T E C H N I C A L
R E P O R T

0698

THE CHIC MODEL : A GLOBAL MODEL FOR COUPLED BINARY DATA

WILDERJANS, T., CEULEMANS, E. and I. VAN MECHELEN

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

The CHIC model: A global model for coupled binary data

Tom Wilderjans

Eva Ceulemans

Iven Van Mechelen

Katholieke Universiteit Leuven

Author notes: The first author is a Research Assistant of the Fund for Scientific Research - Flanders (Belgium). The research reported in this paper was partially supported by the Research Council of K.U.Leuven (GOA/2005/04). We are grateful to Kristof Vansteelandt for providing us with an interesting data set.
Requests for reprints should be sent to Tom Wilderjans, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium. Email: Tom.Wilderjans@psy.kuleuven.be.

Abstract

Often problems result in the collection of coupled data, which consist of different N-way N-mode data blocks that have one or more modes in common. To reveal the structure underlying such data an integrated modeling strategy, with a single set of parameters for the common mode(s), that is estimated based on the information in all data blocks, may be most appropriate. Such a strategy implies a global model, consisting of different N-way N-mode submodels, and a global loss function that is a (weighted) sum of the loss functions associated with the different submodels. In this paper such a global model for an integrated analysis of a three-way three-mode binary data array and a two-way two-mode binary data matrix that have one mode in common is presented. A simulated annealing algorithm to estimate the model parameters is described and evaluated in a simulation study. An application of the model to real psychological data is discussed.

1 Introduction

In different fields of research, coupled data, which consist of different N-way N-mode data blocks that have one or more modes in common, are collected. For example, in contextualized personality psychology one often wants to investigate how individual differences in specific behaviors in specific situations, are related to individual differences in some kind of dispositions. To study this, two data sets are collected: (1) a three-way three-mode person by situation by behavior data array that denotes the extent to which each member of a group of persons displays each behavior from a set of behaviors in each situation from a set of situations and (2) a two-way two-mode person by disposition data matrix. Note that both data blocks have one mode, that is the person mode, in common.

To analyze coupled data two strategies can be followed. A first strategy, which may be called a segmented one, consists of two consecutive steps: First, a multi-way model is fitted to the data block that is of primary interest; second, the quantification of the common mode as obtained in the first step is used in the analysis of the other data blocks. For example, first the person by situation by behavior data array may be analyzed on the basis of a three-way three-mode decomposition model. In a second step a two-way two-mode model may be fitted to the person by disposition data matrix, the parameters for the common person mode being taken from the first analysis. Note that this strategy implies different loss functions, one for each submodel, that are optimized separately and sequentially. In a second strategy, which may be called an integrated one, a global model, which may consist of different submodels - one for each data block - is fitted to the coupled data set. In this global model modes that are common to the different data blocks are represented by a common set of parameters. To estimate this common set of parameters the information in all data blocks is used. For example, a global model consisting of a three-way three-mode and a two-way two-mode submodel may be fitted to the person by situation by behavior data array and the person by disposition data matrix respectively, with a common set of person parameters that is estimated on the basis of the information in both data blocks. Note that the use of a global model implies the optimization of a global loss function, which optionally can be a sum of separate loss functions, one for each submodel. Note further that also a weighted sum of separate loss functions can be used where the weights denote the degree to which each data block is to influence the estimation of the parameters of the common mode(s).

Depending on the research context, each of the two strategies, as explained above, may be appropriate. As such, a segmented strategy may be preferred when the (categorical or dimensional) structure underlying one of the data blocks is of main interest, the other data block being only used to facilitate the interpretation of the main latent structure. For example, when the person by situation by behavior data array is considered to be of primary importance, this data array may be used to induce a latent structure for persons, situations, and behaviors as well as their association; subsequently the person by disposition data matrix may be only used to further characterize the latent person structure as obtained in the first step, with the person by disposition data matrix not playing any role in the determination of that person structure. In general, a disadvantage of a segmented strategy is that only information from one data block is used when the structure for the common mode(s) is derived. As a result, error in this data block may have too large an influence on the derived structure for the common mode (which may further also result in representation errors for the other - non-common - modes). An integrated strategy is most appropriate when the latent structure underlying the common mode(s) in the data is of primary interest. The information in the different data blocks then is used to derive the latent structure for the common mode(s). In this way different sources of information, which may possibly highlight different aspects of the common mode(s), are integrated. For example, when in the case of the personality data the latent person structure is of primary interest, the person by situation by behavior data array and the person by disposition data matrix may be both used to determine the individual differences structure of the persons. In general, an integrated strategy may compensate for the disadvantages of a segmented strategy and may lead to more stable and correct inferences about the structure of the common mode.

In this paper we want to develop a global model for an integrated analysis of coupled data sets that consist of one binary two-way two-mode data matrix and one binary three-way three-mode data array that have one mode in common. More precisely, we want to construct a global model that consists of two submodels, both being members of the family of hierarchical classes (HICLAS) models (De Boeck & Rosenberg, 1988; Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999; Ceulemans, Van Mechelen, & Leenen, 2003) for binary N-way N-mode data. All hierarchical classes models reduce each mode to a limited number of binary variables, called bundles (or components), that imply an overlapping clustering of the elements of the different modes. The new coupled HICLAS model, denoted by the acronym CHIC, represents the three-way three-mode binary data array by an INDCLAS model (Leenen et al., 1999) and the two-way two-mode binary data matrix by a HICLAS model (De Boeck & Rosenberg, 1988). Both submodels are linked (coupled) to each other by imposing the restriction that the overlapping clustering for the common mode should be the same in both submodels.

The remainder of this paper is organized in five main sections: In Section 2 first the HICLAS and the INDCLAS model are reviewed briefly, and next the new CHIC model is presented. In Section 3 a simulated annealing (SA) algorithm to estimate the parameters of the new model is described. In Section 4 the performance of the SA-algorithm is evaluated in a simulation study. An illustrative application of the CHIC model to real psychological data from the personality domain is presented in Section 5. Section 6 finally contains some concluding remarks.

2 Model

2.1 The HICLAS model

A two-way two-mode HICLAS model approximates an $I \times J$ object by attribute binary data matrix $\mathbf{D}$ by an $I \times J$ binary model matrix $\mathbf{M}$ that can be decomposed into an $I \times P$ binary matrix $\mathbf{A}$ and a $J \times P$ binary matrix $\mathbf{B}$, where P denotes the rank of the model. The P columns of $\mathbf{A}$ and $\mathbf{B}$ define P binary variables that are called object bundles and attribute bundles respectively; consequently, the matrices $\mathbf{A}$ and $\mathbf{B}$ are called the object and attribute bundle matrix. The HICLAS model represents three types of structural relations in $\mathbf{M}$.

Association. The association relation is the binary relation between the objects and the attributes as defined by the 1-entries in the model matrix $\mathbf{M}$. An object i is associated with an attribute j iff $m_{ij} = 1$. The HICLAS model represents the association relation between the objects and the attributes by the following decomposition rule:

$$m_{ij} = \bigoplus_{p=1}^{P} a_{ip} b_{jp}, \tag{1}$$

where $\bigoplus$ denotes a Boolean sum. Note that (1) implies a one-to-one correspondence between the respective object and attribute bundles.

