PSYCHOMETRIKA—VOL. 77, NO. 4, 724–740
OCTOBER 2012
DOI: 10.1007/S11336-012-9275-3

THE *SIMCLAS* MODEL: SIMULTANEOUS ANALYSIS OF COUPLED BINARY DATA MATRICES WITH NOISE HETEROGENEITY BETWEEN AND WITHIN DATA BLOCKS

TOM F. WILDERJANS

RESEARCH GROUP OF QUANTITATIVE PSYCHOLOGY AND INDIVIDUAL DIFFERENCES, DEPARTMENT OF PSYCHOLOGY, KU LEUVEN

E. CEULEMANS

DEPARTMENT OF EDUCATIONAL SCIENCES, KU LEUVEN

I. VAN MECHELEN

RESEARCH GROUP OF QUANTITATIVE PSYCHOLOGY AND INDIVIDUAL DIFFERENCES, DEPARTMENT OF PSYCHOLOGY, KU LEUVEN

In many research domains different pieces of information are collected regarding the same set of objects. Each piece of information constitutes a data block, and all these (coupled) blocks have the object mode in common. When analyzing such data, an important aim is to obtain an overall picture of the structure underlying the whole set of coupled data blocks. A further challenge consists of accounting for the differences in information value that exist between and within (i.e., between the objects of a single block) data blocks. To tackle these issues, analysis techniques may be useful in which all available pieces of information are integrated and in which at the same time noise heterogeneity is taken into account. For the case of binary coupled data, however, only methods exist that go for a simultaneous analysis of all data blocks but that do not account for noise heterogeneity. Therefore, in this paper, the *SIMCLAS* model, being a Hierarchical Classes model for the simultaneous analysis of coupled binary two-way matrices, is presented. In this model, noise heterogeneity between and within the data blocks is accounted for by downweighting entries from noisy blocks/objects within a block. In a simulation study it is shown that (1) the *SIMCLAS* technique recovers the underlying structure of coupled data to a very large extent, and (2) the *SIMCLAS* technique outperforms a Hierarchical Classes technique in which all entries contribute equally to the analysis (i.e., noise homogeneity within and between blocks). The latter is also demonstrated in an application of both techniques to empirical data on categorization of semantic concepts.

Key words: data fusion, coupled data, multi-set data, noise heterogeneity, simultaneous clusterings, Hierarchical Classes Analysis, overlapping clustering, hierarchical relations, multivariate binary data.

1. Introduction

In many research domains different pieces of information are collected regarding the same set of objects. Each piece of information constitutes a data block, and all these (coupled) blocks have the object mode in common. For example, in psychiatric diagnosis research, such coupled data are encountered when for a set of patients information is available regarding, on the one hand, their diagnosis (i.e., patient-by-diagnosis data block) and, on the other hand, the symptoms they exhibit (i.e., patient-by-symptom data block).

The first author is a Research Assistant of the Fund for Scientific Research (FWO)—Flanders (Belgium). The research reported in this paper was partially supported by the Research Council of K.U. Leuven (GOA/2005/04 and EF/2005/07, 'SymBioSys') and by IWT-Flanders (SBO 60045, 'Bioframe'). We would like to thank Gert Storms and his collaborators for providing us with an interesting data set.

Requests for reprints should be sent to Tom F. Wilderjans, Research Group of Quantitative Psychology and Individual Differences, Department of Psychology, KU Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium. E-mail: tom.wilderjans@ppw.kuleuven.be

© 2012 The Psychometric Society

To grasp the wealth of information that is present in coupled data, one may uncover the structure underlying each data block, and relate these structures to one another. As such, an overall picture of the structure underlying the whole set of coupled data blocks is acquired. For example, applied to the psychiatric diagnosis case, one may search for the underlying syndromes that may account for both the diagnosis and symptom profiles of the different patients.

A complicating factor herewith may be that the coupled data blocks under study and/or the objects within a single data block may differ regarding their information value. For example, the patient-by-symptom data block may be more reliable than the patient-by-diagnosis block, because it is easier to determine whether or not a symptom is present than to give a full diagnosis (which is a complicated interplay of different symptoms). Moreover, some patients may be easier to diagnose than others, because of the very typical symptom profile they exhibit. Taking these differences in information value into account may result in stronger inferences with respect to the structure underlying the data. However, in most cases no information is available regarding the information value of a block or the objects within a block.

To handle these tasks, an analysis technique needs to be developed in which all available pieces of information are simultaneously integrated and in which noise heterogeneity between and within data blocks is taken into account (see Van Mechelen & Smilde, 2009, 2010). When dealing with real-valued coupled data and noise heterogeneity between blocks, one may rely on the *MxLSCA-P* approach (Wilderjans, Ceulemans, Van Mechelen, & van den Berg, 2011), a stochastic extension of standard simultaneous component analysis (Kiers & ten Berge, 1994; Timmerman & Kiers, 2003; Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen 2009). However, in the case of binary coupled data, which show up rather frequently in psychology, no such technique exists yet. Indeed, although techniques have been proposed, like, for example, *concatenated HICLAS*, which can be conceived as a Hierarchical Classes counterpart of simultaneous component analysis (Millsap & Meredith, 1988; Kiers & ten Berge, 1989; ten Berge, Kiers, & van der Stel, 1992) for binary coupled data, these techniques do not account for noise heterogeneity between and/or within the blocks.

To remedy this, the aim of the present paper is to propose a new model and associated data-analytic strategy to handle between- and within-block noise heterogeneity in binary coupled data. This model belongs to the family of Hierarchical Classes models (De Boeck & Rosenberg, 1988; Van Mechelen, De Boeck, & Rosenberg, 1995; Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999; Ceulemans, Van Mechelen, & Leenen, 2003; Ceulemans & Van Mechelen, 2005) and will therefore be called a *SIMultaneous HICLAS* model (*SIMCLAS*). Specifically, we will introduce two variants of the *SIMCLAS* model. To deal with noise heterogeneity between data blocks we will propose *block-homogeneous SIMCLAS*, whereas *block-heterogeneous SIMCLAS* allows for noise heterogeneity within (and between) data blocks. The data-analytic strategy will imply estimating the amount of noise in each data block (in the *block-homogeneous* variant) or the amount of noise in each object of each data block (in the *block-heterogeneous* variant), and downweighting entries from more noisy data blocks/objects in the analysis.

The remainder of this paper is organized in five main sections. In the first section, we recapitulate *concatenated HICLAS* analysis and introduce *SIMCLAS* analysis (both the *block-homogeneous* and *block-heterogeneous* variants). In the second section, the *SIMCLAS* loss function will be presented, along with an algorithm to fit the *SIMCLAS* model to data. In the third section, by means of an extensive simulation study, the optimization and recovery performance of the *SIMCLAS* technique will be evaluated, and the *SIMCLAS* technique will be compared to the *concatenated HICLAS* technique. In the fourth section, both techniques will be applied to empirical coupled data on categorization of semantic concepts. The fifth section, finally, contains some concluding remarks.

TABLE 1.
Hypothetical two-way two-mode coupled model matrices $\mathbf{M}^1$, $\mathbf{M}^2$, and $\mathbf{M}^3$.

