver based correlation is influenced by biases in the senders which are reflected in the sender effects and their correlations. However, for the data set under consideration, the correlations are not similar at all.

Two other interesting findings relate to the correlation between sender effects and receiver effects. The strength one attributes to others as a sender is negatively correlated $\rho_{\theta_{str},\beta_{str}}$=-0.41 to the strength one is attributed as a receiver by the senders. This means that those who rate others as strong are being rated as not so strong themselves, but note that the correlation is not significant (the posterior standard deviation is 0.25). Further, when looking at the part of popularity which is not explained by strength, it can be noted that those who consider classmates as popular tend to be rated themselves as less popular $\rho_{\theta^*_{pop},\beta^*_{pop}}$=-0.58. This correlation is significant indeed (the posterior standard deviation is 0.22). It is an intriguing finding that seeing others as popular is negatively correlated with being seen by these others as popular. It actually confirms the mimetic desire theory of Girard (1961). The theory implies that being ''desired'' (being popular) increases one's value in the eyes of others, and that ''desiring'' (considering others as popular) diminishes one's value in the eyes of others, including those of the ''desired'' persons. In other words stated, ''desiring'' makes one less ''desired''.

Conclusion and discussion

In this paper, the analysis of multivariate dyadic binary data has been presented. We have shown that the 2sSEM (González et al., 2008) can be used to analyze three-way data in the form of senders by receivers by links.

Following a random effects approach, the 2sSEM simultaneously accounts for the variability among two of the three ways (senders and receivers) of the data, for each link and also permits studying the structure of the relations between the links, both on the sender and on the receiver side. We believe that this latter characteristic of the model is an extension of existing models for social networks because it allows for functional relations between the links.

In a study of strength and popularity, we have found evidence for a clear difference in the structure at the sender and receiver side which supports our theory that the structures at the senders' and receivers' side do not necessarily have to be identical. Of course with a data set containing more links, more elaborate models with complex structural relations could be fitted.

Finally, we focused on the case of binary dyadic data, but the model can be easily extended in several ways. First, ordered-category such as rating scale data, and continuous data may be modeled as well, and, second, the sets of senders and receivers don't need to be the identical. The fitting of models to study the evolution of links over time is also a possible extension.

Appendix A

WinBUGS code used to fit Model 1.

```
model;
{
for(i in 1:22){
for( j in 1:22) {
logit(pstr[i,j]) <- theta[i,1] + theta[j,2] + tau
xstr[i,j] ~ dbern(pstr[i,j])
}
theta[i,1:2] ~ dmnorm(mu.theta[],R[,])
}
mu.theta[1]<-0
mu.theta[2]<-0
R[1:2,1:2]~dwish(O[,],2)
O[1,1]<-1
O[2,1]<-0
O[1,2]<-0
O[2,2]<-1
tau~dnorm(0,0.0001)
SigmaT[1:2,1:2]<-inverse(R[1:2,1:2])
CorrT<-SigmaT[1,2]/sqrt(SigmaT[1,1]*SigmaT[2,2])
}
```

WinBUGS code used to fit Model 3.

```
model;
{
for(i in 1:22){
for( j in 1:22) {
logit(pstr[i,j]) <- theta[i,1] + theta[j,2] + tau[1]
xstr[i,j] ~ dbern(pstr[i,j])
logit(ppop[i,j]) <- w*theta[i,1]+err[i,1] + la*theta[j,2] + err[j,2] + tau[2]
xpop[i,j] ~ dbern(ppop[i,j])
}
theta[i,1:2] ~ dmnorm(mu.theta[],R[,])
err[i,1:2]~dmnorm(mu.err[],S[,])
}
mu.theta[1]<-0
mu.theta[2]<-0
R[1:2,1:2]~dwish(O[,],2)
O[1,1]<-1
O[2,1]<-0
```

```
        O[1,2]<-0
        O[2,2]<-1
        mu.err[1]<-0
        mu.err[2]<-0
        S[1:2,1:2]~dwish(P[,],2)
        P[1,1]<-1
        P[2,1]<-0
        P[1,2]<-0
        P[2,2]<-1
        for(k in 1:2){
        tau[k]~ dnorm(0,0.0001)
        }
        w~ dnorm(0,0.0001)
        la~ dnorm(0,0.0001)
        SigmaT[1:2,1:2]<-inverse(R[1:2,1:2])
        CorrT<-SigmaT[1,2]/sqrt(SigmaT[1,1]*SigmaT[2,2])
        SigmaerrT[1:2,1:2]<-inverse(S[1:2,1:2])
        CorrerrT<-SigmaerrT[1,2]/sqrt(SigmaerrT[1,1]*SigmaerrT[2,2])
        }
data1
xstr[,1] xstr[,2] xstr[,3] xstr[,4] xstr[,5] xstr[,6] xstr[,7] xstr[,8] xstr[,9] xstr[,10]
xstr[,11] xstr[,12] xstr[,13] xstr[,14] xstr[,15] xstr[,16] xstr[,17] xstr[,18]
xstr[,19] xstr[,20] xstr[,21] xstr[,22]
NA 1 0 1 1 1 0 1 1 1 0 1 1 1 1 1 0 1 0 1 1 1
0 NA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 NA 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 NA 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 NA 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0
0 1 1 1 1 NA 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1
0 0 0 0 0 0 NA 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 NA 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 NA 0 0 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 NA 0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 1 0 0 NA 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1 0 0 0 0 NA 0 0 0 0 0 0 0 0
0 0 0 1 0 0 1 1 0 0 0 0 0 NA 0 0 0 0 0 0 0
0 0 0 1 0 0 0 1 0 0 1 1 0 0 NA 0 0 0 0 0 0
0 1 0 0 0 0 1 1 0 0 1 0 0 0 0 NA 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0 0
```

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA 0 0
0 1 1 1 1 0 1 1 1 0 0 0 1 1 0 0 0 1 0 1 NA 0
0 1 0 1 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 NA
END
```

data2
```
xpop[,1] xpop[,2] xpop[,3] xpop[,4] xpop[,5] xpop[,6] xpop[,7] xpop[,8]
xpop[,9] xpop[,10] xpop[,11] xpop[,12] xpop[,13] xpop[,14] xpop[,15]
xpop[,16] xpop[,17] xpop[,18] xpop[,19] xpop[,20] xpop[,21] xpop[,22]
NA 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 1 0 0 1 0
0 NA 0 0 1 0 0 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1
1 1 NA 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 1 1 NA 1 1 1 0 0 1 1 0 1 1 1 1 1 0 1 1 1 1
1 0 1 0 NA 1 1 0 1 1 1 0 1 1 1 1 1 0 1 1 1 0
1 0 1 0 1 NA 1 0 0 1 1 0 1 1 0 0 1 0 1 0 1 0
0 0 0 0 0 1 NA 1 0 0 1 0 0 1 0 0 1 0 1 0 0 0
0 0 1 1 1 1 1 NA 1 1 1 1 1 1 1 1 0 1 0 1 1
1 1 1 0 1 1 1 0 NA 1 1 0 1 1 1 1 1 0 1 1 1 1
0 1 1 1 1 1 0 1 1 NA 1 0 0 1 1 1 0 1 1 0 1 1
0 0 0 0 0 0 1 1 0 0 0 NA 0 1 1 0 0 1 0 1 0 1 0
1 1 1 1 1 1 1 1 1 1 1 NA 1 1 1 1 1 1 1 1 1 1
0 0 1 0 1 1 1 0 1 0 0 0 NA 1 1 1 0 0 0 0 0 0
1 0 1 0 1 1 0 0 0 0 0 0 0 NA 0 1 1 0 1 0 1 0
1 1 1 0 1 1 1 0 1 1 1 0 1 1 NA 1 1 0 1 1 1 1
1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 NA 1 1 1 1 1 1
1 0 1 0 0 0 1 1 0 0 1 0 0 1 0 1 NA 0 1 0 1 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 NA 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 NA 1 1 1
1 0 1 0 1 1 1 0 0 0 1 0 1 1 0 1 1 0 1 NA 1 0
1 0 1 0 1 0 0 1 0 1 0 0 1 0 0 1 1 0 0 1 NA 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 NA
END
```

Appendix B

Posterior predictive p-values

To measure the discrepancy between some features of the model and data, a test quantity $T(\mathbf{x},\vartheta)$ which is calculated on the data and may depend on the model parameters, is used. Then, lack of fit can be assessed by the tail-area probability or *p*-value of the test quantity. Note that the PPC procedure allows one to choose any test quantity which can be used to check a particular relevant aspect of the model.

