A taxonomy of latent structure assumptions for probability matrix decomposition models — Source link

Michel Meulders, Paul De Boeck, Iven Van Mechelen

Institutions: Katholieke Universiteit Leuven

Published on: 01 Jan 2003 - Psychometrika (Springer-Verlag)

Topics: Latent variable model, Latent variable, Latent class model, Probabilistic latent semantic analysis and Gibbs sampling

PSYCHOMETRIKA—VOL. 68, NO. 1, 61–77
MARCH 2003

A TAXONOMY OF LATENT STRUCTURE ASSUMPTIONS FOR PROBABILITY MATRIX DECOMPOSITION MODELS

MICHEL MEULDERS, PAUL DE BOECK, IVEN VAN MECHELEN

KATHOLIEKE UNIVERSITEIT LEUVEN

A taxonomy of latent structure assumptions (LSAs) for probability matrix decomposition (PMD) models is proposed which includes the original PMD model (Maris, De Boeck, & Van Mechelen, 1996) as well as a three-way extension of the multiple classification latent class model (Maris, 1999). It is shown that PMD models involving different LSAs are actually restricted latent class models with latent variables that depend on some external variables. For parameter estimation a combined approach is proposed that uses both a mode-finding algorithm (EM) and a sampling-based approach (Gibbs sampling). A simulation study is conducted to investigate the extent to which information criteria, specific model checks, and checks for global goodness of fit may help to specify the basic assumptions of the different PMD models. Finally, an application is described with models involving different latent structure assumptions for data on hostile behavior in frustrating situations.

Key words: discrete data, matrix decomposition, Bayesian analysis, data augmentation, posterior predictive check, psychometrics.

PMD models were introduced by Maris, De Boeck, and Van Mechelen (1996) to analyze three-way three-mode binary data. The data typically represent associations between two types of elements that are repeatedly observed, for instance, persons who judge whether or not they would display a certain hostile response in a frustrating situation. PMD models have been applied in several substantive contexts such as psychiatric diagnosis (Maris et al., 1996; Meulders, De Boeck, & Van Mechelen, 2001), marketing research (Candel & Maris, 1997), emotion perception in facial expressions (de Bonis, De Boeck, Pérez-Diaz, & Nahas, 1999), and cross-cultural research about the risks of contracting AIDS in different situations (Meulders, De Boeck, Van Mechelen, Gelman, & Maris, 2001).

To explain a rater's judgment about the association of two elements (e.g., a situation and a response), PMD models assume a twofold process: First, it is assumed that, at each new encounter, each of a set of latent features may be linked to an element. Second, it is assumed that the association between two elements results from combining the feature patterns of these elements according to a specific mapping rule. For instance, according to a *disjunctive communality* (DC) rule, the linkage of one common feature to both elements is a sufficient condition for the elements to be associated, whereas, with a *conjunctive dominance* (CD) rule it is necessary that all the features that are linked to one element are also linked to the other element.

The assumption that, at each new encounter, possibly a new pattern of features is linked to each element, will further be denoted the latent structure assumption (LSA) of the model as it describes how observed associations between elements are determined by latent processes. The LSA of the original PMD model (published originally by Maris et al., 1996) is often untenable in practice because of two reasons. First, it implies that all observed associations are statistically

Note: The research reported in this paper was partially supported by the Fund for Scientific Research-Flanders (Belgium) (project G.0207.97 awarded to Paul De Boeck and Iven Van Mechelen), and the Research Fund of K.U. Leuven (F/96/6 fellowship to Andrew Gelman, OT/96/10 project awarded to Iven Van Mechelen and GOA/2000/02 awarded to Paul De Boeck and Iven Van Mechelen). We thank Marcel Croon and Kristof Vansteelandt for commenting on an earlier draft of this paper.

Requests for reprints should be sent to Department of Psychology, Tiensestraat 102, B-3000 Leuven, BELGIUM. E-Mail: Michel.Meulders@psy.kuleuven.ac.be

0033-3123/2003-1/1999-0772-A $00.75/0

independent, and therefore it will also be referred to as the *independence* model. Second, it excludes the existence of individual differences as it implies that persons have the same probability to indicate an association between two elements.

In this paper we will formulate alternative LSAs that do allow to model dependencies in the data and that do allow to capture individual differences. The models that will be presented can formally be regarded as latent class models. For parameter estimation, a combined approach, which involves locating the posterior mode(s) with EM and computing a sample of the posterior distribution with the Gibbs sampler, may be of interest, as both approaches have strengths and weaknesses. More specifically, they differ in how the specification of different types of priors is supported, how standard errors are computed, how convergence is monitored, and how over-parameterization is dealt with. In addition, a strategy for model selection using information criteria as well as global goodness-of-fit tests and specific model checks will be proposed. As will be seen, these are powerful tools which may help to specify the LSA of the PMD model, but they can also help to specify other model aspects, such as the mapping rule and the number of features.

The outline of the remaining parts of the paper is as follows: First, PMD models with alternative LSAs will be presented. Second, parameter estimation will be discussed. Third, tools for model selection will be proposed and their performance will be evaluated in a simulation study. Finally, models involving different LSAs will be applied to real data on hostile behavior in frustrating situations.

A Taxonomy of PMD Models with Different Latent Structure Assumptions

To introduce PMD models with alternative LSAs, we first describe the original PMD model in the context of a specific example. Consider persons $p(p = 1, \ldots, P)$ who judge whether or not they would display a certain hostile response $r(r = 1, \ldots, R)$ (e.g., "curse", "wanting to hit someone") in a certain frustrating situation $s(s = 1, \ldots, S)$ (e.g., "you bang your shins against a park bench", "you just found out that someone opened your personal mail", etc.). Observed judgments are denoted by the random variable Y_p^{sr} which equals 1 if person p would display response r in situation s, and 0 otherwise. Applying the independence PMD model to these data, the following assumptions are made:

1. *Situation-feature association.* It is assumed that persons, at each new encounter, attribute to a situation certain features that pertain to the nature of the frustration in that situation. For example, whether or not the frustrating event is seen as caused intentionally, or whether or not other persons are present in the situation. Formally, the atribution of a feature to a situation is conceived as the realization of a Bernoulli variable $X_{rp}^{sf} \sim \text{Bern}(\sigma_{sf})$ which equals 1 if feature $f(f = 1, \ldots, F)$ is attributed to situation s when person p is judging pair (s, r), and 0 otherwise.
2. *Feature-response association.* It is assumed that, at each new judgment, a response may (or may not) be linked to each of the features. For example, whether or not intentional causality is linked to "cursing" as a response, or whether or not the absence of other persons is linked to cursing. Formally, the linking of a feature to a response is represented as a realization of the Bernoulli variable $Z_{sp}^{rf} \sim \text{Bern}(\rho_{rf})$ which equals 1 if response r is linked to feature f by person p judging pair (s, r), and 0 otherwise.
3. *Situation-response association.* It is assumed that whether or not a person would associate the response to the situation is a function of the features that are attributed to the situation and the features that are linked to the response. For instance, with a *disjunctive communality* rule, attributing a feature to a situation that is also linked to a given response is a sufficient condition for that response to be displayed in that situation. For example, a link between "cursing" and the feature "that no other persons are present in the situation" means that one will curse as soon as this feature is attributed to a frustrating situation (e.g., "you bang your

shins against a park bench"). Formally, the disjunctive communality rule can be expressed as

$$Y_p^{sr} = 1 \Leftrightarrow \exists f : X_{rp}^{sf} = Z_{sp}^{rf} = 1.$$

With a *conjunctive dominance* rule, for a response to be displayed in a certain situation, it is necessary (and sufficient) that all the features that are linked to the response are also attributed to the situation. This means that "features" should be interpreted here as requisites, imposed by the responses. For instance, if the requisites for "cursing" are "intentionally caused" and "absence of others", then one will curse only if both requisites are met by the situation (e.g., "you just found out that someone has opened your personal mail"). Formally, the conjunctive dominance rule reads that

$$Y_p^{sr} = 1 \Leftrightarrow \forall f : X_{rp}^{sf} \geq Z_{sp}^{rf}.$$

Note that this rule allows one to capture in a very natural way the case of responses that occur in all situations of a given set (e.g., a feeling of irritation in frustrating situations). Indeed, such responses will tend to impose no requisites at all and hence, the conjunctive dominance rule will be satisfied for all situations.

