Behavior Research Methods
2009, 41 (4), 1073-1082
doi:10.3758/BRM.41.4.1073

The LMPCA program: A graphical user interface for fitting the linked-mode PARAFAC-PCA model to coupled real-valued data

Tom F. Wilderjans and Eva Ceulemans
University of Leuven, Leuven, Belgium

Henk A. L. Kiers
University of Groningen, Groningen, The Netherlands

and

Kristof Meers
University of Leuven, Leuven, Belgium

In behavioral research, PARAFAC analysis, a three-mode generalization of standard principal component analysis (PCA), is often used to disclose the structure of three-way three-mode data. To get insight into the underlying mechanisms, one often wants to relate the component matrices resulting from such a PARAFAC analysis to external (two-way two-mode) information, regarding one of the modes of the three-way data. To this end, linked-mode PARAFAC-PCA analysis can be used, in which the three-way and the two-way data set, which have one mode in common, are simultaneously analyzed. More specifically, a PARAFAC and a PCA model are fitted to the three-way and the two-way data, respectively, restricting the component matrix for the common mode to be equal in both models. Until now, however, no software program has been publicly available to perform such an analysis. Therefore, in this article, the LMPCA program, a free and easy-to-use MATLAB graphical user interface, is presented to perform a linked-mode PARAFAC-PCA analysis. The LMPCA software can be obtained from the authors at http://ppw.kuleuven.be/okp/software/LMPCA. For users who do not have access to MATLAB, a stand-alone version is provided.

One of the common aims of behavioral research is to reveal the pattern present in three-way three-mode data. Research in emotion psychology, for example, tries to account systematically for individual differences in the emotions that people experience in specific situations (see, e.g., Kuppens, Van Mechelen, & Rijmen, 2008). In the field of behavioral decision making and consumer psychology, researchers aim to disclose how the profile of different purchasing behaviors over brands may differ between individuals (see, e.g., Wood, 2004) or how personal characteristics may influence situation-dependent buyer behavior (see, e.g., Belk, 1974). In many fields outside psychology, such as chemical research, three-way three-mode data are also gathered and analyzed (see, e.g., Smilde, Bro, & Geladi, 2004).

Three-mode component analysis techniques can be very useful for analyzing three-way three-mode data (for an introduction and some applications, see Kroonenberg, 2008). As three-mode generalizations of standard principal component analysis (PCA; Hotelling, 1933; Pearson, 1901), these techniques summarize the data by reducing its three modes to a few components and defining a linking structure among the three sets of components. A popular and well-known three-mode component analysis technique is PARAFAC (Carroll & Chang, 1970; Harshman, 1970). In PARAFAC, the number of components is identical for the three modes, and the linking structure boils down to a one-to-one correspondence. In the literature, PARAFAC was successfully applied to problems concerning, among others, marketing research (Harshman & De Sarbo, 1984), food control (Bro, 1998; Bro & Kiers, 2003), acoustics (Harshman, Ladefoged, & Goldstein, 1977), and chromatography (Bro, Andersson, & Kiers, 1999).

Applying PARAFAC to data on the emotions that people experience in given situations may serve to summarize the individual differences in these data. Similarly, a PARAFAC analysis may result in a concise description of the individual differences in brand purchasing behaviors. However, the PARAFAC analysis does not afford insight into the underlying psychological mechanisms. To capture these mechanisms, researchers often want to relate the PARAFAC results to external information (i.e., covariates) about the people being studied. Research in emotion psychology, for instance, may focus on people's disposi-

T. F. Wilderjans, tom.wilderjans@psy.kuleuven.be

© 2009 The Psychonomic Society, Inc.

tions and traits, such as appraisal tendencies, chronic accessibility of the different emotions, and the "Big Five" personality dimensions (Kuppens et al., 2008; Mischel & Shoda, 1998). Note that revealing the mechanisms underlying emotional experience is not only interesting from a theoretical point of view but also has clinical implications in, among other things, the development of emotion management programs and the prevention of illness; for example, research shows that higher anger levels are related to aggression, heart disease (e.g., Spielberger et al., 1985), and hypertension (e.g., Spielberger et al., 1991).

Relating the covariate information to the PARAFAC results may involve choosing between a segmented and an integrated strategy. The segmented strategy consists of two consecutive steps: first, the three-way three-mode data are analyzed by means of a PARAFAC model; second, the PARAFAC results are related to the covariate information. In the integrated strategy, the three-way three-mode data array and the two-way two-mode data matrix containing the covariate information about the persons are analyzed simultaneously. Since the number of covariates under study is often high, it is appealing to also reduce the covariates to a few components using PCA and to link the PCA and PARAFAC models by constraining the component scores of the persons to be the same in both models. As a consequence, the different emotions experienced in various situations may be linked to personal dispositions. In Wilderjans, Ceulemans, and Van Mechelen (2009), such a coupled PCA and PARAFAC model of this type—henceforth called *linked-mode PARAFAC-PCA*—was developed. This model is a member of the family of linked-mode PARAFAC models (Harshman & Lundy, 1984, 1994), or, more generally, the family of multiway multiblock component models (Smilde, Westerhuis, & Boqué, 2000). Comparing the results of a segmented strategy with those of an integrated strategy, Wilderjans, Ceulemans, and Van Mechelen (2008) showed, by means of simulation studies, that the integrated strategy is to be preferred to the segmented one, since it may lead to more correct and stable results.

Until now, however, no software program has been publicly available to perform a linked-mode PARAFAC-PCA analysis. Therefore, in this article, we present a user-friendly MATLAB graphical user interface called LMPCA, which can be downloaded from http://ppw.kuleuven.be/okp/software/LMPCA (for inexperienced users, or those without access to MATLAB, a standalone version is provided). The main features of the new program are: (1) It is easy to use and can be downloaded free; (2) it is flexible, allowing the user to specify different options for the analysis; (3) it supports the user in selecting an appropriate model that describes the data well without being overly complex; and (4) the results of the analysis can be saved in different formats to enable the user to further process the obtained results—for example, by plotting the components with popular software packages like SPSS and SAS.