We use the χ^2 discrepancy measure which assesses the model fit in a global way (Gelman et al., 2003). In the context of the application, the Pearson χ^2 measure is defined as

$$\chi^2(x,\vartheta) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{(x_{ijk} - E(x_{ijk}|\vartheta))^2}{\mathrm{Var}(x_{ijk}|\vartheta)} \tag{9}$$

where ϑ contains all the parameters of interest. In practice, given a discrepancy measure $T(\mathbf{x},\vartheta)$, the posterior predictive check *p*-value is calculated as follows:

1. Draw a vector $\vartheta^{(k)}$ from the posterior distribution.
2. Simulate a replicated data set $\mathbf{x}_{rep}^{(k)}$ from $f(\mathbf{x}|\vartheta^{(k)})$.
3. Calculate $T(\mathbf{x},\vartheta^{(k)})$ and $T(\mathbf{x}_{rep}^{(k)},\vartheta^{(k)})$.
4. Repeat 1 to 3 *K* times.
5. Count the proportion of replicated data sets $\mathbf{x}_{rep}^{(k)}$ for which $T(\mathbf{x}_{rep}^{(k)},\vartheta^{(k)}) \geq T(\mathbf{x},\vartheta^{(k)})$.

The proportion obtained in 5 is the posterior *p*-value.

REFERENCES

Anderson, C., Wasserman, S., & Crouch, B. (1999). A *p** primer: logit models for social networks. *Social Networks, 21*, 37-66.

Fienberg, S., Meyer, M., & Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association, 80*, 51-67.

Fienberg, S., & Wasserman, S. (1981). Categorical data analysis of single sociometric relations. *Sociological methodology, 12*, 156-192.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). Bayesian data analysis (2nd ed.). London: Chapman and Hall.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 4*, 733-807.

Gill, P., & Swartz, T. (2004). Bayesian analysis of directed graphs data with applications to social networks. *Journal of the Royal Statistical Society, Series C, 53*, 249-260.

Girard, R. (1961). Mensogne romantique et vérité romantique. Paris: Bernard Grasset.

González, J., De Boeck, P., & Tuerlinckx, F. (2008). A double-structure structural equation model for three-mode data. *Psychological Methods, 13*, 337-353.

Hoff, P. (2003). Random effects models for network data. In R. Breiger, K. Carley, & P. Pattison (Eds.), Dynamic social network modeling and analysis: Workshop summary and papers (pp. 303-312). Washington, DC, The National Academies Press.

Holland, P., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association, 77*, 33-50.

McCullogh, C., & Searle, S. (2001). Generalized, linear, and mixed models. New York: John Wiley and Sons.

Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology, 52*, 169-193.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS version 1.4 user manual [Computer software manual]. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf.

Van Duijn, M., Snijders, T., & Zijlstra, B. (2004). p_2: a random effects model with covariates for directed graphs. *Statistica Neerlandica, 58*, 234-254.

Wang, Y., & Wong, G. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association, 82*, 8-19.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. New York: Cambridge University Press.

Wasserman, S., & Galaskiewicz, J. (1984). Some generalizations of p_1: External constraints, interactions, and non-binary relations. *Social Networks, 6*, 177-192.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov random graphs and p^*. *Psychometrika, 60*, 401-426.

Zijlstra, B., Van Duijn, M., & Snijders, T. (2006). The multilevel p_2 model: A random effects model for the analysis of multiple social networks. *Methodology, 2*, 42-47.

Author Note

Jorge González, Francis Tuerlinckx, Paul De Boeck, Department of Psychology, K.U.Leuven, Belgium. The authors have been partly supported by the Fund for Scientific Research⸺Flanders (F.W.O.) Grant no. G.0148.04 and by the K.U. Leuven Research Council Grant no. GOA/2005/04.

We thank Els Brusten and Iven Van Mechelen from the Research Group of Quantitative Psychology and Individual Differences, K.U.Leuven, who kindly provided us the data.

Correspondence should be addressed to Jorge González, who is now at the Centro de Medición MIDE UC, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile. Tel. +56 2 3541730, e-mail: jagonzal@uc.cl

GENERALIZED PROCRUSTES ANALYSIS:
A TOOL FOR
EXPLORING AGGREGATES AND PERSONS

James W. Grice
Oklahoma State University

Kimberly K. Assad
Michigan State University

ABSTRACT

Gower (1975) introduced *Generalized Procrustes Analysis* (GPA) as a multivariate statistical technique for analyzing three-dimensional data matrices. The current paper presents a non-technical introduction to the logic underlying GPA and then presents a completely worked example using genuine data. Specifically, self and peer ratings obtained from students attending a Summer Science Academy are analyzed and discussed. It is shown that GPA offers a powerful set of tools for exploring data at both the aggregate and individual level. A number of issues regarding the current analysis methods are also discussed.

INTRODUCTION

Imagine a female student who rates herself and nine other people whom she knows personally on 30 adjective trait terms (e.g., friendly, outgoing, agreeable) using a 5-point Likert-type scale. A personality psychologist subsequently records her responses in a 30 x 10 2-dimensional matrix:

Person #1

The psychologist next recruits more people to participate in the study, obtaining 49 additional 30 x 10 matrices:

Person #50

5	1		3
2	2		3
			2

Person #2

3	3		2
2	3		4
			5
			1
			2

Person #1

	1	2	...	10
1	1	5		4
2	5	2		1
3	4	1		5
.				
30	4	3		2

Trait Terms (vertical axis) · Rated individuals (horizontal axis)

As shown the data can be represented in a 3-dimensional array comprised of trait terms, rated individuals, and participants. The goal of the study may be to identify common patterns in how the participants organize the 30 trait terms. How would the psychologist analyze this wealth of information; namely, 15,000 data points?

Now imagine a marketing researcher who asks a male participant to describe his favorite breakfast cereals using his own terms (e.g., crunchy, sweet, grows stale quickly). The researcher then constructs rating scales from the terms and asks the man to rate 10 popular breakfast cereals and the 'ideal breakfast cereal' on his personal scales. The researcher repeats this process with other participants, allowing each to use his or her own set of adjectives to describe the cereals. Moreover, the researcher does not require the participants to use the same number of adjectives; hence some participants provide three or four adjectives while others freely elicit twice as many terms to describe the cereals. Thirty-five people participate in the study, yielding data that may be represented as follows:

Person #35

Wholesome	5	1		3
Sweet	2	2		3
Healthy	5	3		2

Person #2

Expensive	3	3		2
Bitter	2	3		4
Flavorful	4	1		5
				1

Person #1

Descriptions	1	2	...	10
Crunchy	4	5		4
Sweet	5	2		1
Stale	1	1		5
Colorful	5	2		

Rated Cereals

Unlike the personality psychologist, the market researcher is confronted with data that are matched on one dimension (cereals) but not matched on the other (the personal descriptors that differ across participants). Nonetheless, the goal of the researcher is to determine which breakfast cereals are alike and most similar to the ideal breakfast cereal. How is the researcher to reach this goal?

