

# SCA with rotation to distinguish common and distinctive information in linked data

Martijn Schouteden · Katrijn Van Deun ·  
Sven Pattyn · Iven Van Mechelen

Published online: 30 January 2013  
© Psychonomic Society, Inc. 2013

**Abstract** Often data are collected that consist of different blocks that all contain information about the same entities (e.g., items, persons, or situations). In order to unveil both information that is common to all data blocks and information that is distinctive for one or a few of them, an integrated analysis of the whole of all data blocks may be most useful. Interesting classes of methods for such an approach are simultaneous-component and multigroup factor analysis methods. These methods yield dimensions underlying the data at hand. Unfortunately, however, in the results from such analyses, common and distinctive types of information are mixed up. This article proposes a novel method to disentangle the two kinds of information, by making use of the rotational freedom of component and factor models. We illustrate this method with data from a cross-cultural study of emotions.

**Keywords** Simultaneous component analysis · Multigroup factor analysis · Rotation · Common information · Distinctive information

In the behavioral sciences, data are often collected in blocks that all contain information about the same set of entities (e.g., tests, items, persons, or situations; Curran & Hussong, 2009; Van Mechelen & Smilde, 2010). An example of data comprising different blocks that contain information about the same set of tests can be found in the field of clinical psychology, where Spikman, Kiers, Deelman, and van Zomeren (2001)

investigated the concept of attention in healthy controls and patients with a closed head injury, by subjecting them to a series of neuropsychological tests. In this way, Spikman et al. obtained a data set consisting of two person-by-test data blocks, one for each group under study, that contained information about the same set of tests (for a graphical representation, see Fig. 1a). An example of data consisting of blocks with information about the same set of persons instead of the same set of items can be found in the field of personality psychology. In that area, Rossier, de Stadelhofen, and Berthoud (2004) compared aspects of personality measured by two personality questionnaires. In this way, Rossier et al. obtained a data set consisting of two person-by-item data blocks, one per questionnaire, with information about the same set of persons (see Fig. 1b). In the remainder of this report, data sets consisting of different data blocks containing information about the same set of entities will be called *multiblock* or *linked* data (for a formal description of linked data and a conceptual framework, see Van Mechelen & Smilde, 2010).

A main challenge in the analysis of linked data is to reveal the mechanisms underlying the different data blocks. However, a still greater challenge is to reveal, on the one hand, the mechanisms underlying all data blocks under study, and, on the other, mechanisms underlying a single data block or a few such blocks only (i.e., common and distinctive mechanisms, respectively).

For example, Spikman et al. (2001) investigated whether patients with a closed head injury rely on the same attentional mechanisms as healthy controls. In the personality example referred to above, Rossier et al. (2004) wanted to investigate which aspects of personality were measured by the two questionnaires in their study, as well as which aspects were measured by only one of the two, and not by the other.

Principal-component and factor-analytic methods are an obvious choice to reveal the mechanisms underlying object-by-variable data. A strategy to apply these methods in a multi-block setting could be first to analyze each of the data blocks

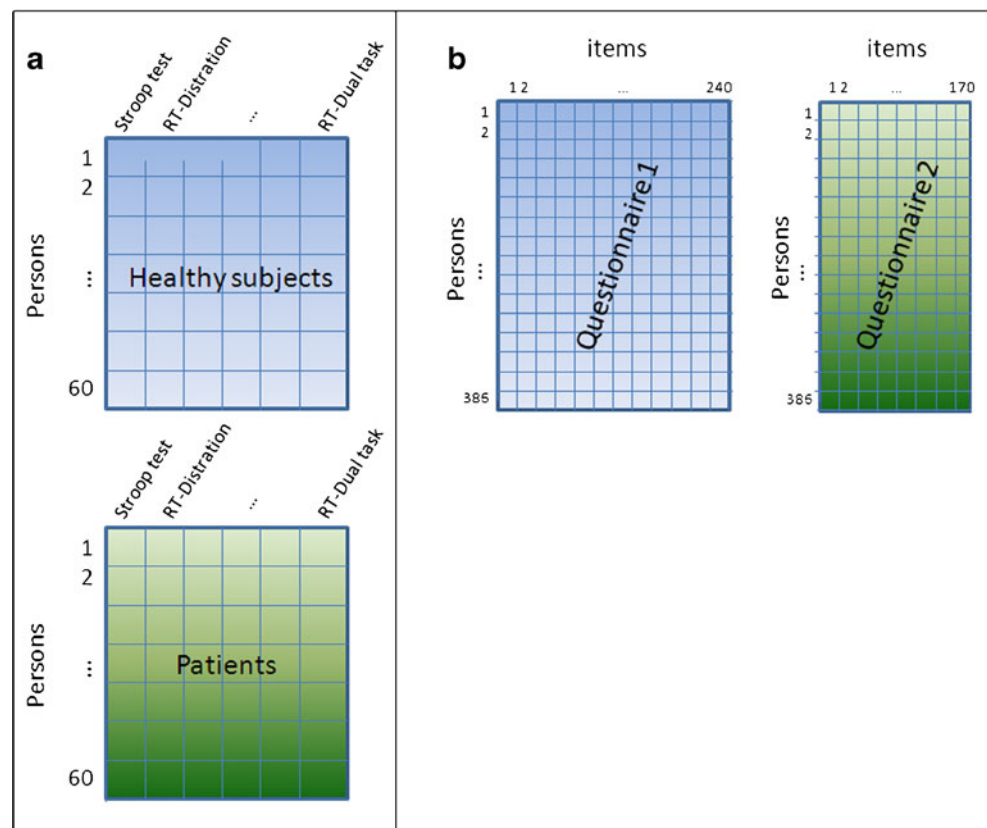
---

M. Schouteden · K. Van Deun · I. Van Mechelen (✉)  
Research Group for Quantitative Psychology and  
Individual Differences, KU Leuven, Leuven, Belgium  
e-mail: iven.vanmechelen@ppw.kuleuven.be

M. Schouteden  
e-mail: martijn.schouteden@psy.kuleuven.be

S. Pattyn  
Department of Developmental, Personality and Social  
Psychology, Universiteit Gent, Ghent, Belgium

**Fig. 1** Graphical representation of data consisting of different data blocks with information about (a) the same set of tests and, respectively, (b) the same set of persons



separately, and subsequently to tie the results of the block-specific analyses together—for example, by means of congruence coefficients or (generalized) Procrustes techniques (ten Berge, 1977). However, the results of such an approach are not unequivocal. For instance, low congruence between the components/factors of a first data block and those of the other data blocks does not guarantee that the components/factors of the first data block do not account for a sizeable amount of variance in the other data blocks. Therefore, a multigroup component or factor-analytic method may be more suitable (Kiers & ten Berge, 1994); in the case that no prior knowledge about the underlying structure is available, an exploratory version of these methods is preferable.

Examples of exploratory multigroup methods are the family of simultaneous-component analysis (SCA) methods (Kiers & ten Berge, 1989; Millsap & Meredith, 1988; ten Berge, Kiers, & Van der Stel, 1992; for a recent review, see Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen, 2009). SCA is a family of component methods that have been developed for the analysis of linked data. SCA methods typically reveal a small number of simultaneous components that maximally account for the variation in the data set. These methods have already been applied in a broad range of domains (see, e.g., Caprara, Barbaranelli, Bermudez, Maslach, & Ruch, 2000; Hagedoorn, Van Yperen, Van De Vliert, & Buunk, 1999; Silva, Martinez-Arias, Rapaport, Ertle, & Ortet, 1997; Spikman et al., 2001).