The remainder of this article is organized in three main sections. In section 1, the theory of linked-mode PARAFAC-PCA analysis is recapitulated. In section 2, making use of an illustrative example, guidance is provided to different choices (i.e., steps) that have to be made when performing a linked-mode PARAFAC-PCA analysis. Finally, section 3 briefly discusses how the LMPCA program can be used (an up-to-date version of this manual section can be found on the Web site mentioned above).

1. Linked-Mode PARAFAC-PCA Analysis

Linked-mode PARAFAC-PCA analysis may be used to analyze a real-valued three-way three-mode data set and a real-valued two-way two-mode data set that have one mode in common. First, the model is described. Next, the data analysis is discussed.

1.1. Model

The linked-mode PARAFAC-PCA model (Wilderjans et al., 2009) is a model for an $I \times J \times K$ real-valued data array $\underline{\mathbf{D}}^{\mathbf{1}}$ and an $I \times L$ real-valued data matrix $\mathbf{D}^{\mathbf{2}}$ that have the first mode in common. $\underline{\mathbf{D}}^{\mathbf{1}}$ is approximated by an $I \times J \times K$ real-valued model array $\underline{\mathbf{M}}^{\mathbf{1}}$ that can be decomposed according to a PARAFAC model with P components. $\mathbf{D}^{\mathbf{2}}$ is approximated by an $I \times L$ real-valued model matrix $\mathbf{M}^{\mathbf{2}}$ that can be decomposed according to a PCA model with P components. Importantly, the component matrix for the first mode is constrained to be the same in both models. Hence, the entries of $\underline{\mathbf{M}}^{\mathbf{1}}$ and $\mathbf{M}^{\mathbf{2}}$ may be computed as follows:

$$m^1_{ijk} = \sum_{p=1}^{P} a_{ip} b^1_{jp} c^1_{kp},$$

$$m^2_{il} = \sum_{p=1}^{P} a_{ip} b^2_{lp},$$

with a_{ip}, b^1_{jp}, c^1_{kp}, and b^2_{lp} denoting the entries of the $\mathbf{A}$ ($I \times P$), $\mathbf{B}^{\mathbf{1}}$ ($J \times P$), $\mathbf{C}^{\mathbf{1}}$ ($K \times P$), and $\mathbf{B}^{\mathbf{2}}$ ($L \times P$) real-valued component matrices.

Relation with other model families. The family of multiway multiblock component models (the linked-mode PARAFAC-PCA model being a specific instance thereof) is closely related to the family of multiway covariate regression models (Smilde & Kiers, 1999), with the latter family of models in Smilde et al. (2000) being also called *multiway multiblock regression models*. The major difference between both model families is that in multiway multiblock regression models the different data blocks have a distinctive role (i.e., one block serves as a predictor block, and the other as a criterion block), whereas in the family of multiway multiblock component models the different data blocks are interchangeable in terms of conceptual status.

Multiway multiblock component analysis is further related to canonical correlation analysis (Hotelling, 1936), in that both methods summarize the data by means of a few components (i.e., linear combinations of the variables). Both methods, however, differ in that (1) the multiway multiblock component analysis can handle three-way coupled arrays, whereas (generalized) canonical correlation

analysis is restricted to coupled two-way matrices; and in that (2) in canonical correlation analysis components are extracted that correlate as high as possible without necessarily accounting for much of the variance in the different data matrices; multiway multiblock component analysis, on the other hand, is specifically aimed at finding components that explain as much variance as possible in all data blocks simultaneously (de Jong & Kiers, 1992; ten Berge, 1993).

Uniqueness. Under mild conditions (for PARAFAC, see Sidiropoulos & Bro, 2000), the component matrices of a linked-mode PARAFAC-PCA solution have no rotational freedom and thus are unique as far as the scaling, reflection, and permutation of the P components are concerned. To identify the solution, the following procedure may be applied: First, the columns of $\mathbf{B^1}$ and $\mathbf{C^1}$ are scaled and reflected so that their sum of squares equals 1 and the number of negative entries is minimized. Next, the latter scaling and reflection is compensated for in $\mathbf{A}$ and $\mathbf{B^2}$. Finally, the columns of all component matrices are simultaneously permuted; the columns of $\mathbf{A}$ are ranked in descending order vis-à-vis their sum of squares. (Note that permuting, scaling, and reflecting the columns of the component matrices does not fundamentally influence their interpretation.)

1.2. Data Analysis

Aim. The aim of a linked-mode PARAFAC-PCA analysis with P components of data ($\underline{\mathbf{D}}^{\mathbf{1}}$,$\mathbf{D^2}$) is to estimate a model ($\underline{\mathbf{M}}^{\mathbf{1}}$,$\mathbf{M^2}$) so that the value of the loss function

$$f = \alpha \times \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(d_{ijk}^{1} - \sum_{p=1}^{P} a_{ip}b_{jp}^{1}c_{kp}^{1}\right)^2 + (1-\alpha)\times\sum_{i=1}^{I}\sum_{l=1}^{L}\left(d_{il}^{2} - \sum_{p=1}^{P} a_{ip}b_{lp}^{2}\right)^2, \quad (0 \le \alpha \le 1),$$

is minimized and $\underline{\mathbf{M}}^{\mathbf{1}}$ and $\mathbf{M^2}$ can be represented by a PARAFAC and a PCA model with P components, respectively. The weights α and $1 - \alpha$ denote the extent to which the entries of the three-way and the two-way data set influence the analysis. Note that $\alpha = 1$ implies that the components for the common mode are determined on the basis of the three-way data only, whereas $\alpha = 0$ implies that only the two-way data are considered in estimating the common components.

Algorithm. To estimate the component matrices of the linked-mode PARAFAC-PCA model, an alternating least-squares algorithm is used. In this algorithm, starting from initial estimates each component matrix is alternately reestimated conditionally on the others until a specified stop criterion is satisfied (for an introduction to alternating least-squares algorithms, see ten Berge, 1993).