	$\mathbf{M}^1$				$\mathbf{M}^2$					$\mathbf{M}^3$					
	C_1^1	C_2^1	C_3^1	C_4^1	C_1^2	C_2^2	C_3^2	C_4^2	C_5^2	C_1^3	C_2^3	C_3^3	C_4^3	C_5^3	C_6^3
R_1	1	0	0	0	0	0	1	1	0	0	1	0	0	1	1
R_2	1	0	0	0	0	0	1	1	0	0	1	0	0	1	1
R_3	0	1	0	1	0	1	0	0	0	1	1	0	0	0	0
R_4	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1
R_5	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1
R_6	0	1	1	0	1	1	0	1	1	0	1	1	0	0	0

2. The *SIMCLAS* Analysis Technique

We assume a data set consisting of N coupled $I \times J_n$ binary data matrices $\mathbf{D}^n$, with J_n differing across data blocks ($n = 1 \ldots N$). Note that, without loss of generality, the rows are the common mode. When concatenating all N data blocks horizontally (Kiers, 2000), an $I \times J^*$ binary matrix $\mathbf{D}^*$ is obtained, with $J^* = \sum_{n=1}^{N} J_n$. First, we will describe the *concatenated HICLAS* approach. Next, we will introduce the *block-homogeneous* and *block-heterogeneous* variants of the *SIMCLAS* technique.

2.1. The Concatenated HICLAS Approach

In the *concatenated HICLAS* approach for N data matrices[1] (for an introduction to a *HICLAS* analysis for a single data matrix, see De Boeck & Rosenberg, 1988), each binary $I \times J_n$ data matrix $\mathbf{D}^n$ is approximated by a binary model matrix $\mathbf{M}^n$ of the same size:

$$d_{ij_n}^n = m_{ij_n}^n + e_{ij_n}^n, \tag{1}$$

with $d_{ij_n}^n$, $m_{ij_n}^n$, and $e_{ij_n}^n$, denoting the entries of $\mathbf{D}^n$, $\mathbf{M}^n$, and the error matrix $\mathbf{E}^n$ ($I \times J_n$), respectively. Further, each model matrix $\mathbf{M}^n$ can be decomposed into an $I \times P$ binary matrix $\mathbf{A}$ and a $J_n \times P$ binary matrix $\mathbf{B}^n$, where P denotes the rank of the model, and $\mathbf{A}$ is the same for all model matrices $\mathbf{M}^n$. The P columns of $\mathbf{A}$ and $\mathbf{B}^n$ define P binary variables, which are called bundles, and imply an overlapping clustering of the elements of $\mathbf{A}$ and $\mathbf{B}^n$.

The *concatenated HICLAS* model represents two types of structural relations in $\mathbf{M}^n$ ($n = 1 \ldots N$): association and quasi-order. To explain these structural relations, we will use the three hypothetical coupled model matrices $\mathbf{M}^1$, $\mathbf{M}^2$, and $\mathbf{M}^3$ in Table 1 as a guiding example. In Table 2, the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{B}^2$, and $\mathbf{B}^3$ of the *concatenated HICLAS* model with three bundles for $\mathbf{M}^1$, $\mathbf{M}^2$, and $\mathbf{M}^3$ (in Table 1) are presented.

Association The association relation is the binary relation between the row and column elements as defined by the 1-entries in $\mathbf{M}^n$. The *concatenated HICLAS* model represents this association relation by the following decomposition rule:

$$m_{ij_n}^n = \bigoplus_{p=1}^{P} a_{ip} b_{j_n p}^n, \tag{2}$$

[1] In the remainder of this paper, the terms 'matrix' and 'block' are used interchangeably.

TABLE 2.
Concatenated HICLAS model with three bundles for the coupled model matrices $\mathbf{M}^1$, $\mathbf{M}^2$, and $\mathbf{M}^3$ in Table 1.

	$\mathbf{A}$				$\mathbf{B}^1$				$\mathbf{B}^2$				$\mathbf{B}^3$		
	Bu_1	Bu_2	Bu_3		Bu_1	Bu_2	Bu_3		Bu_1	Bu_2	Bu_3		Bu_1	Bu_2	Bu_3
R_1	0	1	0	C_1^1	0	1	0	C_1^2	0	0	1	C_1^3	1	0	0
R_2	0	1	0	C_2^1	1	0	1	C_2^2	1	0	1	C_2^3	1	1	1
R_3	1	0	0	C_3^1	0	0	1	C_3^2	0	1	0	C_3^3	0	0	1
R_4	1	1	0	C_4^1	1	0	0	C_4^2	0	1	1	C_4^3	0	0	0
R_5	0	1	1					C_5^2	0	0	1	C_5^3	0	1	0
R_6	0	0	1									C_6^3	0	1	0

where $\bigoplus$ denotes a Boolean sum (i.e., $1 \bigoplus 1 = 1$), and a_{ip} and $b^n_{j_n p}$ the entries of $\mathbf{A}$ and $\mathbf{B}^n$, respectively. In Table 1, for example, R_4 and C_5^3 are associated in $\mathbf{M}^3$ because both elements belong to the second bundle Bu_2 in $\mathbf{A}$ and $\mathbf{B}^3$ (see Table 2), respectively. However, R_4 and C_4^3 are not associated in $\mathbf{M}^3$ (see Table 1) because these elements do not have a bundle in common (see Table 2).

Quasi-order On the columns of $\mathbf{M}^n$, a quasi-order relation $\leq$ is defined as follows: When for column j the set of associated rows in $\mathbf{M}^n$ is denoted by S^j, then column $j \leq$ column j' in $\mathbf{M}^n$ iff $S^j \subseteq S^{j'}$. The *concatenated HICLAS* model represents the quasi-order relation among the columns in $\mathbf{M}^n$ by subset-superset relations among their bundle patterns in $\mathbf{B}^n$. For example, in Table 1, one can see that $C_3^2 \leq C_4^2$ in $\mathbf{M}^2$ because $S^{C_3^2} \subseteq S^{C_4^2}$; therefore, in Table 2 the bundle pattern for C_3^2 is a subset of the bundle pattern for C_4^2. Also on the (common) rows a quasi-order relation $\leq$ is defined, where, for row i, S^i consists of the associated column elements from all N matrices $\mathbf{M}^n$ (e.g., $S^{R_3} = \{C_2^1, C_4^1, C_2^2, C_1^3, C_2^3\}$). For example, in Table 1, $R_3 \leq R_4$ because $S^{R_3} \subseteq S^{R_4}$; consequently, in Table 2, the bundle pattern for R_3 is a subset of the bundle pattern for R_4. Note that the quasi-order relation among the elements of a mode implies a partition of these elements into classes (i.e., elements with an identical bundle pattern) and a hierarchical relation among these classes (for more information, see De Boeck & Rosenberg, 1988).

Note that in a *concatenated HICLAS* analysis the noise in the data is not modeled in an explicit way, as is true for all Hierarchical Classes models. An exception to this, in the case of a single binary data block, is the *probabilistic HICLAS* model of Leenen, Van Mechelen, Gelman, and De Knop (2008). This model may be applied to binary coupled data by concatenating all data blocks into a single block, which implies that the noise is modeled in the same way across all data blocks (i.e., noise homogeneity within and between blocks).

2.2. *Block-Homogeneous SIMCLAS for Noise Heterogeneity Between Blocks*

The *block-homogeneous SIMCLAS* approach allows the modeling of noise heterogeneity between data blocks. Specifically, in a *block-homogeneous SIMCLAS* analysis it is assumed that the absolute values of the noise entries $e^n_{ij_n}$ in (1) follow a Bernoulli distribution with parameter π_n (i.e., $|e^n_{ij_n}| \sim Bern(\pi_n)$). To account for possible noise heterogeneity, the parameters π_n are allowed to differ between data blocks, implying a different amount of noise for each block; the parameters π_n, which are restricted to be smaller than 0.50, equal the probability that a data entry $d^n_{ij_n}$ differs from the corresponding model entry $m^n_{ij_n}$.[2] It is further assumed that all $e^n_{ij_n}$

[2]A value of π_n larger than 0.50 is not realistic, because this implies that the model entries may differ from their corresponding data entries with a probability that is above chance.

are sampled independently (i.e., uncorrelated), resulting in all data entries $d^n_{ij_n}$ being locally independent. Note that in the case of a single data block (i.e., $N = 1$), the *block-homogeneous SIMCLAS* model reduces to the *probabilistic HICLAS* model of Leenen et al. (2008).

