Check for updates

Received: 25 November 2022 | Accepted: 6 November 2024

DOI: 10.1111/bmsp.12372

the british psychological society
promoting excellence in psychology

ARTICLE

Data fusion by T3–PCA: A global model for the simultaneous analysis of coupled three-way and two-way real-valued data

Elisa Frutos-Bernal[1] | **Eva Ceulemans**[2] | **Purificación Galindo-Villardón**[1,3,4] | **Tom F. Wilderjans**[2,5,6]

[1]Department of Statistics, University of Salamanca, Salamanca, Spain

[2]Research Group of Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, Katholieke Universiteit, Leuven, Belgium

[3]Centro de Estudios e Investigaciones Estadísticas, Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Ecuador

[4]Centro de Investigación Institucional (CII), Universidad Bernardo O'Higgins, Santiago de Chile, Chile

[5]Methodology and Statistics Unit, Institute of Psychology, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, The Netherlands

[6]Leiden Institute for Brain and Cognition, Leiden, The Netherlands

Correspondence
Elisa Frutos-Bernal, Department of Statistics, University of Salamanca, Salamanca, Spain.
Email: efb@usal.es

Abstract

In various areas of science, researchers try to gain insight into important processes by jointly analysing different datasets containing information regarding common aspects of these processes. For example, to explain individual differences in personality, researchers collect, for the same set of persons, data regarding behavioural signatures (i.e., the reaction profile of a person across different situations), on the one hand, and traits or dispositions, on the other hand. To uncover the processes underlying such coupled data, to all N-way N-mode data blocks simultaneously a global model is fitted, in which each data block is represented by an N-way N-mode decomposition model (e.g., principal component analysis [PCA], Parafac, Tucker3) and the parameters underlying the common mode are required to be the same for all data blocks this mode belongs to. To estimate the parameters underlying the common mode, a simultaneous strategy is used that pools the information present in all data blocks (i.e., data fusion). In this paper, we propose the T3–PCA model, which represents three- and two-way data with Tucker3 and PCA respectively. This model is less restrictive than the already proposed LMPCA model in which the three-way data block is decomposed according to a Parafac model. To estimate the T3–PCA model parameters, an alternating least-squares algorithm is proposed. The superior performance of the simultaneous T3–PCA strategy over a sequential strategy (i.e., estimating common parameters

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2025 The Author(s). *British Journal of Mathematical and Statistical Psychology* published by John Wiley & Sons Ltd on behalf of British Psychological Society.

using information from the three-way data block only) is demonstrated in an extensive simulation study and an application to empirical coupled anxiety data.

KEYWORDS

alternating least squares, common components, coupled/linked data, data fusion, multiblock multiway data analysis, PCA, sequential strategy, simultaneous strategy, Tucker3

1 | INTRODUCTION

In the simultaneous investigation of different psychological and behavioural processes, often different types of information are available, with each type of information highlighting common, and possibly also specific, aspects of these processes. For example, to study personality (and individual differences therein), researchers often make use of, on the one hand, behavioural data expressing the reactions of a set of subjects to different situations (i.e., behavioural signature) and, on the other hand, trait and dispositional information (e.g., scores on openness and neuroticism) or data regarding cognitive-affective units subjects make use of. When each type of information pertains to the same research units, a coupled dataset is obtained. Of course, whether it makes sense to couple different types of information depends on the degree of comparability/commensurability of the data entries from the different information types. Often, the comparability/commensurability of data entries, within and between data information types, can be obtained by a suitable preprocessing of the data entries (for a more detailed discussion of the challenges underlying data coupling, see Van Mechelen & Smilde, 2010). In this paper, in particular, data measuring the intensity to which subjects show anxiety-related responses (e.g., perspire, heart beats faster) to stressful situations (e.g., final exam and job interview) is coupled to data measuring from the same subjects a set of personal dispositions (e.g., items like "I am a calm person" and "I can deal well with stress"), with these latter dispositions being administered through well-known personality questionnaires (e.g., the State-Trait Anxiety Inventory [STAI] and the Cognitive-Affective processing System questionnaire). In general, a coupled or linked dataset can be defined as a set of N-way N-mode data blocks in which each block has at least one mode in common with at least one other block (Van Mechelen & Smilde, 2010; Wilderjans, Ceulemans, & Van Mechelen, 2009). In the personality example, the coupled dataset consists of a three-way three-mode data array (i.e., behavioural signatures measuring the intensity of subjects' reactions to stressful situations) and a two-way two-mode trait/disposition data matrix (i.e., subjects administered on a set of traits and dispositions) that have one mode (i.e., the subjects) in common (for a discussion on the terminology of modes and ways, see Carroll & Arabie, 1980; Tucker, 1964).

To reveal the common aspects of the processes underlying the coupled data blocks,[1]all the information available in the different data blocks should be integrated and analysed simultaneously (i.e., data fusion). Different terms have been used to denote this type of analysis of coupled data, like *data fusion* (Aerts et al., 2006; Smilde et al., 2005; Van Mechelen & Smilde, 2010), *multiblock data analysis* (Smilde et al., 2003), *analysis of coupled or linked data* (Wilderjans, Ceulemans, Kiers, et al., 2009; Wilderjans, Ceulemans, & Van Mechelen, 2009) and *integrative data analysis* (Curran & Hussong, 2009). One approach is to use a (global) model in which the information in each dataset is summarized by fitting an N-way N-mode decomposition model (e.g., principal component analysis [PCA], Parafac, Tucker3) to each data block. In an N-way N-mode decomposition model, the data are decomposed into N component matrices, one for each of the N modes of the data, and possibly also a linking array that captures the relations

[1]Note that knowing the common aspects further enables researchers to detect functional relations between the different data blocks by linking the pieces of information present in each block to each other (e.g., relating situation-specific behavioural profiles to traits).

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

between the components of the *N* modes (Van Mechelen & Schepers, 2007). For the personality example, this translates into, for the three-way block, deriving individual differences dimensions, response types, and situation types (which are captured in three different component matrices) and describing the main interactions between these different dimensions and response/situation types (through a linking structure or core array) present in the data; for the two-way block, dispositional styles are derived (again with a component matrix). For each common mode, the underlying parameters (i.e., common components) are required to be the same for all data blocks to which the mode in question belongs (Van Mechelen & Smilde, 2010). As a consequence, each common mode is represented by a single set of components. Applied to the personality example, this implies that the individual differences structure regarding the situational responses is set equal to the structure of individual differences in dispositional styles. This allows one to easily link the situational responses to dispositional styles, herewith summarizing the complex relations between these in a few component matrices and a linking array.

When fitting such a model to a particular coupled dataset, two strategies may be adopted; examples of applications of both strategies can be found in Boque and Smilde (1999), Coxe (1986), Smilde et al. (2000), and ten Berge (1986). The first is a *sequential* strategy in which a (multiway) decomposition model is fitted to the main data block of interest, and next, the other data blockor blocks are decomposed, with the quantification of the common mode (i.e., common components) being the same as obtained in the first step. Note that in this strategy, different loss functions, one for each data block, are optimized separately/sequentially. Applied to the personality example, a three-way three-mode decomposition model is fitted to the behavioural-signature data, and in a second step, the disposition data are modelled with a two-way two-mode decomposition model in which the common person components are required to equal those from the first analysis. This implies that the individual differences structure is fully determined by the information in the subject by response by situation data array, with this individual differences structure then being forced upon the subject by the disposition data block. The second strategy, which will be denoted by the term *simultaneous*, consists of fitting a global model to all data blocks simultaneously. The simultaneous strategy implies the optimization of a single global loss function, which may be an (un)weighted sum of partial loss functions, one for each data block, with the weights indicating the extent to which each data block influences the estimation of the common components. Note that these weights may be chosen a priori (e.g., using information on how noisy the data in each data block are; see van den Berg et al., 2009; Wilderjans et al., 2011) or be determined during the data analysis (e.g., by means of cross-validation; see Smilde & Kiers, 1999). In this paper, it will be assumed that these weights are known at the start of the analysis (for a discussion on this, see Section 6). The simultaneous strategy implies that when estimating the parameters of the common mode, the information that is present in all data blocks to which this common mode belongs is used (i.e., data fusion). For our personality example, the behavioural signature and disposition data block are modelled simultaneously with an appropriate decomposition model, and the common components, which are equal for both data blocks, are estimated using the information present in both data blocks. This implies that both information from the subject by response by situation data array and the subject by disposition data matrix is used to obtain the structure of individual differences, with the weights denoting the importance of the information in each data block/array in deriving this structure.

A sequential strategy may be used when the structure underlying one or more data blocks is of primary interest and the other data blocks are only used to guide the interpretation of the main data block. However, when a coupled dataset contains common components and the goal of the analysis is to retrieve these components, a simultaneous strategy may be called for, for the following two reasons. First, by with the fusing fusion of information from different data blocks, a simultaneous strategy may be able to clearly differentiate between common processes underlying all data blocks and important (i.e., explaining a substantial part of the variance in the data) distinctive processes that only play happen in a single block; a sequential strategy, on the in contrarycontrast, may not be capable of delineating common from distinctive processes because such a strategy only focuses on the important processes (both common and distinctive ones) underlying a single data block. Second, as in a sequential strategy, the common components are estimated using only information from a single data block, the obtained solution (both for the block of

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

interest and for the structure underlying the other data blocks) may be blurred by the noise present in that data block. A simultaneous strategy may overcome this by pooling, when estimating the common components, all the information that is present in the data blocks. As such, noise may get filtered out, resulting in more correct and stable inferences regarding the structure underlying the (common and non-common modes of the) coupled dataset. For a discussion of the advantages and disadvantages of both strategies and a comparison of both strategies by means of using an extensive simulation study in the context of hierarchical classes (HICLAS) models (see Wilderjans et al., 2008).

In this paper, we will confine our attention to a coupled dataset consisting of a real-valued three-way data array that shares a single mode with a real-valued two-way data array (e.g., behavioural signature and dispositional data collected for the same set of subjects). Recently, the *linked-mode PARAFAC-PCA* (LMPCA) model was proposed to analyse such coupled data (Wilderjans, Ceulemans, Kiers, et al., 2009; Wilderjans, Ceulemans, & Van Mechelen, 2009). The LMPCA model is a global model in which a three-way array is decomposed according to Parafac (Carroll & Chang, 1970; Harshman, 1970; Hitchcock, 1927) and a two-way matrix with PCA (Hotelling, 1933; Jolliffe, 2002; Pearson, 1901), and the components for the common mode are constrained to be equal for both data blocks. To estimate the parameters of the LMPCA model, the authors proposed a simultaneous fitting strategy in which a weighted loss function is optimized by means of an alternating least squares (ALS) algorithm (de Leeuw, 1994; Kroonenberg & De Leeuw, 1980; ten Berge, 1993). When analysing psychological processes, however, LMPCA may not yield a good description of the underlying structure as the Parafac model underlying the three-way data imposes a set of restrictions that are often not met for psychological data. In particular, Parafac assumes that the number of components is equal for each mode. For the personality example, this would imply that the number of individual differences dimensions equals the number of response types and the number of situational types, which is a rather strict requirement for this empirical dataset. Moreover, components from different modes are only related to one another in a one-to-one fashion, implying that the component loadings for the different slices (i.e., third mode of the array) are assumed to be proportional to each other (i.e., only a limited set of interactions between individual differences dimensions, response types, and situation types is allowed for). Applied to our example, there is no theoretical ground to assume that the variation in the persons, responses, and situations can be adequately captured by the same number of components. Moreover, it is also not tenable to believe that the same behaviour components, although they may differ in strength (i.e., proportionality), play a role in each encountered situation (i.e., limited interactions). To address these limitations, in this paper, we will propose the T3–PCA model, which is a global model for a simultaneous analysis of the type of coupled data we focus on. In this model, the three-way data block is decomposed according to a Tucker3 model (Tucker, 1964, 1966) and the two-way data are modelled with PCA, with the restriction that the common components must be equal for both data blocks. When the three-way data block is decomposed according to a Tucker3 model, a less restrictive global model (than LMPCA) is obtained as Tucker3 allows the number of components to differ across modes and implies a more complex linking structure (than LMPCA) between the components of the different modes. For the personality example, this implies that the number of individual differences dimensions, response types, and situational types can differ and that a complex set of interactions between these dimensions and types is permitted.

The remainder of this paper is organized into five main sections. In Section 2, following a discussion of the data structure, options for preprocessing and recapitulating the Tucker3 model, the T3–PCA model is presented and the relations with related models are discussed. In the third section, the loss function is introduced and an ALS algorithm to estimate the model parameters is described, together with heuristics for rank selection and postprocessing issues. In the following two sections, the performance of the simultaneous T3–PCA strategy is studied and compared with the performance of a sequential strategy and with LMPCA. In the fourth section, this is done by means of an extensive simulation study and in the fifth section by applying T3–PCA to empirical data from the field of contextualized personality psychology. Section 6 contains some concluding remarks.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | MODEL

2.1 | Data structure and preprocessing

2.1.1 | Data structure

A linked-mode T3–PCA analysis requires a (three-way) $I \times J \times K$ object by attribute by source real-valued data array $\mathbf{D}^1$ and a (two-way) $I \times L$ object by covariate real-valued data matrix $\mathbf{D}^2$ that have objects (i.e., the elements of the first mode) in common. For example, in personality research, as can be seen in the graphical presentation of such a data structure in Figure 1, the data array $\mathbf{D}^1$ may contain for a set of participants their behaviour profile across different situations, whereas $\mathbf{D}^2$ may have scores for the same set of participants on a list of personality items (e.g., dispositions). Other examples of these types of data (for an overview, see Acar et al., 2013) can be found in psychometrics (Wilderjans, Ceulemans, Kiers, et al., 2009; Wilderjans, Ceulemans, & Van Mechelen, 2009), neuropsychology (Chiang et al., 2010; Mitchell et al., 2008; Palatucci et al., 2009), community detection (Lin et al., 2009), collaborative filtering (Ermis et al., 2012; Zheng et al., 2010), bioinformatics (Acar et al., 2012), and chemometrics (Smilde et al., 2000). It is important to note that the set of elements for the common mode in $\mathbf{D}^1$ and $\mathbf{D}^2$ is identical (i.e., the same set of participants in the personality psychology example).

2.1.2 | Preprocessing

Preprocessing often involves operations of centring and scaling that aim at removing arbitrary differences between the variables regarding offsets (i.e., scale means) and scale variances. A classical method of preprocessing a two-way data block (as is done in, for example, PCA) is to centre (i.e., mean of zero) and standardize (i.e., variance of one) each variable (Bro & Smilde, 2003). For a three-way data block, however, different centring and scaling options exist (Kroonenberg, 1983, 2008; Smilde et al., 2004). Because we want to reveal the common mechanisms underlying both data blocks, we advise, prior to an actual T3–PCA analysis, removing differences between both data blocks in overall variable means. To this end, in the three-way array, the mean of each attribute-source combination (across the objects) is set to zero, whereas in the two-way matrix, each covariate is centred (across the objects). Further, to account for arbitrary differences in scale variances (Timmerman & Kiers, 2003), in the three-way array, the variance of each attribute (computed across objects and sources) is set to one (i.e., sum of squares of IK), whereas in the two-way matrix, the variance of each covariate is fixed to one (i.e., sum of squares

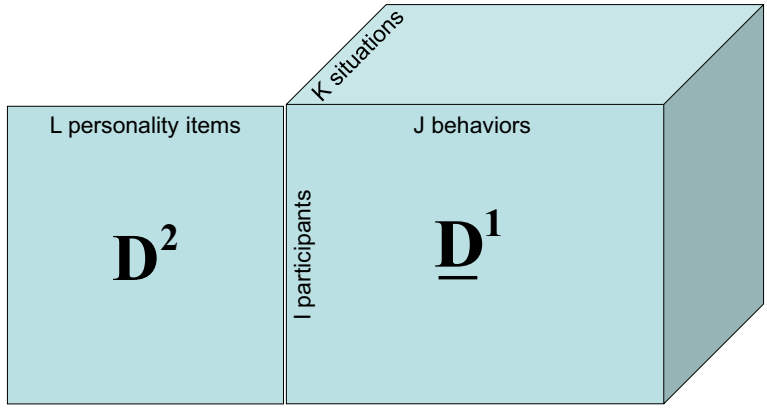


FIGURE 1 Graphical representation of data for a linked-mode T3–PCA analysis.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of *I*). As a consequence, the importance of each variable, considered both for the whole dataset and for each data block separately, is set as equal in the analysis. In the remainder of this paper, we will always assume that the coupled data are centred and scaled as explained in this section.