Gower's (1975) approach to the type of data gathered by the personality and market researchers above involved an inventive combination of Procrustes rotation and Analysis of Variance (ANOVA), which he referred to as *Generalized Procrustes Analysis*. The essence of this analysis can best be described using a simple data set:

		Cereal A	Cereal B	Cereal C	Ideal Cereal
Joe	*Sweet*	-3	-2	2	3
	Crunchy	1	2	2	2
Mary	*Expensive*	-2	-3	-2	-1
	Healthy	-2	-1	2	2

Joe and Mary elicited their own terms to describe breakfast cereals, and then they rated three cereals and the ''ideal cereal'' on 7-point Likert-type scales (scored -3 to +3) constructed from their unique descriptors. The ratings for Joe and Mary can be plotted separately, as shown in Figure 1. For these data the four cereals are matched and the two dimensions are not matched. The goal of Generalized Procrustes Analysis (GPA) is to maximally align the multivariate configurations of the matched figures in the two spaces. Examination of Figure 1 clearly reveals that both Joe and Mary viewed cereals A and B as highly similar to each other and distinct from cereal C and the ideal cereal. These latter cereals, C and the ideal, are themselves rated as highly similar. The key to GPA is to treat the unmatched dimensions as essentially arbitrary, and then rotate the two spaces to maximum similarity using Procrustes rotation (Cliff, 1966; Schönemann, 1966). In this example, rotating the axes in Joe's graph approximately -90° would align the four cereals very closely with the cereals in Mary's 2-dimensional space.

The rotated alignment, however, between the two sets of ratings would not be perfect. The next major feature of GPA thus turns toward quantifying the degree of similarity between the two rotated multivariate configurations of cereals. Gower (1975) accomplished this goal by computing a *consensus matrix*, which is simply an average of the rotated matrices of ratings. With this matrix of means in hand, the techniques underlying ANOVA can then be used to partition the total amount of variability in the rotated ratings into two unique portions: *consensus* and *residual*. If Joe's and Mary's rotated ratings match the consensus matrix closely, the lion's share of the variance will be allocated to the consensus term. If the rotated ratings are quite discrepant, then the residual variance term will be relatively large. In order to quantify the magnitude of effect—that is, agreement among raters regarding the multivariate configuration of matched figures—a *consensus proportion* can furthermore be computed as the ratio of the consensus
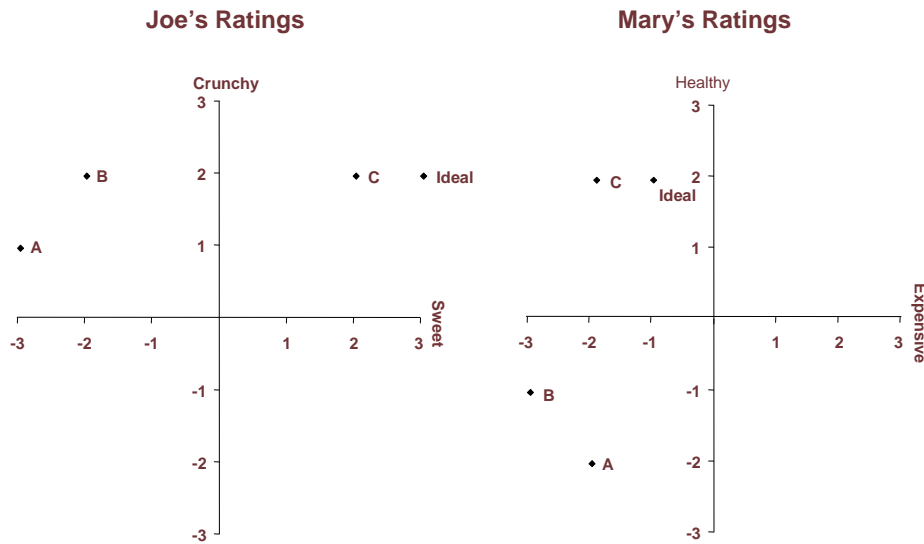
Figure 1. Bi-plots of two individuals' ratings of four cereals.

sum of squares to the total sum of squares in the ANOVA. The consensus proportion is thus kin to the traditional multiple R^2 statistic in regression, which ranges in value from 0 to 1 and indicates the proportion of overlap among the rotated ratings. It can be tested for statistical significance most efficiently using a randomization test (Wakeling, Raats, & Halliday, 1992). The result for Joe and Mary was equal to .99 ($p < .001$), indicating near-perfect similarity in their multivariate configurations of the cereals, as can be clearly seen in Figure 1. In addition to this useful summary statistic, a number of plotting routines can also be used to visualize the configuration of the matched figures in the consensus (average) matrix (see Gabriel, 1971; Gower & Dijksterhuis, 1994; Slater, 1977). These techniques typically revolve round Principal Components Analysis or similar multivariate statistical techniques.

In addition to these analyses and statistics that are based on aggregated data, what is perhaps most impressive about GPA is the ability to focus on the persons in the analysis. This is typically done in two different ways. First, each person's complete matrix of rating values can be related to the consensus matrix, and the degree of similarity between the two quantified. Particular individuals who deviate greatly from the consensus can thus easily be identified. Moreover, the ANOVA procedures discussed above can be refined so that specific points of discrepancies among the raters can be identified. For instance, if Joe and Mary disagreed primarily in their views of 'cereal A', the refined ANOVA results would help the investigator identify this difference in the data. Second, the general and specific points of departure from the consensus matrix can be examined using powerful graphing routines. The multivariate configuration of a particular person's matrix of ratings can be plotted and compared to the consensus configuration to visually examine points of disparity. Additionally, all of the persons' particular dimensions can be plotted simultaneously with the consensus configuration, even if the rating scales are not matched as with Joe (''sweet'', ''crunchy'') and Mary (''expensive'', ''healthy'') above. These plotting tech-

niques again revolve around Principal Components Analysis and another multivariate technique referred to as extension analysis (see Grice, 2007).

In order to demonstrate the aggregate and person-centered features of GPA, we present data below collected from high school students attending a 1-week Summer Science Academy at Oklahoma State University. These students shared a common living space and participated in numerous activities over the course of the academy. At the end of the week we asked the students to rate themselves and their peers on scales constructed from their own personal constructs and scales constructed from adjectives for the Big Five personality traits. The data were therefore similar to the hypothetical studies described at the beginning of this manuscript. Did the high school students reach a consensus with regard to themselves and their peers on the Big Five traits? Did the average, multivariate configuration of the students' views of each other derived from the personal constructs match the configuration derived from the Big Five traits? Were particular students discrepant in their views of themselves and their peers compared to everyone else's views? In the analysis below, these questions will be addressed, and a number of the specific details regarding GPA will be introduced and discussed.

Method

Participants

Twenty-five high school students participated in the Summer Science Academy at Oklahoma State University, a program designed for Oklahoma high school students with the intention of developing interest in psychological research. Of these students, twenty (5 males, 15 females) voluntarily participated in this study after having their parents or guardians sign a consent form. Participants ranged in age from 13 to 16 years ($M = 14.55$, $Mdn = 14.00$, $SD = 1.10$) and were obtained from across Oklahoma. Six participants were Caucasian, 4 African American, 7 Asian, 2 Native American, and 1 was Hispanic. All participants were tested individually on computers, and did not take more than 2 hours to complete the experiment. One participant, however, failed to complete all of the procedures appropriately, and her data were consequently removed from the analyses.