Equivalence. Two equivalence relations are defined on $\mathbf{M}$: one for the object mode and one for the attribute mode. In particular, object i is equivalent to object i' in $\mathbf{M}$ iff both objects are associated with the same set of attributes. Equivalent objects constitute an object class, with those classes implying a partition of the objects. The HICLAS model represents the equivalence relation among the objects by assigning in $\mathbf{A}$ the same set of bundles (i.e., identical bundle patterns) to equivalent objects. The equivalence relation among the attributes, represented in $\mathbf{B}$, is defined similarly.

Hierarchy. Two hierarchical relations are defined, one on each mode of $\mathbf{M}$. An object i is hierarchically below an object i' in $\mathbf{M}$ iff the set of attributes associated with object i is a strict subset of the set of attributes associated with object i'. Note that the hierarchical relation among the objects implies a hierarchical relation among the object classes. The HICLAS model represents the hierarchical relation among the objects by strict subset-superset relations among their bundle patterns in **A**. The hierarchical relation among the attributes is defined similarly and is represented in **B**.

2.2 The INDCLAS model

A three-way three-mode INDCLAS model approximates an $I \times J \times K$ object by attribute by source binary data array $\underline{\mathbf{D}}$ by an $I \times J \times K$ binary model array $\underline{\mathbf{M}}$ that can be decomposed into an $I \times P$ binary object bundle matrix **A**, a $J \times P$ binary attribute bundle matrix **B** and a $K \times P$ binary source bundle matrix **C**, where P denotes the rank of the model. The P columns of **A**, **B** and **C** define P (possibly overlapping) bundles of objects, attributes and sources, respectively. As the HICLAS model, the INDCLAS model represents three types of structural relations in $\underline{\mathbf{M}}$.

Association. An INDCLAS model represents the ternary relation among the objects, attributes and sources, as defined by the 1-entries in $\underline{\mathbf{M}}$, by the following decomposition rule:

$$m_{ijk} = \bigoplus_{p=1}^{P} a_{ip} b_{jp} c_{kp}. \tag{2}$$

Note that (2) implies a one-to-one correspondence among the respective object, attribute and source bundles.

Equivalence. Three equivalence relations are defined; one on each mode of $\underline{\mathbf{M}}$. An element i is equivalent to an element i' in $\underline{\mathbf{M}}$ iff both elements are associated with the same set of pairs of elements of the other two modes. The INDCLAS model represents the equivalence relation in that equivalent elements have an identical bundle pattern in the respective bundle matrices.

Hierarchy. A hierarchical relation is defined on each of the three modes of $\underline{\mathbf{M}}$. An element(class) i is hierarchically below an element(class) i' in $\underline{\mathbf{M}}$ iff the set of pairs of elements of the other two modes that is associated with the first element(class) is a strict subset of the set of pairs of elements of the other two modes that is associated with the second element(class). The INDCLAS model represents the hierarchical relation among the elements of one mode in terms of strict subset-superset relations among the associated bundle patterns.

2.3 A global hierarchical classes model for coupled binary data: CHIC

2.3.1 Model

The combined HICLAS-INDCLAS model for coupled binary data (CHIC) approximates an $I \times J \times K$ object by attribute by source binary data array $\underline{\mathbf{D}}^1$ and an $I \times L$ object by covariate binary data matrix $\mathbf{D}^2$ by an $I \times J \times K$ binary model array $\underline{\mathbf{M}}^1$ and an $I \times L$ binary model matrix $\mathbf{M}^2$, where (1) $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ can be decomposed into an INDCLAS model and a HICLAS model of rank P, respectively, and (2) the common object bundle matrix $\mathbf{A}$ is the same in both models. Note that, without loss of generality, the object mode is considered the common mode. Note further that the term 'covariate', which denotes the distinct mode of $\mathbf{M}^2$, is not to be understood as exogenous, as in the CHIC model both data blocks $\underline{\mathbf{D}}^1$ and $\mathbf{D}^2$ are used to derive the structure underlying the common mode. An important feature of the CHIC model is that it includes a single structure for the common mode in terms of bundles, which makes it possible to relate the underlying structure for $\underline{\mathbf{D}}^1$ to the underlying structure for $\mathbf{D}^2$. This implies that the common objects (object types) can be characterized both in terms of (1) associated attribute-source (-type) pairs and (2) associated covariates (covariate types).

Like the HICLAS and the INDCLAS models, the CHIC model represents three types of structural relations (i.e., association, equivalence and hierarchy) in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$. Below we will successively discuss the representation of each of these three relations more in detail. For this purpose we will make use of the hypothetical three-way three-mode model array $\underline{\mathbf{M}}^1$ and the hypothetical two-way two-mode model matrix $\mathbf{M}^2$ in Tables 1 and 2 as a guiding example; a CHIC model in rank 3 for $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ is presented in Table 3.

Association. The CHIC model represents the ternary relation among the objects, attributes and sources (as defined by the 1-entries in $\underline{\mathbf{M}}^1$), and the binary relation between the objects and the covariates (as defined by the 1-entries in $\mathbf{M}^2$), by the following decomposition rules respectively:

$$m^1_{ijk} = \bigoplus_{p=1}^{P} a_{ip} b^1_{jp} c_{kp}, \tag{3}$$

$$m^2_{il} = \bigoplus_{p=1}^{P} a_{ip} b^2_{lp}, \tag{4}$$

where P denotes the rank of the CHIC model. These decomposition rules are equivalent to:

$$m^1_{ijk} = 1 \Leftrightarrow \exists p : a_{ip} = 1 \wedge b^1_{jp} = 1 \wedge c_{kp} = 1, \tag{5}$$

$$m^2_{il} = 1 \Leftrightarrow \exists p : a_{ip} = 1 \wedge b^2_{lp} = 1. \tag{6}$$

The CHIC decomposition rules imply that (1) an object i is associated with an attribute j and a source k in $\underline{\mathbf{M}}^1$ ($m^1_{ijk} = 1$) iff there is at least one bundle to which object i, attribute j and source k belong, and (2) an object i is associated with a covariate l in $\mathbf{M}^2$ ($m^2_{il} = 1$) iff there exists at least one bundle to which object i and covariate l belong. For example, one can derive from the CHIC model in Table 3 that Object 8, Attribute 4 and Source 1 are associated in $\underline{\mathbf{M}}^1$, because all three elements belong to the second bundle (i.e., OB_2, AB_2 and SB_2). Further, Object 8 and Covariate 5 are associated in $\mathbf{M}^2$, because both elements belong to the third bundle (i.e., OB_3 and CB_3). Note that (3)-(6) imply a one-to-one correspondence among the respective object, attribute, source and covariate bundles.

Equivalence. On each mode of $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ an equivalence relation is defined. An attribute j is equivalent to an attribute j' iff both attributes are associated with the same set of pairs of objects and sources in $\underline{\mathbf{M}}^1$. Similar equivalence relations are defined for the sources and the covariates. Furthermore, an object i is equivalent to an object i' iff both objects are associated with the same set of pairs of attributes and sources in $\underline{\mathbf{M}}^1$ *and* with the same set of covariates in $\mathbf{M}^2$. Note that, as in the HICLAS and INDCLAS models, equivalent elements constitute an element class, with such classes implying a partition of the elements in question. The CHIC model represents the equivalence relation among the attributes, sources, covariates and objects in terms of identical bundle patterns in $\mathbf{B}^1$, $\mathbf{C}$, $\mathbf{B}^2$ and $\mathbf{A}$, respectively. For example, in Table 1 (Table 2) one can see that Attribute 2 (Covariate 1) is equivalent to Attribute 5 (Covariate 6) in $\underline{\mathbf{M}}^1$ ($\mathbf{M}^2$); hence both attributes (covariates) have an identical bundle pattern in Table 3. Further, one can see in Table 1 and Table 2 that Object 1 is equivalent to Object 2 both in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$; hence, both objects have an identical bundle pattern in Table 3.