2.3. Block-Heterogeneous SIMCLAS for Noise Heterogeneity Within and Between Blocks

To take noise heterogeneity within data blocks into account, the *block-homogeneous SIMCLAS* model may be further extended to the *block-heterogeneous SIMCLAS* model by allowing the parameter π to differ among the rows of each data block (and, as a consequence, also among the data blocks).[3] In particular, it is assumed that the noise entries $e^n_{ij_n}$ are uncorrelated and that $|e^n_{ij_n}| \sim Bern(\pi_{in})$, with $\pi_{in} \leq 0.50$ $(i = 1 \ldots I; n = 1 \ldots N)$.

3. Data Analysis

3.1. Aim

Given a set $\tilde{\mathbf{D}}$ of binary coupled data blocks $\mathbf{D}^1, \ldots, \mathbf{D}^n$, and a number of bundles P, the aim of a *block-heterogeneous SIMCLAS* analysis is to estimate $\mathbf{A}$, $\mathbf{B}^n$, and π_{in} such that the value of the (log) likelihood function

$$l = \sum_{i=1}^{I}\sum_{n=1}^{N}\left[\sum_{j_n=1}^{J_n}\left(d^n_{ij_n} - \bigoplus_{p=1}^{P} a_{ip}b^n_{j_n p}\right)^2 \times \log\left(\frac{\pi_{in}}{1-\pi_{in}}\right) + J_n \times \log(1-\pi_{in})\right] \tag{3}$$

is maximized. In the case of a *block-homogeneous SIMCLAS* analysis, (3) reduces to

$$l = \sum_{n=1}^{N}\left[\sum_{i=1}^{I}\sum_{j_n=1}^{J_n}\left(d^n_{ij_n} - \bigoplus_{p=1}^{P} a_{ip}b^n_{j_n p}\right)^2 \times \log\left(\frac{\pi_n}{1-\pi_n}\right) + I J_n \times \log(1-\pi_n)\right]. \tag{4}$$

Note that (4) implies that the extent to which $d^n_{ij_n}$ contributes to the likelihood function is inversely related to the amount of noise that is present in the corresponding $\mathbf{D}^n$; as such, data entries of noisy data matrices are downweighted in favor of entries from less noisy matrices.[4] Analogously, (3) implies that data entries belonging to noisy rows of a data matrix are downweighted in favor of entries from less noisy rows.

3.2. The SIMCLAS Algorithm

To estimate the *SIMCLAS* model parameters (i.e., the bundle matrices and the noise parameters) that maximize the likelihood function, an iterative algorithm is adopted that consists of the following five steps:

[3] Note that allowing π to differ among the columns of each data block would yield a special case of *block-homogeneous SIMCLAS* (i.e., the case in which each data block consists of one variable only). Note further that it may not be a good idea to allow π to be different for each data element, because it would be impossible to estimate such a model (i.e., there are more parameters to fit than there are data points).

[4] When $\pi_1 > \pi_2$, then $c_1 = \log(\frac{\pi_1}{1-\pi_1}) > c_2 = \log(\frac{\pi_2}{1-\pi_2})$, with c_1 and c_2 representing, respectively, the contribution of entries from $\mathbf{D}^1$ and $\mathbf{D}^2$ to the likelihood, which has to be maximized. Note that c_1 and c_2 are negative when $\pi_n \leq 0.50$, which implies that $|c_1| < |c_2|$. As such, entries from more noisy blocks (i.e., larger π_n and smaller $|c_n|$) imply a smaller decrease in the likelihood than entries from less noisy blocks (i.e., smaller π_n and larger $|c_n|$), resulting in entries from more noisy blocks being downweighted.

1. Generate initial estimates for the noise parameters π_n $(n = 1 \ldots N)$, in the case of the *block-homogeneous* variant, or π_{in} $(i = 1 \ldots I;\ n = 1 \ldots N)$ in the case of the *block-heterogeneous* variant. For π_n, three types of initial estimates may be obtained (for a discussion of different types of starting procedures, see Ceulemans, Van Mechelen, & Leenen, 2007): (1) rational, (2) random, or (3) smart-random (also called pseudo-rational). To obtain rational starting values for the noise parameters (in each block), first (1) a *HICLAS* analysis (De Boeck & Rosenberg, 1988) on $\mathbf{D}^*$, or (2) a weighted *HICLAS* analysis on $\mathbf{D}^*$, with the weight for each data matrix being $\frac{1}{IJ_n}$ (i.e., the inverse of the number of entries in $\mathbf{D}^n$), implying that all data matrices contribute equally to the analysis, is performed. Next, for each block, the average squared residual between the data and the model entries (see also Step 3 of the *SIMCLAS* algorithm) is computed. Random starting values may be generated by independently sampling values from a $U(0, 0.5)$ distribution. Smart-random starting values can be obtained by adding randomly sampled numbers to the rational starting values described above, with these numbers being drawn from a $U(-\delta, \delta)$ distribution with δ being equal to the corresponding rational starting value divided by five. Initial starting values for π_{in} can be obtained in a similar way.
2. Estimate $\mathbf{A}$ and $\mathbf{B}^n$, conditionally upon the noise parameters obtained in Step 1, by means of a simulated annealing procedure (*SA*; for an introduction to *SA*, see Kirkpatrick, Gelatt, & Vecchi, 1983; Aarts, Korst, & van Laarhoven, 1997). Note that *SA* has already been applied successfully for parameter estimation of other Hierarchical Classes models (see, Ceulemans et al., 2007; Wilderjans, Ceulemans, & Van Mechelen, 2008). In the Appendix, the principles and the different steps of such an *SA* algorithm are presented, along with the way in which *SA* is implemented in the *SIMCLAS* context. Since an *SA* algorithm may return a suboptimal solution (i.e., local optimum), a multi-start procedure is recommended (see the Appendix).
3. Re-estimate the noise parameters, conditional upon $\mathbf{A}$ and $\mathbf{B}^n$ (obtained in Step 2), as follows:

$$\hat{\pi}_n = \frac{\sum_{i=1}^{I}\sum_{j_n=1}^{J_n}(d^n_{ij_n} - \bigoplus_{p=1}^{P} a_{ip}b^n_{j_n p})^2}{I\,J_n},$$
$$\hat{\pi}_{in} = \frac{\sum_{j_n=1}^{J_n}(d^n_{ij_n} - \bigoplus_{p=1}^{P} a_{ip}b^n_{j_n p})^2}{J_n}. \tag{5}$$

4. Compute the loss (likelihood) function (4) or (3). When it has increased by a value larger than some tolerance value, repeat Steps 2 and 3; otherwise go to Step 5.
5. Perform a closure operation (Barbut & Monjardet, 1970; Birkhoff, 1940) on $\mathbf{A}$ and $\mathbf{B}^n$. This is necessary because the bundle matrices obtained at the end of Step 4 do not yet represent the quasi-order relation in $\mathbf{M}^n$ correctly. This closure operation consists of changing each 0-entry in $\mathbf{A}$ and $\mathbf{B}^n$ to 1 iff this modification does not alter $\mathbf{M}^n$ (and consequently does not change the loss function value).