Procedures

Personal Construct grid. Each participant was initially handed a sheet containing both labeled pictures of 21 students in the Summer Science Academy, and 4 spaces to induce cues of the 4 students who did not have their pictures taken, to insure the participants remembered the other students' names. Each participant was taken individually into a private workspace to complete the experiment. The ''ideal self'' and the names of the 25 students were entered into Idiogrid—software for managing and analyzing repertory grids and other types of self-report, personality data (Grice, 2002). The participants were then administered a sentence completion task developed by Grice, et al. (2004) to elicit their bipolar personal constructs. For instance, the participants were required to complete the following sentence with a single word or short phrase: ''Generally speaking, I am the type of person who is _____ .'' After entering a response, the participants were immediately asked to enter the opposite word or phrase, thus yielding bipo-

lar personal constructs (e.g., happy-sad, shy-outgoing). After eliciting the 10 personal constructs, participants rated (in an individually determined random order) each of the 26 figures (viz., ideal self, and 25 individuals, including themselves) on each construct using a 5-point Likert-type scale ranging from ''Very Inaccurate'' to ''Very Accurate.'' Each item stem was displayed as ''_____ is the type of person who is _____'' (as opposed to _____). The first blank contained the figure being rated, and the other 2 blanks contained the poles of a personal construct; for example, ''Robert is the type of person who is shy (as opposed to outgoing).'' The 260 ratings for each participant were recorded in a 10 x 26 (constructs x figures) matrix, or grid, for further analysis in Idiogrid, and the names of the 25 students were changed to pseudonyms to protect confidentiality in the analyses and results reported below.

Big Five grid. For each participant the 26 figures from the personal construct grid were loaded into Idiogrid. The participants then rated each figure on 20 Big Five trait descriptors. The descriptors were selected from Goldberg's list of 100 lexical marker items (Goldberg, 1999) and altered slightly to fit the item stems. Four items were selected for each of the Big Five Personality traits, with two of the four items for each trait being reverse-keyed. The figures were then rated (in an individually determined random order) by each participant using the same 5-point Likert scale. Each item stem was presented as ''_____ is the type of person who _____.'' The first blank contained the figure being rated, and the second blank contained the Big Five descriptor being considered; for example, ''Robert is the type of person who thinks of others first.'' The 520 ratings were recorded in a 20 x 26 (Big Five descriptor traits x figures) grid for subsequent analysis, and the names of the 25 students were again changed to the same pseudonyms to protect confidentiality.

Results

In the personal construct grids the constructs differed from participant to participant, while 25 of the 26 figures were matched across grids. Recall that each participant rated the other 24 students in the Summer Science Academy as well as himself or herself and the ideal self. Since the ideal self was not strictly matched across grids, it was removed from both the personal construct and Big Five grids. Given the remaining figures were matched across grids the participants views of themselves and their peers could then be examined.

Grid Pre-Scaling

The first decision that a researcher must make when utilizing GPA regards the scaling of the original grid ratings. Three pre-scaling options can be considered that remove systematic sources of variation between raters (for a review see Dijksterhuis & Gower, 1991). The first pre-scaling option is referred to as *dimensional scaling*, and it can be applied when the grids or matrices of ratings differ with respect to their dimensions. If this scaling is applied each grid value will be divided by the square root of the number of non-matched figures in the grid, effectively reducing the variability of the ratings in large grids. In this way larger grids will not have a statistical 'advantage' over smaller grids in the ANOVA portion of the results that is due simply to their size. In this study each participant

rated his or her peers on 10 personal constructs. The non-matched figures (viz., the personal constructs) were therefore equal in number across all of the grids, and dimensional scaling was not necessary.

The second pre-scaling option, referred to as *centering scaling*, essentially removes differences in scale means from the ratings. This scaling was applied to the current grids in Idiogrid; consequently, the grid values for each participant were subtracted from their respective construct means prior to being submitted to the GPA algorithm. The mean of each centered construct was therefore set equal to zero, and potential individual differences in construct means was thus removed.

The third pre-scaling option, *isotropic scaling*, was also applied to the current grids. This scaling technique essentially ''shrinks'' or ''stretches'' the original grid ratings to remove individual differences in scale usage. For example, if one participant endorses many different points on the entire scale (5 scale points) while another participant endorses primarily the mid-point of the scale, this difference may impact the ANOVA portion of the results. Applying isotropic scaling to these two participants' grids would decrease the overall variability of the first grid and increase the overall variability of the second grid. The computation of isotropic scaling values is slightly involved, and descriptions of the procedures can be found in ten Berge (1977) and Gower (1975). The computed isotropic scaling values will be centered around unity, with values greater than unity indicating that the original ratings were stretched (that is, their variability was increased). Values less than unity indicate that the original ratings were shrunk (that is, their variability was decreased). As can be seen in Table 1, the isotropic scaling values for the current 19 participants' personal construct grids show the ratings for Eric were stretched the most compared to the other 18 participants. Examination of Eric's original grid ratings in fact revealed that he never used the end-points of the rating scale. Rhonda's ratings were also stretched because she never endorsed the two lowest points on the rating scale. These individual differences may be of interest in another context, but the current goal is to examine the multivariate configuration of the rated people in the personal construct grids. Consequently, the isotropic scaling option was applied to essentially remove this systematic source of between-participant variability.

The Consensus Matrix

The central feature of GPA is undoubtedly the consensus grid, or matrix, which records the means computed from the rescaled and rotated original grids. The grids are rescaled according to the investigator's chosen options (see above), and then rotated to maximal agreement using the well-known Procrustes rotation technique. The GPA algorithm, as originally outlined by Gower (1975), works in an iterative fashion, passing through the complete set of grids and producing a new consensus matrix at the end of each pass. The most recent consensus matrix is continually compared to the preceding consensus matrix (from the previous pass through the grids) until the two match at a level specified by the investigator; that is, until the quantified difference between the two consensus matrices fails to exceed some *tolerance* criterion (e.g., .001). Once the algorithm converges the final consensus matrix is saved, and the rescaled and rotated grids are then compared to the consensus matrix using standard ANOVA formulas and procedures. In this way individual grids that deviate from the consensus matrix (i.e.,

the average grid) can easily be identified, and particular points of variability in the ratings can also be identified. The ANOVA results also yield an overall index of variability among the rescaled and rotated grids referred to as the *consensus proportion*. When converted to a percentage it will range from 0 to 100, with higher values indicating greater similarity in the organization of the matched figures across raters. In the extremes, a consensus proportion of 0 would indicate ratings that are completely dissimilar (e.g., as one might expect with random numbers) while a consensus proportion of 1.0 (100%) would indicate perfect agreement among raters in the multivariate configuration of the matched figures. Lastly, consensus proportions between pairs of individual grids can also be computed and then submitted to further analysis to identify clusters of raters who are similar to one another.