Hierarchy. On each mode of $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ a hierarchical relation is defined. An attribute j is hierarchically below an attribute j' iff in $\underline{\mathbf{M}}^1$ the set of pairs of objects and sources that is associated with attribute j, is a strict subset of the set of pairs of objects and sources that is associated with attribute j'. Similar hierarchical relations are defined for the sources and the covariates. Furthermore, an object i is hierarchically below an object i' iff (1) in $\underline{\mathbf{M}}^1$ the set of pairs of attributes and sources that is associated with object i, is a subset of the set of pairs of attributes and sources that is associated with object i', (2) in $\mathbf{M}^2$ the set of covariates that is associated with object i, is a subset of the set of covariates that is associated with object i', and (3) at least one of the two subset-superset relations as mentioned above is a strict one. Note that the hierarchical relation among the elements of one mode implies a hierarchical relation among the element classes of that mode. The CHIC model represents the hierarchical relation among the attributes, sources, covariates and objects (classes) by strict subset-superset relations among the bundle patterns in the bundle matrices $\mathbf{B}^1$, $\mathbf{C}$, $\mathbf{B}^2$ and $\mathbf{A}$, respectively. For example, one can see in Table 1 (Table 2) that Attribute 3 (Covariate 2) is hierarchically below Attribute 4 (Covariate 8) in $\underline{\mathbf{M}}^1$ ($\mathbf{M}^2$); hence, the bundle pattern of Attribute 3 (Covariate 2) is a strict subset of the bundle pattern of Attribute 4 (Covariate 8) in Table 3. Further, one can see in Table 1 and Table

2 that Object 3 is hierarchically below Object 7 both in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$; hence, the bundle pattern of Object 3 is a strict subset of the bundle pattern of Object 7 in Table 3.

2.3.2 Graphical representation

An overall graphical representation of the CHIC model can be given that links the graphical representation of the INDCLAS model for $\underline{\mathbf{M}}^1$ (in the lower part of the figure) to the graphical representation of the HICLAS model for $\mathbf{M}^2$ (in the upper part of the figure) by the representation of the underlying bundles of the common (object) mode. As a consequence, a double characterization of the common objects is obtained in terms of (1) attribute-source pairs, and (2) covariates. Further, the overall graphical representation of the CHIC model also accounts for the three structural relations in the model.

As an example, in Figure 1 one may consider the overall graphical representation of the CHIC model in Table 3. The hierarchical classifications of the covariates and the attributes are presented at the top and the bottom of Figure 1 respectively, the hierarchical classification of the attributes being displayed upside down. Classes of equivalent elements are indicated by boxes that enclose the labels of the elements belonging to that class. The hierarchical relation among the element(classe)s of one mode is represented by straight lines that connect the corresponding boxes. Both classifications are linked to each other by the graphical representation of the object bundles, represented by rectangles (i.e., each rectangle corresponds to one object bundle), and the source bundles, represented by hexagons (i.e., each hexagon corresponds to one source bundle); each rectangle (hexagon) further contains the labels of the objects (sources) that belong to the corresponding object (source) bundle.

The association relation between the objects and the covariates is indicated by zigzags that connect corresponding bundle-specific object and covariate classes (i.e., classes of elements that belong to one bundle only). The association relation among the objects, attributes and sources is indicated by zigzags that connect corresponding bundle-specific object and attribute classes, and that include a hexagon, containing the (classes of) sources that belong to the corresponding bundle. The double characterization of the objects in terms of attribute-source pairs and covariates then can be derived immediately from Figure 1 as follows: An object i can be characterized in terms of (1) an attribute-source pair $(j-k)$, and (2) a covariate l, iff there is a path, consisting of straight lines and zigzags, that connects covariate l with attribute j and that contains object i in a rectangle and source k in a hexagon.

For example, in Figure 1 one can see that Object 2 can be characterized in terms of (1) Attribute-Source pair $(5-3)$, and (2) Covariate 7, because there is a path between Covariate 7 and Attribute 5 that contains Object 2 and Source 3. Object 3, however, can not be characterized in terms of (1) Attribute-Source pair $(3-1)$, and (2) Covariate 4, because no path exists between Covariate 4 and Attribute 3 that contains Object 3 and Source 1. Note that the equivalence and hierarchical relations between the objects and the sources cannot be readily derived from Figure 1; as such they are displayed separately in Figure 2 and Figure 3, respectively.

3 Data analysis

3.1 Aim

The aim of a CHIC analysis in rank P of an $I \times J \times K$ binary data array $\underline{\mathbf{D}}^1$ and an $I \times L$ binary data matrix $\mathbf{D}^2$ is to estimate an $I \times J \times K$ binary model array $\underline{\mathbf{M}}^1$ and an $I \times L$ binary model matrix $\mathbf{M}^2$ such that (1) the value of the loss function

$$f(\underline{\mathbf{M}}^1, \mathbf{M}^2) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk} - m^1_{ijk})^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il} - m^2_{il})^2 \tag{7}$$

is minimized, and (2) $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ can be represented by a rank P INDCLAS model and a rank P HICLAS model with a common object bundle matrix $\mathbf{A}$.

3.2 Algorithm

Given an $I \times J \times K$ binary data array $\underline{\mathbf{D}}^1$, an $I \times L$ binary data matrix $\mathbf{D}^2$ and a rank P, the CHIC algorithm goes through two consecutive phases. In the first phase the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$ are estimated such as to minimize loss function (7). In the second phase the bundle matrices are adjusted such as to represent the equivalence and hierarchical relations; this can be done without altering the loss function.

First phase. To estimate the optimal bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$, a simulated annealing (SA) algorithm is used (for a general introduction of SA, see Aarts & Lenstra, 1997). SA is a local search technique that implies a walk through the solution space. The general principle is as follows: Starting from an initial configuration and an initial temperature, a chain of solutions is generated, consisting of a number of subchains. In particular, a subchain is generated by repeating the following procedure until a prespecified subchain stop criterion is satisfied: A new solution is generated from the current solution, with the first current solution of each subchain being the last accepted solution of the previous subchain. If this new solution, in comparison with the current solution, leads to an improvement of the loss function f, this new solution is accepted, which implies that the current solution is replaced by this new solution. To escape from local minima worse new solutions are accepted with a probability that depends on the relative quality of the new solution (i.e., the difference in loss function value f between the current solution and the new solution) and the current temperature. At the end of each subchain the temperature is decreased. Subchains are generated until a prespecified chain stop criterion is met. Finally, the best encountered solution is retained.