Since the *SIMCLAS* algorithm may end up in a local rather than the global optimum, a multi-start procedure is recommended. In such a procedure, the *SIMCLAS* algorithm is run a prespecified number of times, each time from different initial estimates for the noise parameters (see Step 1); subsequently, the solution yielding the maximum value on the likelihood function is retained. Note the difference between the (overall) multi-start procedure for the *SIMCLAS* algorithm as a whole (i.e., all five steps) and the multi-start procedure for the *SA* algorithm (i.e., the second step of the *SIMCLAS* algorithm).

3.3. Model Selection

In general, the (true) number of bundles P of the *SIMCLAS* model is unknown. Therefore, in practice, analyses with increasing numbers of bundles are performed and a heuristic is used to determine the model with optimal balance between model complexity (i.e., the number of bundles), on the one hand, and the fit of the model to the data (i.e., the value on the likelihood function), on the other hand (for more information, see Ceulemans & Van Mechelen, 2005; Wilderjans et al., 2008). In this paper, we will apply for this purpose the generalized scree test, which was proposed by Leenen and Van Mechelen (2001) and which has shown good performance as a model selection heuristic for Hierarchical Classes Analyses (see Leenen & Van Mechelen, 2001; Ceulemans et al., 2003).

4. Simulation Study

4.1. Problem

In this section we present simulation results in which (1) we evaluate the performance of the *SIMCLAS* algorithm, and (2) we compare the *SIMCLAS* modeling technique to the *concatenated HICLAS* approach. At this point, we are interested in two aspects of algorithmic performance: optimization and recovery. With regard to optimization, we will examine the extent to which the *SIMCLAS* algorithm is able to maximize the loss functions (4) and (3). Concerning recovery, we will determine the degree to which both approaches succeed in disclosing the true structure underlying a given coupled data set and whether or not the recovery depends on characteristics of the data.

In this section, we will present three simulation studies. In the first simulation study, we will deal with *block-homogeneous SIMCLAS* for the case of two coupled data blocks. Four data characteristics will be taken into account: (1) the size of the common (row) mode, (2) the size of the data blocks, (3) the total amount of noise in the data, and (4) the way the noise is distributed across data blocks. In the following two simulation studies, we try to generalize the results from the first study to the case of more than two coupled data blocks, and to *block-heterogeneous SIMCLAS*.

4.2. First Simulation Study: Block-Homogeneous SIMCLAS for Two Data Blocks

4.2.1. Design and Procedure

Design We deal with the case of two coupled data matrices $\mathbf{D}^1$ and $\mathbf{D}^2$ with noise heterogeneity between data blocks. Four factors were systematically manipulated in a completely randomized factorial design, with all factors considered random. The first two factors, which pertain to the sizes of the data matrices, are:

(a) the *Size of the common mode*, I, manipulated at three levels: $I = 200$, $I = 100$, $I = 50$;
(b) the *Array/total ratio*, r:

$$r = \frac{I \times J_1}{(I \times J_1) + (I \times J_2)}, \tag{6}$$

which indicates how much the data blocks differ in size. This factor was manipulated at five levels: 0.10 (small $\mathbf{D}^1$ relative to $\mathbf{D}^2$), 0.30, 0.50 (equal size), 0.70, 0.90 (large $\mathbf{D}^1$).

Table 3 presents the sizes of $\mathbf{D}^1$ and $\mathbf{D}^2$ that result from an orthogonal combination of the factors Size of the common mode, I, and Array/total ratio, r. Note that the total $I \times (J_1 + J_2)$ was kept constant at a value of 20,000 entries.

The other two factors, which pertain to the amount of noise in the data matrices, are:

TABLE 3.
The sizes of $\mathbf{D}^1$ $(I \times J_1)$ and $\mathbf{D}^2$ $(I \times J_2)$ that are obtained by an orthogonal combination of the factors Size of the common mode, I, and *Array/total ratio*, r.

I	r	Size of $\mathbf{D}^1$ $(I \times J_1)$	Size of $\mathbf{D}^2$ $(I \times J_2)$
50	0.10	(50×40)	(50×360)
50	0.30	(50×120)	(50×280)
50	0.50	(50×200)	(50×200)
50	0.70	(50×280)	(50×120)
50	0.90	(50×360)	(50×40)
100	0.10	(100×20)	(100×180)
100	0.30	(100×60)	(100×140)
100	0.50	(100×100)	(100×100)
100	0.70	(100×140)	(100×60)
100	0.90	(100×180)	(100×20)
200	0.10	(200×10)	(200×90)
200	0.30	(200×30)	(200×70)
200	0.50	(200×50)	(200×50)
200	0.70	(200×70)	(200×30)
200	0.90	(200×90)	(200×10)

TABLE 4.
The values of ε_1 and ε_2 resulting from an orthogonal combination of the factors Total amount of noise, ε^{tot}, and Relative amount of noise, ε^{rel}.

ε^{tot}	ε^{rel}	ε_1	ε_2
0.20	0	0	0.20
0.20	0.25	0.05	0.15
0.20	0.50	0.10	0.10
0.20	0.75	0.15	0.05
0.20	1	0.20	0
0.40	0	0	0.40
0.40	0.25	0.10	0.30
0.40	0.50	0.20	0.20
0.40	0.75	0.30	0.10
0.40	1	0.40	0

(c) the *Total amount of noise*, ε^{tot}, in the data, defined as $\varepsilon_1 + \varepsilon_2$, with ε_1 (ε_2) being the expected amount of noise in $\mathbf{D}^1$ ($\mathbf{D}^2$). This factor was manipulated at two levels: 0.20 and 0.40. Note the difference between ε (i.e., the expected amount of noise) and π (i.e., the estimated amount of noise);

(d) the *Relative amount of noise*, ε^{rel}, defined as $\frac{\varepsilon_1}{\varepsilon^{\text{tot}}}$. This factor was manipulated at five levels: 0, 0.25, 0.50, 0.75, 1. Note that $\varepsilon^{\text{rel}} = 1$ ($\varepsilon^{\text{rel}} = 0$) implies that only $\mathbf{D}^1$ ($\mathbf{D}^2$) is subject to noise, and that $\varepsilon^{\text{rel}} = 0.50$ implies equal amounts of noise for both data blocks.

In Table 4, the values of ε_1 and ε_2 are displayed that are obtained by orthogonally combining the factors Total amount of noise, ε^{tot}, and Relative amount of noise, ε^{rel}.

Procedure For each combination of the levels of the four manipulated factors, a set $\tilde{\mathbf{D}}$ of coupled data matrices $\mathbf{D}^n$ $(n = 1, 2)$ was constructed as follows: True matrices $\mathbf{A}^{(T)}$ and $\mathbf{B}^{n(T)}$ with four bundles were generated by independently sampling entries from a Bernoulli distribution with a parameter value of 0.50. Next, the true matrices $\mathbf{T}^n$ were calculated by combining

$\mathbf{A}^{(T)}$ and $\mathbf{B}^{n(T)}$ by the *SIMCLAS* decomposition rule in (2). It should be noted that $\mathbf{A}^{(T)}$ and $\mathbf{B}^{n(T)}$ (see above) were generated subject to the constraint that all bundle-specific classes (i.e., a class of elements belonging to one bundle only) were non-empty. This constraint was imposed to ensure that the *SIMCLAS* decomposition of $\mathbf{T}^n$ is unique upon a trivial permutation of the bundles (Ceulemans & Van Mechelen, 2003; Van Mechelen et al., 1995). Note that the true number of bundles was not manipulated here, because a pilot study revealed that this factor has no influence on the results. Next, for each $\mathbf{T}^n$, a data matrix $\mathbf{D}^n$ was constructed by altering the entries of $\mathbf{T}^n$ with a probability ε_n.