The initial estimates of the component matrices are determined either rationally or randomly. Rational initial estimates of $\mathbf{A}$, $\mathbf{B^1}$, and $\mathbf{C^1}$ are obtained by taking the P eigenvectors associated with the P largest eigenvalues of $[\mathbf{D_a^1}|\,\mathbf{D^2}]$, $\mathbf{D_b^1}$, and $\mathbf{D_c^1}$, respectively, with $\mathbf{D_a^1}$, $\mathbf{D_b^1}$, and $\mathbf{D_c^1}$ being the $I \times JK$, $J \times KI$, and $K \times IJ$ matricized versions of $\underline{\mathbf{D}}^{\mathbf{1}}$, and | denoting matrix concatenation (Kiers, 2000). Random initial estimates of $\mathbf{A}$, $\mathbf{B^1}$, and $\mathbf{C^1}$ are generated by sampling the random entries from a $U(-.5, .5)$ distribution and transforming these entries to obtain orthonormal components. Given rationally or randomly generated initial estimates of $\mathbf{A}$, $\mathbf{B^1}$, and $\mathbf{C^1}$, an initial estimate of $\mathbf{B^2}$ is obtained by regressing $\mathbf{A}$ on $\mathbf{D^2}$.

Once initial estimates of $\mathbf{A}$, $\mathbf{B^1}$, $\mathbf{C^1}$, and $\mathbf{B^2}$ have been obtained, the algorithm performs a number of estimation steps, called *iterations*, until a specified stop criterion is met. In each iteration, each component matrix is reestimated conditionally on the others. In particular, first the entries of $\mathbf{A}$ are reestimated while the entries of $\mathbf{B^1}$, $\mathbf{C^1}$, and $\mathbf{B^2}$ are kept fixed. Next, the entries of $\mathbf{B^1}$, $\mathbf{C^1}$, and $\mathbf{B^2}$ are consecutively reestimated while the other matrices remain fixed. The conditional reestimation of the component matrices boils down to solving a regression problem (see Smilde & Kiers, 1999).

Regarding the stop criterion of the alternating least-squares procedure, after each iteration we determine whether or not updating the entries of the component matrices results in a significant drop in the loss function value—that is, a decrease larger than or equal to a prespecified (tolerance) value. When the loss function value decreases significantly, a new iteration is performed. However, when this is not the case, the algorithm stops, since it has converged, and the current estimates of the component matrices are retained. As in PARAFAC analysis, the number of iterations before convergence can sometimes become very large; therefore, when the drop in the loss function value for the last iteration is small and a prespecified (maximal) number of iterations has been performed, it is advisable to stop the algorithm, even when the convergence criterion has not been reached.

Local minima. As is the case for the PARAFAC algorithm (Kroonenberg & de Leeuw, 1980), the linked-mode PARAFAC-PCA algorithm may easily end up in a local instead of a global minimum. To deal with this local minima issue, it is advisable to apply a multistart procedure in which the algorithm is run a prespecified number of times, each time from different randomly generated initial estimates. The solution with the lowest loss function value is retained.

2. Steps in a Linked-Mode PARAFAC-PCA Analysis

In this section, we discuss the five main steps of a linked-mode PARAFAC-PCA analysis: (1) preprocessing the data, (2) choosing the α weight, (3) determining the number of components, (4) studying the stability of the solution, and (5) interpreting the solution. These steps, shown in the flowchart in Figure 1, with arrows indicating the path to follow, will be illustrated by means of the analysis of a hypothetical coupled data set, which will be introduced first.

An important goal in the field of emotion psychology is to reveal the psychological mechanisms that underlie the individual differences in situation-specific emotional

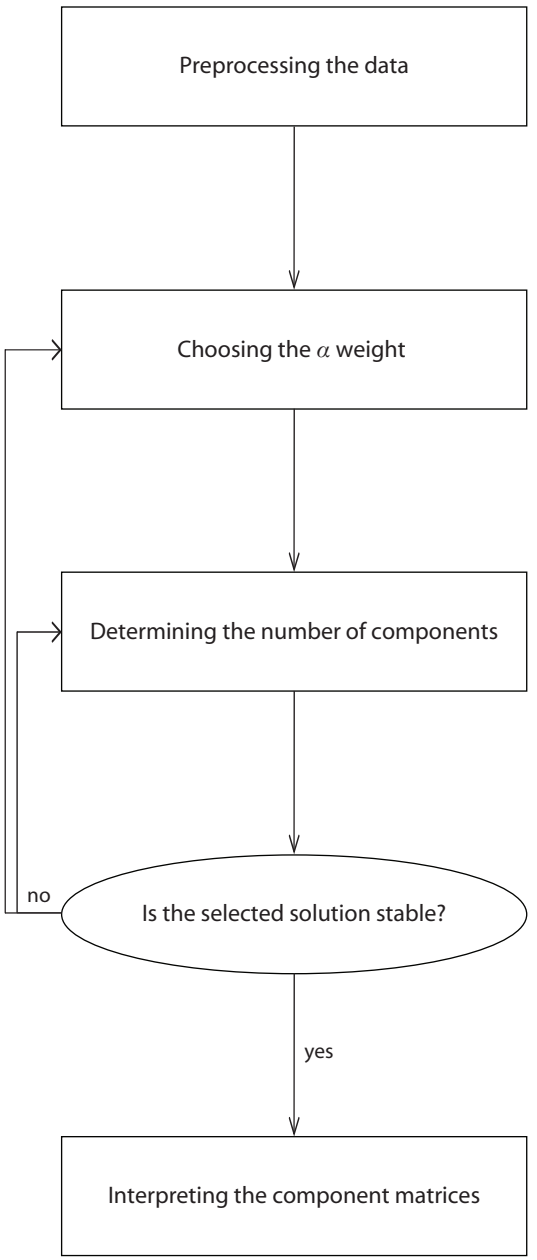


Figure 1. Flowchart of the different steps in a linked-mode PARAFAC-PCA analysis.

experience. To this end, several people may be asked to indicate to what extent they would experience a number of emotions in a set of situations, and to what extent they are characterized by various dispositions. A hypothetical 8 persons × 7 emotions × 6 situations data array and an 8 persons × 10 dispositions data matrix that might have resulted from such a data gathering are shown in the left and middle panels of Figure 2. In the left-hand panel, the rows correspond to the 48 person–situation combinations, the persons being nested within the situations (i.e., the rows in each block of data correspond to the persons, and the blocks correspond to the situations); the columns correspond to the emotions. In the middle panel, the rows of the two-way data matrix correspond to the (same) 8 persons and the columns to the 10 dispositions. Furthermore, labels for the persons, emotions, situations, and dispositions under study are displayed in the right-hand panel of Figure 2. Note that in this example, the persons, emotions, situations, and dispositions constitute the rows, columns, slices, and covariates of the linked-mode PARAFAC-PCA analysis, respectively.