The DC and CD rules are formally related in that applying the DC rule to the complement of the data (i.e., $1 - Y_p^{sr}$) and transforming the situation parameters (i.e., $\sigma_{sf} \leftarrow 1 - \sigma_{sf}$) yields the same results as applying the CD rule to the original data.

The LSA of the PMD model concerns the way the observed data depend on the latent variables X and Z. In particular, the *independence PMD model* assumes that the latent variables $\mathbf{X}_{rp}^{s} = (X_{rp}^{s1}, \ldots, X_{rp}^{sF})$ and the latent variables $\mathbf{Z}_{sp}^{r} = (Z_{sp}^{r1}, \ldots, Z_{sp}^{rF})$ are realized for each triple (s, r, p). This assumption will be denoted as (*srp*, *rsp*), with the first part (i.e., *srp*) referring to the realization of X and the second part (i.e., *rsp*) referring to the realization of Z.

One may construct other LSAs by assuming that the latent variables X and/or Z have a fixed value for certain parts of the data. In this way, different *dependence PMD models* can be constructed. First, one may assume that persons attribute a fixed pattern of features to a situation, independent of the response, that is, $\mathbf{X}_{1p}^{s} = \ldots = \mathbf{X}_{Rp}^{s}$ for each pair (s, p), whereas the linking of features to responses is renewed at each encounter. This assumption, which will be denoted (*sp*, *rsp*), implies the observations $\mathbf{Y}_p^{s} = (Y_p^{s1}, \ldots, Y_p^{sR})$ to be dependent because they are based on the same realization of the latent situation variables. This dependency has a situational origin; it refers to fixed situation meanings within persons.

Second, one may assume that the attribution of features to a situation is renewed at each encounter, but that the linking of features to a certain response is fixed for each person, independent of the situation being judged. This assumption, which will be labeled (*srp*, *rp*) implies the observations $\mathbf{Y}_p^{r} = (Y_p^{1r}, \ldots, Y_p^{Sr})$ to be dependent. This type of dependency originates from the response; it refers to fixed response meanings within persons. As will be explained more in detail in the following paragraphs, models (*sp*, *rsp*) and (*srp*, *rp*) may be considered a three-way extension of the multiple classification latent class model (MCLCM; Maris, 1999).

Finally, one may assume that persons have a fixed opinion about the features that are attributed to a situation and that the linking of features to a response is fixed as well. This model, which will be denoted (*sp*, *rp*), assumes the observations $\mathbf{Y}_p = (Y_p^{11}, \ldots, Y_p^{SR})$ to be dependent. The source of the dependency is then located in the situations and the responses; it refers to fixed situation meanings and fixed response meanings within persons.

A Formal Description of PMD Models with Different Latent Structure Assumptions

As for each person a binary $S \times R$ matrix of responses is observed, we may consider a multinomial model for the frequencies with which these 2^{SR} possible matrices occur. Models with different LSAs can be derived by using different types of independence restrictions in specifying the kernel of this multinomial likelihood. In particular, model (*srp*, *rsp*) assumes all the elements in the $S \times R$ matrix to be independent, model (*sp*, *rsp*) assumes independency within columns while the elements within a row are dependent, model (*srp*, *rp*) assumes independency within rows, while the elements within a column are dependent, and model (*sp*, *rp*) assumes all the elements in the matrix to be dependent.

Furthermore, the different types of independence restrictions imply that each model actually operates on a different marginal table of the $S \times R \times P$ array. In particular, the independence PMD model may be considered a model for the frequencies of persons who display a response in a situation. Model (*sp*, *rsp*) is a model for the frequencies of patterns $\mathbf{y}_p^s$, and likewise, model (*srp*, *rp*) is a model for the frequencies of patterns $\mathbf{y}_p^r$. Finally, model (*sp*, *rp*) is a model for the frequencies with which the patterns $\mathbf{y}_p$ occur. However, we must emphasize that for the models to yield a comparable likelihood, it is necessary to consider them as models for the same frequency table, which is the case if they are considered as restrictions of the same multinomial model.

In the following paragraphs we describe for each model the likelihood of the observed response patterns $\mathbf{y}_p$ rather than the likelihood of the frequencies of these patterns (however, these likelihoods only differ in the normalizing constant of the multinomial likelihood, which is the same for all the models). The latter is convenient as the realization of the latent variables (but not their distribution) may depend on the person. As the DC and the CD rule are formally related, only models involving a DC rule will be described.

It turns out that PMD models with different LSAs can formally be described as extended latent class models (LCM) with constrained parameters. The models are extended LCMs because the distributions of the latent variables may depend on external variables (i.e., situation, response) and they are constrained LCMs as both conditional and marginal latent class probabilities of the LCM are a function of more basic parameters.

Latent Structure Assumption (srp, rsp)

The likelihood of the independence PMD model can be expressed as

$$p(\mathbf{y}|\boldsymbol{\sigma}, \boldsymbol{\rho}) = \prod_s \prod_r \prod_p \sum_{\mathbf{x},\mathbf{z}} p(y_p^{sr}|\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r)\, p(\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r|\boldsymbol{\sigma}_s, \boldsymbol{\rho}_r). \tag{1}$$

The 2^{2F} possible realizations of $(\mathbf{X}_{rp}^s, \mathbf{Z}_{sp}^r)$ may be considered each as a latent class. The 2^{2F+1} *conditional probabilities* $p(y_p^{sr}|\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r)$ (i.e., 2^{2F} for each value of Y_p^{sr}) are constrained to be 0 or 1. This reflects the mapping of latent variables into observed variables. For instance with a DC mapping rule, this probability may be expressed as

$$\begin{aligned} p\big(y_p^{sr}|\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r\big) &= P\big(Y_p^{sr} = 1|\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r\big)^{y_p^{sr}}\, P\big(Y_p^{sr} = 0|\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r\big)^{(1-y_p^{sr})} \\ &= \left[1 - \prod_f \big(1 - x_{rp}^{sf} z_{sp}^{rf}\big)\right]^{y_p^{sr}} \left[\prod_f \big(1 - x_{rp}^{sf} z_{sp}^{rf}\big)\right]^{1-y_p^{sr}}. \end{aligned} \tag{2}$$

In (2), the probability of no association $P(Y = 0|X, Z)$ is derived as the probability that none of the features is at the same time linked to the response and also attributed to the situation. The probability of an association $P(Y = 1|X, Z)$ is then derived by taking the complement, that is, $1 - P(Y = 0|X, Z)$.