Per cell of the design, 20 data sets were generated, resulting in 20 (replications) $\times$ 3 (Size of the common mode) $\times$ 5 (Array/total ratio) $\times$ 2 (Total amount of noise) $\times$ 5 (Relative amount of noise) $=$ 3000 different data sets. These data sets were analyzed with a *block-homogeneous SIMCLAS* analysis given $P = 4$. A multi-start procedure was implemented using 15 starts (with 100 *SA* chains in Step 2 of the algorithm): Two rational starts, five random starts, and eight smart-random starts (see Section *The* SIMCLAS *algorithm*). In addition, in order to compare the *SIMCLAS* technique to the *concatenated HICLAS* approach, a *concatenated HICLAS* analysis with $P = 4$ was performed on each data set, which boils down to a *HICLAS* analysis on $\mathbf{D}^*$. To this end, one may apply the second step of the *SIMCLAS* algorithm to each data set, using 100 starts (i.e., *SA* chains, see the Appendix), with the loss function that has to be minimized being equal to the number of discrepancies between the data and the model.

4.2.2. Results

Goodness-of-Fit Studying the degree to which the *SIMCLAS* algorithm succeeds in maximizing the likelihood function (4) or (3) is problematic because the optimal value for this function is unknown, except in the case of noise-free data. However, since the true data set $\tilde{\mathbf{T}}$ is always a valid *SIMCLAS* model for the coupled data set $\tilde{\mathbf{D}}$, the likelihood for $\tilde{\mathbf{T}}$ is a lower bound for the optimal value of the likelihood function. As a consequence, when the likelihood for $\tilde{\mathbf{T}}$ exceeds the likelihood for $\tilde{\mathbf{M}}$, we are sure that the algorithm got stuck in a local optimum. This was, however, never the case in our simulation.

Goodness-of-Recovery To address the question to which extent the *SIMCLAS* algorithm discloses the true structure underlying coupled data, $\mathbf{A}^{(M)}$ is compared to $\mathbf{A}^{(T)}$ by computing the kappa coefficient[5] (Cohen, 1960), denoted as $\kappa_{\mathbf{A}}$. Since the bundles of a *SIMCLAS* solution can be freely permuted, the maximal $\kappa_{\mathbf{A}}$ was computed across all possible permutations. The $\kappa_{\mathbf{A}}$ statistic equals one when perfect recovery is encountered and zero when recovery is at chance level.

The mean $\kappa_{\mathbf{A}}$ value, across all simulated data sets, equals 0.997 ($SD = 0.01$), implying that the *SIMCLAS* technique recovers the underlying structure very well. To study how the recovery performance varies as a function of the data characteristics, an analysis of variance was performed with $\kappa_{\mathbf{A}}$ as the dependent variable and the data characteristics as independent variables. This analysis, only taking effects with a sizeable intraclass correlation $\hat{\rho}_I$ (Kirk, 1982; Haggard, 1958) into account (i.e., $\hat{\rho}_I \geq 0.10$), reveals that recovery decreases when the size of the common mode increases, with this effect being more pronounced when the amount of noise

[5]The kappa coefficient κ between two dichotomous variables can be computed as follows:

$$\kappa = \frac{(p_{00} + p_{11}) - (p_{0.}p_{.0} + p_{1.}p_{.1})}{1 - (p_{0.}p_{.0} + p_{1.}p_{.1})}, \tag{7}$$

with p_{00} (p_{11}) the proportion of zero-agreements (one-agreements) and $p_{0.}$ and $p_{1.}$ ($p_{.0}$ and $p_{.1}$) the marginal proportion of zeros and ones for the first (second) variable.

in the data is large ($\hat{\rho}_I = 0.15$). This interaction, however, is qualified by two three-way interactions ($\hat{\rho}_I$ around 0.13) and a complicated four-way interaction ($\hat{\rho}_I = 0.16$): The pattern of the two-way interaction is more pronounced when the largest data matrix is subject to more noise than the other one, or when both matrices contain a considerable (but equal) amount of noise. However, perfect recovery is observed when one of both matrices is noise-free.

For each simulated data set, the $\kappa_{\mathbf{A}}$ statistic was also computed for the *concatenated HICLAS* solution. The mean $\kappa_{\mathbf{A}}$ value for *concatenated HICLAS* equals 0.98 ($SD = 0.06$), implying that, on average, *SIMCLAS* outperforms *concatenated HICLAS*. An analysis of variance with the difference in $\kappa_{\mathbf{A}}$ between *SIMCLAS* and *concatenated HICLAS* as the dependent variable and the data characteristics as the independent variables, shows that *SIMCLAS* especially outperforms *concatenated HICLAS* when the largest matrix is most noise-prone and when the data contain a large (total) amount of noise ($\hat{\rho}_I = 0.71$).

4.3. Second Simulation Study: Block-Homogeneous SIMCLAS for More than Two Data Blocks

4.3.1. Design and Procedure In this second simulation, we extended the first study to the case of 10 coupled data blocks. Since increasing the number of data blocks lengthens considerably the computation time, we fixed (1) the *Size of the common mode*, I, at 50, (2) the *Array/total ratio*, r, at 0.90 by setting the size of large and small data blocks at 50×90 and 50×10, respectively, and (3) the *Total amount of noise*, ε_{tot}, which is the sum of $\varepsilon_{\text{larger}}$ and $\varepsilon_{\text{smaller}}$, with $\varepsilon_{\text{larger}}$ and $\varepsilon_{\text{smaller}}$ denoting the noise for the large and the small data blocks, respectively; ε_{tot} is set at 0.40. However, we also manipulated two factors in a completely randomized design. First, the *Dominance of the large data blocks*, defined as the number of large data blocks (50×90) relative to the total number of data blocks (i.e., 10), is manipulated at three levels: 0.2, 0.5, and 0.8. Second, the *Noise level*, in which we manipulated the amount of noise for the larger and smaller data blocks at seven levels: $(0, 0.40)$, $(0.10, 0.30)$, $(0.20, 0.40)$, $(0.20, 0.20)$, $(0.40, 0.20)$, $(0.30, 0.10)$, and $(0.40, 0)$.

To generate 20 data sets per cell of the design, the same procedure was used as before, resulting in 20 (replications) $\times$ 3 (Number of large data blocks) $\times$ 7 (Noise level) = 420 different data sets. Subsequently, a *block-homogeneous SIMCLAS* analysis (with 15 multi-starts overall and 100 *SA* chains in Step 2) and a *concatenated HICLAS* analysis (with 100 *SA* chains) with $P = 4$ was applied to each data set.