When looking for common and distinctive mechanisms underlying multiblock data, one may wish to make an appeal to SCA. Unfortunately, SCA may yield a solution that adequately describes none of the data blocks. In particular, simultaneous components usually reflect a mix of common and distinctive information. Up to now, techniques to disentangle these kinds of information have been lacking. In this report, we present a novel technique to solve this problem, called DISCO-SCA (which is short for “distinctive and common components with simultaneous-component analysis”).

Other possible methods for a simultaneous analysis of linked data include multigroup factor analysis (FA) and structural equation models (SEM), with exploratory multigroup FA and SEM (Asparouhov & Muthén, 2009; Marsh et al., 2010; Marsh et al., 2009) and confirmatory multiple-group FA (Byrne, 2001; Jöreskog, 1971) as important special cases. However, as in SCA, the factors included in these models reflect a mixture of common and distinctive information. In exploratory multigroup FA and SEM, no ready-made solution is available to solve this issue. In their confirmatory counterpart, however, one might consider investigating some form of distinctive factors by allowing the same group of indicator variables to load on a factor in all data blocks under study, and by subsequently testing whether the loadings differ from zero in one data block, unlike in all of the other data blocks. Yet, such an endeavor requires prior

hypotheses about which groups of indicator variables might constitute common and distinctive factors. In the absence of such knowledge, DISCO-SCA could be most useful in identifying such hypotheses.

The remainder of this article is organized as follows: The basic principles of simultaneous-component methods are discussed briefly in the following section. Next, the newly proposed DISCO-SCA technique is presented. Subsequently, DISCO-SCA is applied to data stemming from cross-cultural emotion research. In this application, we also show how DISCO-SCA could contribute to building a meaningful confirmatory multigroup FA model with common and distinctive factors. We conclude with a discussion.

### Simultaneous-component methods

In this section, simultaneous component (SC) methods are briefly discussed (Van Deun et al., 2009). For this purpose, we first introduce the type of data sets to which SC models can be applied, and then we summarize the actual SC model and the associated data analysis.

#### The data

We consider linked data that consist of different data matrices that pertain to a common set of elements. By way of convention, we will further refer to the row elements of the matrices as the *objects* and to the column elements as the *variables*. Two cases can be distinguished, depending on whether the data matrices have the set of objects or the set of variables in common. For instance, in the example of the mood study mentioned in the previous section, the data blocks have the variables (i.e., the items) in common, while in the personality example, the objects (i.e., the persons) constitute the common mode. To ease the explanation, below we focus on SC models for the case of two data matrices with a common variable mode. In the [Extensions](#) section below, we will briefly outline SC models for the case of two data matrices with a common object mode.

#### Model and data analysis

Let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  denote two data matrices with dimensions  $I_1 \times J$  and  $I_2 \times J$ , respectively. Then, an SC model with  $R \leq J$  components takes the following form:

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{bmatrix} \mathbf{P}' + \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix}, \quad (1)$$

subject to  $[\mathbf{T}_1' \mathbf{T}_2'] [\mathbf{T}_1' \mathbf{T}_2']' = \mathbf{I}$ , with  $\mathbf{T}_1$  and  $\mathbf{T}_2$  being  $I_1 \times R$  and  $I_2 \times R$  matrices of component scores;  $\mathbf{P}$ , a  $J \times R$  loading matrix; and  $\mathbf{E}_1$  and  $\mathbf{E}_2$ ,  $I_1 \times J$  and  $I_2 \times J$  matrices of

residuals. Note that in SCA, unlike in separate principal-component analyses of all data blocks, the *same* loading matrix  $\mathbf{P}$  is used for all data blocks. This guarantees that the components are the same in each data block. Equation 1 is further equivalent to

$$\mathbf{X}_{conc} = \mathbf{T}_{conc} \mathbf{P}' + \mathbf{E}_{conc},$$

subject to  $\mathbf{T}_{conc}' \mathbf{T}_{conc} = \mathbf{I}$ , with  $\mathbf{X}_{conc}$  denoting the  $(I_1 + I_2) \times J$  matrix that is obtained by concatenating the two data matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$ ;  $\mathbf{T}_{conc}$ , the  $(I_1 + I_2) \times R$  matrix of component scores resulting from the concatenation of  $\mathbf{T}_1$  and  $\mathbf{T}_2$ ; and  $\mathbf{E}_{conc}$ , the  $(I_1 + I_2) \times J$  matrix of residuals resulting from the concatenation of  $\mathbf{E}_1$  and  $\mathbf{E}_2$  (for a graphical representation of SCA, see Fig. 2, upper panels).

The matrices  $\mathbf{T}_{conc}$  and  $\mathbf{P}$  can be estimated by minimizing the following least squares function:

$$\|\mathbf{X}_{conc} - \mathbf{T}_{conc} \mathbf{P}'\|^2. \quad (2)$$

This can be achieved by means of a singular value decomposition of  $\mathbf{X}_{conc}$ ,