Furthermore, the *marginal probabilities* of the latent classes $p(\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r|\boldsymbol{\sigma}_s, \boldsymbol{\rho}_r)$ are also constrained because the latent variables X and Z are assumed to be independent Bernoulli variables. In other words, the model for these probabilities is not saturated, but is rather an independence model. As a result, the 2^{2F} marginal probabilities $p(\mathbf{x}_{rp}^s, \mathbf{z}_{sp}^r|\boldsymbol{\sigma}_s, \boldsymbol{\rho}_r)$ for a given combination of s and r are based on only $2F$ parameters.

Latent Structure Assumptions (sp, rsp) and (srp, rp)

The likelihood of the model with LSA (*sp*, *rsp*) may be expressed as

$$p(\mathbf{y}|\boldsymbol{\sigma},\boldsymbol{\rho}) = \prod_s \prod_p \sum_{\mathbf{x}} \left[\prod_r p\left(y_p^{sr}|\mathbf{x}_p^s,\boldsymbol{\rho}_r\right) \right] p\left(\mathbf{x}_p^s|\boldsymbol{\sigma}_s\right). \tag{3}$$

The description of model (*srp*, *rp*) is analogous and may be obtained by simply changing the role of situations and responses. The 2^F possible realizations of $\mathbf{X}_p^s$ may be considered each as a latent class. For each response the 2^{F+1} *conditional probabilities* $p(y_p^{sr}|\mathbf{x}_p^s,\boldsymbol{\rho}_r)$ (i.e., 2^F for each value of Y_p^{sr}) are constrained as they are a function of only F response parameters. In general, these conditional probabilities may be derived as

$$p\left(y_p^{sr}|\mathbf{x}_p^s,\boldsymbol{\rho}_r\right) = \sum_{\mathbf{z}} p(y_p^{sr}|\mathbf{x}_p^s,\mathbf{z}_{sp}^r)\, p\left(\mathbf{z}_{sp}^r|\boldsymbol{\rho}_r\right). \tag{4}$$

For instance, with a DC mapping rule (4) can be expressed as

$$p\left(y_p^{sr}|\mathbf{x}_p^s,\boldsymbol{\rho}_r\right) = \left[1 - \prod_f \left(1 - x_p^{sf}\rho_{rf}\right)\right]^{y_p^{sr}} \left[\prod_f \left(1 - x_p^{sf}\rho_{rf}\right)\right]^{1-y_p^{sr}}. \tag{5}$$

In addition, in (3), the *marginal probabilities* $p(\mathbf{x}_p^s|\boldsymbol{\sigma}_s)$ are also restricted as the latent situation variables are assumed to be independent Bernoulli variables. That is, the model for the marginal probabilities is not a saturated model but rather an independence model. As a consequence, the 2^F marginal probabilities $p(\mathbf{x}_p^s|\boldsymbol{\sigma}_s)$ for a given situation s are based on only F parameters.

The PMD model with assumption (*sp*, *rsp*) may also be regarded as a three-way extension of the MCLCM (Maris, 1999) which was originally developed to analyze two-way data. With MCLCM's there are only two types of elements: persons (p) and items (i). The LSA is (p, ip). The latent classes are classes of persons determined by F binary latent variables X_p^f. The marginal probabilities $p(\mathbf{x}_p|\boldsymbol{\sigma}_s)$ do not depend on an external variable. There is, so to speak, only one situation ($S = 1$) and as many binary responses as there are items, and consequently, there is only one set of marginal probabilities.

Latent Structure Assumption (sp, rp)

The likelihood of model (*sp*, *rp*) can be expressed as

$$p(\mathbf{y}|\boldsymbol{\sigma},\boldsymbol{\rho}) = \prod_p \sum_{\mathbf{x},\mathbf{z}} p(\mathbf{y}_p|\mathbf{x}_p,\mathbf{z}_p)\, p(\mathbf{x}_p,\mathbf{z}_p|\boldsymbol{\sigma},\boldsymbol{\rho}). \tag{6}$$

The $2^{(S+R)F}$ realizations $(\mathbf{x}_p, \mathbf{z}_p)$ are considered as latent classes. From (6) it follows that the application of model (*sp*, *rp*) may be problematic from both a theoretical and a practical point of view. A *theoretical* problem is that with model (*sp*, *rp*) the probabilistic process in the latent variables will generally not suffice to explain the observed data. To clarify this, consider for instance a PMD model with F features and a deterministic mapping rule $C(\cdot)$. In that case, it can be that for some observed data $\mathbf{y}_p$ no latent data $(\mathbf{x}_p, \mathbf{z}_p)$ exist such that $C(\mathbf{x}_p, \mathbf{z}_p) = \mathbf{y}_p$. To solve this problem one may add an error component to the model, or construct a probabilistic condensation rule (De Boeck, 1997) so that latent data $(\mathbf{x}_p, \mathbf{z}_p)$ may always lead to observed data $\mathbf{y}_p$ with a nonzero probability. A *practical* problem is that the number of latent classes in the model is generally very large ($2^{(S+R)F}$) so that parameter estimation (with the algorithms described below) becomes impossible in practice.

Estimation

To obtain parameter estimates of the PMD model, one may use a mode-finding algorithm to locate the mode(s) of the likelihood, or one may use a sampling-based approach to compute a sample of the entire likelihood. As will be indicated, both approaches have strengths and weaknesses so that a combined approach may be of interest.

As the augmented likelihood of the PMD model is proportional to a product of Bernoulli variables, that is, $p(\mathbf{x}, \mathbf{z}, \mathbf{y}|\boldsymbol{\sigma}, \boldsymbol{\rho}) = p(\mathbf{x}|\boldsymbol{\sigma})p(\mathbf{z}|\boldsymbol{\rho})p(\mathbf{y}|\mathbf{x}, \mathbf{z}) \propto p(\mathbf{x}|\boldsymbol{\sigma})p(\mathbf{z}|\boldsymbol{\rho})$, it is convenient to use estimation algorithms that especially gain from the tractable form of the augmented likelihood. In particular, one may use an EM algorithm for maximization (Dempster, Laird, & Rubin, 1977) and a Markov chain Monte Carlo (MCMC) method such as the Gibbs sampler in order to compute a sample of the likelihood (Gelfand & Smith, 1990; Geman & Geman, 1984; Tanner & Wong, 1987).

For the independence PMD model, the EM algorithm has been described by Maris et al. (1996) and the Gibbs sampling algorithm has been described by Meulders, De Boeck, Van Mechelen, Gelman, and Maris (2001). For model (*sp*, *rsp*), one may easily adapt the EM algorithm that was developed by Maris (1999) to estimate MCLCMs. The Gibbs sampling algorithm for this model is described in Appendix A. To perform each of these estimation procedures, a Delphi program was written, which may be obtained from the authors upon request. As an alternative, one might consider BUGS (Spiegelhalter, Thomas, Best, & Gilks, 1995) for the fully Bayesian estimation and LEM (Vermunt, 1997) for the ML estimation of the models.

It turns out that maximization and computation of a sample of the posterior each have strengths and weaknesses when they are used for parameter estimation. First, a problem with maximizing the likelihood is that, with PMD models, ML estimates may not exist within the interior of the parameter space (Maris et al., 1996). One may solve this problem by using a strictly concave prior (e.g., a Beta$(\theta_j|2, 2)$ for each parameter θ_j) and by maximizing the *posterior distribution* $p(\boldsymbol{\sigma}, \boldsymbol{\rho}|\mathbf{y})$. On the other hand, with the Gibbs sampler, one may specify any type of prior, or even estimate the form of the prior from the data. This is advantageous because a misspecification of the prior may decrease the fit and may lead to wrong substantive conclusions (see Meulders, De Boeck, Van Mechelen, & Gelman, 2000).