In the CHIC algorithm, a SA chain is implemented as follows: First, an initial configuration is generated by replacing the columns of the bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$ by randomly chosen object, attribute, source and covariate data vectors, respectively. Second, the initial temperature is obtained by running a subchain of solutions and accepting all solutions, and by subsequently dividing the average increase in the loss function f across those links in which worse solutions are accepted, by $ln(0.8)$:

$$T_{initial} = \frac{\text{Average loss function increase of the worse solutions}}{\ln(0.8)}, \tag{8}$$

Third, a new solution is obtained from the current solution by changing the value of one randomly chosen parameter, with each parameter of the CHIC model having an equal probability of being changed. Fourth, a new worse solution is accepted if:

$$p < \exp((f_{current} - f_{new})/T_{current}). \tag{9}$$

where p is a number generated from a uniform(0,1) distribution, $T_{current}$ is the current temperature and $f_{current}$ and f_{new} are the loss function values of the current and the new solution, respectively. Fifth, a subchain stops (1) if a maximum number of $((I + J + K + L) \times 2^P) \times 5$ solutions have been generated, or (2) if 10% of this maximum number have been accepted. Sixth, at the end of each subchain the temperature is decreased as follows:

$$T_{current} = 0.9 \times T_{current}. \tag{10}$$

Seven, a SA chain stops when (1) the current temperature becomes smaller than 0.000001 or (2) for the last five subchains there is no change in the loss function value f of the last accepted solution in each subchain.

Second phase. In the second phase, the bundle matrices obtained at the end of the first phase are transformed so as to represent the equivalence and hierarchical relations in $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$. This is accomplished by performing a closure operation (Barbut & Monjardet, 1970; Birkhoff, 1940). More in particular, each 0-entry in $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$ is changed to 1 iff this modification does not alter $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$.

4 Simulation study

4.1 Problem

In this section a simulation study is presented to evaluate the performance of the CHIC algorithm. Two aspects of the algorithm are of main interest: (1) goodness-of-fit, and (2) goodness-of-recovery of the true association, equivalence and hierarchical relations. With respect to the second aspect, the primary interest lies in the common mode, with focus on the performance of the integrated modeling strategy relative to the segmented one. To clarify this, three types of binary $I \times J \times K$ and $I \times L$ array-matrix couples must be distinguished: (1) the true array-matrix couple ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$) that can be represented by a CHIC model of rank P, (2) the data array-matrix couple ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) that is obtained by adding noise to ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), and (3) the model array-matrix couple ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) that is obtained by applying the CHIC algorithm to the data array-matrix couple ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) and that can be represented by a CHIC model of rank P.

Goodness-of-fit. To address the question whether the CHIC algorithm succeeds in finding the global optimum of the loss function (7), the data ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) are compared to the model ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) as obtained from the algorithm. To this end, a badness-of-fit-statistic, denoted as BOF, is used that is defined as the proportion of discrepancies between ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) and ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$). Because ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), like ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$), can be represented by a CHIC model of rank P, the proportion of discrepancies between ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$) and ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) (denoted as badness-of-data, or BOD) can be considered an upper bound for the BOF of the global optimum in rank P. Therefore, BOF exceeding BOD implies that the algorithm ended in a suboptimal solution. This, however, does not mean that BOD exceeding BOF implies that the algorithm found the global optimum, since there may exist a couple ($\underline{\mathbf{M}}^{*1}$,$\mathbf{M}^{*2}$) that is closer to ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) than ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) is to ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$). Note that when no error is added to ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), and, hence, ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) equals ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), BOD, which then equals zero, cannot exceed BOF. Moreover, in this case, unlike in the case in which error is added, the global optimum is known and equals the CHIC model that represents ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$). As a consequence, for data sets without error, BOF also denotes the degree to which ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) differs from the global optimum, whereas in general BOF only measures the degree to which ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) differs from ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$). This implies that BOF-values for data sets with error cannot be readily compared with their counterparts for errorfree data sets.

Goodness-of-recovery. Studying the goodness-of-recovery implies the question to what extent the data-analytic procedure is capable of uncovering the truth underlying the data. With respect to this truth, a distinction has to be drawn between the true reconstructed data entries (i.e., the true array-matrix couple) and the true parameters (i.e., the true bundle matrices) underlying the data. The former is studied by comparing the entries of the obtained model array-matrix couple ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) to the entries of the true array-matrix couple ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), while the latter is examined by comparing the obtained model parameters to the true parameters in terms of the implied equivalence and hierarchical relations. With respect to the latter two relations, the question may be raised whether the structure of the common mode is better recovered than the structure underlying the non-common modes. Special attention is further to be paid to the question which data-analytic strategy, the integrated or the segmented one, is superior in revealing the underlying truth. Note that this last question can be studied with respect to all three relations.

In the remainder, first, in Subsection 4.2, the design of the simulation study is outlined. Next, the simulation results are presented in Subsection 4.3, drawing a distinction between goodness-of-fit (4.3.1) and goodness-of-recovery of the true association, equivalence and hierarchical relations, with special attention to the recovery of the structure of the common mode (4.3.2).

4.2 Design and procedure

Four factors were systematically manipulated in a completely randomized four-factorial design, all factors considered random:

(a) the *Array/total ratio*, r, of the size of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$,$\underline{\mathbf{M}}^1$) to the total size $\underline{\mathbf{T}}^1$+$\mathbf{T}^2$ ($\underline{\mathbf{D}}^1$+$\mathbf{D}^2$,$\underline{\mathbf{M}}^1$+$\mathbf{M}^2$, respectively):

$$r = \frac{I \times J \times K}{(I \times J \times K) + (I \times L)}. \tag{11}$$

This factor was manipulated at four levels: .50, .90, .95, .99. Note that the absolute size of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$,$\underline{\mathbf{M}}^1$) was kept constant at the level of 27000 entries;

(b) the *Dominance*, d, of the common mode over the other modes of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$,$\underline{\mathbf{M}}^1$), defined as the ratio between the size of the common mode and the overall number of entries in $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$,$\underline{\mathbf{M}}^1$). This factor was manipulated at three levels: .0019, .0011, .0007. Taking into account that the size of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$,$\underline{\mathbf{M}}^1$) was kept constant at 27000 entries, this implies that the size $I \times J \times K$ of $\underline{\mathbf{T}}^1$ ($\underline{\mathbf{D}}^1$,$\underline{\mathbf{M}}^1$) was manipulated at three levels: $50 \times 20 \times 27$, $30 \times 30 \times 30$, $20 \times 50 \times 27$;

(c) the *True rank*, P, of the CHIC model for ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), manipulated at three levels: 3, 4, 5;

(d) the *Error level*, ε, defined as the expected proportion of cells of ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) differing from the corresponding cells of ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$):

$$\varepsilon = E\left[\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(t^1_{ijk} - d^1_{ijk})^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}(t^2_{il} - d^2_{il})^2}{IJK + IL}\right]. \tag{12}$$

This factor was manipulated at four levels: .00, .10, .20, .30.

For each combination of an Array/total ratio, r, a Dominance, d, a True rank, P, and an Error level, ε, bundle matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$ were constructed by independently drawing entries from a Bernoulli distribution with parameter value .50, under the constraint that all bundle-specific classes (i.e., elements belonging to one bundle only) were non-empty. The constraint was imposed to ensure that the true array-matrix couple ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), obtained by combining the true matrices $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$ by the INDCLAS and the HICLAS decomposition rules (3) and (4), could not be perfectly represented by a CHIC model of a lower rank than the true rank. Further, for each true array-matrix couple ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$), a data array-matrix couple ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) was constructed by changing the value in each cell of ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$) with a probability ε. The whole data generation procedure was repeated five times to obtain five replications per cell. As a consequence, 5 (replications) $\times$ 4 (Array/total ratio) $\times$ 3 (Dominance) $\times$ 3 (True rank) $\times$ 4 (Error level) = 720 different array-matrix couples ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$) and ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) were obtained. Subsequently, one CHIC analysis (with 10 SA-chains) in the true rank P was performed on each data array-matrix couple ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$). In addition, in order to compare the integrated to the segmented modeling strategy in terms of the recovery of the true structure underlying the common mode, each data array $\underline{\mathbf{D}}^1$ was subjected to one INDCLAS analysis and each data matrix $\mathbf{D}^2$ to one HICLAS analysis, both in the true rank P. Note that these additional analyses were performed by using the CHIC algorithm (with 10 SA-chains) in which the loss function was replaced by the INDCLAS and HICLAS loss function, respectively.