$$\mathbf{X}_{conc} = \mathbf{U} \mathbf{S} \mathbf{V}',$$

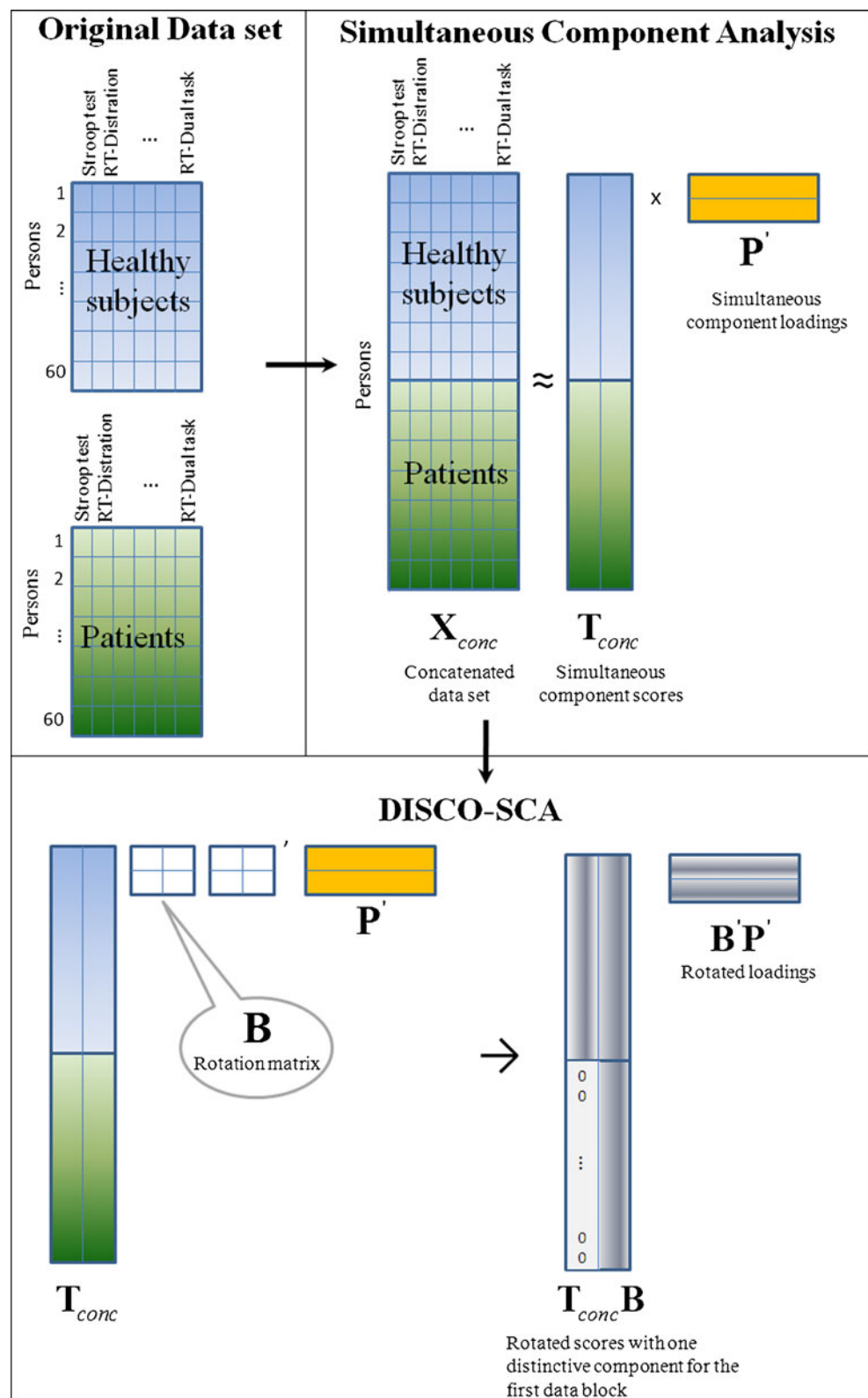
with  $\mathbf{U}$  and  $\mathbf{V}$  denoting, respectively, the left and right singular vectors that are subject to  $\mathbf{U}' \mathbf{U} = \mathbf{I} = \mathbf{V}' \mathbf{V}$ , and  $\mathbf{S}$  being a diagonal matrix with nonnegative real numbers on the diagonal (i.e., the singular values) ranked from largest to smallest. For an  $R$ -component solution, the component score matrix  $\mathbf{T}_{conc}$  and the loading matrix  $\mathbf{P}$  equal

$$\begin{cases} \mathbf{T}_{conc} = \mathbf{U}_R, \\ \mathbf{P} = \mathbf{V}_R \mathbf{S}_R, \end{cases}$$

with  $\mathbf{U}_R$  and  $\mathbf{V}_R$  denoting, respectively, the first  $R$  left and right singular vectors, and  $\mathbf{S}_R$  being a diagonal matrix that contains the first  $R$  diagonal elements of  $\mathbf{S}$ . The scores (or, respectively, loadings) are further usually multiplied (respectively, divided) by the square root of the number of observations.

Prior to an SC analysis, the data are usually preprocessed. This may be done to correct for differences in the offsets and scales of the variables (e.g., due to the fact that the variables are expressed in different measurement units). One possible form of preprocessing is to center the variables within each block and/or to scale them across all data blocks to a sum of squares of 1. Furthermore, in case that the data blocks differ considerably in size, the results may be dominated by the largest data block. A possible strategy then could be to scale each data block to a sum of squares of 1. Another possible strategy could be to divide each data block by its largest singular value, which will correct for both size and redundancy. See Van Deun et al. (2009) for more information about the preprocessing and weighting of data blocks.

**Fig. 2** Graphical representation of DISCO-SCA for the case of variable-wise linked data and two components. The top left panel shows the different data blocks; the top right panel, a simultaneous-component analysis (SCA) on the concatenated data; and the lower panel, a DISCO rotation of the component scores, with rotation matrix **B**, toward a common and distinctive structure, so that the rotated component loadings and scores then equal **PB** and  $T_{conc}B$ , respectively



## DISCO-SCA

As mentioned earlier, simultaneous components usually reflect a mix of common and distinctive information. In this

section, a novel technique is outlined by which the two kinds of information can be disentangled. This technique can be applied when the data blocks have variables in common and when they have objects in common. To ease

the explanation, below we will outline the technique for the case of two data matrices with a common variable mode. In the [Extensions](#) section, we will briefly outline the technique for the case of matrices with a common object mode.

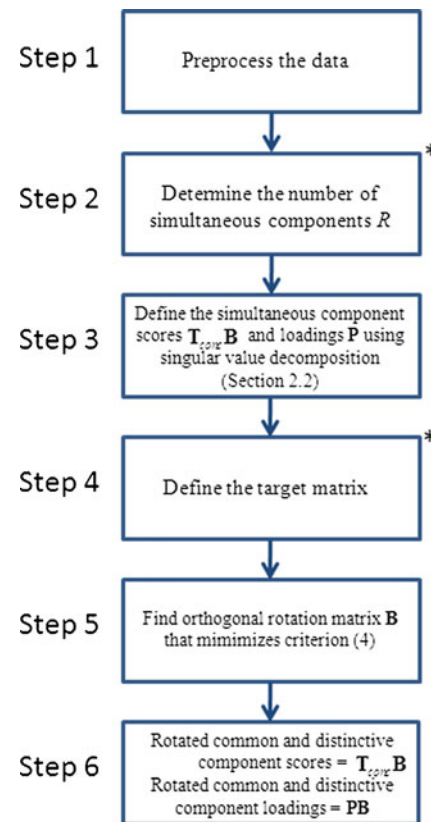
### The model

For two data matrices with a common variable mode, a distinctive mechanism is defined as a component for which the scores equal 0 for the data block in which that distinctive component does not play a part; the scores of the common components do not contain such zero parts. However, the SC scores that result from the minimization of loss function (Eq. 2) usually do not contain such a clear common/distinctive structure. As a result, these components capture a mix of common and distinctive information.

To disentangle these kinds of information, we propose DISCO-SCA. This method takes advantage of the rotational freedom of the simultaneous components and orthogonally rotates the component scores as close as possible toward a clear common/distinctive structure (which will further be called the *target structure*  $\mathbf{T}_{conc}^{target}$ ). For instance, consider a case with two data blocks and three simultaneous components, with the first one being distinctive for the first data block, the second one distinctive for the second data block, and the third one common. In that case, the target matrix  $\mathbf{T}_{conc}^{target}$ , consisting of the concatenated subtarget matrices  $\mathbf{T}_1^{target}$  and  $\mathbf{T}_2^{target}$  that refer to the two data blocks, reads as follows:

$$\mathbf{T}_{conc}^{target} = \begin{bmatrix} \mathbf{T}_1^{target} & & \\ - & - & - \\ & \mathbf{T}_2^{target} & \end{bmatrix} = \begin{bmatrix} * & 0 & * \\ \vdots & \vdots & \vdots \\ * & 0 & * \\ - & - & - \\ 0 & * & * \\ \vdots & \vdots & \vdots \\ 0 & * & * \end{bmatrix}, \quad (3)$$

where \* denotes an unspecified entry. A graphical representation of DISCO-SCA is given in Fig. 2 for the example of variable-wise linked data and two components. Note that, in order to specify a suitable target matrix to rotate the component scores to, the number of components and the characterization of each component (as either distinctive for one or more particular data blocks, or common) have to be specified. Importantly, this specification is part of the actual data analysis (for more information, see the next section, on model selection). This means that no prior knowledge of the number and characterization of the components is required, and that, as such, DISCO-SCA can be considered a fully exploratory method. A flow chart of the full DISCO-SCA process is given in Fig. 3.