Second, computation of a sample of the entire posterior distribution is advantageous because it provides not only point estimates (i.e., posterior mean) but also $(1-\alpha)\%$ posterior intervals without relying on a normal approximation of the posterior distribution. As a result, standard errors will also be accurate for small samples (Tanner & Wong, 1987). In addition, the sample of the posterior distribution may also be used to check the fit of the model with the technique of posterior predictive checks (Rubin, 1984).

Third, monitoring convergence is more straightforward in the context of maximization than when using an MCMC method to approximate the entire posterior. In particular, it is well-known that the EM algorithm has the strong property to increase the posterior density at each iteration, and to converge to a stationary point of the parameter space. Hence, a sufficiently small difference in posterior density between subsequent iterations can be used as an easy criterion to stop iterating. For the Gibbs sampler, it is known that, under some mild regularity conditions, the simulated sequences converge to the true posterior distribution (Gelfand & Smith, 1990), but assessing whether convergence has been attained is a difficult problem which has not completely been solved yet (Cowles & Carlin, 1996). In this paper we will monitor the convergence of the Gibbs sampler with the approach suggested by Gelman and Rubin (1992).

Fourth, although computationally feasible, exploring highly multimodal posteriors with the Gibbs sampler can be complicated if different local maxima are not well-separated. This is often due to a lack of identifiability.

As is generally the case with mixture models, PMD models are not identified in a trivial way because one may permute the labels of the mixture components without changing the likelihood or the posterior. This problem is called label-switching (Stephens, 2000). As a result, the posterior of a PMD model with F features consists of $F!$ identical regions of posterior density. A further complication is that the restrictions that are imposed on the parameters do not imply the posterior of PMD models to have a unique maximum (see also Maris, 1999). When using the EM algorithm for parameter estimation it is guaranteed that the algorithm will always converge to a stationary point. Subsequently, one may evaluate whether the Hessian in this point is negative definite. If this is the case, the model is said to be "locally identifiable" (Formann, 1992; Goodman, 1974).

When applying the Gibbs sampler, the multimodality of the posterior can be problematic if the regions corresponding to different modes are not well-separated, which means that within a chain different modes are visited. As a result, point estimates derived on the basis of the sample (e.g., posterior mean) become difficult to interpret. This kind of multimodality, with connected modes, may often be due to the fact that the model is over-parameterized. In addition, over-parameterization typically causes high posterior correlations between parameters which may seriously retard the way the Gibbs sampler moves through the parameter space and which is usually accompanied by high autocorrelations in this process (Carlin & Louis, 1996, p. 187). In order to solve the problem of over-parameterization one may try to impose extra constraints on the model's parameters.

Model Selection

An important topic in fitting PMD models to empirical data is to specify the three building blocks of the model, that is, the LSA, the number of features, and the mapping rule. Several criteria and model checks may be used to reach this goal. First, *information criteria* are global as they assess the global GOF of the model and they are relative as they are used to compare models. Second, *specific model checks* are not global, but intended to evaluate the fit of specific model assumptions. These checks are relative if they involve a comparison of models and they are absolute if they compare observations to what is expected under the model. Third, *global model checks* are global and absolute as they assess global GOF by comparing observations to what is expected under the model.

Information Criteria

Two well-known information criteria are Akaike's Information Criterion (AIC) (Akaike, 1973, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). Both AIC and BIC take the form of a sum of a badness-of-fit term (minus twice the log likelihood of the fitted model) and a penalty term, which is a measure of the complexity of the model. The model having the lowest value for AIC or BIC is selected. For AIC and BIC the penalty terms equal $2k$ and $\log(N)k$, respectively, with k being the number of free parameters in the model and with N being the total "sample size". The latter is the sum of cell counts rather than the number of cells of the frequency table to which the model is applied (Raftery, 1986). Specifically, for PMD models the sample size equals the number of persons (P).

Spiegelhalter, Best, and Carlin (1998) recently developed the Deviance Information Criterion (DIC). In contrast to AIC and BIC, this criterion is not based on the likelihood of the model at the mode, but rather on the likelihood of the draws of the posterior sample. In particular, it is defined as

$$\text{DIC} = \overline{D} + k_D,$$

with $\overline{D}$ the posterior mean of the deviance, that is, $E_{\boldsymbol{\theta}|\mathbf{y}}[-2 \log p(\mathbf{y}|\boldsymbol{\theta})]$, and with k_D being an estimate of the *effective* number of parameters in the model, namely, $k_D = \overline{D} - D(\overline{\boldsymbol{\theta}})$, with $\overline{\boldsymbol{\theta}}$ being the posterior mean.

Specific Model Checks

In order to *check the validity of the LSA* one may use a specific model check that is absolute in nature. The basic idea in constructing a test statistic for this purpose is the fact that different LSAs imply different subsets of observations to be dependent. For instance, model (*srp*, *rsp*) implies all observations to be independent whereas model (*sp*, *rsp*) assumes observations within

$\mathbf{y}_p^s$ to be dependent. Therefore, one may distinguish between both models by computing, for each situation s, the correlation between responses r and r^* across persons, that is,

$$T_{r,r^*}^{(s)}(\mathbf{y}) = \text{cor}_p(y_p^{sr}, y_p^{sr^*}). \tag{7}$$

Under assumption (*srp*, *rsp*) the expected value of (7) is zero, while under assumption (*sp*, *rsp*) the correlations will reflect dependencies between observations within $\mathbf{y}_p^s$ due to the existence of different latent classes $\mathbf{x}_p^s$. In the same way, one may use correlations (across persons) between pairs of situations for each response to distinguish between models (*srp*, *rsp*) and (*srp*, *rp*).

To *determine the number of features*, we may use a conditional likelihood-ratio (LR) test which is relative in nature as it is used to compare models with different numbers of features. The LR statistic to compare H_0 of F features versus H_1 of $F + r$ features is defined as follows:

$$\text{LR}(\mathbf{y}) = -2 \log \frac{p(\mathbf{y}|\hat{\boldsymbol{\theta}}_F)}{p(\mathbf{y}|\hat{\boldsymbol{\theta}}_{F+r})}. \tag{8}$$

Finally, no specific model checks are proposed for the selection of the *mapping rule*.

Global Model Checks

To evaluate the global GOF of latent structure models for the analysis of categorical data one traditionally uses measures of the power divergence family (Cressie & Read, 1984) such as the Pearson-χ^2 statistic or the unconditional LR statistic (which compares the model under investigation with the saturated model). These statistics may be conceived as measures of divergence between observed frequencies and frequencies that are expected under the model. For instance, the Pearson-χ^2 statistic is defined as

$$\chi^2(\mathbf{y}) = \sum_{j=1}^{J} \frac{\left[O_j(\mathbf{y}) - E_j(\mathbf{y}, \hat{\boldsymbol{\theta}})\right]^2}{E_j(\mathbf{y}, \hat{\boldsymbol{\theta}})}, \tag{9}$$

with O_j and E_j being observed and expected frequencies, respectively.

Von Davier (1997) conducted a Monte Carlo study to investigate the performance of several bootstrapped statistics for global GOF when the data are sparse. His results indicated that the Pearson-χ^2 statistic performed well whereas the unconditional LR statistic accepted underfitting models too frequently. Therefore, we will focus on Pearson-χ^2 measures instead.