2.1. Preprocessing the Data

After such standard preliminary analyses as inspecting frequency distributions, searching and eliminating outliers, dealing with missing data, and so on, and after assessing the need for a three-way analysis for the three-way data block (see Kiers & Van Mechelen, 2001), one may consider preprocessing the three-way and the two-way data. (Preprocessing consists of centering and/or normalizing the data in order to account for neutral points and/or artificial range differences pertaining to the measurement scales of the data.) In most cases, the two-way data are preprocessed by standardizing the variables (see Bro & Smilde, 2003), but different approaches are adopted for the three-way data (see Bro, 1997; Harshman & Lundy, 1994; Kiers & Van Mechelen, 2001; Kroonenberg, 1983, 2008; Smilde et al., 2004). In the remainder of this article, we assume that the data under study are preprocessed, when necessary.

2.2. Choosing the α Weight

In the second step, the weight α is determined. This weight may be chosen on the basis of a priori knowledge—for instance, information concerning the reliability of the data—or on the basis of results from previous studies. In most cases, however, no such a priori knowledge or previous results are available and a reasonable value for α may be determined by trying different alternatives. To choose among these alternatives, for example, cross-validation can be used (see Smilde & Kiers, 1999). In an extensive simulation study, Wilderjans et al. (2009) showed that setting $\alpha = .50$, which implies that all entries of $\underline{\mathbf{D}}^{\mathbf{1}}$ and $\mathbf{D}^{\mathbf{2}}$ contribute equally to the analysis, is a reasonable choice. Our data set will be analyzed with $\alpha = .50$.

2.3. Determining the Number of Components

Since in practice the number of components in a data set is almost never known beforehand, the next step is to determine this number. Therefore, analyses with an increasing number of components are usually performed. Subsequently, one selects the solution that has the best balance between the fit of the model to the data (i.e., the loss function value) and the complexity of the solution (i.e., the number of components). To this end, a model selection heuristic may be used. In the context of component analysis, the scree test of Cattell (1966) and generalized versions thereof (see Ceulemans & Kiers, 2006, 2009) are often used. The scree test consists of plotting the loss function values against the number of components and selecting the solution that lies on the sharpest elbow in

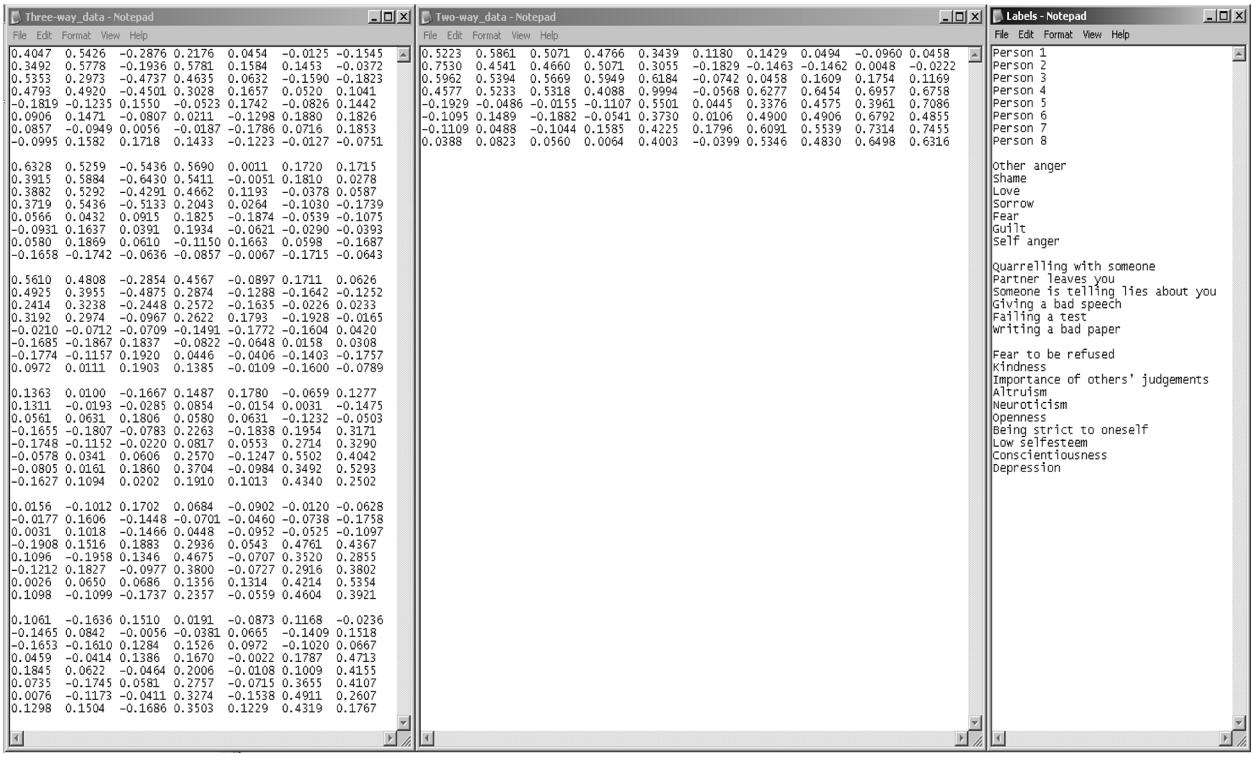


Figure 2. Screenshot of the data files ("Three-way_data.txt" and "Two-way_data.txt") and the file for the labels ("Labels.txt").

this plot. This solution has a good balance between fit to the data and complexity, since extracting one less component implies a large decrease in fit, whereas extracting one more component results in only a small gain in fit. Of course, the interpretability and the stability of the solution (see below) should also be taken into account. In Figure 3, which displays the scree plot for the emotion data set for the number of components ranging from one to eight, the solution with two components lies on the sharpest "elbow," suggesting the selection of the solution with two components.