**Fig. 3** Flow chart of the different steps in DISCO-SCA. Each \* indicates issues that are addressed in the [Model Selection](#) section

The rotation matrix  $\mathbf{B}$  to rotate the component scores as close as possible toward the target structure can be found by making use of the following objective function:

$$\min_{\mathbf{B}} \|\mathbf{W} \circ (\mathbf{T}_{conc} \mathbf{B} - \mathbf{T}_{conc}^{target})\|^2, \quad (4)$$

subject to  $\mathbf{B}' \mathbf{B} = \mathbf{I}$ ; the matrix  $\mathbf{W}$  denotes a weight matrix, with ones in positions that contain the specified zeroes in  $\mathbf{T}_{conc}^{target}$ , and zeroes elsewhere (Browne, 1972), and  $\circ$  denotes the element-wise or Hadamard product. The rotated component loadings, which are the same for both data blocks, then equal  $\mathbf{PB}$ ; the rotated scores equal  $\mathbf{T}_1 \mathbf{B}$  and  $\mathbf{T}_2 \mathbf{B}$ . Denoting the rotated component scores and loadings by  $\mathbf{T}_1^{rot}$ ,  $\mathbf{T}_2^{rot}$ , and  $\mathbf{P}^{rot}$ , DISCO-SCA thus results in a model with

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_1^{rot} \\ \mathbf{T}_2^{rot} \end{bmatrix} \mathbf{P}^{rot} + \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix}. \quad (5)$$

To minimize Criterion 4 above, we use a gradient projection algorithm proposed by Jennrich (2001). The main benefits of this algorithm include that it is fast (see Jennrich, 2001) and that it has been implemented in a variety of different computing environments (see, e.g., [www.stat.ucla.edu/research/gpa](http://www.stat.ucla.edu/research/gpa); Bernaards & Jennrich,



2005). Note that when multiple components have the same status (viz., distinctive for the first data block, distinctive for the second data block, or common), the solution that minimizes Criterion 4 is not unique. DISCO-SCA deals with this identification problem by means of a VARIMAX rotation of the component loadings within each set of components of the same status (also in view of getting closer to a simple structure for the subsets of loadings under study); this VARIMAX rotation is applied after the rotation of the component scores toward the partially specified target matrix.

### Model selection

DISCO-SCA, as outlined above, assumes that the number of components and the characterization of each component as either distinctive or common is known in advance. In order to apply DISCO-SCA, we therefore need to solve a double model selection problem: (1) selecting the number of components, and (2) given the number of components, determining the number of distinctive components for each data block. Below we outline how we deal with these two model selection problems.

*Selecting the number of simultaneous components* To solve the problem of selecting the number of simultaneous components, we rely on a method proposed by Van Deun et al. (2009). In their method, the percentage of variance accounted for by every simultaneous component in each data block is inspected; subsequently, each component that accounts for a sizeable amount of variance in *at least one* of the data blocks is retained. The percentage of variance accounted for by a component  $r$  in data block  $\mathbf{X}_k$  (here, with  $k = 1, 2$ ) is calculated as

$$1 - \frac{\|\mathbf{X}_k - \mathbf{t}_{rk}\mathbf{p}_r'\|^2}{\|\mathbf{X}_k\|^2},$$

with  $\mathbf{t}_{rk}$  and  $\mathbf{p}_r$  being the  $r$ -th columns of  $\mathbf{T}_k$  and  $\mathbf{P}$ , respectively.

Van Deun et al. (2009) did not formalize “sizeable.” Therefore, we propose to define it as “more than the critical noise level.” Critical noise levels for each component in each data block can be obtained using parallel analysis (PA; see, e.g., Buja & Eyuboglu, 1992; Horn, 1965; Peres-Neto, Jackson, & Somers, 2005; Zwick & Velicer, 1986) (given a suitable adaptation to an SCA setting). The general setup of this PA is as follows: (1) construction of noise blocks by randomizing within each data block the values of each variable; (2) preprocessing of the noise blocks resulting from Eq. 1; (3) performing an SCA on the whole of all preprocessed noise blocks and calculating the noise levels as the proportions of variance accounted for by each

simultaneous component in each data block; (4) repeating Steps 1–3 a number of times (e.g., 1,000); and (5) setting the critical noise level for the  $r$ th component in a data block as being equal to the 95th percentile of the corresponding noise-level distribution. For a graphical representation of this method, see Fig. 6 in the Application section.

*Given the number of simultaneous components, select the numbers of common and distinctive components* After selecting the number of simultaneous components, one must determine how many of them are distinctive for each of the two data blocks. For this purpose, we propose a formal model selection rule. To explain this rule, we start from the observation that, for each component, the sum across all blocks and all objects of the squared component scores (before as well as after rotation) is equal to a constant  $c$  (see the *Simultaneous-Component Methods* section above). When denoting the sum across the objects of block  $k$  of the squared scores for component  $r$  by  $sss_{rk}$ , this implies that, for each component  $r$ , it holds that

$$\sum_{k=1}^K \sum_{i=1}^{I_k} t_{rki}^2 = \sum_{k=1}^K sss_{rk} = c, \quad (6)$$

with  $t_{rki}$  denoting the score of the  $i$ th object of the  $k$ th data block on the  $r$ th component.

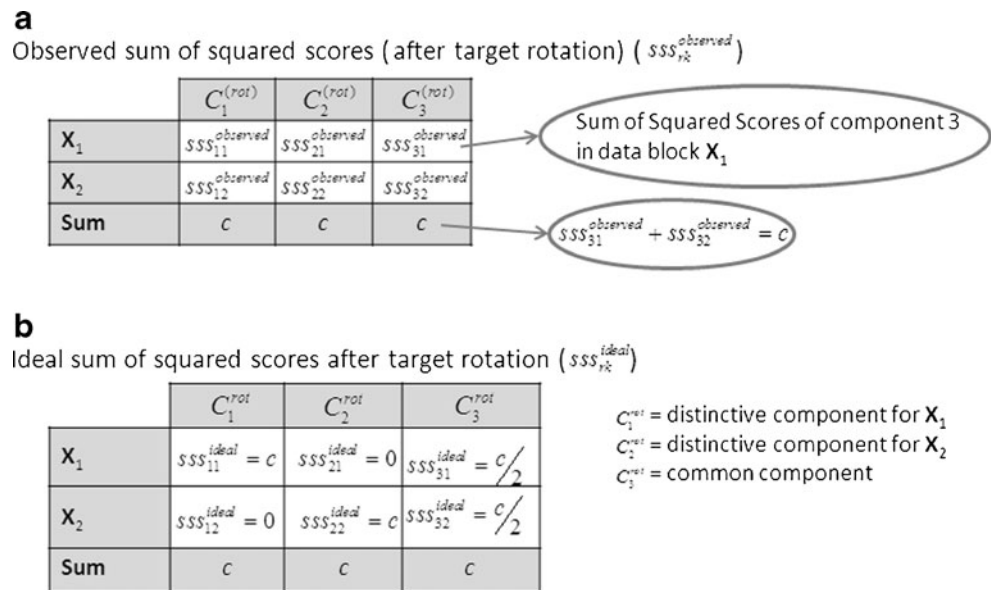
The component scores are then rotated toward *each possible* target matrix. After each rotation, we compare the observed  $sss_{rk}$  (further denoted by  $sss_{rk}^{observed}$ ) with the ideal  $sss_{rk}$  that would have been obtained in the case that the intended target structure had been perfectly achieved (further denoted by  $sss_{rk}^{ideal}$ ). In this ideal case, it holds that a component that is distinctive for the first data block (such as, e.g., the first component in Eq. 3) will have block-specific sums of squared scores that will equal  $c$  for the first data block and 0 for the second one. For a common component, we further postulate that the sums of the squared component scores will take the same value for all data blocks. An illustrative example is given in Fig. 4, with the first and second components being distinctive for the first and second data blocks, respectively, and the third component being common for the two data blocks under study. For each possible target rotation, the difference  $d$  between the observed and ideal block-specific sums of squared scores can then be calculated as

$$d = \sum_{r=1}^R \sum_{k=1}^K (sss_{rk}^{observed} - sss_{rk}^{ideal})^2. \quad (7)$$