Table 1. Isotropic scaling factors for personal construct and Big Five grids

Grid	Personal Construct	Big Five
Tasha	0.98	0.90
Jeff	0.91	0.80
Kayla	0.98	0.89
Tom	1.02	0.88
Jenny	0.97	1.09
Claire	1.06	1.09
Eric	1.51	1.70
Rhonda	1.42	1.16
Debbie	1.05	0.89
Sally	0.88	0.95
Andrea	0.89	0.99
Zack	0.84	0.95
Paige	1.24	1.28
Matt	0.90	0.84
Samantha	0.89	0.92
Sue	1.06	1.19
Jessica	1.01	1.42
Erin	1.02	1.42
Lindsey	0.97	0.83

The results of the GPA for the personal construct grids yielded a consensus proportion of .45, which indicated only moderate average agreement among the 19 students when considering themselves and their peers. This consensus proportion was tested for statistical significance using a randomization test developed by
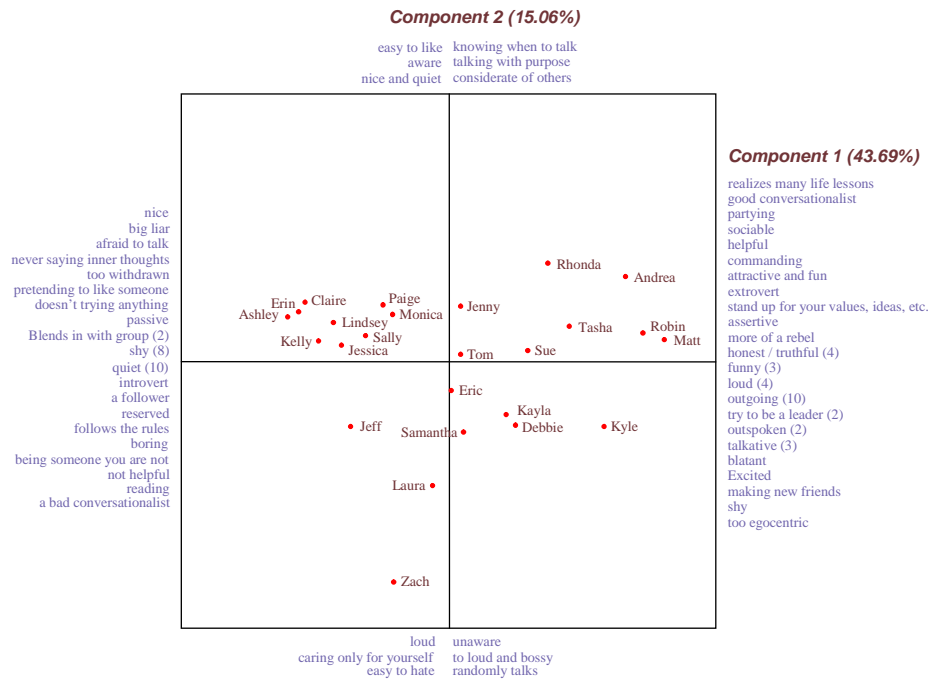
Figure 2. Bi-plot of consensus matrix from personal construct grids. Highly salient personal constructs are also indicated on the two axes.

Wakeling, Raats, and MadFie (1992). The test works by randomizing the matched figures in each of the 19 grids and then conducting the GPA. This process is repeated numerous times and the consensus proportion is computed and recorded for each trial. The observed consensus proportion can then be compared to the distribution of proportions generated from randomized grids. In the current data 500 trials were generated, and not a single value equaled or exceeded the observed consensus proportion. The observed proportion of .45 was therefore judged to be statistically significant at the .05 level (observed p < .002; that is, less than 1/500).

A common strategy is to submit the consensus matrix to a Principal Components Analysis (PCA), the purpose of which is to reduce the dimensionality of the consensus matrix so that the organization of the rated targets (in this case, rated students) can be visually examined in bi-plots. The relationships between the targets and the individual rating dimensions can also be examined in these bi-plots using extension analysis. Grice and his colleagues (Grice, 2007; Grice, Jackson, McDaniel, 2006; Grice, 2004) have shown how extension analysis can be conducted with personality data. For the current data, the first two principal components explained 58.75% of the variance in the consensus matrix values and were used to create the axes in Figure 2. An extension analysis was then conducted in which the personal constructs for the 19 students were mapped into the two-dimensional space formed by the principal components. A total of 190 constructs could potentially be mapped into the space, but the structure coefficients from the PCA were examined to select the most salient constructs shown in Figure 2. As can be seen, the first component differentiates people who are outgoing and talka-

tive versus shy and introverted. Matt, Robin, and Andrea are clearly outgoing whereas Ashley, Erin, and Claire are shy. The second component appears to differentiate people who are nice and considerate (e.g., Rhonda and Andrea) versus those who are selfish and disagreeable (e.g., Zach and Laura). The multivariate configuration in Figure 2 thus represents the aggregate view computed from the 19 personal construct grids.

Analysis of Variance

As mentioned above, GPA is a combination of Procrustes rotation and analysis of variance (ANOVA). The ANOVA results for the personal construct grids are reported in Tables 2 and 3, and they can be examined to identify points of agreement and disagreement among the participants' views of themselves and their peers. The residuals reported in the tables are most useful for this process and can be examined with respect to their relative magnitudes. As can be seen in Table 2, the personal construct residual values are relatively homogeneous, ranging in value from 2.59 to 3.16 for the personal construct grids. Rhonda's grid produced the largest residual, indicating that her ratings deviated most from the consensus matrix; whereas Claire's grid showed the greatest degree of agreement with the consensus matrix. With smaller sets of grids and for different purposes these residuals can be used to identify 'outliers'; that is, people who deviate from the average view of the rated figures. In other contexts, for instance, the discrepant views of a person in a focus group could be identified and explored, or the novel views of a person on a management team could be extracted and investigated.

The ANOVA residuals for the rated peers can also be computed and examined for specific points of departure from the consensus matrix. As shown in Table 3, these residuals ranged in value from 1.56 to 3.37. It appears the 19 participants showed the greatest discrepancies with respect to their ratings of Zach (residual = 3.37) and Debbie (residual = 3.23). As described by Dijksterhuis and Punter (1991), the ANOVA residuals can be broken down even further to identify exactly which participants contributed most to the varied views of Zach and Debbie. Although these specific residuals are not reported here, they indicated that Eric's view of Debbie was most discrepant from the consensus. A bi-plot of Eric's first two principal components (explaining 72.20% of the rating variance) extracted from his grid are shown in Figure 3. It can be seen that Eric viewed Debbie as highly similar to Zach, both of whom were construed as ''immature'', ''dependent'', ''uptight'', and ''serious.'' In the consensus matrix in Figure 2, Debbie was viewed as most like Kayla, and neither were located on the extreme ends of the dimensions in the bi-plot or close to Zach. Comparing these visual plots demonstrates clearly how the views of specific individuals can be related to the aggregate results.

Procrustes statistics

The consensus matrix and consensus proportion are computed from all of the grids in the analysis, and the latter statistic indicates the general degree of similarity among the grids with respect to the configurations of the rated people. The same principles used to compute the consensus proportion can also be applied to pairs of grids. Specifically, the *procrustes statistic* can be computed for

Table 2. Analysis of variance results for the nineteen students' personal construct and Big Five grids

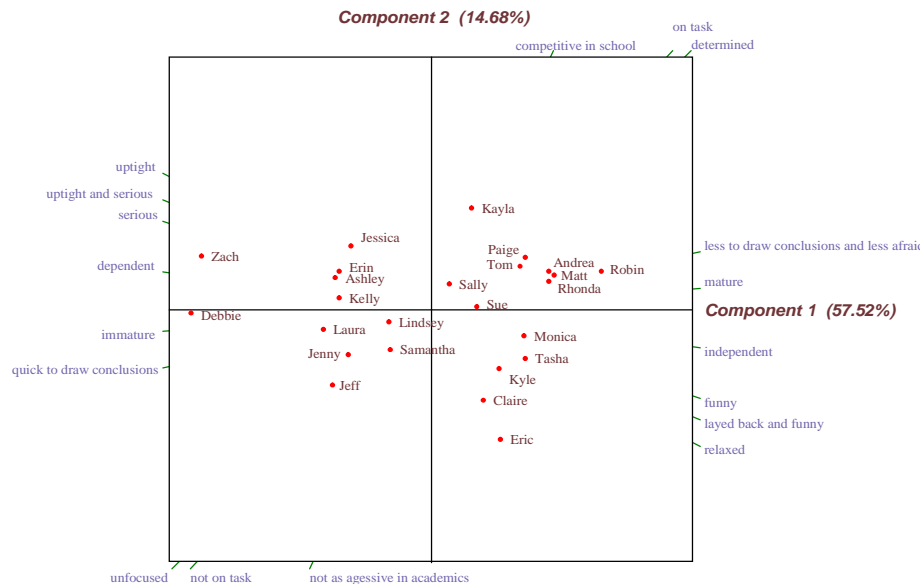| Grid	Personal Construct		Big Five	
	Residual	**Total**	**Residual**	**Total**
Tasha	2.94	6.44	2.23	5.93
Jeff	2.61	3.89	2.30	4.66
Kayla	2.95	4.91	2.19	5.74
Tom	3.12	6.17	2.38	5.21
Jenny	2.91	5.60	2.30	5.71
Claire	2.59	3.07	2.43	5.02
Eric	2.88	4.67	2.32	4.95
Rhonda	3.16	6.35	2.35	5.18
Debbie	2.71	4.53	2.33	4.89
Sally	2.80	5.05	2.35	5.58
Andrea	3.06	6.40	2.24	6.28
Zack	2.78	4.36	2.46	4.27
Paige	2.90	5.41	2.31	5.89
Matt	3.07	7.29	2.30	6.43
Samantha	2.70	3.78	2.40	5.38
Sue	3.03	6.71	2.33	5.73
Jessica	2.97	4.93	2.40	4.65
Erin	2.84	3.79	2.41	4.57
Lindsey	3.11	6.65	2.48	3.92
Total SS	55.11	100.00	44.51	100.00
Consensus Proportion		.45		.55

Figure 3. Bi-plot of a Eric's personal construct grid that was discrepant from the consensus matrix.