2.4. Studying the Stability of the Solution

The next step is to determine whether the solution with the selected number of components is stable over trivial fluctuations in the sample, or over different samples from the same population. (Note that it is not always straightforward to define the population to which one wants to generalize.) To study the stability of a solution, a split-half procedure may be used in which the data are split into two parts of the same size when the entities of one mode are randomly partitioned, after which a linked-mode PARAFAC-PCA analysis is applied to both parts. When both analyses yield very similar component matrices for the other modes, the solution may be considered stable, where similarity of components may, for instance, be measured by means of Tucker's phi coefficient (Tucker, 1951). Another way to determine the stability of a solution is to use resampling techniques (see Efron & Tibshirani, 1993), like the bootstrap (for an application to multiway data, see Kiers, 2004) and the jackknife (for an application to PARAFAC, see Riu & Bro, 2003). When the solution does not seem to be stable, one should try solutions in other complexities (see section 2.3) and/or try different values for α (see section 2.2; note that in the flowchart in Figure 1, this is indicated by arrows that point upward).

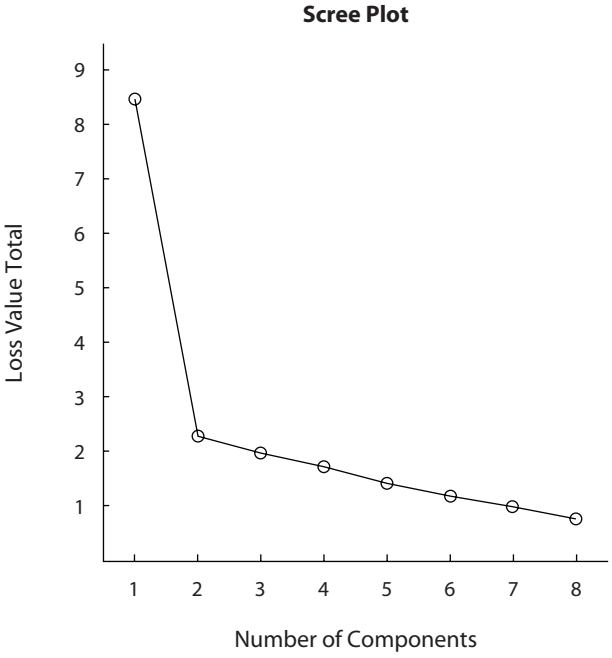


Figure 3. Scree plot for the emotion data set.

We performed a split-half analysis for the emotion data set by separating the odd from the even persons. All phi coefficients were larger than .97, which indicates that the selected solution is very stable.

2.5. Interpreting the Component Matrices

In the final step, the components obtained are interpreted by examining the content of the entities of the different modes that score high on the component under study. In Table 1, the component matrices of the analysis with two components are shown. When looking at the situation and emotion component scores (see "Slice Component Matrix" and "Column Component Matrix" in Table 1, respectively) for the emotion data set, the first component can be interpreted as an *interpersonal dimension*. Situations in which one is confronted with a relationship problem, such as a partner leaving or somebody telling lies about you, score high on this component. These situations elicit emotions, such as shame and anger, that can be experienced only in the (imaginary) presence of others. In contrast, the second component can be conceived of as an *intrapersonal dimension*, because situations that imply a personal failure (e.g., writing a bad paper) and emotions that force a person to confront himself (e.g., guilt and self-contempt) score high on this component. When interpreting the person components (see "Row Component Matrix" in Table 1), one can see that the first three persons score high only on the interpersonal dimension, whereas the last four persons score high only on the intrapersonal dimension; this implies that the former react neutrally to intrapersonal situations and the latter react neutrally to interpersonal situations. (Note that Person 4 shows emotion in both intra- and interpersonal situations.)

To gain insight into the psychological mechanisms that underlie the obtained individual differences in situation-specific emotional experience, one can relate these individual differences to the dispositional information. From the disposition component matrix (see "Covariate Component Matrix" in Table 1) it can be seen that persons who are more other-oriented—that is, altruistic and responsive to others' judgments—mainly react emotionally to interpersonal situations. Furthermore, it seems that situations where one is confronted with oneself elicit intrapersonal emotions from persons who are more self-oriented only (i.e., depression and being strict to oneself).

Table 1
Linked-Mode PARAFAC-PCA Solution With Two Components for the Emotion Data Set

	Component 1	Component 2
Row Component Matrix		
Person 1	1.53	−0.06
Person 2	1.57	−0.24
Person 3	1.48	0.04
Person 4	1.32	1.04
Person 5	−0.16	0.89
Person 6	−0.05	1.01
Person 7	−0.09	1.17
Person 8	−0.05	0.99
Column Component Matrix		
Other anger	0.51	−0.03
Shame	0.55	−0.02
Love	−0.47	0.02
Sorrow	0.46	0.44
Fear	0.04	−0.04
Guilt	0.02	0.62
Self anger	−0.02	0.65
Slice Component Matrix		
Quarrelling with someone	0.57	0.10
Partner leaves you	0.67	−0.11
Someone is telling lies about you	0.47	−0.09
Giving a bad speech	0.05	0.55
Failing a test	0.02	0.63
Writing a bad paper	−0.07	0.52
Covariate Component Matrix		
Fear to be refused	0.41	−0.06
Kindness	0.35	0.09
Importance of others' judgments	0.35	−0.01
Altruism	0.34	0.02
Neuroticism	0.32	0.48
Openness	−0.04	0.04
Being strict to oneself	0.05	0.51
Low selfesteem	0.05	0.51
Conscientiousness	0.05	0.61
Depression	0.05	0.62

3. The LMPCA Program

The LMPCA program consists of the MATLAB figure file "LMPCA_gui" and a set of MATLAB m-files. To start the program, all these files should be stored in the same folder, where the current MATLAB directory should be sent. The program then can be launched in MATLAB by typing *LMPCA_gui* at the command prompt >>LMPCA_gui <ENTER>.