As a model selection rule, we retain the target structure with the lowest  $d$  value.

Finally, note that the number of possible target matrices rapidly increases with increasing numbers of simultaneous

**Fig. 4** Model selection method to determine the status of the components. In panels A and B, respectively, the observed block-specific sum of squared scores (after target rotation) and the ideal block-specific sum of squared scores after target rotation are shown



components and data blocks. The construction of all possible target matrices and the calculation of their corresponding  $d$  values may then become rather cumbersome without appropriate computer software.

### Application: Cross-cultural emotion data

In this section, we present an application of DISCO-SCA to cross-cultural emotion data collected by Diener and colleagues (Kuppens, Ceulemans, Timmerman, Diener, & Kim-Prieto, 2006). The data are self-reported frequency ratings of emotional experience that were taken from a large-scale study, the International College Survey 2001. In this study, an emotion checklist was administered to more than 9,000 college students from 48 countries. The emotions included in the study were love, gratitude, happiness, cheerfulness, pride, worry, stress, anger, shame, guilt, sadness, jealousy, pleasantness, and unpleasantness. Participants were asked to rate how often they had felt each emotion in the past week on a 9-point scale, ranging from 1 (*not at all*) to 9 (*all the time*). This yielded 48 person-by-emotion data matrices, one for each country under study, which were all linked in the variable mode. In this section, we apply DISCO-SCA to the data from two of these countries, Turkey and Hong Kong (the data blocks are graphically represented in Fig. 5). Furthermore, we show how DISCO-SCA may contribute to building confirmatory multigroup FA models with common and distinctive factors.

### DISCO-SCA

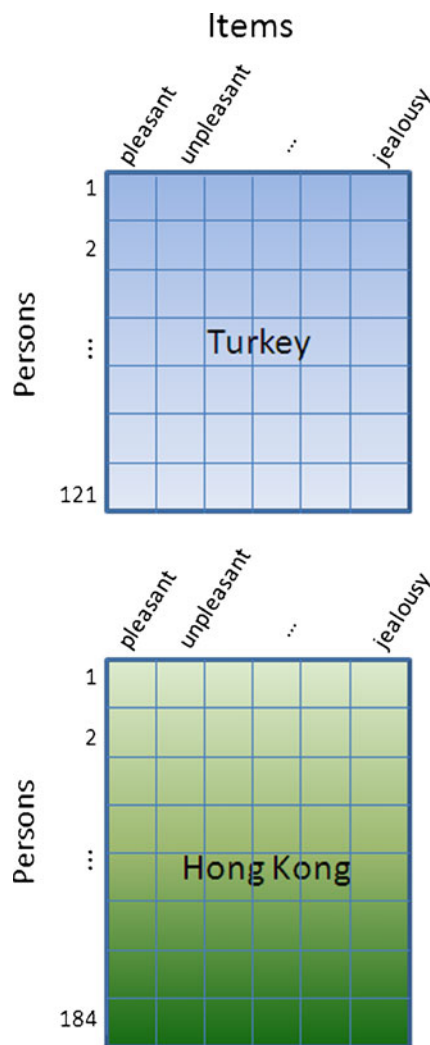
**Preprocessing the data** For the data preprocessing, we used the method proposed by Timmerman and Kiers (2003):

First, to eliminate response tendencies, the raw scores were centered across participants per variable and per country. Second, the (centered) scores were standardized per variable over all participants of the two countries jointly to eliminate artificial scale differences between variables.

**Model selection** To select the number of simultaneous components, we retained each component that accounted for an amount of variance that was higher than its critical noise level in at least one of the data blocks. These amounts are displayed in Fig. 6. From this figure, it appears that a three-component solution should be retained. This solution accounts for up to 55 % of the variance in each data block.

To assess the stability of this result, we performed a bootstrap analysis. The general setup for this analysis was as follows: (1) obtain from each data matrix under study a bootstrap sample by resampling with replacement; (2) preprocess the bootstrapped data matrices; (3) determine the number of underlying simultaneous components by means of the model selection method outlined above; and (4) repeat Steps 1–3 a total of 1,000 times. From this study, it could be concluded that the variability in the estimated number of components was low, and that the distribution of this estimated number showed a large peak for the three-component solution (i.e., in 72 % of the bootstrapped replications, a three-component solution was chosen). This implies that the three-component solution can be considered to be stable (i.e., not depending on sample fluctuations).

Next, we had to decide how many of the selected number of components had to be distinctive for each data block. It appeared that the target structure with one distinctive component for Turkey and two common



**Fig. 5** Graphical representation of data blocks for Turkey and Hong Kong

components had the lowest  $d$  value (bold row in Table 1).<sup>1</sup> To assess the stability of this result, we again performed a bootstrap analysis. The general setup of this analysis was as follows: (1) obtain from each data matrix under study a bootstrap replication by resampling with replacement; (2) preprocess these bootstrapped data matrices; (3) perform an SCA with three simultaneous components; (4) rotate the component solution toward all possible target matrices and determine the target structure with the lowest  $d$  value; and (5) repeat Steps 1–3 a total of 1,000 times. It appears that in a large majority (80 %) of the cases, the solution with one distinctive component for Turkey and two common components had the lowest  $d$  value. This implies that this solution can be considered to be a stable one.

<sup>1</sup> The CPU time for calculating all  $d$  values was less than 1 s.

**Results and discussion** For the finally selected solution, the loadings of the emotions on the three rotated components are reported in Table 2. From this table, it appears that the two common components primarily reflect individual differences in negative (and, respectively, positive) affect. This links up with previous research, in which it has been shown that positive affect and negative affect show up as two universal dimensions underlying intracultural differences in emotional experience (see, e.g., Joiner, Sandin, Chorot, Lostao, & Marquina, 1997; Terracciano, McCrae, & Costa, 2003).

Finally, the emotions love and jealousy have strong loadings on the distinctive component for Turkey. These two emotions are closely linked to the concept of honor in Mediterranean societies. Mediterranean honor is centered on the maintenance of a good family reputation, social interdependence, and feminine (e.g., chastity and virginity) and masculine (e.g., protection of the family honor) honor codes, while on the contrary, Western honor more emphasizes personal attributes and capabilities (see, e.g., Herr, 1969; Rodriguez Mosquera, Manstead, & Fischer, 2002). In the last decades, Turkey has gone through a period of fast urbanization, industrialization, and Westernization, creating a melting pot of Western and Mediterranean values (Özkan & Lajunen, 2005). All of this could give rise to considerable individual differences in feelings and attitudes related to Mediterranean honor.