2.2 | A linked-mode model for coupled three-way and two-way real-valued data: T3–PCA

2.2.1 | Model

We will first briefly review the Tucker3 decomposition model before introducing the linked-mode T3–PCA model.

Tucker3 model

The Tucker3 model (Tucker, 1963, 1964, 1966) can be considered a generalization of PCA to three-way data, for example, in the personality research example, data measuring different subjects' behaviour profiles across various situations (for an illustrative explanation of the Tucker3 model, including a focus on how to interpret its output, see Kiers & Van Mechelen, 2001). In more detail, the Tucker3 model for a three-way real-valued data array **D** estimates components for each of the three modes, resulting in a component loading matrix for each mode (i.e., **A**, **B**, and **C** for the first, second, and third modes, respectively). Applied to the personality research example, these component matrices indicate how subjects, behaviours, and situations can be summarized into individual differences dimensions (subject types), behaviour types, and situational types. Note that the number of components should be chosen by the user and can, as opposed to PCA, differ between modes (i.e., *P*, *Q*, and *R* components are extracted for the first, second, and third mode, respectively). Further, a core array **G** is estimated that represents the strength of the main interactions between the components of the three different modes in the data. In the personality research example, the core values indicate which types of reaction – and to what extent – are typically shown in which situational types by each individual differences dimension. Note that, as opposed to PCA, one situational type can trigger more than one behaviour type and for more than one individual differences dimension. As such, the Tucker3 model allows for a more complex description of the (main interactions in the) data than PCA and than the better known three-mode PARAFAC (also called CP) model, which requires the number of components to be equal across modes and only permits one-to-one interactions between corresponding components across modes (i.e., **G** is constrained to be superdiagonal). Algorithms to fit the Tucker3 model to data are presented in Kroonenberg and De Leeuw (1980) and Smilde et al. (2004), with an adapted algorithm for three-way data with a (very) large number of subjects (i.e., *I* being larger than *JK*) being introduced in Kiers et al. (1992).

T3–PCA model

The linked-mode T3–PCA model approximates an $I \times J \times K$ real-valued data array $\mathbf{D}^1$ that is coupled/linked (through the first mode) with an $I \times L$ real-valued data matrix $\mathbf{D}^2$ by a real-valued model array $\mathbf{M}^1$ (of size $I \times J \times K$) and a real-valued model matrix $\mathbf{M}^2$ (of size $I \times L$), respectively, that have the first mode in common. In particular, $\mathbf{D}^1$ is approximated by model array $\mathbf{M}^1$, which can be decomposed according to a Tucker3 model (Kroonenberg & De Leeuw, 1980; Tucker, 1964, 1966) with *P* object, *Q* attribute, and *R* source components; $\mathbf{D}^2$ is approximated by model matrix $\mathbf{M}^2$, which can be decomposed according to a PCA model with *P* components, with the component matrix for the first mode (i.e., rows) in both $\mathbf{M}^1$ and $\mathbf{M}^2$ being restricted to being identical. Therefore, $\mathbf{M}^1$ and $\mathbf{M}^2$ can be decomposed as follows:

$$m^1_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b^1_{jq} c_{kr} g_{pqr}, \tag{1}$$

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$m_{il}^2 = \sum_{p=1}^{P} a_{ip} b_{lp}^2, \tag{2}$$

where a_{ip}, b_{jq}^1, c_{kr}, and b_{lp}^2 denote the entries of the real-valued component matrices $\mathbf{A}$ ($I \times P$), $\mathbf{B}_1$ ($J \times Q$), $\mathbf{C}$ ($K \times R$), and $\mathbf{B}_2$ ($L \times P$) for the objects, attributes, sources, and covariates, respectively, and P, Q, and R indicate the number of object/covariate, attribute, and source components, respectively. The entries g_{pqr} denote the elements of the real-valued core array $\mathbf{G}$ (of size $P \times Q \times R$) and reflect the relationship between the pth, qth, and rth components in $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$, respectively. Note that $\mathbf{M}^1$ and $\mathbf{M}^2$ are linked to each other by requiring the common object component matrix $\mathbf{A}$ to be identical in both submodels. It should be noted that the parameters of the linked-mode T3–PCA model (i.e., the component matrices and the core array) can only be uniquely identified upon a joint permutation, scaling, reflection, and rotation of the components and/or the core array (Smilde et al., 2004). To partially identify the solution, we will constrain $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$ to be columnwise orthonormal (i.e., $\mathbf{A}'\mathbf{A} = (\mathbf{B}_1)'\mathbf{B}_1 = \mathbf{C}'\mathbf{C} = \mathbf{I}$, which implies that components are orthogonal and have unit length). Note that these constraints, which do not alter the fit of the model, imply that all the variation in the three-way block is carried to the core elements and all the variation in the two-way block to $\mathbf{B}_2$. As a consequence, in general, $\mathbf{B}_2$ will not be orthogonal or normalized and requiring $\mathbf{B}_2$ to be orthonormal implies a quite restrictive model (i.e., the P non-zero eigenvalues of $\mathbf{M}^2$ being equal to one). It should be remarked that additional restrictions are needed to fully identify the model (Smilde et al., 2004). This non-uniqueness, however, as is done in PCA, can be used to obtain a solution that is easier to interpret (see the discussion of postprocessing strategies in Section 3.4).

2.2.2 | Relation to other models

The linked-mode T3–PCA model is related to several existing models both inside and outside the family of multiway multiblock component models. Inside this family, the linked-mode T3–PCA model can be treated as a generalization of the LMPCA model, in the same way as Tucker3 is a generalization of Parafac. In particular, when constraining $P = Q = R$ and $\mathbf{G}$ to be a superidentity matrix (i.e., $g_{pqr} = 1$ if $p = q = r$, and $g_{pqr} = 0$ otherwise), the linked-mode T3–PCA model boils down to the LMPCA model. The family of multiway multiblock component models is further closely related to the family of multiway covariate regression models (also called multiway multiblock regression models; see Smilde et al., 2000; Smilde & Kiers, 1999). The main difference between both model families pertains to the role that is assigned to the different data blocks (i.e., whether or not blocks have exchangeable roles in terms of conceptual status). In a multiway multiblock component model, all data blocks serve the same role, whereas in a multiway covariate regression model, some blocks are considered predictor blocks that are to be regressed on the other blocks that function as criterion blocks.

3 | DATA ANALYSIS

3.1 | A procedure for testing the existence of common components

As it does not always make sense to couple datasets to each other, a procedure for testing whether common components are present in a coupled dataset is proposed. The procedure consists of fitting a Tucker3 model to a three-way dataset $\mathbf{D}^1$ and separately a PCA model to a two-way dataset $\mathbf{D}^2$ and retaining the component matrices for the common mode (i.e., indicated as $\mathbf{A}_{3D}$ for $\mathbf{D}^1$ and $\mathbf{A}_{2D}$ for $\mathbf{D}^2$). Next, after orthogonally rotating (optimally in a least squares sense) $\mathbf{A}_{2D}$ towards $\mathbf{A}_{3D}$, the amount of similarity between these component matrices for the common mode is quantified by means of the Tucker coefficient (Tucker, 1951). A large similarity implies the existence of common components that are present in both $\mathbf{D}^1$ and $\mathbf{D}^2$. With the aim of determining whether the obtained similarity value

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is substantially large to decide for the presence of common components, a permutation test approach is proposed. In this approach, the rows of the two-way data block $\mathbf{D}^2$ are randomly permuted in order to create a dataset without common components, and this is repeated a certain number of times (e.g., 1000 or more), and the similarity value is computed in the same way as explained above for each permuted dataset. Finally, computation is done to determine how extreme the obtained similarity value for the original data is compared to the distribution of similarity values for the permuted data, which represents data without common components (i.e., like a *p*-value). When the similarity value for the original data is larger than, for example, 95% of the values from permuted data, it can be concluded that common components are present in the coupled dataset and, as such, that it makes sense to analyse the coupled data with T3–PCA.

3.2 | Aim

The aim of a linked-mode T3–PCA analysis with rank (P, Q, R) of a coupled dataset $(\mathbf{D}^1, \mathbf{D}^2)$ is to estimate a coupled model $(\mathbf{M}^1, \mathbf{M}^2)$ such that the value of the loss function

$$f = \frac{\alpha}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk})^2} \times \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(d^1_{ijk} - \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b^1_{jq}c_{kr}g_{pqr}\right)^2 + \frac{(1-\alpha)}{\sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il})^2} \times \sum_{i=1}^{I}\sum_{l=1}^{L}\left(d^2_{il} - \sum_{p=1}^{P} a_{ip}b^2_{lp}\right)^2, \tag{3}$$

is minimized (with $0 \leq \alpha \leq 1$), $\mathbf{M}^1$ can be represented by a Tucker3 model with P, Q, and R components (and $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$ being orthonormal), $\mathbf{M}^2$ by a PCA model with P components, and with $\mathbf{A}$ being constrained to be equal in the decomposition of $\mathbf{M}^1$ and $\mathbf{M}^2$. The weights α and $1-\alpha$ denote the extent to which each of the two data blocks influences the analysis (Wilderjans, Ceulemans, Kiers, et al., 2009). When α equals one or zero, a sequential approach is obtained in which the components of the common mode are derived based on the information in one data block only (i.e., the three-way and two-way data blocks, respectively). For intermediate values of α, the common components are estimated using the (weighted) information in both data blocks, which can be conceived as a simultaneous approach. To let the information in both data blocks influence the analysis to the same extent, we normalized the data blocks by dividing the sum of squared differences between data and model for each block by the corresponding sum of squared data entries. When the variables in both data blocks have been centred and scaled (as explained in Section 2.1), this normalization procedure boils down to a correction for the differences in block sizes as each denominator in Equation (3) becomes equal to the number of entries of the corresponding data block (i.e., in that case, $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk})^2 = IJK$ and $\sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il})^2 = IL$). Note that the loss function in Equation (3) can be rewritten as (Vervloet et al., 2013)

$$f_2 = \beta \times \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(d^1_{ijk} - \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b^1_{jq}c_{kr}g_{pqr}\right)^2 + (1-\beta) \times \sum_{i=1}^{I}\sum_{l=1}^{L}\left(d^2_{il} - \sum_{p=1}^{P} a_{ip}b^2_{lp}\right)^2, \tag{4}$$

with $\beta = \frac{\alpha \times \sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il})^2}{\alpha \times \sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il})^2 + (1-\alpha) \times \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk})^2}$ and f_2 in Equation (4) and f in Equation (3) being equal to each other upon a scaling factor that is irrelevant for the optimization (i.e., f and f_2 have the same optimal solution).

It should be noted that the optimization problem implied by loss function (Equation 3) can be reformulated as a scalar multiobjective (in this case biobjective) programming problem (for an overview of multiobjective or multicriterion optimization, see Ehrgott, 2005; Olson, 1992; Ringuest, 1992; Rosenthal, 1985;

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Vervloet et al., 2013). Indeed, $f_1(\mathbf{A}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk} - \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b^1_{jq}c_{kr}g_{pqr})^2$ and $f_2(\mathbf{A}) = \sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il} - \sum_{p=1}^{P} a_{ip}b^2_{lp})^2$ are two objective subfunctions (i.e., criteria), both of which depend on **A**. Further, both objective subfunctions f_1 and f_2 are standardized (by dividing f_1 and f_2 by $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(d^1_{ijk})^2$ and $\sum_{i=1}^{I}\sum_{l=1}^{L}(d^2_{il})^2$, respectively) and the final loss function (Equation 3) is a convex combination using strictly positive weights of the standardized objective subfunctions (i.e., $f = \alpha f_1 + (1-\alpha) f_2$ with $0 \leq \alpha \leq 1$). The goal of multiobjective programming, which is a methodology stemming from operations research that allows analysts to investigate trade-offs between various relevant objective (optimization) criteria (e.g., f_1 and f_2), is to generate the efficient set of solutions (also called the Pareto efficient set, non-dominated set of solutions or the efficient frontier). A solution **A** is dominated by another solution **A′** when $f_1(\mathbf{A}') \leq f_1(\mathbf{A})$ and $f_2(\mathbf{A}') \leq f_2(\mathbf{A})$ and $f_c(\mathbf{A}') < f_c(\mathbf{A})$ for at least one objective subfunction ($c = 1$ and/or $c = 2$), which implies that **A′** does not lead to a worse (i.e., equal or better) loss function value than **A** on each objective subfunction and yields a better loss value than **A** for at least one objective subfunction (for a problem formulation, see Brusco & Stahl, 2001; Brusco & Steinley, 2009). A solution is efficient or non-dominated when there exists no other solution that dominates this solution. The goal of multiobjective programming is to generate the full set of efficient or non-dominated solutions, which for small problems is possible by enumeration but for larger problems only can be approximated by a small subset of efficient solutions. Multiobjective programming is especially interesting for combinatorial optimization problems for which different optimization criteria can be formulated, each of which on its own leads to different interesting solutions. Examples in this regard, for which multiobjective programming has been implemented, are partitioning methods for dissimilarity matrices (Brusco & Cradit, 2005), seriation methods for skew-symmetric matrices (Brusco & Stahl, 2005), and blockmodeling methods for social network analysis (Brusco et al., 2013). To solve the multiobjective problem, three main approaches exist: (1) the scalar optimization or weighted-sum approach, (2) the constraint approach (i.e., optimizing one criterion subject to threshold constraints on the other criteria), and (3) direct search methods to approximate the Pareto efficient frontier (for more information, see Ehrgott, 2005; Ehrgott & Gandibleaux, 2000; Ehrgott & Wiecek, 2005). The approach chosen using loss function (Equation 3) is the weighted-sum approach to a multiobjective problem in which the multiobjective problem is translated into a single-objective problem that consists of a weighted sum – resulting in a scalar function – of the objective subfunctions (Ehrgott & Wiecek, 2005). Although commonly adopted, the weighted-sum approach suffers from two main limitations. First, it is challenging to select an appropriate set of weights to use in the weighted sum. Second, whereas exact solutions to the single-objective problem produce supported Pareto-efficient solutions to the multiobjective problem, many interesting unsupported Pareto-efficient solutions to the multiobjective problem may be missed by the weighted-sum approach as they do not comply with – and thus cannot be identified using – any set of weights (for more information, see Brusco et al., 2013; Brusco & Steinley, 2012; Ehrgott & Gandibleaux, 2000; Ulungu & Teghem, 1994).

3.3 | Algorithm

To estimate the parameters of the linked-mode Tucker3-PCA model, an ALS algorithm is used (Kroonenberg & De Leeuw, 1980). In this algorithm, after obtaining starting values for all model parameters, the component matrices **A**, $\mathbf{B}_1$, **C**, and $\mathbf{B}_2$ and the core array **G** are alternatingly re-estimated conditionally upon the other parameters until a prespecified convergence criterion is satisfied (see ten Berge, 1993).