As a result, a graphical user interface (GUI; see Figure 4, with the boxes of the GUI already containing the information regarding the guiding emotion data set of the previous section) appears that is subdivided in three compartments—data description and data files, analysis options, and output files—which allow the user to control the analysis. For users not experienced with MATLAB, or without access to it, a stand-alone application is provided (for installing this application, see the instructions file "ReadMe_Standalone.txt" at the Web site). After the user double-clicks on the *LMPCA_gui.exe* icon, the same GUI appears (Figure 4). In the subsequent sections, the three compartments will be outlined, closing off with error handling.

3.1. Data Description and Data Files

Data description. In this panel, the user has to enter information regarding the size of the three-way three-mode data array and the two-way two-mode data matrix that have to be analyzed. In particular, the user has to specify the number of rows, columns, slices, and covariates. For instance, to analyze our hypothetical data set, we specify (see Figure 4) that there are 8 rows (persons), 7 columns (emotions), 6 slices (situations), and 10 covariates (dispositions). In the LMPCA program, each mode may maxi-

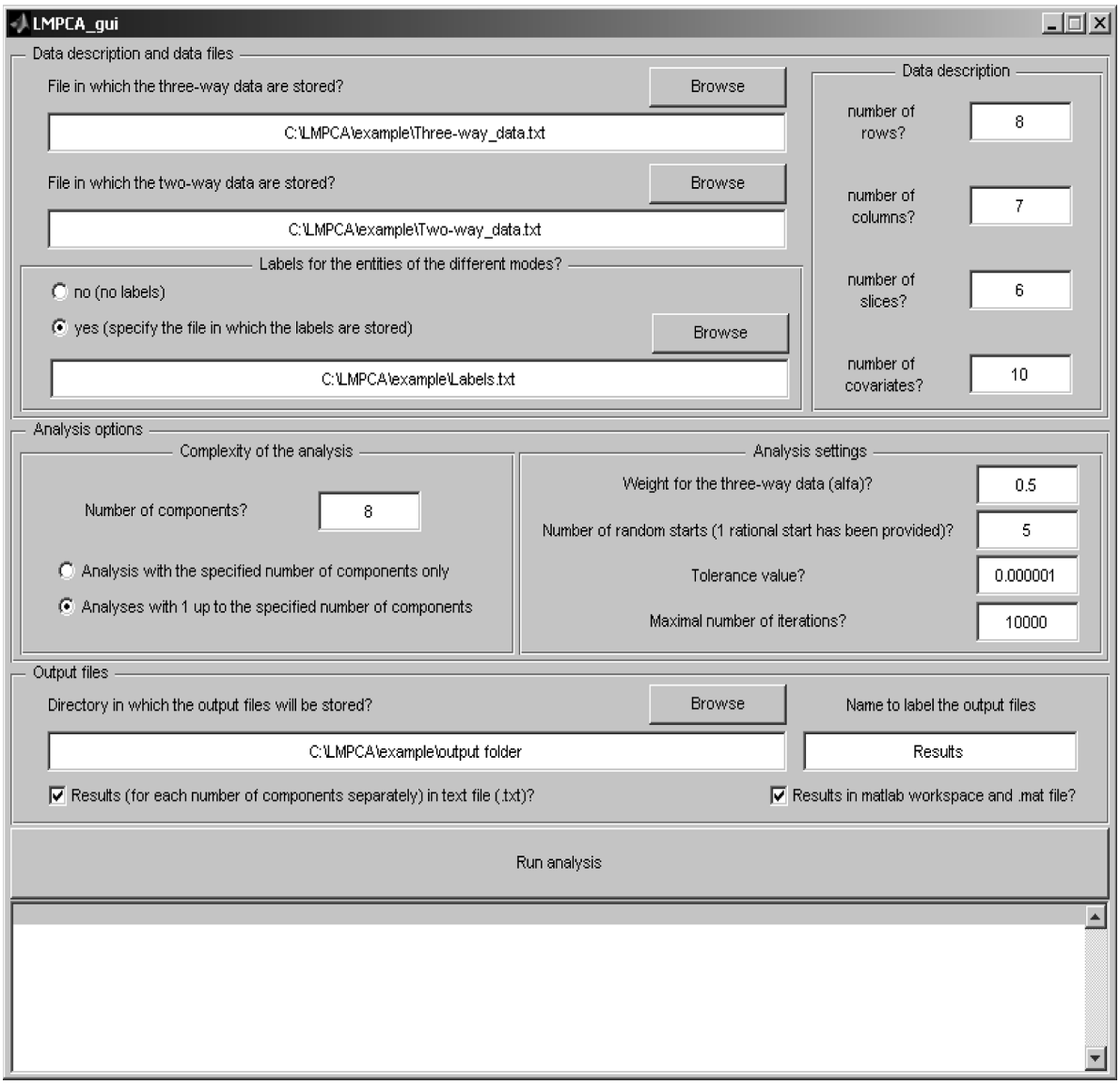


Figure 4. Screenshot of the graphical user interface of the LMPCA program.

mally contain 10,000 entities, which is large enough for data from the behavioral sciences.