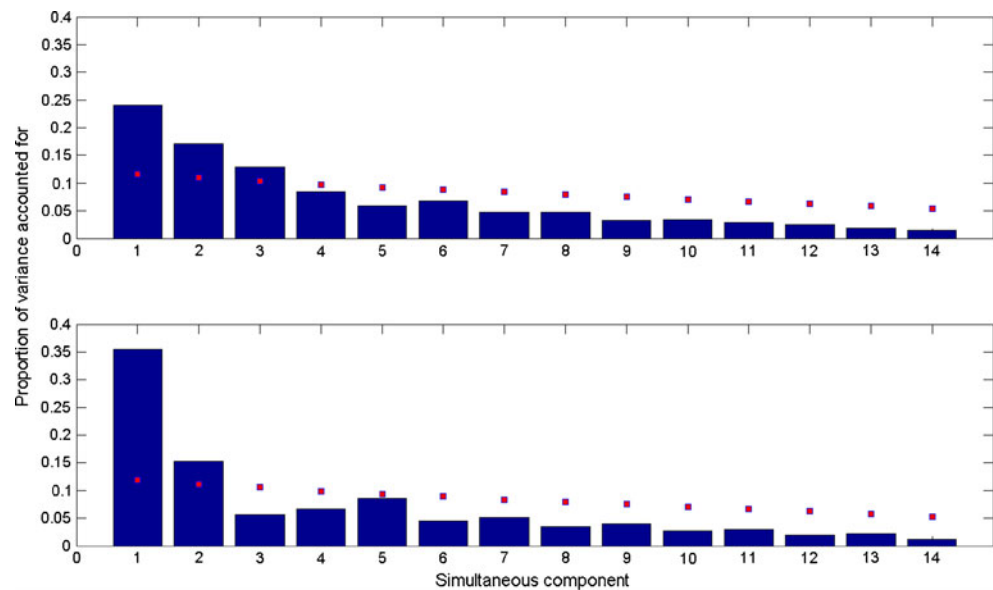
#### Multiple-group factor analysis

As mentioned before, some form of common and distinctive factors could also be included in a multiple-group confirmatory factor-analytic model (MG-CFA). Technically, for a distinctive factor, this can be achieved by allowing the same group of indicator variables to load on that factor in all data blocks under study, and by subsequently testing whether the loadings differ from zero in the data block for which the factor is supposed to be distinctive, unlike in all other data blocks. A common factor could further be obtained if, for each indicator variable of a group of indicator variables that are allowed to load on the factor in all data blocks, the loadings in all data blocks differ from zero and have the same sign. Figure 7 gives a schematic example for the case of two data blocks with one common factor and one distinctive factor for the first data block.

However, as mentioned before, to construct this type of MG-CFA model, prior hypotheses on which groups of indicator variables might constitute common and distinctive factors are needed. In the absence of such knowledge, DISCO-SCA could be most useful in formulating such hypotheses. To illustrate, we have analyzed the Turkey and Hong Kong data set with MG-



**Fig. 6** Proportions of variance accounted for by each simultaneous component in each block of the cross-cultural emotion data (upper panel, Turkey; lower panel, Hong Kong). The red squares indicate the critical noise levels obtained with parallel analysis (provided a suitable adaptation to an SCA setting)



CFA,<sup>2</sup> making use of the results obtained with DISCO-SCA. This led to the specification of a three-factor model for both countries and with the sets of indicator variables for the three countries being defined by the DISCO-SCA loadings of .45 or higher (see Table 2). To make the model identifiable, we further allowed the sadness variable to load on the third factor, and we constrained the variances of the factors to 1. We initially estimated this model with all factor loadings free, except the loadings on the third factor (i.e., the distinctive factor for Turkey), for which the loadings in Hong Kong were constrained to be zero. In line with the procedure outlined by Byrne (2001), this model was then compared with the base model with no parameter constraints. A chi-square difference of 0.85 ( $df = 3$ ) indicated that the loadings on the distinctive factor did not differ significantly from zero in Hong Kong ( $p = .8374$ ). All other unconstrained loadings differed significantly from zero and implied the same interpretation for the common factors as in the DISCO-SCA model. Figure 8 depicts the results of the constrained model.

## Concluding remarks

### Method and scope

Multigroup component and factor-analytic methods are most suitable to reveal the mechanisms underlying linked

data. However, the mechanisms resulting from these methods will often reflect a mixture of information that is common to all data blocks and information that is specific for one data block and not for the other(s). In this article, we have proposed a novel method, called DISCO-SCA, to disentangle the two kinds of information. This method was further illustrated with data from a cross-cultural study on emotions; as such, it revealed both culture-overarching and culture-specific emotional mechanisms.

Obviously, DISCO-SCA may also be applied in several other contexts. For instance, when looking for the cognitive-ability dimensions underlying two or more intelligence tests, one may wonder which abilities are measured by all of the tests and which are test-specific (e.g., Estabrook, 1984). In this case, a data block refers to the person-by-item data pertaining to one particular intelligence test, with the same set of persons being involved in all blocks. A second example can be found in the domain of developmental psychology (Allik, Laidra, Realo, & Pullmann, 2004; McCrae et al., 2002; Zimprich et

**Table 1** The  $d$  values for all possible DISCO-SCA solutions

# dist for $X_1$	# dist for $X_2$	# common	$d$
3	0	0	0.85
2	1	0	0.85
2	0	1	0.38
1	2	0	1.25
1	1	1	0.57
<b>1</b>	<b>0</b>	<b>2</b>	<b>0.09</b>
0	3	0	2.37
0	2	1	1.36
0	1	2	0.64
0	0	3	0.17

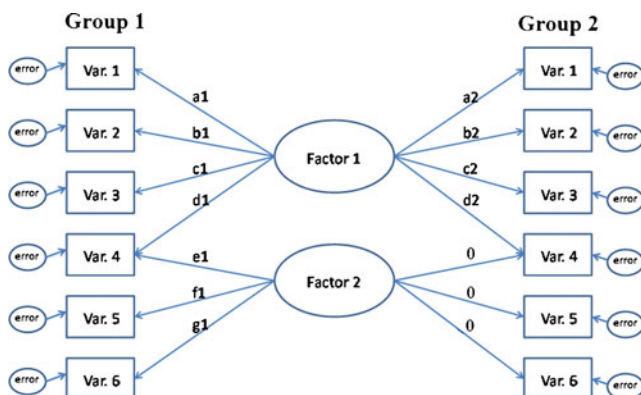
<sup>2</sup> Mplus or IBM SPSS Amos can be used to estimate a MG-CFA model; note, however, that a full hybrid analysis with Mplus or SPSS Amos is not possible, as these packages do not leave room for score rotations.