When using (9) to assess the global GOF of PMD models, O_j and E_j can be defined according to the LSA of the model. For instance, with the independence PMD model the O_j's are frequencies of persons who would (or would not) display a response in a certain situation. In general, there are $2SR$ of such frequencies observed. Furthermore, the E_j's are calculated as $P \times p(\mathbf{y}_p^{sr}|\hat{\boldsymbol{\theta}})$. With LSA (*sp*, *rsp*) the O_j's are frequencies of persons having a specific response pattern $\mathbf{y}_j^s$ $(j = 1, \ldots, 2^R)$; the E_j's are calculated as $P \times p(\mathbf{y}_j^s|\hat{\boldsymbol{\theta}})$. On the other hand, with LSA (*srp*, *rp*) the O_j's are frequencies of persons having response pattern $\mathbf{y}_j^r$ $(j = 1, \ldots, 2^S)$ and the E_j's are calculated as $P \times p(\mathbf{y}_j^r|\hat{\boldsymbol{\theta}})$.

For PMD models, a classical approach to testing can only be used in a limited number of cases. First, the reference distribution of the statistic to check the LSA is generally unknown. Second, when testing H_0 of F features versus H_1 of $F + r$ features, H_0 is at the boundary of the parameter space. In this case, the regularity conditions of LR tests break down (McLachlan & Basford, 1988) so that the asymptotic distribution is unknown. Third, when the data are sparse the χ^2 approximation to the Pearson-χ^2 statistic may be poor, so that simulation of the reference distribution might be a useful alternative.

Having obtained a sample of the posterior distribution, it is straightforward to use the technique of posterior predictive checks (PPCs) to evaluate measures of specific and global fit (see

Gelman, Carlin, Stern, & Rubin, 1995; Meng, 1994; Rubin, 1984). Gelman, Meng, and Stern (1996) define PPC p-values for different types of measures and describe related computational procedures. In particular, we may distinguish between two types of measures: *statistics* that only depend on the data, denoted as $T(\mathbf{y})$ and *discrepancy measures* that depend on the data and the parameters, denoted as $T(\mathbf{y}, \boldsymbol{\theta})$. In some cases the computation of a statistic also involves a posterior mode (which is just a complex function of the data). In this way, discrepancy measures can actually be turned into a statistic.

For instance, (7) is a statistic, (9) is obtained by turning the discrepancy measure $\chi^2(\mathbf{y}, \boldsymbol{\theta})$ into a statistic, and (8) may also be regarded a statistic that is a function of the posterior mode of the restricted and the unrestricted model.

In case of a statistic, the PPC p-value can be computed by generating new data sets $\mathbf{y}^{\text{rep}}$ (using the draws from the posterior) and by computing the proportion of replicated data sets in which $T(\mathbf{y}^{\text{rep}}) \geq T(\mathbf{y})$. In case of a statistic based on posterior mode(s), the procedure is equivalent, except that, to compute the statistic, one has to locate the posterior mode(s). One may note that these Bayesian procedures only differ from the classical parametric bootstrap procedure (Efron & Tibshirani, 1993) in the way the data are replicated: the former uses draws from the posterior to replicate data whereas the latter uses the maximum-likelihood estimate. In case of a discrepancy measure, the p-value is computed as the proportion of replicated data sets in which realized discrepancies $(T(\mathbf{y}^{\text{rep}}, \boldsymbol{\theta}))$ exceed or equal observed discrepancies $(T(\mathbf{y}, \boldsymbol{\theta}))$. This procedure is computationally less demanding than the one using a statistic (derived from a discrepancy measure) because realized and observed discrepancies are computed using a draw from the posterior (which was used to replicate the data), rather than an estimate of the posterior mode for each replicated data set.

The different types of PPC p-values can all be interpreted as measures of surprise or incompatibility, that is, extreme p-values indicate a failure of the model to capture the aspect measured by $T(\cdot)$. PPC p-values typically are not uniformly distributed under the null model. In particular, they are often conservative, which means that their asymptotic distributions are more concentrated around $1/2$ than a uniform (Bayarri & Berger, 2000; Meng, 1994; Robins, Van Der Vaart, & Ventura, 2000).

Simulation Study

A simulation study is conducted to investigate the extent to which information criteria and specific/global model checks can help to specify the different aspects of PMD models. As this paper focuses on the LSA of the PMD model, results concerning this aspect will be described in detail whereas results concerning the mapping rule and the number of features will only be briefly summarized (for a more elaborate description, see Meulders, 2000).

We consider 12 types of models by crossing all levels of three factors: LSA ((*srp*, *rsp*) or (*sp*, *rsp*)), the number of features (1, 2, 3), and the mapping rule (disjunctive communality, DC, or conjunctive dominance, CD). These are called the *generation* models. For each type of generation model, 10 data sets with a particular number of situations and responses (S = 20, R = 6) for each of three sample sizes ($P = 100$, 300, 1000) are generated using uniformly distributed starting values.

For each data set the EM algorithm is used to locate the posterior mode(s) for each of the 12 model types, now called *analysis* models. In addition, for data sets with the smallest sample size ($P = 100$), Gibbs sampling is used to compute a (local) sample of the posterior for analysis models having LSA (*srp*, *rsp*) and having a smaller (or the same) number of features than the generation model. The reason for not applying the Gibbs sampler to models that assume more features than the generation model is that such models are often overparameterized, which may complicate the computation of a local sample. In the same way, models having LSA (*sp*, *rsp*) are not estimated with Gibbs because the posterior is also often multimodal.

Specification of the LSA

The results of the simulation study show that the likelihood of models (evaluated at the posterior mode) with a correct LSA is *always higher* than the likelihood of models with a wrong LSA. Hence, mode-based information criteria (AIC and BIC) have an excellent performance to select the LSA of the PMD model.

To evaluate the capacity of the correlational statistic (7) to distinguish between LSAs (*srp*, *rsp*) and (*sp*, *rsp*), we bootstrapped the 95% confidence interval (CI) of this statistic for each situation s and for all pairs of responses r and r^*. This yields for each analysis a set of $SR(R-1)/2 = 300$ CIs. Note that we use a parametric bootstrap procedure instead of PPCs here because the EM algorithm, unlike the Gibbs sampler, is applied to all the cells of the simulation design. However, the posterior predictive intervals that are obtained with the PPC approach are very similar to the CIs obtained with the bootstrap procedure and lead to the same conclusions.

The proportion of observed correlations which are below or above their CI are denoted as p_l and p_u, respectively. Table 1 shows, for models that were generated and analyzed with a specific LSA the means and the 95% CI of p_l and p_u. Note that the results presented in Table 1 are obtained by aggregating over all levels of the other factors (number of features, mapping rule, and sample size) in the simulation design.

First, consider the case in which the data are generated with the independence model. When using the correct LSA to analyze the data, the 95% CIs of p_l and p_u are rather small ([.007, .043] and [.007, .047], respectively) and their means (.023 and .024, respectively) approximate the expected value of .025 rather well. On the other hand, when using LSA (*sp*, *rsp*), p_l is higher than expected (95% CI of [.037, .353] with mean value of .145) and p_u is lower than expected (95% CI of [.000, .023] with mean value of .009). Hence, correlations induced by model (*sp*, *rsp*) are higher than those in the data generated with the independence model. Moreover, we may note that especially the distribution of p_l deviates more from the expected value of .025 if the sample size increases and that the power of is this test is relatively independent of the (mis)specification of other model aspects (not shown in Table 1).