4.3 Results

4.3.1 Goodness-of-fit

To evaluate the degree to which the CHIC algorithm succeeds in minimizing loss function (7), the following *badness-of-fit*-statistic (BOF), taking values between 0 (i.e., perfect fit) and 1 (i.e., no fit at all), was used:

$$BOF = f(\underline{\mathbf{M}}^1, \mathbf{M}^2) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk} - m^1_{ijk})^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il} - m^2_{il})^2}{IJK + IL}. \tag{13}$$

Further, also the $BOF-BOD$-statistic was computed. Note that a negative $BOF-BOD$ implies that the algorithm overfits the data.

The mean $BOF-BOD$ value across all observations equals -0.0014 (with a standard deviation of 0.0051), implying that, on average, the CHIC algorithm slightly overfits the data. Out of the 720 analyses, 712 (98.99%) result in a solution with a BOF value smaller than or equal to the

BOD value and 239 analyses result in a $BOF - BOD$ value equal to 0, implying that in only 1% of the analyses it is sure that the CHIC algorithm ends in a suboptimal solution. An analysis of variance with BOF as dependent variable shows that almost all variance in BOF is explained by the main effect of Error level ($\hat{\rho}_I = .99$), with fit increasing when the Error level decreases. From the study of $BOF - BOD$-statistics, calculated for the separate INDCLAS and HICLAS parts of the CHIC model, it appears that the CHIC algorithm in a large number of cases (i.e., 480 out of 720) overfits the HICLAS part of the model to a small degree, while in almost all cases the INDCLAS part is neither underfitted nor overfitted. Note that these findings are in line with earlier simulation results for the INDCLAS (see Leenen et al., 1999) and HICLAS models (see Leenen & Van Mechelen, 2001).

4.3.2 Goodness-of-recovery

When evaluating the extent to which the CHIC algorithm is capable of uncovering the true structure underlying the data, one has to differentiate between reconstructing (1) the true association relation, and (2) the true equivalence and hierarchical relations. Further attention is to be paid to the recovery of the structure of the common mode in comparison with the other modes. Finally, the integrated modeling strategy is to be compared to the segmented one.

Association relation. To evaluate the recovery of the true association relation of the CHIC model, the following *goodness-of-recovery*-statistic (GOR) was computed:

$$GOR = 1 - \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(t^1_{ijk} - m^1_{ijk})^2 + \sum_{i=1}^{I}\sum_{l=1}^{L}(t^2_{il} - m^2_{il})^2}{IJK + IL}. \qquad (14)$$

This statistic yields a value between 1 (i.e., perfect recovery) and 0 (i.e., no recovery at all).

The mean GOR across all observations equals 0.9928 (the standard deviation equals 0.0155), which means that the model array-matrix couple ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) differs on average 0.72% from the true array-matrix couple ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$). The association relation is perfectly recovered in 221 analyses. Note that perfect recovery is only possible for these data sets in which BOF equals BOD (i.e., 239 out of 720; see *Goodness-of-fit*). An analysis of variance with GOR as the dependent variable (in which only effects with $\hat{\rho}_I \geq .10$ are considered) shows that recovery increases when the Array/total ratio increases ($\hat{\rho}_I = .17$) and the Error level decreases ($\hat{\rho}_I = .12$). Both main effects, however, are qualified by a strong Array/total ratio-Error level interaction ($\hat{\rho}_I = .45$): The increase in recovery with increasing Array/total ratio is more pronounced for large Error levels than for small Error levels, while in the errorfree conditions recovery decreases with increasing Array/total ratio. The main effects of True rank and Dominance can be neglected ($\hat{\rho}_I \leq .02$). Studying GOR for the INDCLAS and the HICLAS parts of the CHIC model separately further shows that the true INDCLAS association relation is far better recovered than the true HICLAS association relation.

Equivalence and hierarchical relation. To quantify the degree to which the equivalence relation is recovered, the corrected Rand index (Hubert & Arabie, 1985) between the partition of the set of objects (resp., attributes, sources, covariates) in the CHIC model for ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$) and its counterpart in the CHIC model for ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) was computed (CRI). This index takes a value between 1 (indicating that both partitions are identical) and 0 (indicating that both partitions do not correspond more than expected by chance). The performance of the CHIC algorithm with respect to the recovery of the true hierarchical relation, was measured by computing the object hierarchy matrix associated with ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) (respectively ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$)), defined as the $I \times I$ binary matrix $\mathbf{U}^{(\underline{\mathbf{M}}^1,\mathbf{M}^2)}$ (resp. $\mathbf{U}^{(\underline{\mathbf{T}}^1,\mathbf{T}^2)}$) with $u_{ii'}^{(\underline{\mathbf{M}}^1,\mathbf{M}^2)} = 1$ (resp. $u_{ii'}^{(\underline{\mathbf{T}}^1,\mathbf{T}^2)} = 1$) iff object i is hierarchically below object i' in ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) (respectively ($\underline{\mathbf{T}}^1$,$\mathbf{T}^2$)); subsequently, a *goodness-of-hierarchy recovery*-statistic ($GOHR$) was calculated for the objects, defined as the proportion of concordancies between $\mathbf{U}^{(\underline{\mathbf{M}}^1,\mathbf{M}^2)}$ and $\mathbf{U}^{(\underline{\mathbf{T}}^1,\mathbf{T}^2)}$:

$$GOHR = 1 - \frac{\sum_{i=1}^{I}\sum_{i'=1}^{I}\left(u_{ii'}^{(\underline{\mathbf{M}}^1,\mathbf{M}^2)} - u_{ii'}^{(\underline{\mathbf{T}}^1,\mathbf{T}^2)}\right)^2}{I^2}. \tag{15}$$

$GOHR$ yields a value between 1 (i.e., perfect recovery) and 0 (i.e., no recovery at all). Similarly, a $GOHR$-statistic for the attributes, sources and covariates was computed. Finally, a combined CRI-statistic ($cCRI$) and a combined GOHR-statistic ($cGOHR$) were computed by, respectively, averaging the CRI and $GOHR$ for the object, attribute, source and covariate equivalence and hierarchical relation, weighted by the number of elements in the respective modes.

The mean $cCRI$ equals 0.8317 (standard deviation of 0.2061) and the mean $cGOHR$ equals 0.9556 (standard deviation of 0.0585), implying a very good recovery of the true equivalence and hierarchical relations. The true equivalence relation is recovered perfectly in 229 analyses, while 221 analyses resulted in perfect recovery of the true hierarchical relation. Note again that perfect recovery is only possible in 239 analyses. Separate analyses of variance with $cCRI$ and $cGOHR$ as dependent variables show similar results: The recovery of the true equivalence and hierarchical relations increases when the Error level decreases ($\hat{\rho}_I = .36$), when the Array/total ratio increases ($\hat{\rho}_I = .18$) and when the Dominance of the common bundle matrix increases ($\hat{\rho}_I = .09$), the main effect of True rank being very small ($\hat{\rho}_I = .02$). The main effects of Error level and Array/total ratio, however, are qualified by an Error level by Array/total ratio interaction ($\hat{\rho}_I = .22$): The increase in recovery when the Array/total ratio increases is more pronounced for large Error levels than for small Error levels, while in conditions without error recovery decreases when the Array/total ratio increases.