**Table 2** Loadings on the three rotated simultaneous components

	$C_{c1}$	$C_{c2}$	$C_T$
Pleasantness	−.34	<b>.81</b>	−.10
Unpleasantness	<b>.64</b>	−.58	.19
Happiness	−.47	<b>.86</b>	.05
Cheerfulness	−.39	<b>.82</b>	.06
Sadness	<b>.64</b>	−.45	.37
Anger	<b>.65</b>	−.04	.29
Pride	.12	<b>.64</b>	.04
Gratitude	.21	<b>.68</b>	−.34
Love	.10	<b>.47</b>	<b>.82</b>
Guilt	<b>.84</b>	.15	−.21
Shame	<b>.76</b>	.20	−.31
Worry	<b>.81</b>	−.03	.06
Stress	<b>.78</b>	−.12	.33
Jealousy	.24	.28	<b>.90</b>

$C_{c1}$  and  $C_{c2}$  denote the first and second common components, respectively;  $C_T$  denotes the distinctive component for Turkey. Loadings with absolute values  $>.40$  have been put in bold

al., 2008), where children from different age groups may be presented the same personality questionnaire. This results in a set of child-by-item data blocks, with each data block pertaining to a specific age group and with the different data blocks having the questionnaire items in common. In that case, one may wish to look for both general personality dimensions and dimensions that are specific for a certain developmental stage. A third example may be taken from the field of person perception (Bierhoff, 1989; Fransella, 2003, 2005; Funder, 1999; Lee, McCauley, & Draguns, 1999; Tagiuri, 1969), where persons may be asked to rate their significant others on different attributes in view of retrieving both nomothetic and idiographic underlying meaning dimensions. This may be achieved by applying DISCO-SCA to a set of significant-other-by-attribute data blocks, one for each participant under study, that all have the same set of attributes in common. As a



**Fig. 7** Schematic example of a common (Factor 1) and a distinctive (Factor 2) factor in the multiple-group factor analysis framework

fourth and final example, DISCO-SCA may also be applied to three-way data sets (e.g., when the same participants are being subjected to the same questionnaires on the same occasions). In particular, a DISCO rotation could be included in a PCA-SUP or Tucker-1 analysis of such data (Kiers, 1991).

To be sure, one may note that not all types of linked data matrices can be meaningfully modeled by means of DISCO-SCA analyses. As an example, one may consider data matrices with multivariate time series stemming from different participants, with variables constituting the linking mode and individual participants acting as the sub-grouping variable. To capture lagged relationships in such data, models other than DISCO-SCA (such as, e.g., dynamic factor models and state space models) would be needed.

## Extensions

In this article, DISCO-SCA was introduced as a method to analyze two data matrices that share the same variable mode. However, the DISCO-SCA methodology can also be extended to other settings.

One such setting is data blocks that are linked in the object mode. (Note that an example of such a setting was given above, in the example of two or more intelligence tests taken by the same group of persons.) Analyzing object-wise linked data can be done in a similar way to analyzing variable-wise linked data, except that the target must now be specified for the loadings rather than for the scores. To clarify this, let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  denote two  $I \times J_1$  and  $I \times J_2$  object-wise linked data matrices. In this case, an SC model with  $R$  components looks like

$$[\mathbf{X}_1 \mathbf{X}_2] \approx \mathbf{T} [\mathbf{P}'_1 \mathbf{P}'_2]', \quad (8)$$

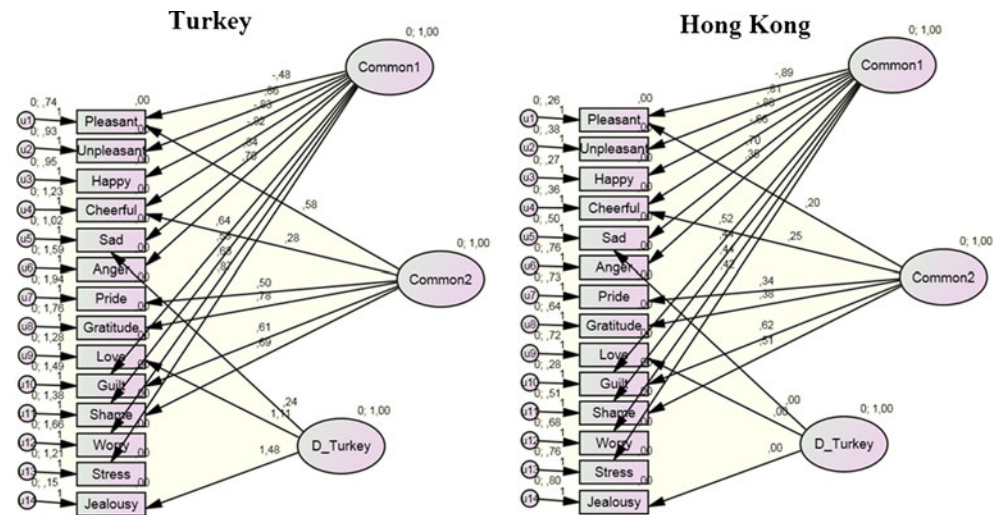
or, equivalently,

$$\mathbf{X}_{conc} \approx \mathbf{T} \mathbf{P}'_{conc}, \quad (9)$$

subject to  $\mathbf{T}' \mathbf{T} = \mathbf{I}$ , with  $\mathbf{T}$  denoting an  $I \times R$  matrix with component scores, and  $\mathbf{P}_1$  and  $\mathbf{P}_2$ , two  $J_1 \times R$  and  $J_2 \times R$  data-block-specific component loadings. In that case, a distinctive mechanism can be defined as a component for which the loadings equal zero for the data block in which that distinctive component does not play a part. To disentangle common and specific information in this setting, the component loadings  $\mathbf{P}_{conc}$  will be rotated as close as possible toward a partially specified concatenated target matrix in which part of the loadings of the distinctive components have been put equal to zero.

A second possible setting to which DISCO-SCA can be extended is that of  $K > 2$  data matrices that all have the same mode in common. In such a setting, one may consider for

**Fig. 8** Results of the constrained multiple-group factor-analytic model, based on the DISCO-SCA results



each nonempty subset of the  $K$  data blocks a type of component that accounts for variance in each of the data blocks within the subset and not in the data blocks outside the subset in question. As there are in total  $2^K - 1$  nonempty subsets, this implies that now  $2^K - 1$  different types of components can be distinguished. This obviously generalizes the case of two data blocks with  $2^2 - 1 = 3$  types of components (viz., one common and two distinctive types).

#### Related methods

In the particular case of data consisting of a set of data blocks pertaining to a common set of objects (as in the example of different intelligence tests that are administered to the same set of participants), the DISCO principle of rotating a latent variable solution in order to tease apart common and distinctive structural information could also be transferred to the exploratory factor analytic (EFA) case. In that case, too, the loadings of the EFA model are to be rotated toward a suitable partially specified target matrix. If the data blocks pertain to a common set of variables, however, a transfer of the DISCO-SCA methodology to the EFA case is not possible, as the methodology then implies a rotation of scores that is at odds with the nature of the factor-analytic model.

Finally, one might also consider using partial common principal-component analysis (PCPC; Flury, 1987) to uncover common and distinctive mechanisms underlying linked data. Within the PCPC approach, a common component is obtained by restricting its loadings to be equal across all data blocks, and a distinctive component by leaving the loadings free. However, one should note that a PCPC in which all components are common coincides with a classic instance of ordinary SCA (i.e., SCA-P; Kiers & ten Berge, 1994). From this, it follows that the “common” components in PCPC usually will represent a mix of common and distinctive information. Moreover, PCPC might be misleading, in that it only looks at the

distinctive components in the data blocks for which they are distinctive; yet it is perfectly possible that some components could also account for a meaningful amount of variance in other data blocks. DISCO-SCA remedies both of these problems.