The initial estimates of the component matrices can be determined either rationally or randomly (for a discussion of different types of starting configurations, see Ceulemans et al., 2007). Rational initial estimates of **A**, $\mathbf{B}_1$, and **C** are obtained by taking the eigenvectors associated with the P, Q, and R largest eigenvalues of $[\mathbf{D}^1_a|\mathbf{D}^2][\mathbf{D}^1_a|\mathbf{D}^2]'$, $\mathbf{D}^1_b(\mathbf{D}^1_b)'$, and $\mathbf{D}^1_c(\mathbf{D}^1_c)'$, respectively, where $\mathbf{D}^1_a$, $\mathbf{D}^1_b$, and $\mathbf{D}^1_c$ are the $I \times JK$, $J \times KI$, and $K \times IJ$ matricized versions of $\mathbf{D}^1$ and | denotes matrix concatenation

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Kiers, 2000). Note that this procedure ensures that the initial $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$ are orthonormal. Random initial estimates for the component matrices are obtained, first, bysampling at random entries from a standard normal distribution and, next, to obtain orthonormal initial component matrices, to compute an orthonormal basis for the range of each of the initial component matrices. Once initial values are obtained for $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$, initial estimates for $\mathbf{B}_2$ and $\mathbf{G}$ can easily be obtained by means of linear regression (see further).

From initial estimates of the model parameters, the algorithm performs a number of iterations, until a specified stop criterion is met. In each iteration, the component matrices and the core array are re-estimated conditionally upon the other parameters. In particular, first, fixing the entries of $\mathbf{B}_1$, $\mathbf{C}$, $\mathbf{G}$, and $\mathbf{B}_2$, the entries of $\mathbf{A}$ are re-estimated. Next, the entries of $\mathbf{B}_1$ (conditionally upon $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{G}$) and $\mathbf{C}$ (conditionally upon $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{G}$) are updated. Finally, $\mathbf{G}$ (conditionally upon $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$) and $\mathbf{B}_2$ (conditionally upon $\mathbf{A}$) are re-estimated. Updating the component matrix $\mathbf{B}_2$ and the core array $\mathbf{G}$ conditionally upon the other model parameters boils down to a multivariate linear regression problem for which a closed-form solution is available (see Andersson & Bro, 1998; Bro & Andersson, 1998; Kroonenberg & De Leeuw, 1980; Smilde et al., 2004). The conditional re-estimation of the component matrices $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$ can be performed by solving a constrained (i.e., orthogonality) multiple linear regression problem. In particular, the update for $\mathbf{A}$, which is shared by the Tucker3 model for the three-way data and the PCA model for the two-way data, equals $\mathbf{UV}'$, where $\mathbf{U}$ and $\mathbf{V}$ are the P left and right singular vectors, respectively, associated with the P largest singular values of a matrix $\mathbf{Q} = \mathbf{YZ}'$, with $\mathbf{Y} = [\sqrt{\alpha} \times \mathbf{D}_a^1 \mid \sqrt{(1-\alpha)} \times \mathbf{D}^2]$, $\mathbf{Z} = [\sqrt{\alpha} \times \mathbf{G}_a (\mathbf{C} \otimes \mathbf{B}_1)' \mid \sqrt{1-\alpha} \times \mathbf{B}_2']$, and $\mathbf{G}_a$ denotes the (row mode) matricized version of $\mathbf{G}$; $\mathbf{B}_1$ and $\mathbf{C}$ are updated in a similar fashion. In particular, in the case of $\mathbf{B}_1$, $\mathbf{Y} = \mathbf{D}_b^1$ and $\mathbf{Z} = \mathbf{G}_b (\mathbf{A} \otimes \mathbf{C})'$, whereas in the case of $\mathbf{C}$, $\mathbf{Y} = \mathbf{D}_c^1$ and $\mathbf{Z} = \mathbf{G}_c (\mathbf{B}_1 \otimes \mathbf{A})'$. Here, $\otimes$ denotes the Kronecker product, also called the matrix direct product,[2]between two matrices (Zehfuss, 1858) and $\mathbf{G}_a$ ($\mathbf{G}_b$, $\mathbf{G}_c$) is the $P \times QR$ ($Q \times RP$, $R \times PQ$) matricized version of $\mathbf{G}$. A more detailed description of the algorithm can be found in Appendix 1.

After each iteration, the loss function value is re-computed using the updated parameter estimates. When updating the component matrices and the core array results in a substantial decrease in the loss function (i.e., a decrease larger than or equal to a certain proportion, e.g., .0000001, of the loss function value of the previous iteration), a new iteration is performed; otherwise, the algorithm stops (i.e., it reached convergence), and the current estimates of the model parameters are retained as the final estimates.

As updating the model parameters during the iterative process can never lead to an increase in the loss function value and the loss function is bounded below by zero, the T3–PCA algorithm is guaranteed to converge (de Leeuw, 1994; Krijnen, 2006). This, however, does not imply that the algorithm will always identify the global optimal parameter estimates, as the loss function may suffer from the presence of (many) locally optimal solutions. To minimize the risk of the T3–PCA algorithm ending up in a local optimal solution, it is advised to adopt a multistart procedure. In such a procedure, the algorithm is run multiple times, each time starting with different randomly or pseudo-randomly obtained initial parameter values, and the solution from the run yielding the lowest loss function value is retained. Pseudo-random initial values can be obtained by slightly perturbating a rationally obtained starting solution (see Ceulemans et al., 2007; Wilderjans & Ceulemans, 2013). Code to perform a T3–PCA analysis can be downloaded from https://github.com/efb2711/T3PCA_programs.

[2]The Kronecker product $\otimes$ between matrices $\mathbf{A}$ (N × M) and $\mathbf{B}$ (U × V) results in matrix $\mathbf{Z}$ (NU × MV), which equals:

$$\mathbf{Z} = \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & x_{12}\mathbf{B} & x_{13}\mathbf{B} & \dots & x_{1M}\mathbf{B} \\ x_{21}\mathbf{B} & x_{22}\mathbf{B} & x_{23}\mathbf{B} & \dots & x_{2M}\mathbf{B} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N1}\mathbf{B} & x_{N2}\mathbf{B} & x_{N3}\mathbf{B} & \dots & x_{NM}\mathbf{B} \end{bmatrix}.$$

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.4 | Model selection

The model selection task for the T3–PCA model consists of two subtasks, which can be handled by using a simultaneous or a sequential strategy: (1) identifying the optimal model rank (P, Q, R) and (2) determining the optimal $\boldsymbol{\alpha}$ weight underlying a coupled dataset at hand.

3.4.1 | Selecting the T3–PCA model rank

In general, no a priori information is available regarding the optimal rank (P, Q, R) underlying a T3–PCA model for a dataset at hand. As a way out, researchers may perform different analyses with varying complexities, in which the rank of the model goes from (1, 1, 1) to (P_{max}, Q_{max}, R_{max}), and select a model that has a good balance between model fit and model complexity; model (mis)fit may be quantified by the optimal loss function value, whereas different options exist to define model complexity, like the total number of components (i.e., $P+Q+R$), the total number of fitted parameters (i.e., $IP+JQ+KR+PQR+LP$), and the number of free parameters (i.e., $IP+JQ+KR+PQR+LP-P^2-Q^2-R^2$; see Ceulemans & Kiers, 2006; Kroonenberg & ten Berge, 2011). It should be remarked that models for which $Q > PR$ or $R > PQ$ should not be considered as it can be shown that there exist models that are more parsimonious (i.e., have a smaller number of fitted parameters) that fit the data equally well (i.e., have the same loss function value; see Ceulemans & Kiers, 2006; Wansbeek & Verhees, 1989). Next, the optimal rank of a T3–PCA model may be identified using the CHull procedure (Ceulemans & Kiers, 2006, 2009; Wilderjans et al., 2013), which is an automated heuristic for model selection.[3]Note that CHull is a flexible method as alternative choices for the (mis)fit and complexity measure can be made, which enables users to choose those measures that best accomplish their study goal.

Starting from a fit f_i (e.g., T3–PCA loss) and complexity c_i (e.g., $P+Q+R$) value for all valid T3–PCA solutions with a rank between (1, 1, 1) and (P_{max}, Q_{max}, R_{max}), the CHull method goes through the following two steps (for a detailed description of the CHull procedure, see Wilderjans et al., 2013): (1) determining the models m_i ($i = 1, \ldots, M$) that are located on the (lower) boundary of the convex hull of the c_i by f_i complexity-by-fit plot of all the valid models and (2) identifying the model on the boundary of the convex hull that optimally balances model (mis)fit and model complexity. Regarding the first step, first, for each level of complexity c_i, only the model with the best fit (i.e., the model with the largest goodness of fit or smallest misfit) is retained. When two (or more) models yield the same optimal fit, both models are retained (but only one of them is used in further computations). Next, models are removed that are not promising because, when compared to a less complex model, their fit increase is rather small relative to their complexity increase. Regarding the second step, only using the models retained after step 1 (i.e., models on the boundary of the convex hull), one may use an automated procedure by computing for each model m_i the corresponding st-ratio,

[3]Cross-validation (CV) could be considered an interesting alternative to CHull for selecting the optimal T3–PCA model rank (P, Q, R). The idea behind CV-based model selection is that the model with the largest predictive power is the proper model to separate the systematic part of the data (i.e., the relevant components) from the residuals (i.e., noise components). Including noise components (i.e., overestimation of the dimensionality of the model) or discarding relevant components (i.e., underestimation) leads to lower predictive power. Studies, however, show that CV is often not a very successful method for model selection in the context of component models like T3–PCA. For PCA, for example, several CV methods have been proposed (which already shows that CV cannot directly be translated from a supervised to an unsupervised context like component analysis), most of which perform rather poorly (Bro et al., 2008). Further, regarding the use of CV for model – rank – selection in the context of the Tucker3 model, research has demonstrated that just unfolding the three-way array and applying (two-way) CV-based PCA model selection methods does not yield satisfactory results, especially not for noise-prone three-way arrays (Louwerse et al., 1999). Finally, in the context of the PCovR model, which also implies a component model, it has been shown that CV overestimates the number of components as CV selects both the strong (i.e., relevant) and weak (including noise-driven) components. As such, CV is not an optimal model selection method as it erroneously extracts weak/noise components. CHull, however, tends to focus on only the strong and relevant components, resulting in often selecting a good model, especially when a proper threshold is applied, as recommended by Wilderjans et al. (2013), in which all models are excluded for which the fit is less than a certain percentage (e.g., 1 %) better than the fit of a less complex model (for a similar approach, see Timmerman & Kiers, 2000).

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$st_i = \frac{(f_{i-1} - f_i)/(c_i - c_{i-1})}{(f_i - f_{i+1})/(c_{i+1} - c_i)},$$

and selecting the model with the largest *st*-value (which, when multiple models have identical fit f_i and complexity c_i values, is more than a single model). This solution optimally balances model fit and model complexity as increasing the complexity has only a slight effect on the fit measure (i.e., a small denominator), whereas lowering complexity implies a substantial drop in fit (i.e., a large numerator). Note that for the first and last models (i.e., m_1 and m_M) the *st*-ratio is not defined and, as a consequence, these models cannot be selected. For a discussion regarding which models to include in the procedure (see Wilderjans et al., 2013).

3.4.2 | Determining the α weight

Two main options exist to determine the optimal α for a given coupled dataset. First, an optimal α value could be computed by means of maximal likelihood. To this end, a minimal stochastic extension of the T3–PCA model is needed, which can be obtained, for example, by making assumptions regarding the – distribution of the – noise entries of each data block (for similar approaches, see the formulas mentioned in Gurden, n.d.; Wilderjans, Ceulemans, & Van Mechelen, 2009). For the PCovR model, which, like T3–PCA, is a decomposition model for coupled data that includes an α weight, a maximum likelihood estimator for α was derived, and research has demonstrated that an α value between .90 and 1 yields the best recovery results (Vervloet et al., 2013); a similar conclusion was obtained for the PCovR2 model (Gvaladze et al., 2021) for multivariate regression models (Kiers & Smilde, 2007) and for LMPCA (Wilderjans, Ceulemans, & Van Mechelen, 2009). Further, studies have shown that the optimal α depends on the amount of noise in both data blocks and the number of variables in the data (Vervloet et al., 2013, 2016). A second option consists of using cross-validation to identify the α value that maximizes the predictive power (i.e., the performance of the model in terms of predicting new data). For models for coupled data, like PCovR, it was demonstrated that CV, which aims to make the best predictions about future data, often selects a large α value ($\alpha > .90$; see Gurden, n.d.; Heij et al., 2007). Further, good results in terms of model selection were obtained with CV in models for sparse PCA (de Schipper & Van Deun, 2021). It can be concluded that when accurate information on the amount of noise in data blocks is present, an optimal α can be determined by maximum likelihood, whereas CV is preferred when this information is not available, an approach which was also advocated by Smilde and Kiers (1999) in the context of a multiway extension of PCovR. It should be noted that selecting the optimal weight(s) is also a challenging task for multiobjective programming problems (Section 3.2). In this context, the goal is to reveal the shape of the Pareto-efficient frontier, which is often done by evaluating a range of weights and further refining this range in subsequent steps (for more information, see Brusco et al., 2013; Brusco & Stahl, 2001; Ehrgott, 2005).

3.4.3 | Simultaneous and sequential model selection strategies for T3–PCA

The two T3–PCA model selection tasks (i.e., selecting the model rank and the α weight) can be performed simultaneously, but this can become computationally demanding as all possible combinations of T3–PCA model ranks and α values need to be evaluated. Especially in combination with CV (i.e., splitting the sample in several parts and refitting the model several times), a simultaneous model selection procedure can become computationally prohibitive. As an alternative, one can use a sequential procedure for model selection in which one task is performed first and the second task is performed in a second step using the optimal model selection parameters obtained in the first step/task (for an example, see Gvaladze et al., 2021). In the illustrative application (Section 5), for example, first, α is fixed at .50 based on the results of the simulation study, and in a second step

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the optimal model rank is determined, considering only solutions with $\alpha = .50$. Sequential model selection approaches have been demonstrated to work well (adopting CV) for the PCovR (Vervloet et al., 2016) and the SCDCovR model (Park et al., 2021) and a logistic regression version of these models (Park et al., 2023). For a discussion of simultaneous and sequential model selection strategies in other decomposition models for coupled data, one can consult (Park et al., 2021, 2023; Vervloet et al., 2015, 2016).

3.5 | Postprocessing and rotational freedom

As both the Tucker3 and PCA decomposition of the three-way and two-way data blocks, respectively, have rotational freedom (even after imposing the orthonormality constraint on $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$), the T3–PCA decomposition is also not unique (see also Section 3.1). In particular, the component matrices $\mathbf{A}$, $\mathbf{B}_1$, and $\mathbf{C}$ can be independently rotated as long as the core array $\mathbf{G}$ (and, in case of $\mathbf{A}$, also $\mathbf{B}_2$) is appropriately counterrotated. Moreover, one can rotate $\mathbf{A}$ or $\mathbf{B}_2$ as long as this is compensated for in the core array and the component matrix that has not been rotated (i.e., $\mathbf{B}_2$ or $\mathbf{A}$, respectively).

The rotational freedom can be used to obtain a solution that is easier to interpret. To this end, different strategies can be followed (for an overview, see Kroonenberg, 2008): (1) rotate the component matrices, for example, by means of a varimax rotation, in order to obtain a simple structure (Kaiser, 1958) and afterwards adjust the core array, and (2) rotate the core array (Kiers, 1992, 1998a, 1998b; Kroonenberg, 1983) and afterwards counterrotate the component matrices (for an overview, see Smilde et al., 2004). The first option yields components that are easy to interpret but probably at the expense of a complicated core array containing many large numbers (i.e., complex relations between the components). Contrarily, the second option results in a parsimonious core (i.e., only a few large values indicating the most important interactions between the components), but the components may be hard to interpret. Kiers (1997) proposes a more general approach in which the component matrices and the core array are rotated simultaneously in such a way that a criterion is optimized that balances between core simplicity and simplicity of the component matrices. It should be noted that after fixing the rotational freedom, the T3–PCA model is still only identified upon a permutation and/or reflection of the components. As this non-uniqueness does not have any definitive for the interpretation of the solution, it is not discussed here (for more information, see Smilde et al., 2004).