Data files. In this panel, the user has to specify the files that contain the three-way and the two-way data that have to be analyzed. This is done by clicking the appropriate "Browse" button and selecting the files. The LMPCA program accepts only ASCII files (i.e., .txt files) that are organized as follows: The three-way data file (the three-way data file for the emotion data set is displayed in the left-hand panel of Figure 2) consists of a number of data blocks that may be separated by one or more empty lines. Each of these data blocks corresponds to a slice and should consist of as many rows and columns as specified in the data description panel, with no empty lines being allowed (i.e., separated by line breaks only). The two-way data file (see the middle panel in Figure 2 for the file that contains the two-way data of the emotion data set) should contain as many rows and covariates as specified in the data description panel, with no empty lines being allowed (i.e., line breaks only). In both files, the data elements may be separated by one or more spaces, commas, semicolons, or tabs, or any combination thereof. Each data element should be an integer or a real number, with decimal separators being a period, *not* a comma! The LMPCA program cannot handle complex numbers and missing values.

Label file. Optionally, user-defined labels for the entities of all modes may be provided. This is done by checking the option "yes (specify the file in which the labels are stored)" and by subsequently browsing for the file containing the labels. This file should again be an ASCII file (.txt) that contains four blocks of labels (i.e., row labels,

column labels, slice labels, and covariate labels, respectively), that may be separated by one or more empty lines. Within each block, the labels are separated by line breaks. Obviously, the number and ordering of the labels within the block have to correspond to the number and ordering of the entities in the data files. The labels are character strings that may contain any kind of symbol; for instance, the right-hand panel of Figure 2 shows the label file for the emotion data set.

If the user does not want to provide labels for the entities of the different modes, "no (no labels)" should be checked. As a consequence, the program will generate default row labels of the form "Row1", "Row2", and so on. Similar labels are created for the columns, slices, and covariates.

3.2. Analysis Options

Complexity of the analysis. In this panel, the user has to specify how many components should be (maximally) extracted. This number should be an integer between one and *min*(20,*IJ*,*IK*,*JK*), which is large enough, since we are usually interested in only a few components. Furthermore, the user should decide whether only an analysis of the specified complexity is performed, or whether different analyses with the number of components going from one to the specified number of components should be run. The former can be achieved by checking the "Analysis with the specified number of components only" option, and the latter is done by selecting the "Analyses with 1 up to the specified number of components" option. Since, in general, the user does not know how many components are in the data, the latter option will usually be selected, as in Figure 4 (see section 2.3). Note that the computation time of the algorithm will increase with the complexity of the analysis.

Analysis settings. The user needs to specify the weight for the three-way data (i.e., α). This weight, which by default is .5, should be a real number between 0 and 1 (see section 2.2). Furthermore, to deal with the local minima issue, the LMPCA program has a built-in multistart procedure consisting of a rational start and a number of random starts (see "Algorithm" under section 1.2). In the "Analysis settings" panel, the user has to choose how many random starts should be included. This number should be an integer between 0, implying that only a rational start is performed, and 10,000. The user may further manipulate the stopping rule of the alternating least-squares algorithm by setting the tolerance value, which should be a real number between 0 and .1, and the maximal number of iterations (see "Algorithm" under section 1.2). Note that the algorithm may not have converged when the specified (maximal) number of iterations is reached. One should take into account that increasing the number of random starts, increasing the maximal number of iterations, and decreasing the tolerance value will improve the quality of the obtained solution (i.e., will lead to lower loss function values), but will also lengthen the computation time of the algorithm. In the LMPCA program, the default number of random starts is 5, the default tolerance value equals 0.000001, and the default maximal number of iterations amounts to 10,000.

3.3. Output Files

In the "Output files" compartment (see Figure 4) the user specifies the directory where all the standard output files (in .mht format) of the LMPCA program will be stored. To this end, the user clicks the "Browse" button and selects a folder on the computer. Furthermore, the user has to choose a string with which the name of all output files will start. In order to further process the results, the user can, optionally, ask for extra output files in .txt format, which can be easily loaded into any popular software package such as SAS and SPSS. The user can also ask to store the results in .mat format and into the MATLAB workspace. These optional output files are also stored in the selected output folder. In Figure 4, it is obvious that in the guiding example, the folder *C:\LMPCA\example\output folder* was selected as the output folder, and the name of all output files will start with "Results".

Standard output files (.mht format). In the output folder, .mht files are created where names start with the string specified in the "Output files" compartment and contains the number of components. For instance, the results for the guiding example can be found in the files "Results_1component.mht", "Results_2components.mht", and so on. In these files the component matrices for the different modes are displayed, together with some fit information. If analyses of varying complexities were obtained ("Analyses with 1 up to the specified number of components" option in the "Analysis options" compartment), an overview file of the results of the different analyses is created (for the emotion data set, this file is labeled "Results_overview.mht"). In particular, fit information for the different analyses is provided, together with a scree plot.

Extra output files (.txt and .mat format). When the user asks for extra output files in .txt format, the obtained solutions are stored in files whose names start with the string specified and end with "_1component.txt", "_2components.txt", and so on (e.g., "Results_1component .txt", etc.). When the user selected the "Results in Matlab workspace and .mat file?" option, an object with the name "LMPCAsolution" is stored in the MATLAB workspace and saved in a .mat file with the specified name (e.g., "Results.mat"). If the analysis were run in one complexity only—the "Analysis with the specified number of components only" option—*LMPCAsolution* would be a structure with the following fields: (1) *DataDescription*, (2) *AnalysisOptions*, (3) *ComponentMatrices*, and (4) *FitInformation*. If analyses of varying complexities were obtained ("Analysis with 1 up to specified number of components" option in the "Analysis options" compartment), *LMPCAsolution* would be a cell array with as many cells as the specified number of components. Each cell contains a structure, consisting of the fields mentioned above, that can be accessed by typing in MATLAB "Solution.NumberOfComponents{*desired number of components*}".

3.4. Status of the Analysis and Error Handling

After specifying the data description, the data files, the analysis options, and the output files, the user clicks the

"Run analysis" button in order to start the analysis. During the analysis, information concerning the status of the analysis will be displayed in the box at the bottom of the GUI screen. A screen will notify the user when the analysis is done. After clicking the "OK" button, the user can consult the results in the output files stored in the selected folder.

When the data or label files are incorrectly specified, or do not comply with the format mentioned above, one or more error screens will appear with information about the problems encountered. When an error message appears, the analysis is deferred and, after clicking the "OK" button(s), the user can correct the input files or the analysis specifications. To this end, the content of the error messages will also be displayed in the box at the bottom of the GUI screen. After correcting the information, the user should again click the "Run analysis" button.