Second, consider the case in which the data are generated with LSA (*sp*, *rsp*). When using the correct LSA to analyze the data, the 95% CIs of p_l and p_u are rather large ([.007, .253] and [.000, .793], respectively) and the mean values (.067 and .162, respectively) are larger than the expected value of .025. A further inspection of the distribution of the values of p_l and p_u in the different cells of the simulation design reveals that p_l and p_u are especially large if the number of features and/or the mapping rule are wrongly specified. For analysis models with LSA (*sp*, *rsp*) that are completely correctly specified, the 95% CIs of p_l and p_u ([.007, .055] and [.005, .037], respectively) are much smaller and the means (.022 and .017, respectively) are much closer to the expected value of .025. On the other hand, when using the independence model for the analysis, p_l is lower than expected (95% CI of [.00, .007] with mean value of .001) and p_u is higher than expected (95% CI of [.303, .973] with mean value of .731). Hence, observed correlations of data generated with model (*sp*, *rsp*) are generally higher than predicted by the independence model. Moreover, we may note that, especially the distribution of p_u deviates more from the expected

TABLE 1.
Mean and 95% CI of p_l and p_u as a function of the LSA of the generation and the analysis model

LSA		p_l		p_u	
Generation Model	Analysis Model	Mean	CI	Mean	CI
(*srp*, *rsp*)	(*srp*, *rsp*)	.023	[.007,.043]	.024	[.007,.047]
(*srp*, *rsp*)	(*sp*, *rsp*)	.145	[.037,.353]	.009	[.000,.023]
(*sp*, *rsp*)	(*srp*, *rsp*)	.001	[.000,.007]	.731	[.303,.973]
(*sp*, *rsp*)	(*sp*, *rsp*)	.067	[.007,.253]	.162	[.000,.793]

value of .025 with increasing sample sizes and that the power of this test is relatively independent of the (mis)specification of other model aspects (not shown in Table 1).

Finally, we note that the proposed specific model check is more powerful to reject LSA (*srp*, *rsp*) when the data stem from LSA (*sp*, *rsp*) than to reject LSA (*sp*, *rsp*) when the data are generated with the independence model. This is plausible as model (*sp*, *rsp*) is a model for the correlations between pairs of responses and as such, it can capture zero correlations to some extent.

Specification of the Mapping Rule

The likelihood of models with a correct mapping rule is always higher than the likelihood of models with a wrong mapping rule. Hence, mode-based information criteria have an excellent performance to select this model aspect. On the other hand, with complex models, a sample-based information criterion (DIC) sometimes selects the wrong mapping rule.

Specification of the Number of Features

The results of the simulation study show that information criteria, global model checks, and a specific model check such as the conditional LR test may lead to different conclusions concerning the number of features that is needed.

Mode-based information criteria (AIC and BIC) always select the correct number of features if the data are generated with LSA (*sp*, *rsp*). If the data are generated with the independence model, the BIC, and to a lesser extent the AIC sometimes select models that assume less features than the generation model. However, if the sample size increases both AIC and BIC perform better. In particular, for sample sizes equal to 100, 300, and 1000, the AIC selects a smaller number of features than the generating model in 7%, 0% and 0% of the cases, respectively; for the BIC this is respectively the case in 22%, 7% and 3% of the model selections that were made.

On the other hand, the results of the DIC (based on the cells of the simulation design in which Gibbs sampling was applied) are promising in that it always selects the correct number of features. The better performance of the DIC, compared to AIC and BIC, could be explained by the fact that it includes a measure of model complexity which is based on the data at hand, whereas AIC and BIC involve an approximation that is based on asymptotic theory.

Finally, a comparison of different PPCs (based on the cells of the simulation design in which Gibbs sampling is applied) indicates that under a correct specification of the LSA, more expensive procedures (in terms of computational costs) generally have more power to reject models that assume less features than the generation model. In particular, the conditional LR test is the most expensive procedure and it is also the most powerful one, the Pearson-χ^2 discrepancy measure is cheapest, and it fails to reject parsimonious models most often, and finally, the Pearson-χ^2 statistic takes an intermediate position with respect to computational costs and power to reject models that assume less features than the generation model.

Example

Data

As an illustration, PMD models with different LSAs are applied to data that were gathered by Vansteelandt (1999) in a study of individual differences in hostile behavior. In this study 316 subjects indicated on a three-point scale to what extent they would display each of 21 hostile responses in each of 14 frustrating situations (0 = not, 1 = limited, 2 = strong). Situations and responses were selected from an S-R inventory of hostility (Endler & Hunt, 1968). In this paper a subset of 6 situations and 4 responses is analyzed (see Table 2) because an exploratory analysis indicated that too many features were needed to obtain a fitting model for the entire data set. To

TABLE 2.
Description of situations and responses

Set	Element
situation	your instructor unfairly accuses you of cheating on an examination
	you have just found out that someone has told lies about you
	you are driving to a party and suddenly your car has a flat tire
	you are waiting at the bus stop and the bus fails to stop for you
	someone has opened your personal mail
	you accidentally bang your shins against a park bench
response	become tense
	feel irritated
	curse
	want to strike something or someone

Note: From "S-R Inventories of Hostility and Comparisons of the Proportions of Variance from Persons, Behaviors, and Situations for Hostility and Anxiousness" by N.S. Endler and J.M. Hunt, 1968, *Journal of Personality and Social Psychology*, *9*, pp. 310–311. Copyright 1968 by the American Psychological Association. Adapted by permission of the author.

apply PMD models, the raw data are dichotomized (0 vs. 1 or 2), given that we, like Vansteelandt (1999), were interested in the occurrence of hostile responses, and not in their degree.

Analysis

Conjunctive and disjunctive models with LSAs (*srp*, *rsp*), (*sp*, *rsp*), or (*srp*, *rp*) and assuming 1 up to 6 features are fitted. Figure 1 shows AIC values for the fitted models: Except for the one-feature model, models involving LSA (*srp*, *rp*) yield a lower AIC than models involving the independence assumption or LSA (*sp*, *srp*). Furthermore, differences between disjunctive and conjunctive models are rather small for models assuming more than 2 features. As the independence model is actually a model for the frequencies in the $S \times R$ table, it is saturated for $F > 2$. As a result, the likelihood of this model cannot decrease for $F > 2$. In order to check the results without being hampered by the saturation problem, we have applied each of the models to a larger data set (i.e., $S = 14$, $R = 8$). These analyses also indicated that the independence PMD model has a substantially lower likelihood than models with other LSAs and that it should be rejected on the basis of a specific model check. As disjunctive models are often more easy to interpret, we will further investigate the fit of the disjunctive model with LSA (*srp*, *rp*) using specific and global model checks.

Specific model check for the validity of the LSA. The LSA (*srp*, *rp*) may be checked by computing the 95% posterior predictive interval of correlations between *pairs of situations* within a response. This yields $RS(S - 1)/2 = 60$ posterior intervals. The results indicate that, for models assuming 1 up to 6 features, none of the observed correlations is below the corresponding posterior interval (except for the one-feature model where 22% of the correlations are lower) and that the proportion of observed correlations above their posterior interval decreases if more features are added. Furthermore, models assuming 4, 5, and 6 features have a reasonable fit as in these cases 93%, 97%, and 98% of the correlations are within their 95% posterior interval. When correlations of the same type are checked for the independence model and for LSA (*sp*, *rsp*), for models assuming 1 up 6 features, only 5% of the correlations could be captured by their posterior interval.