4 | SIMULATION STUDY

4.1 | Problem

In this section, we will present a simulation study to evaluate the performance of the T3–PCA algorithm, herewith focusing on the following two performance aspects: (1) optimization of the loss function and (2) parameter recovery. Regarding the former, we will study how sensitive the algorithm is to the problem of local minima. Concerning recovery, the simulation study has the following two aims: (1) examine the extent to which the T3–PCA algorithm reveals the underlying component structure and (2) determine the degree to which T3–PCA, which is a simultaneous strategy, outperforms a sequential strategy (i.e., extracting the components for $\mathbf{A}$ from the three-way data array and then decomposing the two-way data matrix with the restriction that $\mathbf{A}$ should be the same as in the decomposition of the three-way data) and LMPCA.

To address the latter aim, it is important to distinguish between components for the common mode (in matrix $\mathbf{A}$) that explain a sizeable amount of variance in both data blocks $\mathbf{D}^1$ and $\mathbf{D}^2$ (i.e., common components) and components that only underly one of the two data blocks (i.e., distinctive components, see Schouteden et al., 2013, 2014; Van Deun et al., 2013). When only common

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

components underly a coupled dataset, it can be expected that the sequential strategy would already perform quite well, leaving (almost) no room for the simultaneous strategy to perform better. The reason for this is that Tucker3 imposes a quite strict structure on the three-way data, resulting in an adequate estimation of the common component matrix (i.e., the Tucker3 algorithm filters out the noise in the data quite well; see Wilderjans et al., 2008; Wilderjans, Ceulemans, & Van Mechelen, 2009). However, when the common mode of the three-way data block contains a strong distinctive component (in terms of amount of explained variance) and one or more (small) common components, someone using a sequential strategy may be tempted to retain the distinctive component at the expense of the (small) common ones because the information from the two-way block is not used and model selection strategies are aimed at mainly selecting strong components (Ceulemans & Kiers, 2009). As a simultaneous strategy takes the information in both three-way and two-way data blocks into account, we predict that this strategy will be able to ignore the strong but distinctive component and will retain the (less strong) common component(s).

To test our prediction, we will set up a simulation study in which both data blocks contain an extra component (above the common ones) that is distinctive, and we will investigate how recovery performance is affected by the strength of this distinctive component. Furthermore, we will also study how the recovery performance depends on the following four (other) data characteristics: (1) the size of the three-way data block, (2) the ratio of the size of the three-way data block to the size of the two-way block, (3) the rank of the model and (4) the amount of noise in the data. Regarding these manipulated factors, we expect that the performance of the T3–PCA algorithm will decrease when the distinctive component becomes stronger and when the data become more complex (i.e., a larger rank and more noise, see van den Berg et al., 2009; Wilderjans & Ceulemans, 2013; Wilderjans, Ceulemans, & Kuppens, 2012; Wilderjans, Ceulemans, & Van Mechelen, 2012). Further, we expect the simultaneous strategy to outperform the sequential strategy to a greater extent when the distinctive component explains more variance in the data. Finally, we will also explore how recovery performance is related to the choice of α.

In the simulation study, we will also evaluate a procedure to determine whether the three-way and two-way data blocks have common components, which warrants the use of the coupled T3–PCA model.

4.2 | Design and procedure

The factors that were introduced above were systematically manipulated in a completely randomized five-factorial design in which all factors were considered random:

1. The size ($I \times J \times K$) of the true three-way data array $\underline{\mathbf{T}}^1$, at two levels: ($20 \times 20 \times 20$) and ($20 \times 20 \times 100$);
2. The ratio r of the size of the true three-way data array $\underline{\mathbf{T}}^1$ to the size of the true two-way data matrix $\mathbf{T}^2$:

$$r = \frac{I \times J \times K}{I \times L}.$$

This factor was manipulated at two levels: 1 and 4. When the three-way array has size ($20 \times 20 \times 20$), the size of the two-way matrix is (20×400) for $r = 1$ and (20×100) for $r = 4$. When the three-way array has size ($20 \times 20 \times 100$), the size of the two-way matrix is (20×2000) for $r = 1$ and (20×500) for $r = 4$;
3. The rank of the true model ($P \times Q \times R$) at two levels: (2, 2, 2) and (2, 3, 4);
4. The amount of noise in the data, ε, at two levels: .15 and .30;
5. The amount of variance that the (extra) distinctive component for the common mode (in matrix $\mathbf{A}$) in the two- and three-way data blocks (i.e., $\mathbf{D}^2$ and $\mathbf{D}^1$) explains. This factor was manipulated at three levels: 5% (weak distinctive component), 15% (intermediate strong), and 30% (strong). In Table 1, in which the amount of explained variance for each component is displayed, one can see that the ratio of explained variances for the other two (common) components is kept constant at $\frac{70\%}{30\%}$.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Proportion of explained variance of the components underlying the common mode (in matrix $\mathbf{A}$) of the three-way $\mathbf{D}^1$ and the two-way $\mathbf{D}^2$ data blocks for each level of the amount of variance that the distinctive component explains.

	Common		Distinctive
Level	Component 1	Component 2	Component 3
Weak (5%)	.665	.285	.050
Intermediate (15%)	.595	.255	.150
Strong (30%)	.490	.210	.300

For each combination of the levels of the five manipulated factors, component matrices $\mathbf{A}^{(T)}$, $\mathbf{B}_1^{(T)}$, $\mathbf{C}^{(T)}$, and $\mathbf{B}_2^{(T)}$ were constructed by independently drawing entries from a standard normal distribution; in a second step, these matrices were centred and columnwise orthonormalized. To ensure that the components underlying the two-way data block had the required proportions of explained variance (Table 1), the columns of $\mathbf{B}_2^{(T)}$ were appropriately scaled. The entries for $\mathbf{G}^{(T)}$ were independently drawn from a standard normal distribution and then rescaled such that the sums of squared entries of the rows of $\mathbf{G}^{(T)}$ were in accordance with the required proportions of explained variance of the components for the common mode (in matrix $\mathbf{A}$; see Table 1). For Tucker3, when the component matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are orthonormal, the squared entries of the core array $\mathbf{G}$ are proportional to the amount of variation explained by each combination of components. As such, the sums of squared entries of the rows of $\mathbf{G}$ are related to the amount of explained variance of each component in $\mathbf{A}$ (Kiers & Van Mechelen, 2001; Smilde et al., 2004). To add a distinctive component to each data block, we augmented $\mathbf{A}^{(T)}$ with a different column (yielding $\mathbf{A}_{3D}^{(T)}$ and $\mathbf{A}_{2D}^{(T)}$, which only differ in the last column) such that each augmented $\mathbf{A}^{(T)}$ had been centred and columnwise orthonormalized, the correlation between the distinctive component in $\mathbf{A}_{3D}^{(T)}$ and $\mathbf{A}_{2D}^{(T)}$ was zero and the amounts of explained variance of the components was as required (Table 1). Next, the true block couple ($\mathbf{T}^1$, $\mathbf{T}^2$) was computed by combining the true parameters in $\mathbf{A}^{(T)}$, $\mathbf{B}_1^{(T)}$, $\mathbf{C}^{(T)}$, $\mathbf{B}_2^{(T)}$, and $\mathbf{G}^{(T)}$ by the Tucker3 and PCA decomposition rules in Equations (1) and (2). Further, for each true block couple ($\mathbf{T}^1$, $\mathbf{T}^2$), a data block couple ($\mathbf{D}^1$, $\mathbf{D}^2$) was constructed by adding error matrices $\mathbf{E}^1$ ($I \times J \times K$) and $\mathbf{E}^2$ ($I \times L$) to $\mathbf{T}^1$ and $\mathbf{T}^2$, respectively; $\mathbf{E}^1$ and $\mathbf{E}^2$ were randomly generated, sampling from a standard normal distribution for each error entry, and rescaled in order to obtain the required amount of noise ϵ, which was held constant for both data blocks. In particular, $\mathbf{E}^1$ and $\mathbf{E}^2$ are rescaled – before adding them to $\mathbf{T}^1$ and $\mathbf{T}^2$ – such that $\frac{||\mathbf{E}^1||_F^2}{||\mathbf{D}^1||_F^2} = \epsilon$ and $\frac{||\mathbf{E}^2||_F^2}{||\mathbf{D}^2||_F^2} = \epsilon$, which can be obtained by multiplying the randomly generated $\mathbf{E}^1$ and $\mathbf{E}^2$ with $\sqrt{\frac{\epsilon}{1-\epsilon}\frac{||\mathbf{T}^1||_F^2}{||\mathbf{E}^1||_F^2}}$ and $\sqrt{\frac{\epsilon}{1-\epsilon}\frac{||\mathbf{T}^2||_F^2}{||\mathbf{E}^2||_F^2}}$, respectively, with $||\ldots||_F$ being the Frobenius norm (i.e., the square root of the sum of squared entries) of a data array or matrix (for more information, see Wilderjans, Ceulemans, & Van Mechelen, 2009). The whole data generation procedure was repeated 20 times to obtain 20 replications per cell of the design.

Fully crossing all design factors resulted in 2 (size three-way data) × 2 (ratio) × 2 (rank) × 2 (amount of noise) × 3 (strength of distinctive component) × 20 (replication) = 960 different true block couples ($\mathbf{T}^1$, $\mathbf{T}^2$) and data block couples ($\mathbf{D}^1$, $\mathbf{D}^2$). After centring the data (as described in Section 2.1), the T3–PCA algorithm was applied to each data block couple ($\mathbf{D}^1$, $\mathbf{D}^2$), using the true rank ($P \times Q \times R$) and repeating the analysis with 10 different α-values: .050, .100, .300, .500, .600, .700, .800, .900, .950 and .999.

To compare the simultaneous with the sequential strategy, each (preprocessed) generated coupled dataset was further subjected to the following sequential procedure. First, a Tucker3 analysis (in the true rank) was applied to $\mathbf{D}^1$; next, a PCA model was fitted to $\mathbf{D}^2$ with $\mathbf{A}$ fixed to the optimal component

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

matrix $\mathbf{A}$ from the Tucker3 analysis. Further, the generated datasets with an equal amount of components across modes (i.e., $P = 2$, $Q = 2$, and $R = 2$) were also analysed with LMPCA in order to show the benefit of modelling the three-way data array $\mathbf{D}^1$ with Tucker3 versus Parafac.[4]

With the aim of evaluating the performance of the procedure for testing the existence of common components, the procedure proposed in Section 3.1 was applied to the simulated data. To reduce computational costs, and also because we believe results would not differ much in other simulation conditions, only a subset of the simulated data was re-analysed, consisting of all 20 datasets from each of three selected conditions (i.e., 60 datasets in total). In particular, all datasets with $\mathbf{D}^1$ being of size $20 \times 20 \times 20$, $\mathbf{D}^2$ being of size 20×100, the T3–PCA rank being (2, 2, 2), and the amount of noise being .15 were selected (i.e., all three levels of the amount of variance explained by the distinctive component were included). Moreover, for all analyses, a permutation test approach with 100 permutations of the rows of $\mathbf{D}^2$ was adopted. Note that we would generally recommend using a larger number of permutations (e.g., 1000) but that we did not implement that here because of computational limitations. Note further that the value of α has no effect on the proposed procedure for testing the existence of common components as $\mathbf{D}^1$ and $\mathbf{D}^2$ are analysed separately in this procedure. For the purpose of determining the sensitivity of the proposed procedure, we also tested the procedure on newly generated data in which no common components were present in the coupled data. To this end, we generated 60 datasets with the same features as the 60 selected datasets discussed above but now generating the data such that no common components are present in the coupled dataset (i.e., with only distinctive components). In particular, the same procedure for generating data as explained above was used but now two separate matrices (called $\mathbf{A}_{3D}$ and $\mathbf{A}_{2D}$) for $\mathbf{A}$ were randomly generated and used in the computation of the true block couple $(\mathbf{T}^1, \mathbf{T}^2)$, with $\mathbf{A}_{3D}$ being used for $\mathbf{T}^1$ and $\mathbf{A}_{2D}$ for $\mathbf{T}^2$. To increase the probability of the T3–PCA and the Tucker3 algorithm not retaining a local optimal solution, a multistart procedure with one rational and 100 random starts was adopted. The simulation study was performed in R (version 3.0.2) on a cluster of computers from the Flemish Supercomputer Center (VSC). Code to generate the data as described in this paper can be downloaded from https://github.com/efb2711/T3PCA_simulation.

4.3 | Results

4.3.1 | Computation time

The average computation time of a T3–PCA analysis – computed across all generated datasets and adopted α levels – equals 39 s, with a standard deviation of 90 s. Computation time increases when the size of the three-way data array gets larger (i.e., mean computation times of 22 and 56 s for arrays of size $20 \times 20 \times 20$ and $20 \times 20 \times 100$ respectively), the ratio of the size of $\mathbf{D}^1$ to $\mathbf{D}^2$ rises (i.e., mean computation time of 36 vs. 43 s for an r ratio of .25 vs. 1 respectively), the rank of the T3–PCA analysis increases (i.e., mean computation time of 34 and 45 s for analyses with rank (2, 2, 2) and (2, 3, 4), respectively), and when the strength of the distinctive component for the common mode enlarges (i.e., mean computation times of 14, 25, and 78 s for weak, intermediate strong, and strong distinctive components respectively). The computation time markedly increases for combinations of these factors. In particular, the smallest computation time (of 9 s) is encountered when all these factors take their smallest factor level, whereas the largest computation time (of 148 s) is observed in conditions in which all these factors take their largest factor level.

[4]We decided not to analyse the simulated datasets with rank ($P = 2$, $Q = 3$, $R = 4$) with LMPCA as LMPCA does not allow for a solution with a different number of components across modes. Imposing this restriction (e.g., analysing these datasets with LMPCA with four components) is expected to lead to even worse results than for the LMPCA analysis in the correct rank for the ($P = 2$, $Q = 2$, $R = 2$) datasets.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.3.2 | Optimization of loss function: sensitivity to local minima

The goal of this subsection is to investigate whether the T3–PCA algorithm is capable of identifying the global optimum of the loss function in Equation (3). From the moment, however, that noise is added to the data, the optimal solution is always unknown (see Wilderjans et al., 2008). A way out consists in defining a proxy for the global optimum, which can be conceived of as our best guess for the global optimal solution, and evaluating whether or not the T3–PCA algorithm yields a solution that has a loss function value that equals or is less than the loss function value of the proxy. We define the proxy as the solution with the lowest loss function value encountered among the following two T3–PCA solutions: (1) the true solution $\mathbf{A}^{(T)}$, $\mathbf{B}_1^{(T)}$, $\mathbf{C}^{(T)}$, $\mathbf{G}^{(T)}$, and $\mathbf{B}_2^{(T)}$ underlying the data and (2) the solution obtained by using the sequential procedure. We consider a T3–PCA solution to be suboptimal when its loss value exceeds the loss value of the proxy. This appeared to be the case for only 14 out of the 9600 analyses (.15%), with all these cases belonging to the analyses with α being equal to .999.