4.3.2. Results With respect to goodness-of-fit, for all simulated data sets the likelihood for $\tilde{\mathbf{M}}$ exceeds the likelihood for $\tilde{\mathbf{T}}$. To compare *block-homogeneous SIMCLAS* to *concatenated HICLAS* in terms of recovery performance, we computed the $\kappa_{\mathbf{A}}$ value for the solutions obtained with both techniques. It appears that *block-homogeneous SIMCLAS* (mean of 0.98) outperforms *concatenated HICLAS* (mean of 0.94) to a larger extent than in the first study. Moreover, an analysis of variance revealed that this effect is more pronounced ($\hat{\rho}_I = 0.28$) when (1) the large data blocks contain all the noise (i.e., the seventh level of *Noise level*), or (2) the large data blocks are more noise-prone and the total amount of noise in the data is large (i.e., the fifth level of *Noise level*). Interestingly, as can be seen in Figure 1, the latter effect becomes larger when the number of large data blocks relative to the total number of blocks increases ($\hat{\rho}_I = 0.49$). Finally, when the small blocks are most noise-prone or in the case of equal noise, both techniques perform equally well (and almost perfectly, i.e., $\kappa_{\mathbf{A}} > 0.99$).

4.4. Third Simulation Study: Block-Heterogeneous SIMCLAS for Two Data Blocks

4.4.1. Design and Procedure In the third simulation study, we extended the first study to the case of two coupled data blocks with noise heterogeneity within and between blocks. In order to limit the number of conditions, the *Size of the common mode*, I, was fixed at 50, and

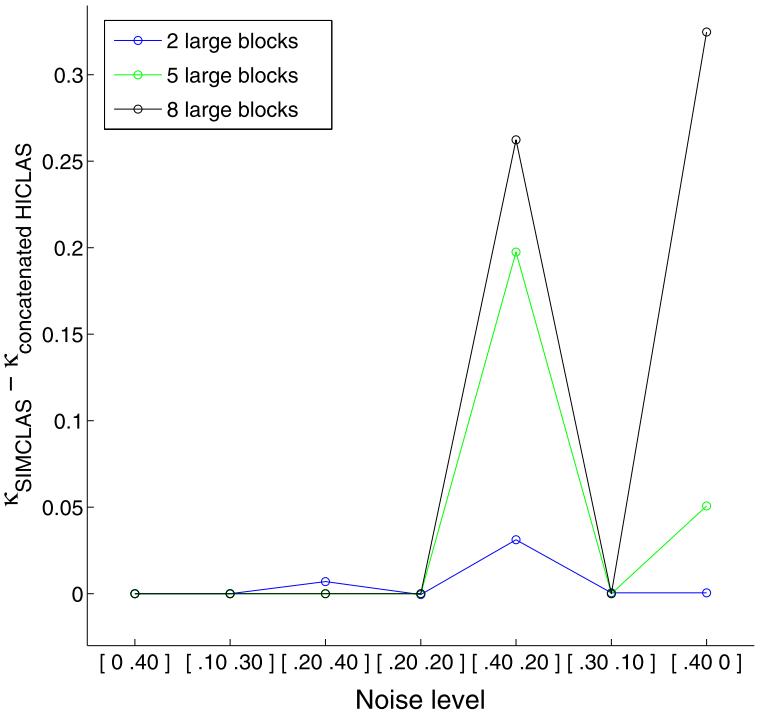

FIGURE 1.
Mean difference in κ between *SIMCLAS* and *concatenated HICLAS* as a function of the Noise level and the Dominance of the large data blocks.

the *Array/total ratio*, r, at 0.90, resulting in one large and one small block of size 50×90 and 50×10, respectively. The *Noise level* was manipulated at the same seven levels as in the second study.

To simulate the data we used the same procedure as in the other two simulation studies, except the way in which noise was added to $\tilde{\mathbf{T}}$. Specifically, to obtain heterogeneous noise within data blocks, we selected 25 rows in $\mathbf{T}^1$ and $\mathbf{T}^2$ randomly to which a low amount of noise ($\varepsilon^n_{\text{low}}$) was added; the other 25 rows were perturbed with a larger amount of noise ($\varepsilon^n_{\text{high}}$); for ε^n levels of 0.05, 0.10, 0.20, 0.30, and 0.35, $\varepsilon^n_{\text{low}}$ and $\varepsilon^n_{\text{high}}$ were respectively chosen as follows: (0.025, 0.075), (0.050, 0.150), (0.100, 0.300), (0.150, 0.450), and (0.250, 0.450). As such, conditions were obtained in which noise heterogeneity is present within and between data blocks. For each cell of the design, 20 replications were used, resulting in 20 (replications) $\times$ 7 (Noise levels) $=$ 140 simulated data sets. Subsequently, a *block-heterogeneous SIMCLAS* analysis (with 15 multi-starts overall and 100 *SA* chains in Step 2) and a *concatenated HICLAS* analysis (with 100 *SA* chains) with $P = 4$ was applied to each simulated data set.

4.4.2. Results Regarding goodness-of-fit, for all data sets the likelihood for $\tilde{\mathbf{T}}$ does not exceed the likelihood for $\tilde{\mathbf{M}}$. In order to compare the recovery performance of both techniques, we calculated the $\kappa_{\mathbf{A}}$ value for each obtained *block-heterogeneous SIMCLAS* and *concatenated HICLAS* solution. The results show that *SIMCLAS* (mean of 0.90) clearly outperforms *concatenated HICLAS* (mean of 0.81) and this to a larger extent than in the first and second study. Further, this effect is more pronounced in these conditions where the largest block is subject to the largest amount of noise. In particular, the average difference in $\kappa_{\mathbf{A}}$ value between *SIMCLAS* and *concatenated HICLAS* in the fifth (i.e., $\varepsilon_1 = 0.40$ and $\varepsilon_2 = 0.20$), sixth (i.e., $\varepsilon_1 = 0.30$ and

$\varepsilon_2 = 0.10$), and seventh (i.e., $\varepsilon_1 = 0.40$ and $\varepsilon_2 = 0$) condition of *Noise level* amounts to 0.14, 0.18, and 0.32; the mean *SIMCLAS* $\kappa_{\mathbf{A}}$ values in these conditions equal 0.61, 0.86, and 0.95. In the other conditions (i.e., equal amount of noise or the smallest block being most noise-prone), both strategies perform equally well.

4.5. Discussion of the Results

To understand why the *SIMCLAS* approach outperforms the *concatenated HICLAS* approach, note that the latter (deterministic) approach is equivalent to a stochastic version of model (2) in which it is assumed that the noise entries $|e^n_{ij_n}| \sim Bern(\pi)$ $(n = 1 \ldots N)$. As a consequence, when coupled data matrices are subject to noise heterogeneity between and/or within data blocks, the concatenation approach, unlike *SIMCLAS*, is based on a misspecified and a too simple model. This also explains why *SIMCLAS* especially outperforms the concatenation approach in those conditions where the difference in noise between and within the data matrices is large, because in those conditions the misspecification of the concatenation approach becomes worse.

5. Illustrative Application

In this section we will illustrate the superior performance of the *block-homogeneous SIMCLAS* technique over the *concatenated HICLAS* approach in terms of disclosing the 'true' structure underlying empirical coupled data. To this end, we will apply both techniques to a data set from the field of categorization of semantic concepts. In this field, one of the main research questions pertains to whether elements (i.e., exemplars) of different categories can be distinguished on the basis of their properties (i.e., features). The data set that we will analyze is a coupled exemplar-by-feature data set from the Animal domain, which belongs to the Leuven Natural Concept Database (De Deyne, Verheyen, Ameel, Vanpaemel, Dry, Voorspoels, & Storms 2008). The first data set was obtained by asking four persons whether or not 60 animals (i.e., 30 exemplars of the category of mammals and 30 of the category of birds) are characterized by 225 features, where these features resulted from a feature generation task in which participants had to list typical features for each category of animals. In the second data set, four persons rated the same 60 animals on a different set of 764 features that also were obtained by a feature generation task, but now participants had to list features that are typical for each animal. As such, the two data sets are coupled via the exemplar (i.e., animal) mode. The data were dichotomized using a majority rule (i.e., a 1 was recorded when two or more participants judged that a feature characterizes an exemplar, and 0 was recorded otherwise).