The common mode. When comparing the recovery of the common bundle matrix (in terms of the implied equivalence and hierarchical relations) to the recovery of the other bundle matrices, it appears that, on average, the structure of the common mode is better recovered than the structure of the other modes (see Table 4 in which the mean CRI and $GOHR$ for $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$ are displayed). As an aside, one may note the poor recovery of the structure of the non-common mode of $\mathbf{D}^2$ that can be attributed to the overfitting of the HICLAS part of the CHIC model by the CHIC algorithm.

To evaluate the performance of the integrated modeling strategy relative to the segmented one, the recovery of the true association (GOR), equivalence (CRI) and hierarchical relations ($GOHR$) for the CHIC model is compared to the recovery of these relations for separate INDCLAS and HICLAS models, as obtained from a segmented modeling strategy. In Table 5 mean recovery values are presented both for the CHIC model and for separate INDCLAS and HICLAS models. It appears that the integrated modeling strategy outperforms the segmented one: When using the integrated strategy the recovery of all three relations is increased. Note that using the integrated modeling strategy leads to a better recovery of the structure of both the common and the non-common modes.

5 Illustrative application

In this section, the CHIC model is applied to a coupled data set, gathered by Vansteelandt and Van Mechelen (1998), in a study from the domain of contextualized personality psychology. Within this domain one tries to capture individual differences beyond differences in general dispositions (e.g., traits). A key concept in this regard is that of a behavioral signature, which is the profile across situations of the extent to which a specific behavior is displayed in each situation. A major challenge is to capture the structure of individual differences in behavioral signatures, and to relate this structure to individual differences in dispositional variables that reflect cognitive-affective processes underlying the signatures in question. In their study, Vansteelandt and Van Mechelen (1998) asked 54 persons to indicate on a 3-point scale the degree to which they would display 15 hostile behaviors in each of 23 frustrating situations (0=you would not display this behavior in this situation, 1=to a limited extent, 2=to a strong extent). Further, they asked the same persons to rate themselves on 7 dispositional process variables making use of a 7-point scale (1=not applicable at all, 7=applicable to a strong extent). The $54 \times 15 \times 23$ person by behavior by situation data array $\underline{\mathbf{D}}^1$ was dichotomized by recoding scores of 1 and 2 to 1 and 0 to 0, while the 54×7 person by dispositional variable data matrix $\mathbf{D}^2$ was dichotomized by performing a median split on each variable (with a score on and above the median being recoded to 1 and a score under the median to 0).

CHIC analyses in rank one to six (with 1000 SA-chains for each rank) were performed on the dichotomized data ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$). Based on a scree plot of the number of discrepancies between ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) and ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$), rank three was retained. The CHIC algorithm, however, yielded a large number of optimal solutions in this rank. When comparing these solutions to each other, it appeared that almost all differences pertained to the dispositional variables. More in particular, these solutions included identical bundle matrices for the situations and responses, and almost identical bundle matrices for the persons; with regard to the dispositional variables, however, considerable differences showed up. This implies that the structure for the person by behavior by situation data array is univocal, while a large amount of uncertainty exists about (at least part of) the structure for the person by dispositional variable data matrix. To handle this problem, only dispositional variables will further be taken into account that belong to the same class across all different optimal solutions, as shown in Figure 4, which displays the overall graphical representation of the retained CHIC model for the hostility data. Note that in this figure the situations and the responses are represented by their respective classes, with the number of situations and responses belonging to each class denoted between parentheses.

Clearly, the three situation classes form a Guttman scale; as such, they can be conceived as points on an underlying quantitative dimension (Gati & Tversky, 1982). This dimension can be interpreted as a frustration dimension, varying from lowly frustrating (e.g., accidently banging your shins against a park bench), over moderately frustrating (e.g., talking to someone who does not answer you) to highly frustrating (e.g., being unfairly accused of cheating). With respect to the responses, five classes can be distinguished, varying from weak (e.g., grimace and turning away) to strong responses (e.g., enragement). With regard to the persons, almost all of them belong to one of three person types P_1, P_2 and P_3 (that contain respectively 22, 13, and 43 % of the persons); therefore, only these three person types will be further discussed and displayed in Figure 4. From Figure 4 one may derive that all three person types do not display hostile behaviors in lowly frustrating situations. Further, persons belonging to P_1 differ from persons belonging to P_2 and P_3 in that the former react only to highly frustrating situations in terms of tension and enragement, while the latter display hostile behaviors in both moderately and highly frustrating situations. The latter two person types, however, differ from one another in that persons belonging to P_3, unlike those belonging to P_2, differentiate in terms of their behaviors between moderately and highly frustrating situations, with enragement being displayed by them in highly frustrating situations only. These differences in behavioral signatures can, at least in part, be explained by individual differences in dispositions. For example, Persons from P_1 and P_3 differ from persons from P_2 in that the former, unlike the latter, find it important to react to frustration in an assertive way. In this way standing for one's rights may be a source of enragement.

6 Concluding remarks

In this paper the CHIC model, a global model for an integrated analysis of coupled data, has been proposed. More in particular, CHIC can be used to simultaneously model a binary three-way three-mode data array and a binary two-way two-mode data matrix that have one mode in common. This type of data is often encountered in psychology, for example, when subjects rate a set of objects with respect to a set of attributes and when at the same time the same objects are characterized in terms of objective features or when dispositional information is available about the raters. In the context of psychiatric diagnoses, for example, psychiatrists may be asked to rate a set of patient vignettes on a list of symptoms, while at the same time information on the psychiatric diagnoses of the patients under study (in terms of a list of syndromes) is available. In this paper the gain of CHIC as an integrated modeling strategy, as compared to a segmented one, has been demonstrated by means of a simulation study and by means of an illustrative application. More in particular, the results of our simulation study revealed that the CHIC algorithm succeeds well in minimizing the CHIC loss function. Further, it was shown that the true structure underlying the different modes is recovered well, with the structure of the common mode being recovered best. Moreover, the integrated modeling strategy was shown to outperform a segmented one in terms of (1) the recovery of the true association relation of the separate INDCLAS and HICLAS parts of the CHIC model, and of (2) the recovery of the true equivalence and hierarchical relation of the common as well as the non-common modes. In the illustrative application, it has been shown that an integrated modeling strategy may lead to a deeper insight into a phenomenon under study in terms of how its different parts are connected to one another. More in particular, in the application the CHIC model revealed how the structure of individual differences in situation-specific hostile behavior relates to individual differences in process-based dispositional variables.