4.3.3 | Recovery of model parameters

To study the recovery performance of the T3–PCA algorithm, herewith focusing on the recovery of the common component matrix **A**, a goodness-of-recovery (GOR) statistic was computed for each component matrix. In particular, for **A**, GOR was defined as follows:

$$\mathrm{GOR}_{\mathbf{A}} = \frac{\sum_{p=1}^{P} |\phi(\mathbf{a}_p^{(T)}, \mathbf{a}_p^{(M)})|}{P},$$

with $|\phi()|$ being the absolute value of the Tucker phi-coefficient (Tucker, 1951), $\mathbf{a}_p^{(T)}$, and $\mathbf{a}_p^{(M)}$ the pth component of the true loading matrix $\mathbf{A}^{(T)}$ and the loading matrix $\mathbf{A}^{(M)}$ resulting from the T3–PCA analysis respectively. To account for the rotational freedom of the components (Section 3.4), before computing $\mathrm{GOR}_{\mathbf{A}}$, $\mathbf{A}^{(M)}$ was optimally (in a least squares sense) orthogonally rotated towards $\mathbf{A}^{(T)}$. A similar GOR statistic was computed for $\mathbf{B}_1$, **C**, and $\mathbf{B}_2$. The GOR statistic yields values between 0 and 1, with 1 indicating perfect recovery and 0 that recovery is at chance level. The recovery performance at the level of the core array was determined by calculating the mean squared difference (i.e., a badness-of-recovery measure that will be indicated by $\mathrm{BOR}_{\mathbf{G}}$) between the true and estimated core elements in $\mathbf{G}^{(T)}$ and $\mathbf{G}^{(M)}$ respectively. To address the rotational freedom of the core array, the estimated core array $\mathbf{G}^{(M)}$ was first optimally rotated towards $\mathbf{G}^{(T)}$, herewith minimizing the sum of squared differences between $\mathbf{G}^{(T)}$ and the rotated $\mathbf{G}^{(M)}$. Note that the T3–PCA model is unchanged under the following orthogonal rotation(s) of the core array by means of orthogonal rotation matrices **R**, **S**, and **T** as long as these orthogonal core rotations are compensated for in the component matrices **A**, $\mathbf{B}_1$, and **C**: $\mathbf{G}_a^{(M)^{\mathrm{opt}}} = \mathbf{R}\mathbf{G}_a^{(M)}(\mathbf{T} \otimes \mathbf{S})'$; here, $\mathbf{G}_a^{(M)^{\mathrm{opt}}}$ and $\mathbf{G}_a^{(M)}$ are the mode-A matricized version of $\mathbf{G}^{(M)^{\mathrm{opt}}}$ and $\mathbf{G}^{(M)}$, and $\otimes$ denotes the Kronecker product of two matrices (see Golub & Van Loan, 2013). To identify the optimally rotated core array $\mathbf{G}^{(M)^{\mathrm{opt}}}$, the matrices **R**, **S**, and **T** were sought such that $||\mathbf{G}_a^{(T)} - \mathbf{G}_a^{(M)^{\mathrm{opt}}}||^2$ is minimal, with $\mathbf{G}_a^{(T)}$ being the mode-A matricized version of $\mathbf{G}^{(T)}$ and $||\mathbf{Z}||^2$ denoting the sum of squared elements of matrix **Z**. To this end, an alternating algorithm was used in which **R**, **S**, and **T** were updated successively, with the other matrices kept fixed. It can be shown that these conditional updates boil down to solving a multivariate multiple regression problem with an orthogonality constraint. In particular, minimizing $||\mathbf{G}_a^{(T)} - \mathbf{R}\mathbf{G}_a^{(M)}(\mathbf{T} \otimes \mathbf{S})'||^2$ over **R** with $\mathbf{R}'\mathbf{R} = \mathbf{I}$ and **S** and **T** being fixed can be done by performing a singular value decomposition on $\mathbf{Q} = \mathbf{G}_a^{(M)}(\mathbf{T} \otimes \mathbf{S})'(\mathbf{G}_a^{(T)})'$ and taking $\hat{\mathbf{R}} = \mathbf{V}\mathbf{U}'$, with

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

V and **U** being matrices with the left and right singular vectors of **Q** in its columns. A similar procedure can be followed to update **S** and **T** conditional upon the other matrices. This algorithm leads to a monotonic decreasing loss function which is bounded from below by zero.

The T3–PCA algorithm

Table 2 presents the mean GOR value for **A**, $\mathbf{B}_1$, **C**, and $\mathbf{B}_2$ and the mean BOR value for **G**, computed across all analyses and for each α level separately. From this table we can see that, on average, $\mathbf{B}_1$ and **C** are recovered almost perfectly and that **A** and $\mathbf{B}_2$ are disclosed to a very large extent (i.e., a ϕ value larger than .85 indicates good recovery and a ϕ value exceeding .95 implies excellent recovery; see Lorenzo-Seva & ten Berge, 2006). Also, for the elements of **G** the recovery is good. In particular, the square root of $\overline{\text{BOR}}_{\mathbf{G}}$ averaged across all datasets equals .15, which can be considered quite small given that the true core elements range between -3 and 3. When inspecting the results for the different values of α, it appears that the best recovery results are obtained when $\alpha = .500$. Further, the recovery of all parameters increases when α increases from .050 to .500. For $\alpha > .500$, however, the recovery of **A**, $\mathbf{B}_2$, and **G** decreases when α gets larger, whereas recovery remains excellent for $\mathbf{B}_1$ and **C**.

To study how the recovery performance varies as a function of the five design factors and the α value, we performed a repeated-measures analysis of variance (RMANOVA) for each GOR/BOR measure, with the recovery measure as the dependent variable and the five factors (i.e., between factors) and the α value (i.e., within factor) as independent variables. From these analyses, only considering sizeable effects (i.e., with $\eta^2_G \geq .25$, see Bakeman, 2005), it appears that most of the variation in $\text{GOR}_{\mathbf{A}}$, $\text{GOR}_{\mathbf{B}_2}$, and $\overline{\text{BOR}}_{\mathbf{G}}$ is explained by a very large main effect of the strength of the (extra) distinctive component ($\eta^2_G = .77$, $\eta^2_G = .68$, and $\eta^2_G = .59$ respectively) and a large main effect of the α value ($\eta^2_G = .37$, $\eta^2_G = .55$, and $\eta^2_G = .14$ respectively): recovery performance decreases when the distinctive component becomes more important in terms of explained variance and when one of the data blocks dominates the analysis to a larger extent (i.e., α getting further away from .50), with the latter effect being stronger for the two-way data block (when α goes to 0) than for the three-way data block (when α goes to 1). Further, there exists a considerable interaction ($\eta^2_G = .55$, $\eta^2_G = .72$, and $\eta^2_G = .25$ respectively) between both factors. In particular, as one can see in Figure 2, when the distinctive component is not very strong (i.e., 5% and 15% of explained variance), although still being excellent, recovery decreases a bit when α increases. Moreover, the difference in recovery between datasets with a small versus intermediate strong distinctive component increases when α gets larger. However, when there is a strong distinctive component (i.e., explaining 30%), recovery is optimal for intermediate α values and decreases when α goes to 0 (i.e., all weight on the two-way data block) and 1 (i.e., only the information in the three-way data block has an influence on the analysis), with the decrease being more pronounced for the former than for the latter. For $\text{GOR}_{\mathbf{B}_1}$, because of limited variability, there are no sizeable effects, whereas no interesting effects are observed for $\text{GOR}_{\mathbf{C}}$. In particular, for $\text{GOR}_{\mathbf{C}}$ an effect of α ($\eta^2_G = .27$) is observed, which is further involved in an interaction between α and the strength of the distinctive component ($\eta^2_G = .43$) and a three-way interaction between α, the strength of the distinctive component and the rank of the true model ($\eta^2_G = .30$). This three-way interaction is mainly explained by $\text{GOR}_{\mathbf{C}}$ being around .82 in a few specific conditions (i.e., strong distinctive component and the true model having a larger rank and the two-way data block strongly dominating the analysis, $\alpha \leq .10$) and $\text{GOR}_{\mathbf{C}}$ being larger than .95 in all other conditions (even $\geq .99$ in most conditions).

Recovery difference between simultaneous T3–PCA and sequential strategy

To compare the simultaneous T3–PCA strategy to the sequential strategy (Section 4.2), we computed, for each recovery measure and for each analysis, the difference in recovery value between the simultaneous and the sequential strategy (e.g., for **A**, $\text{GOR}^{\text{diff}}_{\mathbf{A}} = \text{GOR}^{\text{T3-PCA}}_{\mathbf{A}} - \text{GOR}^{\text{sequential}}_{\mathbf{A}}$, and for **G**,

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Mean recovery value for component matrices (GOR) and core array (BOR), computed across all datasets and analyses and for each α level separately.

α Level	GOR_A	GOR_{B_1}	GOR_C	GOR_{B_2}	$\overline{BOR}_G$
.050	.84	.99	.96	.83	.08
.100	.84	.99	.96	.83	.07
.300	.94	.99	.99	.92	.02
.500	.98	.99	.99	.98	.02
.600	.96	.99	.99	.97	.03
.700	.94	.99	.99	.97	.04
.800	.91	.99	.99	.96	.05
.900	.89	.99	.99	.95	.06
.950	.88	.99	.99	.95	.06
.999	.87	.99	.99	.94	.06
Overall	.91	.99	.99	.93	.05

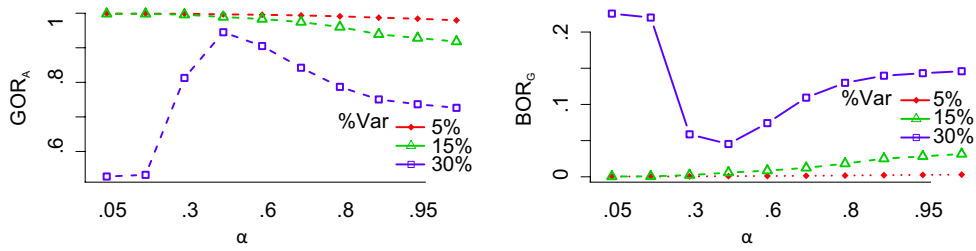


FIGURE 2 Mean GOR_A (left-hand panel) and $\overline{BOR}_G$ (right-hand panel) as a function of α and the strength of the distinctive component.

$\overline{BOR}_G^{diff} = \overline{BOR}_G^{T3-PCA} - \overline{BOR}_G^{sequential}$). Note that the sequential strategy is outperformed by the simultaneous one when GOR^{diff} is positive and BOR^{diff} is negative.

The mean recovery difference value (with 95% confidence intervals in parentheses) for all recovery measures is displayed in Table 3, computed across all analyses and for each level of α separately. From this table it can be seen that the underlying parameters are better disclosed by T3–PCA than by the sequential strategy for intermediate (i.e., a more or less balanced influence of both data blocks on the analysis) and larger values of α. Although mean difference values are rather small, the confidence intervals (for intermediate and larger α values) suggest that these differences are larger than zero and may be large. Moreover, from Table A1, which displays for which percentage of the simulated datasets T3–PCA outperforms the sequential strategy in terms of recovery (i.e., $GOR^{diff} > 0$ and $BOR^{diff} < 0$), it appears that for intermediate and large α values T3–PCA almost always – in more than 90% of the cases – better recovers **A**, $\mathbf{B}_2$, and **G**than the sequential strategy. The largest recovery differences between both strategies are observed when both data blocks have an equal amount of influence on the analysis (i.e., $\alpha = .50$). This is especially true for the recovery of the common mode **A** and to a lesser extent for $\mathbf{B}_2$ and **G** (i.e., T3–PCA outperforming the sequential strategy in 95% or more of the datasets; see Table A1), whereas no recovery differences are encountered for $\mathbf{B}_1$ and **C**. Note that the sequential strategy outperforms T3–PCA – although confidence intervals suggest that differences are not larger than zero – when the information in the two-way data block dominates the analysis (i.e., $\alpha \leq .10$), whereas both strategies perform equally well when the three-way data block dominates the analysis (i.e., $\alpha \geq .95$). Table A1,

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Mean recovery difference between simultaneous T3–PCA and sequential strategy for component matrices (GOR^{diff}) and core array (BOR^{diff}) with 95% confidence intervals in parentheses, computed across all datasets and analyses and for each level of α separately.

α Level	GOR_{A}^{diff}	$GOR_{B_1}^{diff}$	GOR_{C}^{diff}	$GOR_{B_2}^{diff}$	$\overline{BOR}_{G}^{diff}$
.050	−.03 [−.36, .25]	−.01 [−.04, .00]	−.03 [−.25, .00]	−.12 [−.46, .06]	.02 [−.14, .23]
.100	−.03 [−.36, .25]	−.00 [−.03, .00]	−.03 [−.24, .00]	−.11 [−.46, .06]	.01 [−.14, .22]
.300	.06 [−.19, .35]	−.00 [−.01, .00]	−.01 [−.07, .00]	−.02 [−.37, .14]	−.04 [−.22, .04]
.500	.10 [.00, .43]	.00 [.00, .00]	−.00 [−.01, .00]	.03 [.00, .29]	−.04 [−.24, .00]
.600	.09 [.00, .40]	.00 [.00, .00]	−.00 [−.01, .00]	.03 [.00, .27]	−.03 [−.19, .00]
.700	.06 [.00, .33]	.00 [.00, .00]	.00 [.00, .00]	.03 [.00, .22]	−.02 [−.12, .00]
.800	.04 [.00, .19]	.00 [.00, .00]	.00 [.00, .00]	.02 [.00, .14]	−.01 [−.07, .01]
.900	.02 [.00, .08]	.00 [.00, .00]	.00 [.00, .00]	.01 [.00, .05]	−.00 [−.03, .00]
.950	.01 [.00, .04]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .03]	−.00 [−.01, .00]
.999	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]
Overall	.03 [−.26, .33]	.00 [.00, .00]	−.01 [−.15, .00]	−.01 [−.43, .16]	−.01 [−.15, .11]

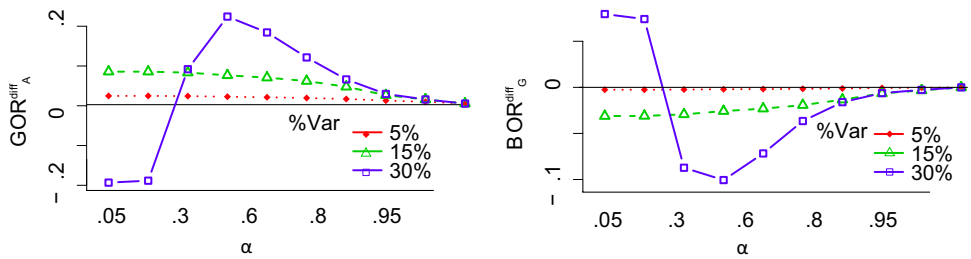


FIGURE 3 Mean GOR_{A}^{diff} (left-hand panel) and $\overline{BOR}_{G}^{diff}$ (right-hand panel) as a function of α and the strength of the distinctive component.

however, shows that even when $\alpha \leq .10$, T3–PCA outperforms the sequential strategy in terms of $\mathbf{A}$, $\mathbf{B}_2$, and $\mathbf{G}$ recovery in 60%–70% of the datasets.

To determine the influence of α and the design factors on the recovery difference between the simultaneous and the sequential strategy, we performed five RMANOVAs with the recovery difference as the dependent variable and the five factors and the α value as independent variables. When only focusing on sizeable effects (i.e., $\eta_G^2 \geq .25$; Bakeman, 2005), it appears that T3–PCA outperforms the sequential strategy in terms of disclosing the parameters in $\mathbf{A}$, $\mathbf{B}_2$, and $\mathbf{G}$ to a larger extent when the information in both data blocks becomes (more or less) equally weighted ($\overline{\eta}_G^2 = .31$, .51, and .17 respectively). As can be seen in Table 3, the largest GOR_{A}^{diff} and $GOR_{B_2}^{diff}$ (and smallest $\overline{BOR}_{G}^{diff}$) are encountered for intermediate values of α (i.e., around .50). Moreover, the recovery difference decreases when one data block starts to dominate the analysis, with this effect being stronger for the two-way data matrix (i.e., with even the sequential strategy outperforming T3–PCA when $\alpha \leq .10$) than for the three-way data array. This main effect, however, is qualified by an interaction between the strength of the distinctive component and α (η_G^2 is .49, .68, and .30 respectively). From Figure 3, one can see that T3–PCA always uncovers the common mode better than the sequential strategy when $\alpha \geq .30$, with the difference between both strategies being larger when the distinctive component is more influential. For low values of α, however, T3–PCA only outperforms the sequential strategy when the distinctive component is not too strong (and this to the same extent for all low α values), whereas the sequential strategy is clearly better when the distinctive

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Mean recovery difference between simultaneous T3–PCA and simultaneous LMPCA strategy for component matrices (GOR^{diff}) with 95% confidence intervals in parentheses, computed overall across the datasets with rank ($P = 2$, $Q = 2$, $R = 2$) and for each level of α separately.