In conclusion, the DISCO-SCA method proposed in the present article constitutes a unique and powerful tool to unveil common and distinctive mechanisms underlying linked data. The method is versatile, in that it is applicable to a broad range of contexts inside and outside psychology, and it can be extended in several directions.

**Acknowledgments** This work was supported by IWT-Flanders (IWT/060045/SBO Bioframe), the Research Fund of the University of Leuven (EF/05/007 and GOA/2005/04), and by Belgian Federal Science Policy (IAP P7/06).

#### References

- Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality*, 18, 445–462. doi:10.1002/per.524
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397–438. doi:10.1080/10705510903008204
- Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65, 676–696. doi:10.1177/0013164404272507
- Bierhoff, H. (1989). *Person perception and attribution*. New York, NY: Springer.
- Browne, M. W. (1972). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 115–120.
- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, 27, 509–540. doi: s15327906mbr27042/s15327906mbr27042
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.



- Caprara, G., Barbaranelli, C., Bermudez, J., Maslach, C., & Ruch, W. (2000). Multivariate methods for the comparison of factor structures in cross-cultural research—An illustration with the big five questionnaire. *Journal of Cross-Cultural Psychology*, 31, 437–464. doi:10.1177/0022022100031004002
- Curran, P., & Hussong, A. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14, 81–100. doi:10.1037/a0015914
- Estabrook, G. E. (1984). A canonical correlation analysis of the Wechsler intelligence scale for children-revised and the Woodcock-Johnson tests of cognitive ability in a sample referred for suspected learning disabilities. *Journal of Education Psychology*, 76, 1170–1177. doi:10.1037/0022-0663.76.6.1170
- Flury, B. K. (1987). Two generalizations of the common principal component model. *Biometrika*, 74, 59–69. doi:10.2307/2336021
- Fransella, F. (2003). *An international handbook of personal construct psychology*. Chichester, UK: Wiley.
- Fransella, F. (2005). *The essential practitioner's handbook of personal construct psychology*. Chichester, UK: Wiley.
- Funder, D. (1999). *Personality judgment: A realistic approach to person perception*. London, UK: Academic Press.
- Hagedoorn, M., Van Yperen, N. W., Van De Vliert, E., & Buunk, B. P. (1999). Employees' reactions to problematic events: A circumplex structure of five categories of responses, and the role of job satisfaction. *Journal of Organizational Behavior*, 20, 309–321. doi:10.1002/(SICI)1099-1379(199905)20:3<309::AID-JOB895>3.0.CO;2-P
- Herr, R. (1969). Review: Jean G. Peristiany, "Honour and shame: The values of Mediterranean society" (Book review). *Journal of Social History*, 3, 89–92.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. doi:10.1007/BF02289447
- Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, 66, 289–306. doi:10.1007/BF02294840
- Joiner, T. E., Sandin, B., Chorot, P., Lostao, L., & Marquina, G. (1997). Development and factor analytic validation of the SPANAS among women in Spain: (More) cross-cultural convergence in the structure of mood. *Journal of Personality Assessment*, 68, 600–615. doi:10.1037/0022-3514.68.3.600
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426. doi:10.1007/BF02291366
- Kiers, H. A. L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, 56, 449–470. doi:10.1007/BF02294485
- Kiers, H. A. L., & ten Berge, J. M. F. (1989). Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices in 2 or more populations. *Psychometrika*, 54, 467–473. doi:10.1007/BF02294629
- Kiers, H. A. L., & ten Berge, J. M. F. (1994). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple structure. *British Journal of Mathematical and Statistical Psychology*, 47, 109–126.
- Kuppens, P., Ceulemans, E., Timmerman, M. E., Diener, E., & Kim-Prieto, C. (2006). Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology*, 37, 491–515. doi:10.1177/0022022106290474
- Lee, Y., McCauley, C. R., & Draguns, J. G. (1999). *Personality and person perception across cultures*. Mahwah, NJ: Erlbaum.
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22, 471–491. doi:10.1037/a0019227
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16, 439–476. doi:10.1080/10705510903008220
- McCrae, R. R., Costa, P. T., Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses. *Journal of Personality and Social Psychology*, 83, 1456–1468. doi:10.1037/0022-3514.83.6.1456
- Millsap, R. E., & Meredith, W. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika*, 53, 123–134. doi:10.1007/BF02294198
- Özkan, T., & Lajunen, T. (2005). Masculinity, femininity, and the Bem Sex Role Inventory in Turkey. *Sex Roles*, 52, 103–110. doi:10.1007/s11199-005-1197-4
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, 49, 974–997. doi:10.1016/j.csda.2004.06.015
- Rodriguez Mosquera, P. M., Manstead, A. S. R., & Fischer, A. H. (2002). Honor in the Mediterranean and Northern Europe. *Journal of Cross-Cultural Psychology*, 33, 16–36. doi:10.1177/0022022102033001002
- Rossier, J., de Stadelhofen, F. M., & Berthoud, S. (2004). The hierarchical structures of the NEO PI-R and the 16 PF 5. *European Journal of Psychological Assessment*, 20, 27–38. doi:10.1027/1015-5759.20.1.27
- Silva, F., Martinez-Arias, R., Rapaport, E., Ertle, A., & Ortet, G. (1997). Dimensions of interpersonal orientation: Cross-cultural studies regarding the structure of the "DOI Kit. *Personality and Individual Differences*, 23, 973–985. doi:10.1016/S0191-8869(97)00127-X
- Spikman, J., Kiers, H., Deelman, B., & van Zomeren, A. (2001). Construct validity of concepts of attention in healthy controls and patients with CHI. *Brain and Cognition*, 47, 446–460. doi:10.1006/brcg.2001.1320
- Tagiuri, R. (1969). Person perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (The individual in a social context, Vol. 3, pp. 395–449). Reading, MA: Addison-Wesley.
- ten Berge, J. M. F. (1977). *Optimizing factorial invariance*. Unpublished doctoral dissertation, University of Groningen, the Netherlands.
- ten Berge, J. M. F., Kiers, H. A. L., & Van der Stel, V. (1992). Simultaneous components analysis. *Statistica Applicata*, 4, 377–392.
- Terracciano, A., McCrae, R., & Costa, P. (2003). Factorial and construct validity of the Italian Positive and Negative Affect Schedule (PANAS). *European Journal of Psychological Assessment*, 19, 131–141. doi:10.1027/1015-5759.19.2.131
- Timmerman, M. E., & Kiers, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, 68, 105–121. doi:10.1007/BF02296656
- Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. A. L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, 10, 246–261. doi:10.1186/1471-2105-10-246
- Van Mechelen, I., & Smilde, A. K. (2010). A generic linked-mode decomposition model for data fusion. *Chemometrics and Intelligent Laboratory Systems*, 104, 83–94. doi:10.1016/j.chemolab.2010.04.012
- Zimprich, D., Martin, M., Kliegel, M., Dellenbach, M., Rast, P., & Zeintl, M. (2008). Cognitive abilities in old age: Results from the Zurich longitudinal study on cognitive aging. *Swiss Journal of Psychology*, 67, 177–195. doi:10.1024/1421-0185.67.3.177
- Zwick, R. W., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442. doi:10.1037/0033-2909.99.3.432