The proposed CHIC model is a novel model, but it bears interesting relationships to several existing models both inside and outside the hierarchical classes family. Inside the hierarchical classes family, the CHIC model can be conceived as an integrated model that includes two members of this family as component models (viz., the INDCLAS and the HICLAS model). On the other hand, the CHIC model can also be considered a hierarchical classes counterpart of an integrated multiway covariate regression model consisting of a CANDECOMP/PARAFAC submodel for a three-way three-mode (real-valued) data array and a PCA submodel for a two-way two-mode (real-valued) data matrix, with the two data blocks having one mode in common (see Smilde & Kiers, 1999) [1]. In both models a three-mode/two-mode coupled data structure ($\underline{\mathbf{D}}^1$,$\mathbf{D}^2$) is

[1]Note that a distinction can be drawn between two variants of the multiway covariate regression model: A symmetric variant in which both data blocks play an identical role, and an asymmetric variant in which the three-way data block is being considered a predictor block and the two-way block a block of criterion variables, with the CHIC model being closer related to the former variant.

represented by a coupled model structure ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) where: (1) each mode of ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) is reduced to a limited number of components, with the common mode having the same component scores in each submodel, and (2) ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$) can be fully reconstructed on the basis of the different sets of component scores and a linking rule. Both models, however, differ in the following aspects: (1) the CHIC model includes binary component matrices, and as such implies a categorical rather than a dimensional reduction of the modes involved in the data; (2) the CHIC model is based on a Boolean decomposition of ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$), while the multiway covariate regression model is based on a standard real-valued decomposition; and (3) in the CHIC model the component matrices are restricted to represent the quasi-order relations $\leq$ between the elements of the corresponding modes of ($\underline{\mathbf{M}}^1$,$\mathbf{M}^2$).

Finally, the CHIC model, as proposed in the present paper, can be extended in various ways. In this regard, one may first note that the INDCLAS-model, which is included in CHIC to represent the binary three-way three-mode data block, is a rather restrictive model, because (1) each mode is reduced to the same number of bundles, and (2) the linking structure among the object, attribute and source bundles is restricted to a one-to-one correspondence. From a substantive point of view, indeed, in several cases there may be no a priori reason for the object, attribute and source bundle matrices to have the same rank, let alone for a one-to-one correspondence among the respective bundles. As a way out, the CHIC model may be extended to a global model in which the three-way three-mode binary data block is represented by a more general (i.e., less restrictive) model than INDCLAS, like, for example, the TUCKER3-HICLAS (Ceulemans et al., 2003) or the TUCKER2-HICLAS model (Ceulemans & Van Mechelen, 2004). The latter two models can be considered generalizations of the INDCLAS-model in that the relation between the different bundle matrices is allowed to be more complex than a one-to-one correspondence and in that the numbers of bundles for the different modes are allowed to differ. With regard to a second possible generalization one may note that in its present form CHIC is a model for binary data. In psychology, however, often rating-valued data show up (e.g., as obtained from ratings on Likert-scales). The CHIC model could be extended to accommodate such data, by making use of HICLAS models for rating-valued data that have been developed recently (see Van Mechelen, Lombardi, & Ceulemans, in press). Otherwise CHIC extensions could also be developed to accommodate coupled data sets in which one data block is binary and the other one rating-valued.

References

Aarts, E. H., & Lenstra, J. K. (1997). *Local search in combinatorial optimization.* Chichester, UK: Wiley.

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

Birkhoff, G. (1940). *Lattice theory.* Providence: American Mathematical Society.

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

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

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

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

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

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.

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

Van Mechelen, I., Lombardi, L., & Ceulemans, E. (in press). Hierarchical classes modeling of rating data. *Psychometrika*.

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

Table 1: Hypothetical three-way three-mode model matrix $\underline{\mathbf{M}}^1$

Source 1						Source 2						Source 3					
	Attributes						Attributes						Attributes				
Objects	A_1	A_2	A_3	A_4	A_5	Objects	A_1	A_2	A_3	A_4	A_5	Objects	A_1	A_2	A_3	A_4	A_5
O_1	1	1	1	1	1	O_1	1	1	0	1	1	O_1	1	1	1	1	1
O_2	1	1	1	1	1	O_2	1	1	0	1	1	O_2	1	1	1	1	1
O_3	0	0	0	0	0	O_3	1	0	0	0	0	O_3	1	0	0	0	0
O_4	1	1	0	1	1	O_4	1	1	0	1	1	O_4	1	1	0	1	1
O_5	0	0	1	1	0	O_5	0	0	0	0	0	O_5	0	0	1	1	0
O_6	1	1	1	1	1	O_6	1	1	0	1	1	O_6	1	1	1	1	1
O_7	0	0	1	1	0	O_7	1	0	0	0	0	O_7	1	0	1	1	0
O_8	1	1	0	1	1	O_8	1	1	0	1	1	O_8	1	1	0	1	1

Table 2: Hypothetical two-way two-mode model matrix $\mathbf{M}^2$

	Covariates							
Objects	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
O_1	1	1	1	1	0	1	1	1
O_2	1	1	1	1	0	1	1	1
O_3	0	1	0	0	1	0	1	1
O_4	0	0	1	1	0	0	1	1
O_5	1	1	1	0	0	1	0	1
O_6	1	1	1	1	1	1	1	1
O_7	1	1	1	0	1	1	1	1
O_8	0	1	1	1	1	0	1	1

Table 3: CHIC model in rank 3 for the model matrices $\underline{\mathbf{M}}^1$ and $\mathbf{M}^2$ in Table 1 and Table 2

$\mathbf{A}$				$\mathbf{B}^1$				$\mathbf{C}$				$\mathbf{B}^2$			
	OB_1	OB_2	OB_3		AB_1	AB_2	AB_3		SB_1	SB_2	SB_3		CB_1	CB_2	CB_3
O_1	1	1	0	A_1	0	1	1	S_1	1	1	0	C_1	1	0	0
O_2	1	1	0	A_2	0	1	0	S_2	0	1	1	C_2	1	0	1
O_3	0	0	1	A_3	1	0	0	S_3	1	1	1	C_3	1	1	0
O_4	0	1	0	A_4	1	1	0					C_4	0	1	0
O_5	1	0	0	A_5	0	1	0					C_5	0	0	1
O_6	1	1	1									C_6	1	0	0
O_7	1	0	1									C_7	0	1	1
O_8	0	1	1									C_8	1	1	1

Table 4: mean CRI and $GOHR$ value for $\mathbf{A}$, $\mathbf{B}^1$, $\mathbf{C}$ and $\mathbf{B}^2$

	CRI	$GOHR$
$\mathbf{A}$	0.9973	0.9993
$\mathbf{B}^1$	0.9932	0.9983
$\mathbf{C}$	0.9945	0.9987
$\mathbf{B}^2$	0.6612	0.9122

Table 5: Mean GOR, CRI and $GOHR$ for the CHIC model and separate INDCLAS and HICLAS models

		CHIC	separate
GOR	INDCLAS	0.9996	0.9992
	HICLAS	0.9593	0.9135
CRI	$\mathbf{A}^{IND}$	0.9973	0.9896
	$\mathbf{B}^1$	0.9932	0.9895
	$\mathbf{C}$	0.9945	0.9877
	$\mathbf{A}^{HI}$	0.9973	0.6945
	$\mathbf{B}^2$	0.6612	0.5925
$GOHR$	$\mathbf{A}^{IND}$	0.9993	0.9972
	$\mathbf{B}^1$	0.9983	0.9971
	$\mathbf{C}$	0.9987	0.9972
	$\mathbf{A}^{HI}$	0.9993	0.9108
	$\mathbf{B}^2$	0.9122	0.8892