4. Summary

In this article, we have discussed and illustrated the different steps in a linked-mode PARAFAC-PCA analysis and presented a program called LMPCA to perform such an analysis. The LMPCA program is a MATLAB graphical user interface (a standalone version is also provided) that is freely available and easy to use, and allows the user to flexibly steer the analysis. Furthermore, the program assists the user in determining the number of components present in the data. Finally, the solutions obtained and the associated fit information can be stored in different file formats, so that the results can be easily loaded into popular statistical programs like SAS and SPSS for further processing.

AUTHOR NOTE

T.F.W. is a research assistant at the Fund for Scientific Research, Flanders (Belgium). The research reported in this article was partly supported by the Research Council of the University of Leuven (GOA/2005/04 and EF/2005/07, "SymBioSys") and by IWT-Flanders (SBO 60045, "Bioframe"). Correspondence concerning this article should be addressed to T. F. Wilderjans, Department of Psychology, University of Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium (e-mail: tom.wilderjans@psy.kuleuven.be).

REFERENCES

Belk, R. W. (1974). An exploratory assessment of situational effects in buyer behavior. *Journal of Marketing Research*, **11**, 156-163.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics & Intelligent Laboratory Systems*, **38**, 149-171. doi:10.1016/S0169-7439(97)00032-4

Bro, R. (1998). *Multi-way analysis in the food industry: Models, algorithms, and applications*. Unpublished doctoral dissertation, University of Amsterdam.

Bro, R., Andersson, C. A., & Kiers, H. A. L. (1999). PARAFAC2: Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics*, **13**, 295-309.

Bro, R., & Kiers, H. A. L. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, **17**, 274-286. doi:10.1002/cem.801

Bro, R., & Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, **17**, 16-33. doi:10.1002/cem.773

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. doi:10.1007/BF02310791

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

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 & Statistical Psychology*, **59**, 133-150. doi:10.1348/000711005X64817

Ceulemans, E., & Kiers, H. A. L. (2009). Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical & Statistical Psychology*, **62**, 601-620.

de Jong, S., & Kiers, H. A. L. (1992). Principal covariates regression: Part I. Theory. *Chemometrics & Intelligent Laboratory Systems*, **14**, 155-164.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.

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

Harshman, R. A., & De Sarbo, 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. (1984). 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.

Harshman, R. A., & Lundy, M. E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis*, **18**, 39-72.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417-441.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321-377.

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

Kiers, H. A. L. (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics*, **18**, 22-36. doi:10.1002/cem.841

Kiers, H. A. L., & Van Mechelen, I. (2001). Three-way component analysis: Principles and illustrative application. *Psychological Methods*, **6**, 84-110. doi:10.1037/1082-989X.6.1.84

Kroonenberg, P. M. (1983). *Three-mode principal component analysis*. *Theory and applications*. Leiden: DSWO Press.

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

Kroonenberg, P. M., & de Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, **45**, 69-97. doi:10.1007/BF02293599

Kuppens, P., Van Mechelen, I., & Rijmen, F. (2008). Toward disentangling sources of individual differences in appraisal and anger. *Journal of Personality*, **76**, 969-1000. doi:10.1111/j.1467-6494.2008.00511.x

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Reviews in Psychology*, **49**, 229-258.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, **2**, 559-572.

Riu, J., & Bro, R. (2003). Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemometrics & Intelligent Laboratory Systems*, **65**, 35-49. doi:10.1016/S0169-7439(02)00090-4

Sidiropoulos, N. D., & Bro, R. (2000). On the uniqueness of multilinear decomposition of *n*-way arrays. *Journal of Chemometrics*, **14**, 229-239.

Smilde, A. K., Bro, R., & Geladi, P. (2004). *Multi-way analysis with applications in the chemical sciences*. Chichester, U.K.: Wiley.

Smilde, A. K., & Kiers, H. A. L. (1999). Multi-way covariates regression models. *Journal of Chemometrics*, **13**, 31-48.

SMILDE, A. K., WESTERHUIS, J. A., & BOQUÉ, R. (2000). Multiway multiblock component and covariates regression models. *Journal of Chemometrics*, **14**, 301-331.

SPIELBERGER, C. D., CRANE, R. S., KEARNS, W. D., PELLEGRIN, K. L., RICKMAN, R. L., & JOHNSON, E. H. (1991). Anger and anxiety in essential hypertension. In C. D. Spielberger, I. G. Sarason, Z. Kulcsar, & G. L. Van Heck (Eds.), *Stress and emotion: Anxiety, anger, and curiosity* (Vol. 14, pp. 265-283). Washington, DC: Hemisphere.

SPIELBERGER, C. D., JOHNSON, E. H., RUSSELL, S. F., CRANE, R. J., JACOBS, G. A., & WORDEN, T. J. (1985). The experience and expression of anger: Construction and validation of an anger expression scale. In M. A. Chesney & R. H. Rosenman (Eds.), *Anger and hostility in cardiovascular and behavioral disorders* (pp. 5-30). Washington, DC: Hemisphere.

TEN BERGE, J. M. F. (1993). *Least squares optimization in multivariate analysis*. Leiden: DSWO Press.

TUCKER, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Rep. No. 984). Washington, DC: Department of the Army.

WILDERJANS, T. F., CEULEMANS, E., & VAN MECHELEN, I. (2008). The CHIC model: A global model for coupled binary data. *Psychometrika*, **73**, 729-751. doi:10.1007/S11336-008-9069-9

WILDERJANS, T. F., CEULEMANS, E., & VAN MECHELEN, I. (2009). Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics & Data Analysis*, **53**, 1086-1098. doi:10.1016/j.csda.2008.09.031

WOOD, L. M. (2004). Dimensions of brand purchasing behaviour: Consumers in the 18–24 age group. *Journal of Consumer Behaviour*, **4**, 9-24.

(Manuscript received February 24, 2009;
revision accepted for publication April 17, 2009.)