α Level	GOR^{diff}_{A}	$GOR^{diff}_{B_1}$	GOR^{diff}_{C}	$GOR^{diff}_{B_2}$
.050	.14 [−.01, .52]	.29 [.01, .82]	.31 [.01, .81]	.24 [.01, .51]
.100	.15 [−.03, .52]	.30 [.01, .80]	.30 [.01, .78]	.25 [.01, .51]
.300	.18 [−.06, .52]	.30 [.01, .78]	.30 [.01, .76]	.23 [.01, .54]
.500	.21 [.01, .53]	.30 [.01, .75]	.31 [.01, .75]	.21 [.01, .53]
.600	.20 [.01, .51]	.31 [.01, .74]	.32 [.01, .76]	.20 [.01, .51]
.700	.20 [.00, .48]	.30 [.01, .74]	.31 [.01, .75]	.21 [.01, .50]
.800	.18 [.00, .48]	.29 [.01, .70]	.31 [.01, .74]	.21 [.01, .50]
.900	.18 [−.02, .47]	.29 [.01, .69]	.31 [.01, .74]	.21 [.01, .50]
.950	.17 [−.03, .47]	.29 [.01, .69]	.31 [.01, .72]	.20 [.01, .50]
.999	.17 [−.04, .46]	.29 [.01, .68]	.31 [.01, .73]	.20 [.01, .50]
Overall	.18 [−.02, .49]	.30 [.01, .74]	.31 [.01, .76]	.22 [.01, .51]

component is strong (and $\alpha \leq .10$). No interesting effects are observed for $GOR^{diff}_{B_1}$ and GOR^{diff}_{C}. No sizeable effects were observed for $GOR^{diff}_{B_1}$, and in most conditions GOR^{diff}_{C} equals almost zero. Only in some (limited number of) specific conditions (i.e., a strong distinctive component and a larger rank for the true model and $\alpha \leq .10$) GOR^{diff}_{C} is negative.

Recovery difference between T3–PCA and LMPCA strategy

To compare T3–PCA to LMPCA, the same GOR^{diff} measures as above but now between the T3–PCA and the LMPCA recovery were computed for the component matrices. Note that T3–PCA outperforms LMPCA when GOR^{diff} is positive and that no BOR^{diff}_{G} measure is computed as LMPCA has no (free) core array. The mean recovery difference values (with 95% confidence intervals between parentheses) for all recovery measures are displayed in Table 4, computed across all datasets with rank ($P = 2$, $Q = 2$, $R = 2$) overall and for each level of α separately. Overall, T3–PCA largely outperforms LMPCA (i.e., average GOR^{diff} value of .25), with this effect being the strongest for the non-common modes $\mathbf{B}_1$ and $\mathbf{C}$ of $\mathbf{D}^1$ (average GOR^{diff} around .30) and to a lesser extent – but still substantially large – for the common mode $\mathbf{A}$ (.18) and the non-common mode $\mathbf{B}_2$ of $\mathbf{D}^2$ (.22). This result is expected as the difference between T3–PCA and LMPCA lies in the submodel used for $\mathbf{D}^1$ and, thus, mostly affects the component matrices (solely) underlying $\mathbf{D}^1$. The common component matrix $\mathbf{A}$ is less affected as also information from $\mathbf{D}^2$ (and $\mathbf{D}^2$ is correctly modelled with PCA) is used to estimate $\mathbf{A}$. Interestingly, using a too restrictive submodel for $\mathbf{D}^1$ also substantially affects the component matrix $\mathbf{B}_2$, which solely underlies $\mathbf{D}^2$ (with, again, $\mathbf{D}^2$ being correctly modelled with PCA). This pattern of results is very similar across different α levels, with the results for intermediate α levels being more pronounced. Interestingly, for GOR^{diff}_{A}, large α levels ($\alpha = .999$) yield larger recovery differences between T3–PCA and LMPCA than small α levels ($\alpha = .050$), whereas the opposite is encountered for $GOR^{diff}_{B_2}$ (and ends up being symmetrical around $\alpha = .50$ for $GOR^{diff}_{B_1}$ and GOR^{diff}_{C}). Finally, the confidence intervals support the conclusion that T3–PCA outperforms LMPCA, especially for intermediate α levels (i.e., only for GOR^{diff}_{A} and small and large α levels, the lower boundary of the confidence interval is negative). Moreover, in Table A2, which shows the percentage of datasets for which T3–PCA outperforms LMPCA, one can see that T3–PCA almost always (i.e., 99% or more) better recovers the component matrices than LMPCA – and for each α level. Only for $\mathbf{A}$ and small α levels is the effect slightly smaller (i.e., T3–PCA outperforms LMPCA in 95% of cases for $\mathbf{A}$ in general, which decreases to 85% for $\alpha \leq .10$).

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Evaluating procedure for testing existence of common components

The analysis reveals that the observed similarity between $\mathbf{A}_{3D}$ and $\mathbf{A}_{2D}$ is extremely large for each of the 60 tested datasets compared to the similarity values across the 100 permutations. In particular, in 48 of the 60 datasets (80%), for all permuted data the similarity value was smaller than the observed similarity value for the data. On average, across the 60 datasets, .76 permuted datasets had a similarity value larger than the observed data, with this number ranging between 0 and 7. In 54 datasets (90%), less than five permuted datasets have a similarity value larger than the observed similarity. Regarding the sensitivity of the proposed procedure, it appears that, for each dataset, the similarity value from the observed data is smaller than the similarity value from each of the 100 permuted datasets. It can be concluded that this procedure is a valid tool to decide whether it makes sense to analyse a coupled dataset with the T3–PCA model.

5 | ILLUSTRATIVE APPLICATION

A challenging question in behavioural research pertains to the relation between how persons react in a particular situation and which dispositions (sometimes called "traits") they exhibit (Van Coillie et al., 2006). To get some insight into this relation, we will perform a T3–PCA analysis on a coupled dataset regarding anxiety in which self-reported (anxiety-related) behaviours across stressful situations, yielding a person-specific behavioural signature, are coupled with traditional dispositions (i.e., self-reported traits). The goal of the analysis is to disclose how the individual difference structure in behavioural signatures is related to individual differences in dispositions.

To this end, we will analyse data from 128 participants for which at the same time information regarding their behavioural profiles and their dispositions is available. The behaviour profiles (i.e., three-way data) pertain to the scores of the 128 participants on the S-R Inventory of Anxiousness (Endler & Hunt, 1968), which measures 14 anxiety-related responses (e.g., feeling paralysed and wanting to avoid the current situation, see matrix $\mathbf{B}_1$ in Table 5) for 11 different stressful situations (e.g., going for a job interview and being alone at night in the woods; see matrix $\mathbf{C}$ in Table 6), using a 5-point scale ranging from exhibiting the response *not at all* to *very much* (for more information on the behavioural data, see Kiers & Van Mechelen, 2001; Van Mechelen & Kiers, 1999). Further, two personality questionnaires measuring personal dispositions were administered to each participant: (1) the Cognitive-Affective processing System questionnaire (CAPS; Mischel & Shoda, 1995) and (2) the State-Trait Anxiety Inventory (STAI), which is a commonly used measure of anxiety (Spielberger et al., 1983). The CAPS-questionnaire consists of 63 items (measured on a 7-point scale) that are intended to measure various types of cognitive-affective variables (e.g., encodings and expectancies) in regard to different types of stress (e.g., social evaluation and physical danger). The STAI aims at measuring state anxiety (i.e., anxiety about an event) and trait anxiety (i.e., anxiety level as a personal characteristic) by means of 40 items that have to be rated on a 4-point Likert scale. For our analysis, we will rely on the Dutch translation of the STAI (Van der Ploeg, 2000) and use only the 20 items relating to trait anxiety. By combining for all 128 participants the selected items from the CAPS and STAI (i.e., 83 items in total), a (two-way) person-by-disposition data matrix is obtained that is coupled to the (three-way) behavioural signature data block by means of the common person mode. As research has shown that quantitative variables with five (and even four) different possible values (e.g., Likert scales), as long as there are no strong signs of absence of normality, can adequately be treated as continuous variables in a factor5or component analysis (Dolan, 1994), we believe that it is appropriate to couple the three-way behaviour profile data to the two-way dispositional data, at least

[5]Research in the context of factor analysis and SEM showed that normal theory maximum likelihood, which assumes continuous variables, works very well when the data consist of categorical variables with six to seven categories and even with five categories when the sample size is not too large (Rhemtulla et al., 2012). Moreover, estimators specially designed for analysing categorical data (e.g., WLSMV) only clearly perform better than maximum likelihood when the categorical variables have two or three categories (Beauducel & Herzberg, 2006).

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Component values for response scales (matrix $\mathbf{B}_1$) resulting from simultaneous T3–PCA with $\alpha = .5$ (left part), sequential (middle part), and LMPCA (right part) strategies with $P = 6$ person/disposition, $Q = 2$ response, and $R = 3$ situation components for the coupled dataset (with values exceeding .20 in absolute values indicated in bold).

	T3–PCA		Sequential		LMPCA	
Response	**Component 1**	**Component 2**	**Component 1**	**Component 2**	**Component 1**	**Component 2**
Need to urinate frequently	**.40**	−.18	**.39**	−.18	**−.31**	**.30**
Mouth gets dry	**.39**	−.12	**.38**	−.10	**−.26**	**.31**
Need to defecate	**.36**	−.18	**.34**	−.14	**−.30**	**.26**
Feel paralysed	**.35**	.05	**.37**	.03	−.09	**.35**
Full feeling in stomach	**.34**	.01	**.32**	.02	−.11	**.32**
Perspire	**.29**	.13	**.29**	.12	.02	**.31**
Feel nausea	**.28**	.11	**.30**	.05	−.01	**.30**
Emotions disrupt action	**.24**	**.22**	**.25**	**.20**	.12	**.30**
Heart beats faster	**.23**	.09	**.22**	.12	−.01	**.25**
Feel exhilarated and thrilled	−.08	**.45**	−.08	**.45**	**.45**	.09
Enjoy a challenge	−.04	**.44**	−.05	**.47**	**.42**	.13
Seek experiences like this	−.04	**.43**	−.05	**.47**	**.42**	.13
Do not want to avoid situation	.09	**.40**	.10	**.38**	**.33**	**.24**
Uneasy feeling	.18	**.30**	.19	**.28**	**.21**	**.28**

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Component values for situations (matrix **C**) resulting from simultaneous T3–PCA with $\alpha = .5$ (left part), sequential (middle part), and LMPCA (rigth part) strategies with $P = 6$ person/disposition, $Q = 2$ response, and $R = 3$ situation components for the coupled dataset (with values exceeding .20 in absolute value indicated in bold).

	T3–PCA			Sequential			LMPCA		
Situation	**Component 1**	**Component 2**	**Component 3**	**Component 1**	**Component 2**	**Component 3**	**Component 1**	**Component 2**	**Component 3**
Final exam	**.49**	−.02	**.22**	**.43**	−.05	**.28**	**.28**	**.29**	**.35**
Speech before large group	**.48**	−.08	−.08	**.53**	−.01	−.03	.04	**.44**	**.23**
Job interview	**.45**	−.01	−.12	**.49**	−.02	.04	−.03	**.40**	**.24**
Match in front of audience	**.33**	.18	**−.24**	**.37**	.16	−.02	**−.24**	**.25**	**.28**
Consult counseling bureau	**.27**	.00	.06	**.22**	−.06	**.22**	.10	.18	.19
Ledge high on mountainside	−.14	**.77**	.14	−.13	**.75**	.11	**−.23**	**−.56**	**.52**
Sail boat on rough sea	.01	**.53**	−.10	.02	**.56**	−.01	**−.32**	**−.23**	**.37**
New date	.18	.16	**−.50**	**.26**	**.22**	**−.24**	**−.48**	**.25**	.11
Auto trip	.11	.17	**−.35**	.14	.19	−.11	**−.36**	.13	.10
Alone in woods at night	**.28**	.16	**.67**	−.09	.03	**.88**	**.57**	−.13	**.47**
Psychological experiment	.03	.05	.14	.02	.01	.12	.11	−.06	.10

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

after suitable preprocessing of both data blocks in order to obtain the comparability/commensurability of data entries within and between data blocks (see below).

When analysing behavioural data collected in various situations (i.e., three-way data), it is important to (1) account for unknown neutral points for the response scales and (2) remove arbitrary differences in the use of the range of scales. Therefore, the three-way behavioural data were preprocessed by the following two-step procedure (for more information regarding preprocessing of three-way data, see Kiers & Van Mechelen, 2001; Kroonenberg, 1983, 2008; Smilde et al., 2004). First, for each combination of a situation and a response, the scores were mean-centred across participants (i.e., accounting for neutral points). Next, the scores for each response were normalized (over situations and participants) by setting the variance of these scores equal to one. In particular, each element $d^1_{ijk} - d^1_{.jk}$ was divided by $\sigma_j = \sqrt{\frac{\sum_i \sum_k (d^1_{ijk} - d^1_{.jk})^2}{IK}}$ (with $d^1_{.jk}$ being the mean score across persons for response j in situation k), resulting in each response scale influencing the analysis to the same extent. The two-way disposition data were preprocessed by standardizing (i.e., mean of zero and variance of one) the individual items.

T3–PCA analyses were performed on the preprocessed coupled dataset for all valid ranks (Section 3.3) between (1, 1, 1) and (6, 6, 6). Since in the simulation study equally weighting both data blocks in the analysis appeared to be the best option, we used $\alpha = .50$. To select among the many estimated models a solution that optimally balanced model fit and model complexity, we applied the CHull model selection procedure (Ceulemans & Kiers, 2006; Wilderjans et al., 2013) with the sum of squared residuals as (mis) fit value and the total number of fitted components (i.e., $P + Q + R$) as complexity value. Based on the CHull plot, which is displayed in Figure 4, we decided to retain the (6, 2, 3) solution, which has six participant/disposition (matrix $\mathbf{A}/\mathbf{B}_2$), two behavioural (matrix $\mathbf{B}_1$), and three situation components (matrix $\mathbf{C}$) and $\alpha = .50$. When using other values of α, it appears that for intermediate α values, CHull always selects the (6, 2, 3) solution, whereas the (6, 3, 2) and (6, 4, 4) solution is selected for low ($\alpha \leq .10$) and high ($\alpha \geq .95$) values of α respectively. Further, to compare the simultaneous T3–PCA to the sequential strategy, the sequential strategy in which the three-way data are fitted first was also applied to the dataset, adopting the selected (6, 2, 3) model rank and an α value of .50. Next, to compare T3–PCA to the (simultaneous) LMPCA strategy, LMPCA with 6, 2, and 3 components and $\alpha = .50$ was fitted to the dataset. Finally, the procedure for detecting common components in the coupled dataset (Section 3.1) was applied to the dataset, using only rank (6, 3, 2) and adopting, due to computational limitations, only 100 permutations.

Regarding computation time, the T3–PCA analysis for the (6, 2, 3) solution of the illustrative dataset took on average (across all α values) 132 s. In particular, the T3–PCA analysis took 98, 93, 105, 111, 135, 373, 130, 76, and 72 s for $\alpha = .001, .10, .30, .40, .50, .60, .70, .90$, and .999 respectively. Taking all analyses (i.e., across all valid ranks and α values) into account, a T3–PCA analysis takes on average 151 s, with a minimum of 24 s for the (1, 1, 1) solution and a maximum of 239 s for the (5, 5, 5) solution.