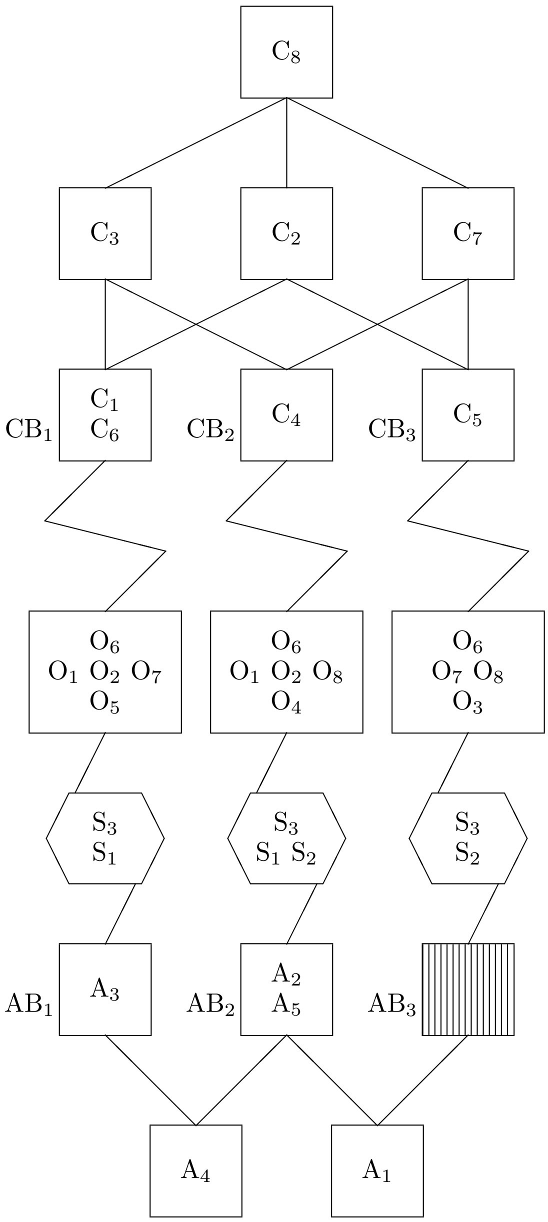

Figure 1: Overall graphical representation of the CHIC model in Table 3.

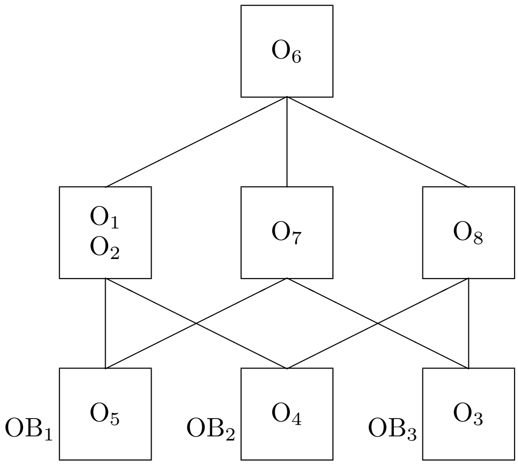

Figure 2: Object hierarchy of the CHIC model in Table 3.

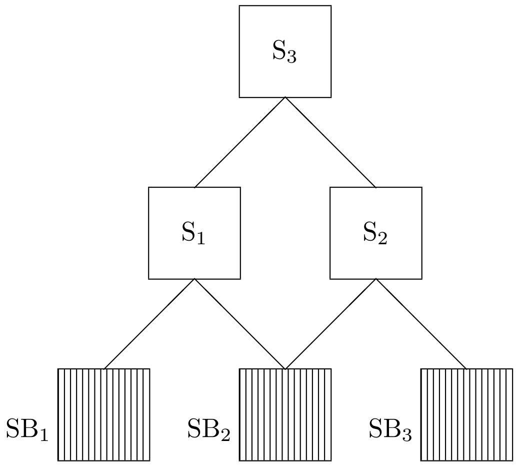

Figure 3: Source hierarchy of the CHIC model in Table 3.

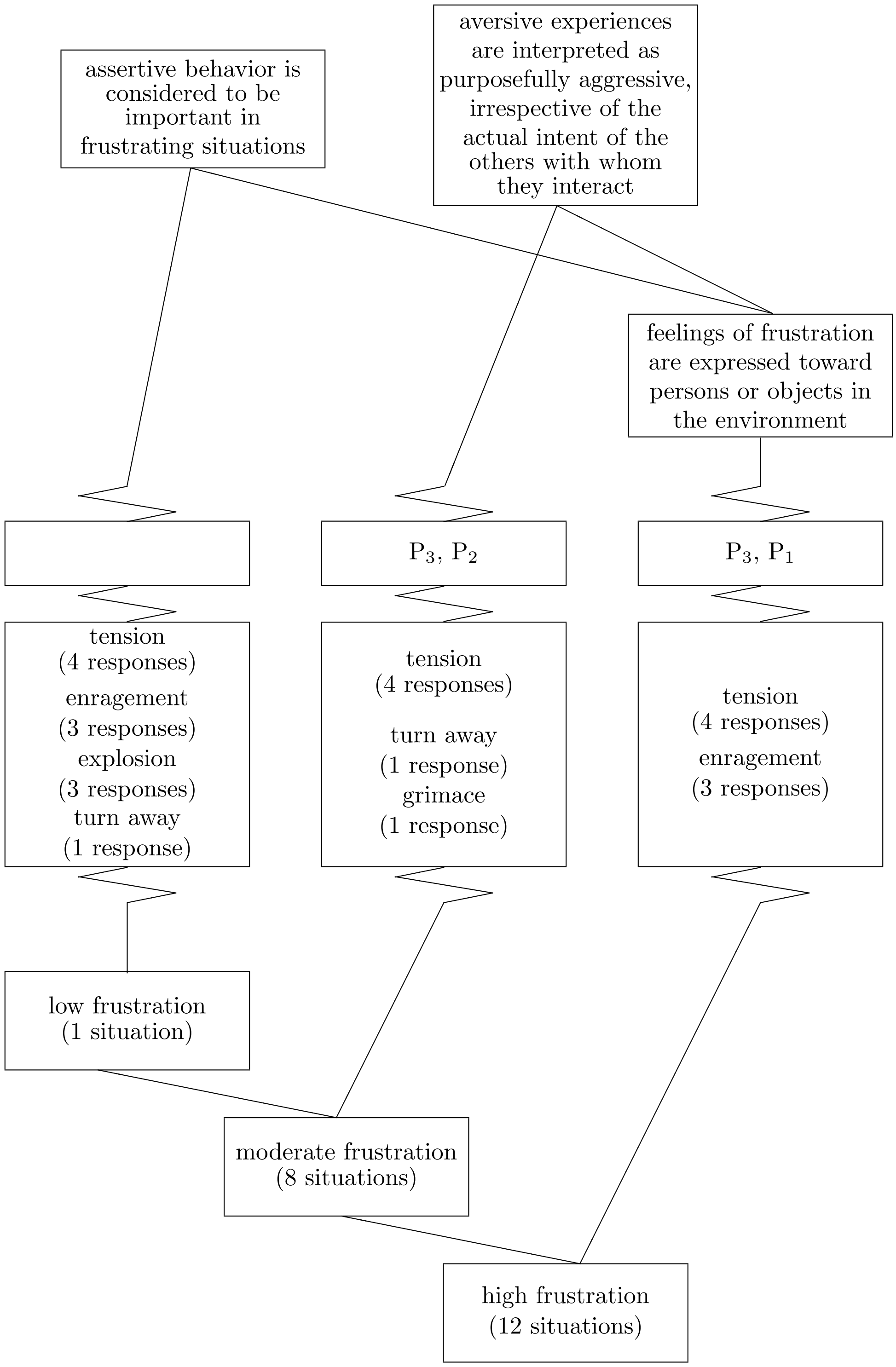

Figure 4: Overall graphical representation of the CHIC model for the hostility data.