Table 3. Analysis of variance results for the twenty-five rated people in the personal construct and Big Five grids

	Personal Construct		Big Five	
Rated Person	**Residual**	**Total**	**Residual**	**Total**
Tasha	2.09	3.64	1.64	3.47
Jeff	2.02	3.46	1.41	3.69
Kayla	2.38	4.56	2.28	4.56
Tom	1.90	2.68	1.64	3.15
Kelly	1.59	2.99	1.38	2.69
Ashley	1.85	4.00	1.46	2.89
Jenny	2.30	2.94	1.75	2.84
Claire	2.58	5.05	1.87	4.22
Eric	1.61	2.72	1.66	3.09
Rhonda	2.02	3.59	1.90	4.49
Debbie	3.23	5.37	2.44	5.15
Kyle	1.99	4.72	1.43	5.27
Sally	1.96	2.85	1.49	2.29
Andrea	2.07	5.46	1.84	5.45
Zach	3.37	6.59	1.99	6.98
Paige	2.49	3.56	1.93	4.04
Matt	2.65	6.10	1.88	6.57
Robin	2.43	5.86	1.73	4.67
Laura	2.89	4.45	1.79	4.04
Monica	1.87	2.31	1.66	2.23
Samantha	2.10	3.40	1.90	3.33
Sue	2.29	3.59	1.96	3.91
Jessica	1.56	2.61	1.71	3.10
Erin	1.56	3.57	1.76	4.31
Lindsey	2.32	3.94	2.02	3.58

each pair of rescaled and rotated grids to indicate the extent to which any two participants differ with respect to the configurations of the rated people. Converse to the consensus proportion, high values for the procrustes statistic indicate greater dissimilarity between the grids, and while there is no limit on its magnitude a value of zero indicates perfect agreement between two grids.

Given that the statistic can be computed for all pairs of grids it is common practice to report the procrustes statistics in matrix form. The matrix appears much like a correlation matrix, with the grids comprising the rows and columns and the procrustes statistics reported in the off-diagonal cells of the table. The main diagonal of the table consists of zeros, indicating that each grid is identical to itself. This matrix can then be submitted to a Singular-Value-Decomposition or
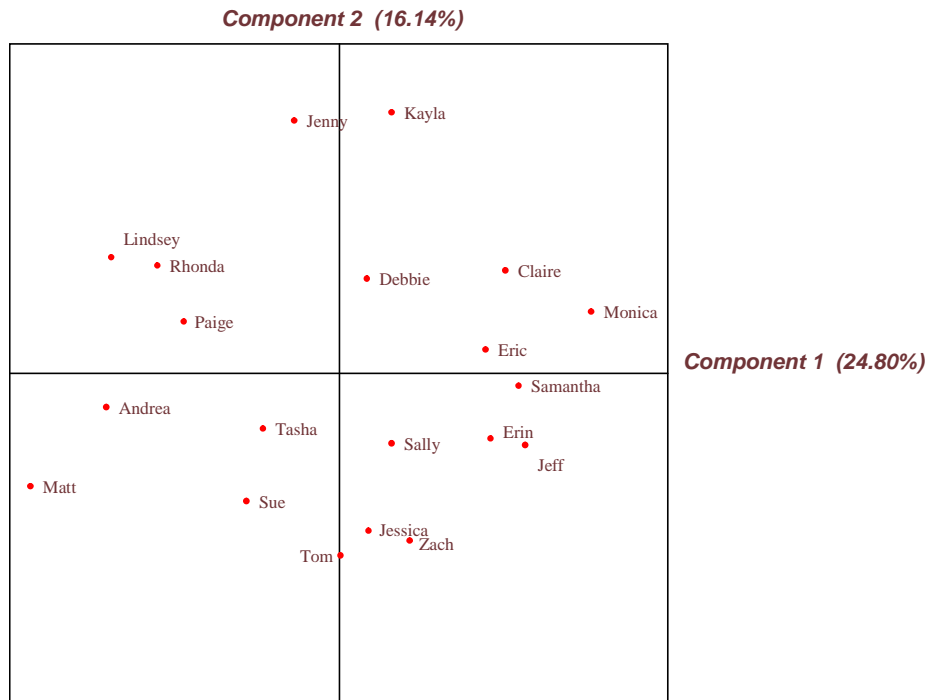
Figure 4. Bi-plot of procrustes statistics computed from personal construct grids.

a Principal Components Analysis and the results plotted in a bi-plot. The latter analysis was chosen for the procrustes statistics computed for the personal construct grids. The first two components, explaining 40.94% of the variance, are plotted in Figure 4. It can be seen that Jenny and Kayla were more similar to one another in their views than they were to the other participants. Similarly, Zach, Tom, and Jessica were relatively similar to one another in their ratings of themselves and their peers. Examination of such a plot can therefore aid in the identification of clusters of individuals who show relative agreement and disagreement in their ratings. For the current analyses, however, it should be noted that the procrustes statistics (not reported) were relatively homogeneous, ranging in value from .21 to .41, which is consistent with the homogeneous residuals reported in Table 2. Consequently, the clusters of people in Figure 4 may not represent substantive differences in their views. For example, the procrustes statistic for Jenny and Kayla (0.22), who appear close to one another in Figure 4, is not much lower than the procrustes statistic for Jenny and Tom (0.37), who appear quite discrepant in the figure. In order to investigate the practical relevance of these differences bi-plots created from Principal Components Analyses conducted on Jenny's, Kayla's, and Tom's individual grids could be constructed and examined.

Big Five grids

   The Big Five grids in which the participants rated themselves and their peers on 20 marker items for the Big Five personality traits were also analyzed with GPA (sans the ideal self). The same pre-scaling options used for the personal construct grids were chosen, and the isotropic scaling factors are reported in Table 1. Eric's ratings were again stretched more than the other 18 students as he again chose not to use the end-points of the rating scale. As stated above, however, these individual differences were not of primary importance in this study.