To obtain a solution that is more easy to interpret, we postprocessed the retained solution as follows. First, to obtain a simple structure (Cattell, 1978; Thurstone, 1947), the response (matrix $\mathbf{B}_1$), situation (matrix $\mathbf{C}$), and disposition (matrix $\mathbf{B}_2$) component matrices were orthogonally rotated by means of varimax (Kaiser, 1958). Next, the person component matrix $\mathbf{A}$ and the core array $\mathbf{G}$ were appropriately counterrotated (Section 3.4). Finally, the disposition component matrix $\mathbf{B}_2$ was scaled (i.e., by dividing each entry by the square root of the number of persons), and this scaling was compensated for in the person component matrix $\mathbf{A}$ and the core array $\mathbf{G}$. The postprocessed component matrices for the response scales ($\mathbf{B}_1$), situations ($\mathbf{C}$), and (part of the) dispositions ($\mathbf{B}_2$) are given in the left part of Tables 5, 6, and 8 (with descriptions for the disposition labels in Table 9). Note that because of its excessive length, for each disposition component (in matrix $\mathbf{B}_2$), only the highest loading dispositions are presented (note also that the dispositions are grouped per component).[6] The postprocessed core array $\mathbf{G}$ is presented in the upper part of Table 7 and part of the person

[6]The full table can be obtained from the authors upon request.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

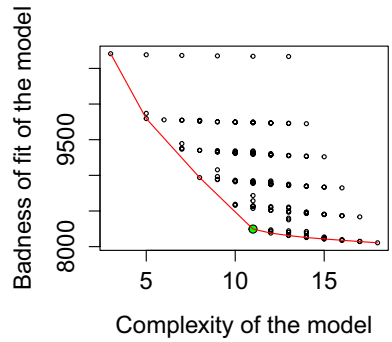


FIGURE 4 CHull plot of total number of extracted components (i.e., complexity measure) versus the loss function value (i.e., badness of fit measure) for all estimated T3–PCA models with $\alpha = .50$ for the coupled dataset. Note that the models located on the convex hull are connected by a red line and that the selected solution is indicated by a green circle.

component matrix **A**, which is not shown in full because of its excessive length (see further), can be found in the left part of Table 10. Note that the core array **G** in Table 7 is presented as a matrix in which the rows correspond with the person components and the columns with combinations of a response and a situation component.

Inspection of the response component matrix $\mathbf{B}_1$ (Table 5) reveals that the first component refers to a set of bodily responses that are a collection of more autonomic physiological reactions (e.g., perspire, mouth gets dry, and need to urinate/defecate), on the one hand, and responses related to illness (e.g., feel nausea and full feeling in stomach), on the other. The second response component clearly refers to an approach-avoidance response dimension (e.g., not wanting to avoid a situation and feeling exhilarated and thrilled). Also the situation components (matrix **C**) can be given a straightforward interpretation: The first component represents situations that involve a (threatening) evaluation by others (e.g., final exam and speech before a large group), whereas the second component pertains to inanimate dangers (i.e., sailing on a rough sea and being on the ledge of a mountainside). Finally, the third situation component comprises adrenaline rushes that can be very threatening (e.g., being alone in the woods) or very exciting (e.g., going on a new date and taking a road trip). The person components (matrix **A**) can be considered as underlying individual difference dimensions, each being associated with a distinct prototypical behavioural signature (which is reflected in the core array **G**). The person scores (shown for prototypical persons in Table 10) indicate the position of each person on these dimensions. When discarding small person scores, it appears that most people load high on one or two dimensions. The core array **G** summarizes the main interactions in the data. Only looking at the largest (in absolute value) core values, it appears that people with a (high) positive score on the second person component can be treated as risk seekers since, although they experience extreme bodily sensations, they more actively seek out situations involving inanimate dangers than the average person. People with a positive score on the sixth person component show no pronounced reactions to the presented situations as all associated core values are close to zero. People associated with the other four person components all exhibit stronger (i.e., positive score) or weaker (i.e., negative score) bodily sensations in evaluative situations than an average person. Some of these person dimensions also relate to displaying stronger/weaker bodily sensations in contexts involving inanimate dangers (i.e., person component 3 and 5), whereas some (other) person dimensions seem to be related to actively seeking out situations involving an evaluation (i.e., person component 1 and 5) to a larger or smaller extent (than the average person in the study). Regarding adrenaline-related situations, no significant differences among the person dimensions seem to exist.

To link the individual differences in behavioural signatures to differences in dispositions, we must inspect the disposition components (matrix $\mathbf{B}_2$). The first disposition component represents a tendency to lose control in and not deal well with situations involving a failure, a threat, or stress. The second component focuses on the physical damage a person can suffer in a situation (e.g., in a physically

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 Core array **G** resulting from simultaneous T3–PCA with $\alpha = .5$ (upper part) and sequential (lower part) strategy with $P = 6$ person/disposition, $Q = 2$ response, and $R = 3$ situation components for the coupled dataset (with values exceeding 1 in absolute value indicated in bold and values exceeding 2 in absolute value being underlined).

	Situation component 1		Situation component 2		Situation component 3	
Person component	Response component 1	Response component 2	Response component 1	Response component 2	Response component 1	Response component 2
1	**1.49**	**1.49**	.71	.89	.77	.57
2	.09	−.17	**1.92**	**2.09**	.20	.54
3	**3.29**	−.16	**2.01**	.23	.01	−.17
4	**1.40**	.96	.82	−.27	.20	.68
5	**2.73**	**1.69**	**1.27**	.66	.20	.18
6	.38	−.04	−.04	−.22	.54	.62
1	**2.96**	**2.33**	**1.00**	.76	**1.80**	.70
2	**1.21**	−.55	**2.24**	.72	.65	−.82
3	.58	.18	**1.63**	**1.88**	**1.36**	**1.81**
4	**1.29**	−.82	.15	−.90	**2.06**	.72
5	**2.16**	.47	**1.18**	−.65	.35	.78
6	**1.11**	−.78	.05	.65	−.24	−.69

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 8 Component values for selected dispositions (matrix $\mathbf{B}_2$) resulting from simultaneous T3–PCA with $\alpha = .5$ (left part), sequential (middle part), and LMPCA (right part) strategy with $P = 6$ person/disposition, $Q = 2$ response, and $R = 3$ situation components for coupled dataset (with values exceeding .40 in absolute value indicated in bold).

	T3–PCA						Sequential						LMPCA					
Disp	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
D14	**.73**	−.00	.03	.11	.28	−.18	**.53**	−.04	.22	.02	.06	−.01	.27	**.52**	.36	−.15	−.38	−.15
D25	**.42**	.00	−.01	.05	**.56**	−.01	**.40**	−.10	.33	.10	.20	.02	.28	.36	.38	−.31	−.13	.16
D31	**−.56**	−.32	−.06	−.02	−.25	−.07	−.39	.00	**−.41**	.00	−.07	.02	−.24	−.18	**−.53**	.17	.16	.23
D43	**−.77**	−.03	.09	.03	−.24	.30	**−.43**	.07	−.26	.10	−.01	.03	−.26	**−.44**	−.29	.24	**.56**	.19
D82	**−.41**	.09	−.22	−.04	−.09	**.41**	−.34	−.11	−.00	−.02	.01	−.14	−.19	**−.47**	.08	.15	.20	.26
D8	−.02	**.42**	−.14	.21	.31	−.18	.10	.02	.38	−.08	.14	−.09	**.57**	−.16	.13	−.09	.01	.02
D40	.08	**.40**	−.20	.26	.33	−.08	.15	.03	.32	−.13	.16	−.12	**.55**	−.12	.25	−.04	−.05	.07
D50	−.10	−.03	**−.47**	.21	.03	−.09	−.12	−.27	−.02	−.20	.06	−.06	.22	−.17	−.08	.19	−.25	.31
D52	−.09	.07	**−.42**	.22	.01	−.15	−.13	−.07	−.01	−.22	.03	−.08	.29	−.20	−.09	.18	−.23	.21
D3	.02	−.05	−.14	**.50**	−.21	−.07	−.07	−.00	−.05	.01	.02	−.30	.20	.08	−.08	**.51**	−.08	.05
D26	−.05	.02	−.11	**.45**	−.29	−.02	−.24	.06	−.05	.04	.07	−.05	.15	−.02	−.09	**.52**	−.02	−.02
D53	−.02	.10	−.06	**.70**	−.12	−.18	.07	.00	.02	−.12	.22	−.12	**.48**	.11	−.09	**.55**	.04	.00
D5	.22	.00	.01	.08	**.56**	**−.41**	**.43**	−.03	.18	−.11	.23	−.09	**.49**	.35	.00	−.38	−.13	.12
D13	.05	.19	−.07	.11	**.62**	−.32	.30	.08	.22	−.08	.23	−.04	**.60**	.13	.06	−.38	−.02	.18
D17	.29	.09	−.03	−.02	**.66**	−.25	**.51**	.04	.24	−.07	.09	−.08	**.45**	.28	.20	**−.48**	−.14	.14
D37	.18	−.07	−.06	−.11	**.74**	−.32	**.50**	.03	.08	−.10	.10	−.11	.39	.31	.06	**−.60**	−.14	.28
D67	.24	−.03	.04	.14	**.47**	−.37	.33	.10	.10	−.01	.23	−.02	**.45**	.38	.00	−.27	−.12	.09
D68	.04	.03	−.04	.03	**.44**	**−.57**	.34	−.02	.13	−.12	−.00	−.03	**.52**	.20	−.24	−.37	−.14	.07
D73	−.08	−.01	−.03	.03	−.18	**.75**	−.23	−.01	−.04	.21	−.02	−.04	**−.44**	−.22	**.43**	.31	.24	.15
D79	−.11	−.01	.09	.05	−.20	**.77**	−.18	.03	−.05	.20	.05	−.04	**−.46**	−.18	**.41**	.32	.35	.11

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 9 Labels for (selected) dispositions (matrix $\mathbf{B}_2$) used in coupled dataset.

Label	Description
D14	I find it hard to stay calm in a stressful and uncertain situation
D25	When I fail, I am completely thrown off balance
D31	I can cope with a physically threatening situation
D43	I can deal well with stress
D82	I am a calm person
D8	In a physically threatening situation, I readily expect that my health will be harmed
D40	I easily experience a situation in which I may suffer physical injury as risky
D50	I think it's important to live without stress
D52	I attach great importance to having as little stress as possible
D3	I think it is important to be healthy
D26	I think it is important to keep my body fit
D53	I tend to pay much attention to physiological changes in my body
D5	I often expect that things will end badly
D13	In an unclear situation, I immediately prepare myself for the worst
D17	When I am obliged to do something, I easily think that I am not able to do it
D37	In a performance-related task, I easily expect that I will fail
D67	I can cope very difficultly with a setback
D68	I feel I have shortcomings in almost everything
D73	I am happy
D79	I am satisfied

threatening situation where there is an expection of harm to one's health). The third component is associated with a disposition to have a life with no stress (e.g., "I think it is important to live without stress"). The fourth component is related to a tendency to take care of one own's body and health (i.e., "I find it irresponsible to neglect my health"). The fifth component comprises a disposition to have negative expectations (i.e., negative forecasting) in various situations (e.g., "I always prepare myself for the worst"). Finally, the sixth component is a tendency to feel good and to be happy (e.g., I am happy). As each disposition component is only linked to a single person component, it can be derived that people who do not show pronounced responses feel good in general. Further, negative forecasting goes hand in hand with stronger bodily sensations in evaluative and inanimate danger situations and more actively avoiding such situations. Finally, individual differences in the tendency to lose control are (positively) related to individual differences in exhibiting bodily sensations in evaluative situations.

When comparing the T3–PCA results with those obtained with the sequential strategy, it appears that the simultaneous T3–PCA model fits the data better (loss value of 8246.36) than the sequential strategy (loss of 9799.94). Moreover, this comparison also nicely illustrates the interpretational advantage of T3–PCA over the sequential analysis (see middle part of Tables 5–10). As could be expected based on the simulation results, only small differences between both solutions are encountered for the response and situation components (i.e., $\mathbf{B}_1$ and $\mathbf{C}$), whereas large(r) differences are obtained for the core array and the person and the disposition components (i.e., $\mathbf{G}$, $\mathbf{A}$ and $\mathbf{B}_2$). Note that the latter differences are the most important ones as they reveal the systematic links between the processes underlying both data blocks. Regarding the core array $\mathbf{G}$, the sequential strategy focuses more on differences between persons in the adrenaline-seeking situations, whereas T3–PCA ignores these differences as they cannot be linked in a systematic way to differences in dispositions. In particular, for the sequential solution, no clear set of dispositions is associated with the first, third, and fourth person components, which exhibit decreased (i.e., negative score) or pronounced (i.e., positive score) responses to high-adrenaline situations. Finally, T3–PCA more clearly discloses the important differences between the persons than the sequential strategy.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 10 Component values for selected persons (matrix **A**), resulting from the simultaneous T3–PCA with $\alpha = .5$ (left part), the sequential (middle part), and LMPCA (right part) strategies with $P = 6$ person/disposition, $Q = 2$ response, and $R = 3$ situation components for coupled dataset (with person scores larger than 2.5 in absolute value indicated in bold).

	T3–PCA						Sequential						LMPCA					
Person	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6
6	−.90	.16	**2.80**	1.40	2.32	.96	1.39	−2.27	−.16	−2.09	1.46	−.78	1.51	2.13	.39	−.05	**3.16**	.19
11	.84	−.93	1.19	1.49	**−3.06**	−.99	−1.34	−.53	1.40	−.42	1.11	−.76	−1.43	−.33	−1.59	**2.89**	.13	−1.59
13	**2.60**	−.46	.12	−.02	.47	.28	1.22	−.22	−.72	−.94	−.41	.68	−1.51	1.76	1.03	.22	−.81	−.34
28	**−3.02**	−.81	−.78	.76	−.46	−.10	−1.89	−.78	1.68	−.59	−1.03	.33	1.51	−1.54	−1.87	.38	.55	1.62
68	−.64	−.81	−.00	**2.53**	−.66	.34	.61	−.27	1.64	1.57	1.52	−.10	.69	.24	−.76	2.41	.47	.94
69	−.54	−1.12	2.13	**−2.71**	−1.07	−.19	−.94	−.43	.97	.05	.11	−2.20	−2.13	−.78	−1.50	−1.77	1.70	−1.08
75	.50	**2.86**	1.03	.88	−1.78	.14	−.47	**−3.37**	−1.55	1.54	−.08	1.49	.50	−1.32	1.68	1.64	.66	−2.35
94	−.24	**2.84**	.35	−1.62	1.37	−.37	−.40	−.32	**−2.86**	−.07	.15	−1.82	1.40	−.51	1.78	−2.24	.21	−1.55
100	−.87	2.26	**2.75**	1.26	.64	−.03	.14	**−3.78**	−1.37	−.54	1.27	−.01	1.98	.43	.87	.50	**2.60**	−1.87
126	**2.96**	1.61	−.12	.74	−1.73	1.23	−.94	−.54	−2.11	−.38	.12	.77	−1.61	−.17	**2.69**	2.08	−.76	−1.29

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Comparing the T3–PCA solution to the LMPCA solution (see right part of Tables 5–10), it appears that the LMPCA components have a less clear interpretation than the T3–PCA components, with the former components often being a (less interpretable) combination of the latter. In particular, for the responses (matrix $\mathbf{B}_1$ in Table 5), the first LMPCA component largely captures an approach-avoidance response dimension (i.e., the second T3–PCA component) but also includes some bodily sensations like dry mouth (i.e., the first T3–PCA component), whereas the second LMPCA component represents bodily sensations but also has some aspects of approach-avoidance (e.g., not wanting to avoid a situation). Looking at the situation component matrix ($\mathbf{C}$ in Table 6), the first LMPCA component combines inanimate dangers with adrenaline rush situations, herewith lumping together the second and third T3–PCA components. The second and third LMPCA components are quite similar and capture mainly situations involving an evaluation by others (i.e., the first T3–PCA component) and inanimate dangers (i.e., second T3–PCA component), with the sign of the inanimate danger loadings differing between both LMPCA components (i.e., negative on the second and positive on the third LMPCA component). Again, the LMPCA components seem to be a (hard-to-interpret) blend of the T3–PCA components, herewith confusing the different dimensions existing in the data, which are neatly disentangled by T3–PCA. Also for the subject (matrix $\mathbf{A}$ in Table 10) and disposition loadings (matrix $\mathbf{B}_2$ in Table 8), it is true the LMPCA conflates the existing dimensions in the data, whereas T3–PCA nicely tears them apart.