Specific and global model checks to select the number of features. Information criteria and specific/global model checks lead to different conclusions concerning the number of features that

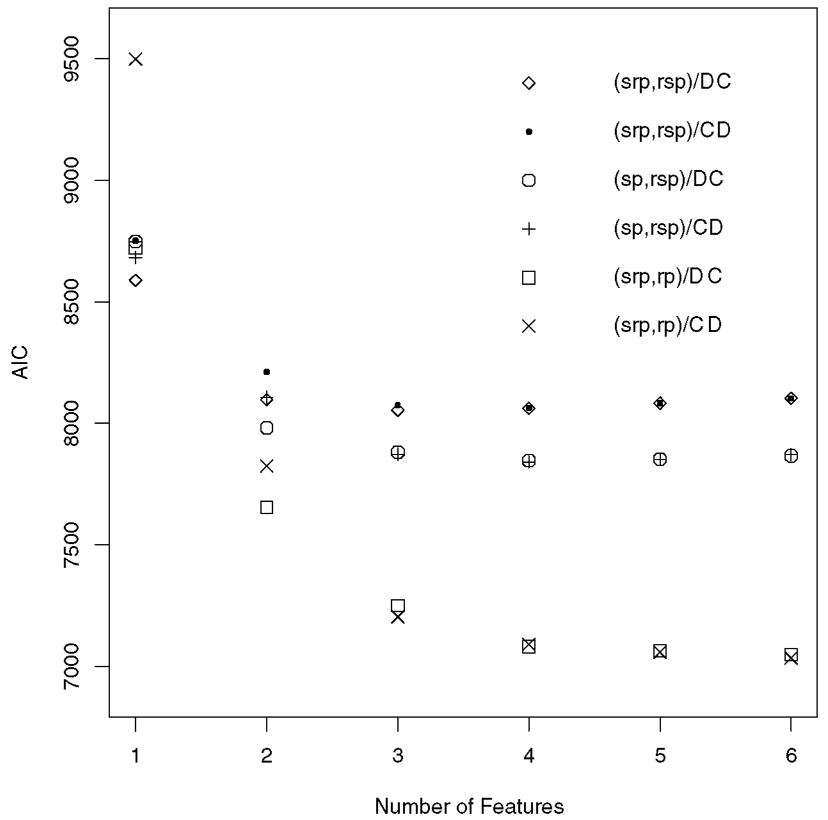

FIGURE 1.
AIC values for disjunctive and conjunctive models having LSA (*srp*, *rsp*), (*sp*, *rsp*), or (*srp*, *rp*) and assuming 1 up to 6 features.

is needed in the present application. In particular, the BIC would select 4 features, the AIC and the Pearson-χ^2 discrepancy measure would select 6 features and according to the conditional LR test and the Pearson-χ^2 statistic the data even provide evidence in support for a solution with more than 6 features. Finally, according to the DIC at least 6 features are needed.

In the following section we will discuss the parameters of the four-feature model more in detail. This is the simplest model that has a reasonable fit for the LSA test and it is the best model according to the BIC. Also, the AIC did not decrease much after four features.

Interpretation of Selected Model

Table 3 shows the posterior means and 95% posterior intervals for the parameters of the selected model. The first feature has a high probability of being attributed to most situations and must therefore be considered a *general frustration feature*. Only the situation in which "you bang your shins against a park bench" has a low probability of being associated to this feature (.22). This first feature is very much linked to "feelings of irritation" as a response (.76).

The second feature has a high probability of being attributed to situations in which expressed aggressive reactions from the part of the frustrated person are perceived as inappropriate. This is very clear and direct in the situation of "being unfairly accused" (.90). Aggression may in this situation be inhibited by the presence of the examination instructor, who is usually a person of higher status, such as a professor. This second feature may be labeled the *aggression-is-inappropriate feature*. Typical responses that are linked to this feature are "feelings of irrita-

TABLE 3.
Posterior mean and 95% posterior interval (PI) of the feature probabilities for a disjunctive four-feature model with LSA (*srp*, *rp*)

		Feature 1		Feature 2		Feature 3		Feature 4	
Set	Element	Mean	95% PI	Mean	95% PI	Mean	95% PI	Mean	95% PI
situation	unfairly accused	.85	[.71,.94]	.90	[.84,.95]	.05	[.00,.15]	.14	[.06,.23]
	lies	.98	[.92,1.00]	.44	[.34,.54]	.44	[.32,.57]	.38	[.28,.48]
	flat tire	.64	[.50,.75]	.51	[.42,.59]	.89	[.82,.94]	.14	[.08,.21]
	bus stop	.81	[.70,.91]	.46	[.36,.54]	.85	[.79,.91]	.14	[.08,.22]
	mail	.95	[.88,1.00]	.31	[.21,.42]	.51	[.41,.60]	.23	[.15,.31]
	bang your shins	.22	[.11,.33]	.01	[.00,.04]	.85	[.79,.91]	.21	[.14,.29]
response	become tense	.34	[.25,.45]	**.94**	[.90,.98]	.12	[.05,.19]	.11	[.01,.29]
	feel irritated	**.76**	[.67,.83]	**.67**	[.48,.82]	.33	[.22,.44]	**.73**	[.36,.98]
	curse	.14	[.08,.21]	.04	[.00,.12]	**.73**	[.65,.80]	**.73**	[.55,.89]
	want to strike	.06	[.03,.10]	.08	[.02,.15]	.01	[.00,.02]	.49	[.40,.59]

tion"(.67) and "becoming tense" (.94), both of which are internal reactions. On the other hand, external aggressive reactions such as "cursing" and "wanting to strike something or someone" have a low probability to be linked to this feature (.04 and .08, respectively).

The third feature has a high probability of being attributed to situations without a frustrating person that could possibly hear anything from the part of the frustrated person. (e.g., "you are waiting at the bus stop and the bus fails to stop for you" (.85); "you are driving to a party and suddenly your car has a flat tire" (.89); "you accidentally bang your shins against a park bench" (.85)). The third feature may therefore be called the *no-frustrating-person-present* feature. The response that is linked to this feature is a verbal aggressive reaction, namely "cursing" (.73), and to a lesser extent it is also linked to irritation (.33).

Finally, the fourth feature is attributed with a moderate probability to situations in which someone shows lack of respect for you (e.g., "someone has told lies about you" (.38); "someone has opened your personal mail" (.23)). This is a *lack-of-respect feature*. This feature is the only one that is linked to "want to strike" (.49), and it is very much linked to "feeling irritated" (.73) and "cursing" (.73).

In sum, *feelings of irritation* occur in all frustrating situations, unless the irritation can be directly and freely expressed because no frustrating person is present; *becoming tense* is specific for situations in which hostility may not be shown, *cursing* occurs when nobody can feel offended, and *wanting to strike* is rather rare and restricted to serious cases of lack of respect. Irritation is a rather general response, and if it occurs, it can be accompanied by tension (not showing or expressing aggression), or by cursing or wanting to strike (expressing one's aggression).

General Discussion

In this paper, the independence assumption of the PMD model is questioned and a taxonomy of alternative LSAs is proposed. Going beyond the independence assumption in a particular application is advantageous for several reasons:

First, from a *substantive point of view*, different LSAs imply different psychological processes and related individual differences for the way in which the data were generated. For instance, whether or not persons differ in what a situation means to them, as expressed in the features that are attributed to a situation, has implications for the LSA. Second, from a *methodological point of view*, a correct specification of the LSA generally yields more plausible models (i.e., with a higher likelihood), and allows to account for correlations in the data.

For parameter estimation of PMD models, both a mode-finding algorithm (EM) and a sampling-based algorithm (Gibbs) are described. As each algorithm has strong and weak points, a combined approach is recommended. With respect to model selection, it is shown that information criteria, specific model checks, and checks for global GOF are powerful tools to specify the basic assumptions of PMD models, that is, the LSA, the mapping rule, and the number of features.