Earlier analyses of these data (Ceulemans & Storms, 2010) showed that the two categories under study (i.e., mammals and birds) clearly can be distinguished based on the feature profiles. Moreover, the classification of the different animals into mammals and birds conforms the biological knowledge about the hierarchical classification of the animal world. To determine whether *SIMCLAS* or *concatenated HICLAS* best unveils this 'true' underlying structure, both techniques were applied with two bundles to the coupled data set (*SIMCLAS* with 15 multi-starts and 500 *SA* chains in Step 2 of the algorithm, and *HICLAS* with 500 *SA* chains; see Section *The SIMCLAS algorithm*). The obtained bundle matrix for the animals was compared with the true bundle matrix (i.e., partition of the animals) by computing the kappa coefficient κ_{animals}. For both techniques, the κ_{animals} value amounts to 0.97, indicating that the true structure is recovered almost perfectly. Specifically, except for one mammal that is erroneously categorized as a bird, both techniques classify all animals correctly. It can be concluded that for this data set *SIMCLAS* does not outperform *concatenated HICLAS*. This result, however, is not surprising in this

case as *concatenated HICLAS* already recovers the underlying structure almost perfectly, leaving *SIMCLAS* no room for improvement (i.e., a ceiling effect). Moreover, the same amount of noise appears to be present in both data sets (i.e., the estimated π is 0.075 and 0.068 for the first and second data set, respectively). In the simulation study it was shown that in such conditions (i.e., equal noise) both strategies perform equally well.

Therefore, we created a new coupled data set as follows: First, we computed for each feature, based on a *HICLAS* analysis of the separate data sets, the proportion of discrepancies between the data and the model, which may give an indication of the amount of noise in the feature scores. Next, we selected in the first data set the features with a proportion of discrepancies smaller than 0.005 (i.e., small amount of noise), whereas in the second data set the features with a proportion of discrepancies larger than 0.10 (i.e., large amount of noise) were retained; this resulted in 29 and 167 features being selected, respectively. As such, we mimicked the simulation conditions in which *SIMCLAS* outperforms *concatenated HICLAS* to a large extent (i.e., largest block being most noise-prone). Such a data set may be obtained, for example, when a health psychologist collects for the same subjects information regarding, on the one hand, a small number of health parameters which can be measured very reliably (e.g., whether or not the blood pressure or cholesterol level are within a range that can be considered healthy), and, on the other hand, a large number of health behaviors as reported by the subjects (with these subjective reports probably not being very reliable). To this new coupled data set, we applied a *SIMCLAS* and a *concatenated HICLAS* analysis (with the same number of starts as above) with two bundles, and we computed the kappa coefficient for the obtained animal bundle matrix; this resulted in *SIMCLAS* ($\kappa_{\text{animals}} = 0.97$) clearly outperforming *concatenated HICLAS* ($\kappa_{\text{animals}} = 0.72$) in terms of disclosing the true partition of the animals. Specifically, *concatenated HICLAS* misclassifies 9 out of the 60 animals, whereas *SIMCLAS* misclassifies only two exemplars.

6. General Discussion

In this paper, the *SIMCLAS* technique, together with an associated data-analytic strategy, was introduced for an integrated analysis of coupled binary data matrices that are subject to noise heterogeneity between and within data blocks. The key idea behind *SIMCLAS* is to downweight in the analysis entries from noisy data matrices (*block-homogeneous SIMCLAS*) or entries from noisy rows within a data matrix (*block-heterogeneous SIMCLAS*). In a simulation study, it was demonstrated that the *SIMCLAS* algorithm has no clear problems with optimizing the loss function (in that a suboptimal solution was never retained for sure), and the algorithm recovers the true structure underlying the coupled data well. Moreover, *SIMCLAS* was shown to outperform *concatenated HICLAS*, in which all data entries contribute equally to the analysis. This effect is more pronounced (1) when the largest data matrices are more noise-prone, (2) when the number of large data matrices is large relative to the number of small matrices, and (3) when the noise is heterogeneous within data matrices. It can be expected that in the case of more than two coupled data blocks with noise heterogeneity within and between blocks (or with less structured noise within blocks), *SIMCLAS* will outperform *concatenated HICLAS* to a larger extent.

In the remainder of this section, we discuss some generalizations of the *SIMCLAS* technique for, on the one hand, the same type of coupled data, and, on the other hand, other types of coupled data.

Generalizations for the Same Type of Data In the *block-homogeneous SIMCLAS* model, the probability that a model entry differs from a data entry is constant within data blocks, whereas in the *block-heterogeneous SIMCLAS* model, this probability is allowed to differ within blocks, but only in a specific way (i.e., between rows within a block). In some applications, however,

other types of noise heterogeneity within data blocks may be encountered. One possibility is that the probability of obtaining a false positive (i.e., $d^n_{ij_n} = 1$ while $m^n_{ij_n} = 0$) may differ from the probability of a false negative (i.e., $d^n_{ij_n} = 0$ while $m^n_{ij_n} = 1$). For instance, in the psychiatric diagnosis example, it may be the case that some psychiatrists easily (or, in contrast, are rather reluctant to) assign a symptom to a patient; this may be because these psychiatrists believe that it is important not to miss relevant symptoms (or, in contrast, that it is harmful to erroneously assign a symptom). To account for this phenomenon within the context of a *HICLAS* analysis of a single two-mode binary data matrix, the *two-error probabilistic HICLAS* model was proposed by Leenen et al. (2008). A similar generalization could be considered for the *SIMCLAS* model, both for the homogeneous and the heterogeneous variant. To implement this, a separate parameter may be introduced for false positives and false negatives, with again these parameters being allowed to differ across data matrices (for the homogeneous variant) or across rows within the same data matrix (in the case of *block-heterogeneous SIMCLAS*). However, estimating the associated noise parameters is not a trivial task.

Generalizations for Other Types of Data The *SIMCLAS* technique may also be extended to the analysis of two (or more) three-way three-mode coupled data blocks that are subject to noise heterogeneity between and/or within blocks. As an example, take an emotion researcher who wants to study individual differences in situation-specific emotional reactions. To this end, the researcher may, on the one hand, register for a set of persons which physiological reactions they display in a number of situations, and, on the other hand, collect for the same sets of persons and situations information regarding to the appraisals that are activated by the situations. Note that this implies two three-way three-mode data blocks that share two modes (i.e., the persons and the situations). In this case, noise heterogeneity between blocks may occur when, for example, the appraisal data are more noise-prone than the physiological measures, because the former measure is more subjective than the latter. To analyze such data, one may adopt an analysis technique that generalizes the *SIMCLAS* approach to the case of coupled three-way data blocks. In particular, a new model may be introduced, which consists of coupled three-way *HICLAS* models (see Leenen et al., 1999; Ceulemans et al., 2003; Ceulemans & Van Mechelen, 2004, 2005) with a noise parameter for each data block that is allowed to differ across these blocks. Analogous extensions may be considered for the case of noise heterogeneity within data blocks.