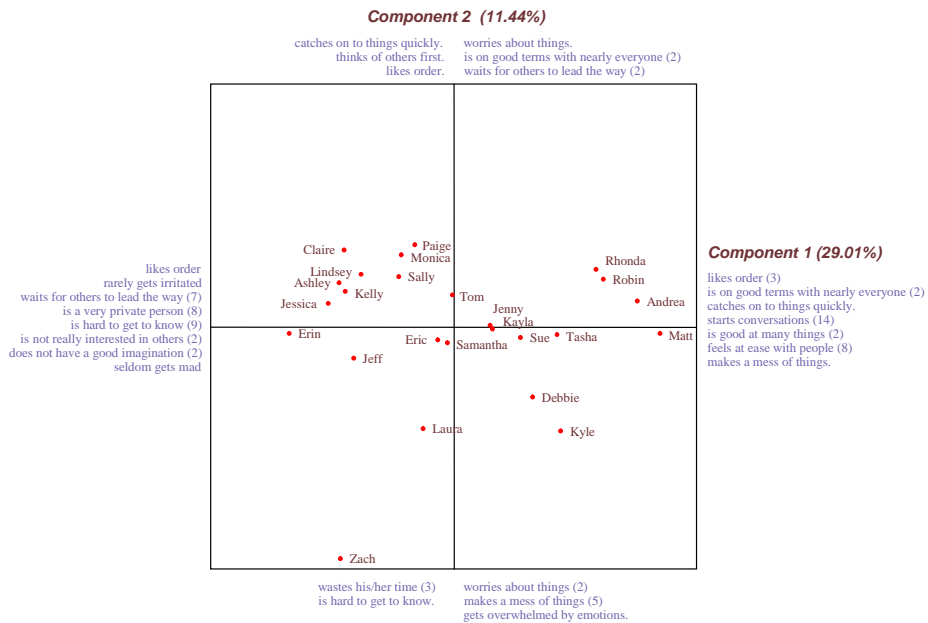

Figure 5. Bi-plot of consensus matrix from Big Five grids. Highly salient traits terms are also indicated on the two axes.

   The consensus proportion for the Big Five grids was equal to .55 ($p < .002$, 500 randomized trials) and was slightly higher than the result for the personal construct grids. A Principal Components Analysis was conducted on the consensus matrix, and the first two components that explained 40.45% of the variance in the consensus matrix were plotted in Figure 5. Also shown in the figure are the 25 rated people and the Big Five marker items with salient structure coefficients. The first component clearly differentiates people who ''start conversations'' and ''feel at ease with people'' from those who are ''hard to get to know'' and ''very private.'' The second component differentiates people who think ''...of others first'' and who are ''...on good terms with nearly everyone.'' Matt, Andrea, Robin, and Rhonda are viewed as similar to one another and opposite to Erin, Jessica, and others in Figure 5. The former individuals are, on average, viewed as being at ease with others and able to start conversations, while the latter are viewed as disinterested in others and hard to get to know. Zach is again relatively disparate from the other rated people, and is viewed, on average,

as someone who wastes time and makes a mess of things. The constructs located on the opposite end of the second component also apply to Zach in a negative way, as he is not on good terms with others and does not think of others first. With respect to the ANOVA results, Table 2 again revealed a great deal of homogeneity among the residuals for the participants' Big Five ratings. In other words, no single student's grid deviated distinctly from the consensus matrix compared to the other students in the study. With respect to the views of particular students, Table 3 indicated that Debbie was again rated in a manner that was relatively inconsistent with the consensus view. Examination of the specific residuals (not shown) revealed that Eric viewed Debbie more negatively than he did Zach, whereas in the consensus matrix (see Figure 5), Zach alone was viewed as highly disparate from most of the other students.

When comparing the personal construct (Figure 2) and Big Five (Figure 5) bi-plots, it becomes obvious that the 25 people are arranged in a similar pattern in both graphs. While the descriptive terms associated with the poles of the components may not match perfectly across the two figures, the organization of the rated people is highly similar. In fact, a GPA conducted on the two consensus matrices produced an extremely high consensus proportion of .97 ($p < .002$, 500 randomized trials). In terms of the aggregated data, then, the participants grouped themselves and their peers into highly similar clusters using either their own personal constructs or the Big Five Marker items.

Discussion

Gower's *Generalized Procrustes Analysis* (GPA; 1975) is an inventive combination of Procrustes rotation and analysis of variance. In the results above it was shown that GPA is a flexible and powerful technique for modeling three-dimensional matrices of data. Specifically, a number of questions were addressed regarding the intra- and inter-personal perceptions of students attending a Summer Science Academy. First, did the high school students reach a consensus with regard to their views of themselves and their peers? For the personal construct ratings the *consensus proportion* from the GPA was equal to .45, and for the Big Five ratings the value was equal to .55. The commonality found among the multivariate configurations of the students' ratings of themselves and their peers, although statistically significant, was therefore modest in both the personal construct and Big Five grid grids. Second, did the average, multivariate configuration of the students derived from the personal construct grids match the configuration derived from the Big Five trait grids? To address this question, we computed the consensus matrix from each rating task. As described above, a consensus matrix is an average grid that is computed from all of the participants' rescaled and rotated original grids. These two consensus matrices were themselves then submitted to a GPA, and the resulting consensus proportion was extremely high, .97, indicating very high similarity between the two configurations. Examination of the principal component bi-plots of the consensus matrices also indicated a high degree of similarity in the multivariate configurations of the rated students in the personal construct and Big Five consensus matrices. Third, did particular students emerge from the analysis as highly discrepant in their ratings of themselves and the other students when

compared to the consensus? This question was answered on a general level with respect to each student's overall grid responses, and on a more specific level with respect to the ratings of particular students in each participant's grid. Several students were identified, for instance, to deviate generally from the consensus, although the extent of the differences were not noteworthy when compared to all of the participants in the study. On a more specific level, points of disagreement with regard to particular students (Zach and Debbie) in the group were clearly identified in both the Personal Construct and Big Five grids. These specific points of disagreement were furthermore attributed primarily to one student rater, Eric, as the differences between his views of Zach and Debbie and those expressed in the consensus matrix could be clearly seen in the comparison of the aggregate and individual bi-plots.

The analyses addressing this third question in fact reveal what is perhaps the most intriguing feature of GPA; namely, the capability of deftly moving from the aggregate to the individual in the analysis. Despite analyzing 4750 total ratings in the personal construct grids and 9500 total ratings in the Big Five grids, the individuals in the study were never lost from view nor were they swept away completely into the aggregate. Through the specific analysis of variance results and the visual examination of the bi-plots particular points of similarity and dissimilarity among the persons were discovered and examined. Recent years have seen renewed interest in sophisticated analyses that place considerably more focus on the individual relative to the aggregate (e.g., Molenaar, 2004, 2005), and Grice and his colleagues (Grice, 2004, 2007; Grice et al., 2004; Grice et al., 2006) have explored techniques that connect person-centered and aggregate-based results. In a similar vein, the recent emphasis on reporting and interpreting effect sizes will necessitate connecting aggregate patterns of results to the lives of individuals who are impacted by published research. While the aggregate may reveal important features of the data that can be meaningfully tied to variables and to hypotheses, the testing ground of most psychological theories will be the person, whether considered in isolation or considered as embedded in a social situation. With GPA a researcher can place his or her focus on the person and the aggregate in a single set of analyses. Consequently, GPA could prove particularly fruitful in the study of groups, such as families, classrooms, or organizations. It could also prove fruitful for studying how an individual changes over time in his or her views of self and others, or how a person's perceptions change across contexts. In the analyses above, rating matrices were obtained from 20 different students, but matrices obtained from the same person on different occasions or from the same person rating different thoughts, emotions, or actions in different situations could also be analyzed. The flexibility of GPA can in fact be seen in the variety of domains for which it has been used; for instance, in the study of individuals' perceptions of genetic engineering (Frewer, Howard, & Shepherd, 1997), meat (Russell & Cox, 2003), commercial ports (Williams & Langron, 1984), and their views of the 'personalities' of pigs (Wemelsfelder, Hunter, Mendl, & Lawrence, 2000).

Technical Issues

Beginning with Gower's (1975) original article, a number of technical issues have consistently been raised with respect to GPA. Perhaps foremost is the

long-standing criticism of the sensitivity of Procrustes rotation. It has been shown, for instance, that even when two matrices of data are generated randomly, the Procrustes transformation can rotate the matrices to a level of agreement that can be surprisingly high (e.g., see Horn, 1967). In the context of GPA, this property of Procrustes rotation would suggest that the consensus proportion may be inflated for some data sets. For instance, a consensus proportion of .70 may seem high given the possible range of values from 0 to 1; however, this result may simply be an artifact of the rotation's capitalization on chance patterns in the data. The randomization test developed by Wakeling, et al. (1992) is therefore an essential part of any GPA, since it allows the researcher to compare the observed consensus proportion to a distribution of proportions created from randomized versions of the original data matrices. Traditional cut-points for statistical significance (viz., .05 and .01) can then be used to determine if the observed consensus proportion is statistically significant.