Finally, some topics need further research. The first topic is the *dependency structure* of the model. The estimation of a model which accounts for correlations between pairs of responses and between pairs of situations at the same time (i.e., (*sp*, *rp*)) is hampered by theoretical as well as practical problems. Although the problem is currently unsolved, we do have some suggestions about how to proceed. A first approach, which deals with the theoretical problem, but not with the practical one, is to use a probabilistic mapping rule. A second approach, is to build a model which assumes that persons differ in the subset of features they are sensitive to for the whole set of situation-response pairs that are being judged (Meulders, De Boeck, & Van Mechelen, 2002). Such a model can explain correlations between pairs of situations and between pairs of responses at the same time, but only in a moderate way, and less well than models with LSA (*sp*, *rsp*) or (*srp*, *rp*) which focus on correlations between reponses and situations, respectively.

The second topic is the *multimodality of the posterior distribution*. Using the Gibbs sampler for parameter estimation is not straightforward if the posterior distribution is highly multimodal and if the different modes are not well-separated. In particular, it may be the case that several modes are visited within one chain which makes summarizing the information in the posterior sample less straightforward. To deal with this problem one may use clustering-like tools to classify the draws of the posterior depending on the mode they are associated with (Celeux, Hurn, & Robert, 2000). If multimodality is due to label-switching, one may use a relabelling algorithm (Stephens, 2000).

Appendix A: Gibbs Sampling Algorithm

This section describes the different steps of the Gibbs sampler to compute a sample of the posterior distribution $p(\boldsymbol{\sigma}, \boldsymbol{\rho}|\mathbf{y})$ for model (sp, rsp). A conjugate prior $\text{Beta}(\theta_j|\alpha, \beta)$ is assumed for each parameter θ_j.

In each iteration, the algorithm involves the following steps:

1. For each pair (s, p) draw the vector $\mathbf{x}_p^s$ from

$$p(\mathbf{x}_p^s|\mathbf{y}_p^s, \boldsymbol{\sigma}_s, \boldsymbol{\rho}) \propto \prod_r p(y_p^{sr}|\mathbf{x}_p^s, \boldsymbol{\rho}_r)p(\mathbf{x}_p^s|\boldsymbol{\sigma}_s).$$

2. For each triple (s, r, p) draw the vector $\mathbf{z}_{sp}^r$ from

$$p(\mathbf{z}_{sp}^r|y_p^{sr}, \mathbf{x}_p^s, \boldsymbol{\rho}_r) \propto p(y_p^{sr}|\mathbf{x}_p^s, \mathbf{z}_{sp}^r)p(\mathbf{z}_{sp}^r|\boldsymbol{\rho}_r).$$

3. For each pair (s, f) draw σ_{sf} from

$$\text{Beta}\left(\alpha + \sum_p x_p^{sf}, \beta + \sum_p \left(1 - x_p^{sf}\right)\right).$$

4. For each pair (r, f) draw ρ_{rf} from

$$\text{Beta}\left(\alpha + \sum_s \sum_p z_{sp}^{rf}, \beta + \sum_s \sum_p \left(1 - z_{sp}^{rf}\right)\right).$$

The values of $\boldsymbol{\sigma}$ and $\boldsymbol{\rho}$ in subsequent iterations form a Markov chain that converges, under some mild regularity conditions, to the true observed posterior distribution (Tanner & Wong, 1987).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 271–281). Budapest: Academiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Bayarri, M. J., & Berger, J. O. (2000). P-values for composite null models. *Journal of the American Statistical Association*, *95*, 1127–1142.

Candel, M. J. J. M., & Maris, E. (1997). Perceptual analysis of two-way two-mode frequency data: Probability matrix decomposition and two alternatives. *International Journal of Research in Marketing*, *14*, 321–339.

Carlin, B.P., & Louis, T.A. (1996). *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall.

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*, 957–970.

Cowles, K., & Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, *91*, 883–904.

Cressie, N.A.C., & Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, *46*, 440–464.

De Boeck, P. (1997). Feature-based classification models with a dominance rule. In *Proceedings of the Biennial Sessions of the Bulletin of the International Statistical Institute* (42nd Session, Book 2, pp. 389–392). Istanbul: International Statistical Institute.

de Bonis, M., De Boeck, P., Pérez-Diaz, F., & Nahas, M. (1999). A two-process theory of facial perception of emotions. *Life Sciences*, *322*, 669–675.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

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

Endler, N.S., & Hunt, J.M. (1968). S-R inventories of hostility and comparisons of the proportions of variance from persons, behaviors, and situations for hostility and anxiousness. *Journal of Personality and Social Psychology, 9*, 309–315.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*, 476–486.

Gelfand, A.E., & Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 398–409.

Gelman, A., & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1995). *Bayesian data analysis*. London: Chapman & Hall.

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

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*, 187-212.

Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika, 61*, 7–29.

McLachlan, G.J., & Basford, K.E. (1988). *Mixture models*. New York, NY: Marcel Dekker.

Meng, X.L. (1994). Posterior predictive p-values. *The Annals of Statistics, 22*, 1142–1160.

Meulders, M. (2000). *Probabilistic feature models for psychological frequency data: A Bayesian approach*. Unpublished doctoral dissertation, University of Leuven, Belgium.

Meulders, M, De Boeck, P., Van Mechelen, I. (2001). Probability matrix decomposition models and main-effects generalized linear models for the analysis of replicated binary associations. *Computational Statistics and Data Analysis, 38*, 217–233.

Meulders, M., De Boeck, P., & Van Mechelen, I. (2002). Rater classification on the basis of latent features in responding to situations. In W. Gaul & G. Ritter (Eds.), *Classification, automation, and new media*. Proceedings of the 24th Annual Conference of the Gesellschaft für Klassifikation, University of Passau (pp. 453–461). Berlin: Springer-Verlag.

Meulders, M., De Boeck, P., Van Mechelen, I., & Gelman, A. (2000) *Hierarchical extensions of probability matrix decomposition models*. Manuscript submitted for publication.

Meulders, M., De Boeck, P., Van Mechelen, I., Gelman, A., & Maris, E. (2001). Bayesian inference with probability matrix decomposition models. *Journal of Educational and Behavioral Statistics, 26*, 153–179.

Raftery, A.E. (1986). A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society, Series B, 48*, 249–250.

Robins J.M., Van Der Vaart, A., & Ventura, V. (2000). Asymptotic distribution of p-values in composite null models. *Journal of the Statistical Association, 95*, 1143–1172.

Rubin, D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics, 12*, 1151–1172.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6*, 461–464.

Spiegelhalter, D.J., Best, N.G., & Carlin, B.P. (1998). *Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models*. Manuscript submitted for publication.

Spiegelhalter, D.J., Thomas, A., Best, N., & Gilks, W.R. (1995). *BUGS: Bayesian inference using Gibbs sampling, Version 0.50* (Tech. Rep.). Cambridge, U.K.: Cambridge University, Institute of Public Health, Medical Research Council Biostatistics Unit.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B, 62*, 795–809.

Tanner, M.A., & Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82*, 528–540.

Vansteelandt, K. (1999). A formal model for the competency-demand hypothesis. *European Journal of Personality, 13*, 429–442.

Vermunt J.K. (1997). *Log-linear models for event histories*. Thousand Oaks, CA: Sage.

Von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research Online, 2*, 29–48.

Manuscript received 20 OCT 1999
Final version received 24 JAN 2002