Appendix: Simulated Annealing to Estimate the Bundle Matrices, Conditional on the Noise Parameters

To estimate, in Step 2 of the *SIMCLAS* algorithm (see section *The SIMCLAS algorithm*), the binary bundle matrices $\mathbf{A}$ and $\mathbf{B}^n$ that maximize the loss function, conditionally upon the noise parameters, a simulated annealing procedure is adopted. Simulated annealing is a local search technique that implies a walk through the solution space. In particular, a chain of solutions, consisting of several subchains, is generated by each time creating a candidate solution based on the current solution. Next, the loss function values of the current and the candidate solution are compared. If the candidate solution has a better loss function value f, it is accepted, which implies that the current solution is replaced by the candidate solution. If the candidate solution, however, has a worse loss function value, it is accepted with a probability that depends on its relative quality (i.e., the difference in loss function value f between the current solution and the candidate one) and the current temperature, a quantity that controls the acceptance probability. At the end of each subchain the temperature is decreased. Subchains are generated until a prespecified subchain stop criterion is met. Finally, the best encountered solution in the chain is retained.

Initialize(S_{current}, L_{current}, T_{initial}, CL, L_{best}, L_{previous});
$T_{\text{current}} := T_{\text{initial}}$;
repeat
  $i_{\text{gen}} := 0$;
  $i_{\text{acc}} := 0$;
  while ($i_{\text{gen}} < CL$) and ($i_{\text{acc}} < (.1 \times CL)$) **do**
    $i_{\text{gen}} := i_{\text{gen}} + 1$;
    generate S_{trial} and associated L_{trial}, based on S_{current};
    draw h from $U(0, 1)$
    if ($L_{\text{trial}} > L_{\text{current}}$) or ($h < \exp(\frac{L_{\text{trial}} - L_{\text{current}}}{T_{\text{current}}})$) **then**
      $S_{\text{current}} := S_{\text{trial}}$;
      $L_{\text{current}} := L_{\text{trial}}$;
      $i_{\text{acc}} := i_{\text{acc}} + 1$
      if $L_{\text{trial}} > L_{\text{best}}$ **then**
        $S_{\text{best}} := S_{\text{trial}}$;
        $L_{\text{best}} := L_{\text{trial}}$
      end if
    end if
  end while
  $T_{\text{current}} := \alpha * T_{\text{current}}$;
  if $L_{\text{current}} = L_{\text{previous}}$ **then**
    $i_{\text{id}} := i_{\text{id}} + 1$
  else
    $i_{\text{id}} := 1$;
    $L_{\text{previous}} := L_{\text{current}}$
  end if
until ($T_{\text{current}} \leq T_{\text{stop}}$) or ($i_{\text{id}} = max_{i_{\text{id}}}$);
return S_{best};

Algorithm A1: Pseudo-code for generating a single *SA* chain in Step 2 of the *SIMCLAS* algorithm.

Based on the results of a pilot study and on the *SA* implementations that have been used for other Hierarchical Classes models (see, Ceulemans et al., 2007), we implemented the procedure for generating a single *SA* chain (see Algorithm A1 for pseudo-code) in the *SIMCLAS* algorithm as follows:

1. An initial solution S_{current} and associated initial loss value L_{current} is obtained by replacing the P columns of each bundle matrix by P data vectors sampled at random (i.e., for $\mathbf{A}$, column vectors are drawn from the different $\mathbf{D}^n$, whereas for each $\mathbf{B}^n$, row vectors are chosen from the corresponding $\mathbf{D}^n$).
2. The initial temperature T_{initial} is obtained by running a subchain of solutions and accepting all solutions; subsequently, the average increase in the likelihood function across those links in which worse solutions are accepted, is divided by ln(0.8); as such, during the first subchains in which the algorithm is still far from the optimal solution, worse solutions are accepted with a high probability (see Kirkpatrick et al., 1983; Aarts et al., 1997; Ceulemans et al., 2007).
3. A candidate solution S_{trial}, and associated loss value L_{trial}, is obtained from the current solution S_{current} by altering the value of a randomly chosen cell of a randomly chosen bundle matrix, with each cell of each bundle matrix having the same probability of being changed.

4. A worse candidate solution is accepted if: $p < \exp((L_{\text{trial}} - L_{\text{current}})/T_{\text{current}})$, with p being a number generated from a uniform (0, 1) distribution.
5. A subchain stops (1) if the number of generated solutions i_{gen} equals the maximum number of solutions $CL = ((I + \sum_{n=1}^{N} J_n) \times 2^P) \times 5$, or (2) if the number of accepted solutions i_{acc} equals $CL \times 0.10$.
6. At the end of each subchain, the temperature is decreased by a factor $\alpha = 0.90$, implying a smaller acceptance probability for worse solutions: $T_{\text{current}} = 0.9 \times T_{\text{current}}$.
7. An *SA* chain stops when (1) the current temperature becomes smaller than $T_{\text{stop}} = 0.000001$, or (2) the number of subsequent subchains i_{id} with an identical loss value L_{current} for the last accepted solution in each subchain (i.e., $L_{\text{current}} = L_{\text{previous}}$) equals $max_{i_{\text{id}}}$, which is set to five.
8. The retained solution is the best encountered solution S_{best} across all subchains.

To lower the risk of ending in a suboptimal solution (i.e., local optimum), a multi-start procedure may be advised, which consists of running 100 *SA* chains, each time with a different initial solution and initial temperature (see Steps 1 and 2), and retaining the best encountered solution across all chains (see Ceulemans et al., 2007).

References

Aarts, E.H.L., Korst, J.H.M., & van Laarhoven, P.J.M. (1997). Simulated annealing. In E.H.L. Aarts & J.K. Lenstra (Eds.), *Local search in combinatorial optimization* (pp. 91–120). Chichester: Wiley.

Barbut, M., & Monjardet, B. (1970). *Ordre et classification : Algèbre et combinatoire*. Paris: Hachette.

Birkhoff, G. (1940). *Lattice theory*. Providence: Am. Math. Soc.

Ceulemans, E., & Storms, G. (2010). Detecting intra- and inter-categorical structure in semantic concepts using HICLAS. *Acta Psychologica*, *133*, 296–304.

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

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

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

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

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2007). The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika*, *72*, 377–391.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

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

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M., Voorspoels, W., & Storms, G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavioral Research Methods*, *40*, 1030–1048.

Haggard, E.A. (1958). *Intraclass correlation and the analysis of variance*. New York: Dryden.

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

Kiers, H.A.L., & ten Berge, J.M.F. (1989). Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices for all populations. *Psychometrika*, *54*, 467–473.

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 simultaneous structure. *British Journal of Mathematical & Statistical Psychology*, *47*, 109–126.

Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont: Brooks/Cole.

Kirkpatrick, S., Gelatt, C.D.J., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, *220*, 671–680.

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

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

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

Millsap, R.E., & Meredith, W. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika*, *53*, 123–134.

ten Berge, J.M.F., Kiers, H.A.L., & van der Stel, V. (1992). Simultaneous components analysis. *Statistica Applicata*, *4*, 377–392.

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.

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.

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

Van Mechelen, I., & Smilde, A.K. (2009). *A generic model for data fusion*. Paper presented at the 6th meeting of TRICAP (Three-way methods in chemistry and psychology), June 14–19, Vall de Núria, Spain.

Van Mechelen, I., & Smilde, A.K. (2010). A generic linked-mode decomposition model for data fusion. *Chemometrics and Intelligent Laboratory Systems*, *104*, 83–94.

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

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I., & van den Berg, R.A. (2011). Simultaneous analysis of coupled data matrices subject to different amounts of noise. *British Journal of Mathematical & Statistical Psychology*, *64*, 277–290.

Manuscript Received: 13 APR 2010
Final Version Received: 13 JUL 2011
Published Online Date: 30 MAY 2012