A second issue regarding GPA is the application of the centering, dimensional, and isotropic pre-scaling transformations. Specifically, when are these transformations to be applied to the original data? Unfortunately, no simple answer can be offered to this question, but we generally recommend running a series of analyses with any given data set. When conducting a GPA, the goal of most researchers is to match the multivariate configurations of the rated targets (e.g., people, things, places, etc.) across raters. Removing potential, systematic differences in scale usage or dimensionality of the rating matrices appears to be consistent with this goal. The primary analysis should therefore be based on ratings that have been transformed with each of the three scaling options. Of course, if the rating matrices are all of the same dimensionality, as with the repertory grids in this study, then dimensionality scaling is superfluous. After the primary results are interpreted, the researcher should conduct the GPA again without one of the scaling options. The results can then be compared to the completely scaled results and the impact of the particular scaling option can be thoroughly assessed.

A third issue involves analyses based on grids of different dimensionality. Consider, for instance, the following two sets of ratings:

		Cereal A	Cereal B	Cereal C	Ideal Cereal
Joe	*Sweet*	-3	-2	2	3
	Crunchy	1	2	2	2
	Pricey	2	-1	-3	2
Mary	*Expensive*	-2	-3	-2	-1
	Healthy	-2	-1	2	2

Joe has now rated the four cereals on three dimensions whereas Mary has rated the cereals on only two dimensions. As described above GPA can still be used to analyze such data by treating the dimensions as arbitrary and 'padding' Mary's ratings with a third row of zeros. As described above, the overarching goal of

109

GPA is to align the multivariate configurations of the cereals. Joe's ratings comprise a 3-dimensional space whereas Mary's comprise only a 2-dimensional space. Mary's ratings, however, can be considered to comprise a 3-dimensional space in which the cereals do not extend into the third dimension. In other words, the ratings for the cereals are all zero on the third dimension. In computer programs such as Idiogrid (Grice, 2002), small matrices will automatically be padded with zeros prior to the analysis, and the consensus matrix will be labeled in such a way as to indicate the arbitrary nature of the dimensions. The consensus matrix for Joe's and Mary's ratings above, for instance, appears as follows:

	Cereal A	Cereal B	Cereal C	Ideal Cereal
Dim_1	-0.25	-1.14	-0.10	1.49
Dim_2	-1.87	-0.76	1.60	1.03
Dim_3	0.00	0.00	0.00	0.00

The labels 'Dim_1', 'Dim_2', and 'Dim_3' refer to the arbitrary dimensions, and the zeros in the third row indicate that the two matrices could not be matched in the third dimension. If other raters who used three dimensions were included in the analysis, the third row in the consensus matrix above would report values other than zero. It is thus by padding smaller matrices with zeros that matrices of different dimensionality can be analyzed in GPA. Of course, the matched figures must be the same and in the same order across matrices. In other words, both Joe and Mary needed to rate the same cereals that are recorded in identical columns in their two matrices.

Finally, the graphical methods associated with GPA play a critical role in interpreting and reporting the results. In many cases these methods involve bi-plots created from the results of Principal Components Analyses. The figures created for the results above show how useful they can be for summarizing the aggregate results as well as relating individual data to the aggregate. Examples of similar plots can be found in Monteleone, Raats, and Mela (1997), Oreskovich, Klein and Sutherland (1991), and Russell and Cox (2004). Furthermore, Gower and Dijksterhuis (1994) have discussed creating bi-plots from matrices with both quantitative and qualitative variables. Similar developments on both the applied and analytical fronts should contribute to the continued growth and popularity of Generalized Procrustes Analysis.

REFERENCES

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*, 33-42.
Dijksterhuis, G. B., & Gower, J. C. (1991/2). The interpretation of generalized procrustes analysis and allied methods. *Food Quality and Preference, 3*, 67-87.

Dijksterhuis, G., & Punter, P. (1991). Interpreting generalized procrustes analysis 'analysis of variance' tables. *Food Quality and Preference, 2*, 255-265.

Frewer, L. J., Howard, C., & Shepherd, R. (1997). Public concerns in the United Kingdom about general and specific applications of genetic engineering: Risk, benefit, and ethics. *Science, Technology, & Human Values, 22*, 98-124.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika, 58*, 453-467.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe: Vol. 7* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*, 33-51.

Gower, J. C., & Dijksterhuis, G. B. (1994). Multivariate analysis of coffee images: A study in the simultaneous display of multivariate quantitative and qualitative variables for several assessors. *Quality & Quantity*, 165-184.

Grice, J. W. (2007). Person-centered structural analyses. In R. Robins, C. Fraley, and R. Krueger (Eds.) *Handbook of research methods in personality psychology* (pp. 557-572). New York: Guilford Press.

Grice, J. W. (2004). Bridging the idiographic-nomothetic divide in ratings of self and others on the Big Five. *Journal of Personality, 72*, 203-242.

Grice, J. W. (2002). Idiogrid: Software for the management and analysis of repertory grids. *Behavior Research Methods, Instruments, & Computers, 34*, 338-341.

Grice, J., Burkley, E., Burkley, M., Wright, S., & Slaby, J. (2004). A sentence completion task for eliciting personal constructs in specific domains. *Personal Construct Theory & Practice, 1*, 60-75.

Grice, J. W., Jackson, B., & McDaniel, B. (2006). Bridging the idiographic-nomothetic divide: A follow-up study. *Journal of Personality, 74*, 1191-1218.

Horn, J. L. (1967). On subjectivity in factor analysis. *Educational and Psychological Measurement, 27*, 811-820.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*, 201-218.

Molenaar, P. C. M. (2005). How generalization works through the single case: A simple idiographic process analysis of an individual psychotherapy. *International Journal of Idiographic Science, 1*, 1-20.

Monteleone, E., Raats, M., & Mela, D. (1997). Perceptions of starchy food dishes: Application of the repertory grid technique. *Appetite, 28*, 255-265.

Oreskovich, D. C., Klein, B. P., & Sutherland, J. W. (1991). Procrustes analysis and its applications to free-choice and other sensory profiling. In Lawless and Klein (Eds.) *Sensory Science Theory and Applications in Foods* (pp. 353-393). New York: Marcel Dekker.

Russell, C. G., & Cox, D. N. (2004). Understanding middle-aged consumers' perceptions of meat using repertory grid methodology. *Food Quality and Preference, 15*, 317-329.

Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika, 31*, 1-10.

Slater, P. (1977). *The measurement of intrapersonal space by grid technique: Vol. 2. Dimensions of intrapersonal space*. London: Wiley.

Ten Berge, J. M. F. (1977). Orthogonal procrustes rotation for two or more matrices. *Psychometrika, 42*, 267-276.

Wakeling, I. N., Raats, M. M., Halliday, J. H. (1992). A new significance test for consensus in generalized procrustes analysis. *Journal of Sensory Studies, 7*, 91-96.

Wemelsfelder, F., Hunter, E. A., Mendl, M. T., & Lawrence, A. B. (2000). The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science, 67*, 193-215.

Williams, A. A., & Langron, S. P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture, 35*, 558-568.

Notes

Correspondence concerning this article should be addressed to James W. Grice, Department of Psychology, 215 North Murray, Oklahoma State University, Stillwater, OK 74078. Electronic mail may be sent to james.grice@okstate.edu.