Finally, we also applied the procedure for testing for the presence of common components in the coupled data to the data from the illustrative application. Using 100 permutations, the results of this analysis show that the similarity value for the original data is larger than the similarity value from each permuted dataset, implying that the coupled illustrative dataset contains common components.

6 | CONCLUDING REMARKS

In this article, we proposed T3–PCA, a global model for conducting a simultaneous analysis of a coupled dataset consisting of a real-valued three-way three-mode data array that shares a single mode with a real-valued two-way two-mode data matrix. In T3–PCA, the three-way data array is modelled with Tucker3 and the two-way data matrix with PCA. The common mode is represented by a single set of common components that are estimated by using the information present in both data blocks (i.e., data fusion). The T3–PCA model is less restrictive than LMPCA (Wilderjans, Ceulemans, Kiers, et al., 2009; Wilderjans, Ceulemans, & Van Mechelen, 2009), which decomposes the three-way data according to Parafac, because the number of components is not required to be equal across the three modes and more complex interactions between components of different modes are allowed. In an extensive simulation study and an application to empirical psychological data from the domain of contextualized personality psychology, it was demonstrated that T3–PCA performed well in terms of optimizing the loss function and revealing the structure underlying the coupled dataset. Regarding the latter aspect, the simultaneous T3–PCA strategy, moreover, appears to clearly outperform LMPCA and a sequential strategy in which common components are determined using the data available in a single data block only. In particular, when information from both data blocks is used, the simultaneous strategy better recovers the common components and ignores components that are specific to one dataset (i.e., distinctive components) than a sequential strategy and, as a consequence, better recovers the structure underlying the non-common modes of the coupled dataset. These effects are most apparent when both data blocks have the same amount of influence on the analysis (i.e., $\alpha = .50$) and when there is a strong distinctive component underlying the data. Further, for a coupled dataset on anxiety, it was shown that T3–PCA better discloses the common components than LMPCA and the sequential strategy, which enables researchers to better link individual differences in behavioural signatures to individual differences in traits and dispositions.

In the remainder of the article, generalizations of the proposed simultaneous modelling strategy are presented, making a distinction between modifications at the level of the model and at the level of the data analysis.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

At the *level of the model*, an extension of T3–PCA could be considered that pertains to a different type of coupled data in which, besides additional information for the units of a single mode, also extra information for the other modes of the three-way data is available. For example, a researcher may collect data on behavioural signatures and at the same time information regarding characteristics of the persons (e.g., traits), responses (e.g., the extent to which an emotion is intra- or interpersonal), and situations (e.g., the extent to which a situation is experienced as threatening and/or whether or not the situation involves other people) under study. To reveal the processes underlying such coupled data, a generalization of T3–PCA may be developed in which the three-way data are modelled with Tucker3 and each of the two-way matrices (i.e., one for the persons, responses and situations) with PCA. Note that in this model, each mode of the three-way data is a common mode. An ALS algorithm for this extension of the T3–PCA model could be constructed in which the common modes and the non-common PCA modes are updated alternatingly. To estimate each common mode, the information in the three-way array is analysed simultaneously with the information of the associated two-way matrix.

At the *level of the data analysis*, two possible improvements could be considered. A first improvement pertains to the ALS algorithm to estimate the T3–PCA model parameters. In the context of Parafac, Acar et al. (2011) showed that an ALS algorithm did not recover the underlying structure well when an overly complex model (i.e., with too many components, also denoted by overfactoring) was selected. To overcome this problem, Acar et al. (2011) proposed CPOPT, a gradient-based optimization approach in which all parameters are updated at once instead of per parameter set as in ALS, and showed that CPOPT was more robust against overfactoring than ALS. Since selecting the model with an optimal complexity is a non-trivial task for T3–PCA (i.e., the optimal number of components should be determined for each mode separately), having an algorithm that is robust against overfactoring may ease this task a lot. Therefore, it may be interesting to implement an all-at-once optimization approach for T3–PCA. To this end, one may extend the CMTF–OPT approach of Acar et al. (2013), which is an all-at-once approach for LMPCA, to T3–PCA.

Second, in T3–PCA, it is assumed that the weight $\boldsymbol{\alpha}$ that is used in loss function (Equation 3) to weight both data blocks is known beforehand. In many cases, however, it is hard to make an adequate choice for the value of $\boldsymbol{\alpha}$. One of the (main) factors determining which $\boldsymbol{\alpha}$ values are optimal pertains to the amount of noise (e.g., measurement error[7]) that is present in each data block. Simulation studies in the context of coupled two-way data blocks, with each block being modelled with PCA, demonstrate that when data blocks have different amounts of noise (i.e., noise heterogeneity), the underlying structure is best recovered when noisy data blocks are downweighted in favour of less noisy ones (van den Berg et al., 2009; Wilderjans et al., 2011). For the type of coupled data considered in this paper, differences in amount of noise can be expected when the information in one data block is measured more reliably than the information in another data block (e.g., one block pertaining to physiological measures and the other one to psychological questionnaires, with the former being measured more reliably than the latter). To deal with noise heterogeneity, the T3–PCA strategy may be extended by estimating the amount of noise in each data block during the analysis and choosing $\boldsymbol{\alpha}$ in such a way that noisy data blocks are downweighted. To this end, an alternating algorithm may be developed in which the underlying structure (i.e., component matrices and core) and the noise are estimated successively. Assuming residuals are independent and normally distributed, an optimal estimate, from a maximum likelihood point of view, for the amount of noise in each data block can be obtained by computing for each block the average of the squared residuals (see De Roover et al., 2014; Wilderjans et al., 2011). It should be noted that differences in the amount of noise may not be limited to the block level but may also be present at the level of the variables within each block. As such, a similar extension of T3–PCA to the case of noise heterogeneity among the variables within a block may also be considered (for a similar approach, see Wilderjans, Ceulemans, & Van Mechelen, 2012).

[7]Note that, besides measurement error, noise may also be present because of weak components that only explain a small amount of variance and that are not relevant for the processes under study (Ceulemans & Kiers, 2009).

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting this study are available from the corresponding author upon reasonable request.

ORCID

Elisa Frutos-Bernal https://orcid.org/0000-0001-7108-481X
Eva Ceulemans https://orcid.org/0000-0002-7611-4683
Purificación Galindo-Villardón https://orcid.org/0000-0001-6977-7545
Tom F. Wilderjans https://orcid.org/0000-0002-1677-4938

REFERENCES

Acar, E., Dunlavy, D. M., & Kolda, T. G. (2011). A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics*, *25*(2), 67–86.

Acar, E., Plopper, G., & Yener, B. (2012). Coupled analysis of in vitro and histology tissue samples to quantify structure–function relationship. *PLoS One*, *7*(3), e32227.

Acar, E., Rasmussen, M. A., Savorani, F., Naes, T., & Bro, R. (2013). Understanding data fusion within the framework of coupled matrix and tensor factorizations. *Chemometrics and Intelligent Laboratory Systems*, *129*, 53–63.

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L. C., De Moor, B., Marynen, P., Hassan, B., Carmeliet, P., & Moreau, Y. (2006). Gene prioritization through genomic data fusion. *Nature Biotechnology*, *24*, 537–544.

Andersson, C. A., & Bro, R. (1998). Improving the speed of multi-way algorithms: Part I. Tucker3. *Chemometrics and Intelligent Laboratory Systems*, *42*, 93–103.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379–384.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203.

Boque, R., & Smilde, A. K. (1999). Monitoring and diagnosing batch processes with multiway covariates regression models. *AIChE Journal*, *45*, 1504–1520.

Bro, R., & Andersson, C. A. (1998). Improving the speed of multiway algorithms: Part II. Compression. *Chemometrics and Intelligent Laboratory Systems*, *42*, 105–113.

Bro, R., Kjeldahl, K., Smilde, A. K., & Kiers, T. C. Y. J. A. V. P. (2008). Cross-validation of component models: a critical look at current methods. *Analytical and bioanalytical chemistry*, *390*, 1241–1251.

Bro, R., & Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, *17*(1), 16–33.

Brusco, M. J., & Cradit, J. D. (2005). Bicriterion methods for partitioning dissimilarity matrices. *British Journal of Mathematical and Statistical Psychology*, *58*(2), 319–332.

Brusco, M. J., Doreian, P., Steinley, D., & Satornino, C. B. (2013). Multiobjective blockmodeling for social network analysis. *Psychometrika*, *78*(3), 498–525.

Brusco, M. J., & Stahl, S. (2001). An interactive approach to multiobjective combinatorial data analysis. *Psychometrika*, *66*(1), 5–24.

Brusco, M. J., & Stahl, S. (2005). Bicriterion seriation methods for skew-symmetric matrices. *British Journal of Mathematical and Statistical Psychology*, *58*(2), 333–343.

Brusco, M. J., & Steinley, D. (2009). Cross-validation issues in multiobjective clustering. *British Journal of Mathematical and Statistical Psychology*, *62*(2), 349–368.

Brusco, M. J., & Steinley, D. (2012). A note on the estimation of the Pareto efficient set for multiobjective matrix permutation problems. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 145–162.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 438–457.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*, 283–319.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. Plenum.

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, *59*, 133–150.

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

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

Chiang, J., Wang, Z. J., & McKeown, M. J. (2010). Multiblock PLS model for group corticomuscular activity analysis in Parkinson disease. In M. B. Matthews (Ed.), *Conference Record of the Forty-Fourth Asilomar Conference on Signals, Systems and*

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Computers (ASILOMAR), Pacific Grove, California, US, 7-10 November 2010 (pp. 1375–1379). The Institute of Electrical and Electronics Engineers (IEEE), Inc.

Coxe, K. L. (1986). Principal components regression analysis. In N. L. Johnson & S. Kotz (Eds.), *Encyclopedia of statistical science* (Vol. 7, pp. 181–186). Wiley.

Curran, P., & Hussong, A. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100.

de Leeuw, J. (1994). Block-relaxation algorithms in statistics. In H. Bock, W. Lenski, & M. M. Richter (Eds.), *Information systems and data analysis* (pp. 308–325). Springer Verlag.

De Roover, K., Timmerman, M. E., Van Diest, I., Onghena, P., & Ceulemans, E. (2014). Switching principal component analysis for modeling means and covariance changes over time. *Psychological Methods*, *19*, 113–132.

de Schipper, N., & Van Deun, K. (2021). Model selection techniques for sparse weight-based principal component analysis. *Journal of Chemometrics*, *35*(2), e3289.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *41*, 309–326.

Ehrgott, M. (2005). *Multicriteria optimization* (2nd ed.). Springer.

Ehrgott, M., & Gandibleaux, X. (2000). A survey and annotated bibliography on multiobjective combinatorial optimization. *OR-Spektrum*, *22*, 425–460.

Ehrgott, M., & Wiecek, M. M. (2005). Multiobjective programming. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 667–722). Springer.

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.

Ermis, B., Acar, E., & Cemgil, A. (2012). Link prediction via generalized coupled tensor factorisation. In *ECML/PKDD workshop on collective learning and inference on structured data*.

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations (Johns Hopkins studies in the mathematical sciences)* (4th ed.). The John Hopkins University Press.

Gurden, S. (n.d.). *Multiway covariates regression: Discussion facilitator.*

Gvaladze, S., Vervloet, M., Van Deun, K., Kiers, H. A. L., & Ceulemans, E. (2021). PCovR2: A flexible principal covariates regression approach to parsimoniously handle multiple criterion variables. *Behavior Research Methods*, *53*, 1648–1668.

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

Heij, C., Groenen, P. J. F., & Van Dijk, D. (2007). Forecast comparison of principal component regression and principal covariate regression. *Computational Statistics and Data Analysis*, *51*, 3612–3625.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematical Physics*, *6*, 164–189.

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

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). *Springer series in statistics*. Springer-Verlag.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200.

Kiers, H. A. L. (1992). Tuckals core rotations and constrained Tuckals modelling. *Statistica Applicata*, *4*, 659–667.

Kiers, H. A. L. (1997). Three-mode orthomax rotation. *Psychometrika*, *62*, 579–598.

Kiers, H. A. L. (1998a). Joint orthomax rotation of the core and component matrices resulting from three-mode principal component analysis. *Journal of Classification*, *15*, 245–263.

Kiers, H. A. L. (1998b). Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Computational Statistics and Data Analysis*, *28*, 307–324.

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

Kiers, H. A. L., Kroonenberg, P. M., & ten Berge, J. M. F. (1992). An efficient algorithm for TUCKALS3 on data with large numbers of observation units. *Psychometrika*, *57*, 415–422.

Kiers, H. A. L., & Smilde, A. K. (2007). A comparison of various methods for multivariate regression with highly collinear variables. *Statistical Methods and Applications*, *16*, 193–228.

Kiers, H. A. L., & Van Mechelen, I. (2001). Three-way component analysis: Principles and illustrative application. *Psychological Methods*, *6*, 84–110.

Krijnen, P. (2006). Convergence of the sequence of parameters generated by alternating least squares algorithms. *Computational Statistics & Data Analysis*, *51*(2), 481–489.

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

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

Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*, 69–97.

Kroonenberg, P. M., & ten Berge, J. M. F. (2011). The equivalence of Tucker3 and Parafac models with two components. *Chemometrics and Intelligent Laboratory Systems*, *106*, 21–26.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lin, Y., Sun, J., Castro, P., Konuru, R., Sundaram, H., & Kelliher, A. (2009). MetaFac: Community discovery via relational hypergraph factorization. In *KDD'09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 527–536).

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64.

Louwerse, D. J., Smilde, A. K., & Kiers, H. A. L. (1999). Cross-validation of multiway component models. *Journal of Chemometrics*, *13*(5), 491–540.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195.

Olson, D. L. (1992). Review of empirical studies in multiobjective mathematical programming: Subject reflection of nonlinear utility and learning. *Decision Sciences*, *23*, 1–20.

Palatucci, M., Pomerleau, D., Hinton, G., & Mitchell, T. M. (2009). Zero-shot learning with semantic output codes. *Advances in Neural Information Processing Systems*, *22*, 1410–1418.

Park, S., Ceulemans, E., & Van Deun, K. (2021). Sparse common and distinctive covariates regression. *Journal of Chemometrics*, *35*, e3270.

Park, S., Ceulemans, E., & Van Deun, K. (2023). Logistic regression with sparse common and distinctive covariates. *Behavior Research Methods*, *55*, 4143–4174.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series*, *6*(2), 559–572.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373.

Ringuest, J. L. (1992). *Multiobjective optimization: Behavioral and computational considerations*. Kluwer Academic Publishers.

Rosenthal, R. E. (1985). Principles of multiobjective optimization. *Decision Sciences*, *16*, 133–152.

Schouteden, M., Van Deun, K., Pattyn, S., & Van Mechelen, I. (2013). SCA with rotation to distinguish common and distinctive information in linked data. *Behavior Research Methods*, *45*, 822–833.

Schouteden, M., Van Deun, K., Wilderjans, T. F., & Van Mechelen, I. (2014). Performing DISCO-SCA to search for distinctive and common information in linked data. *Behavior Research Methods*, *46*, 576–587.

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

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

Smilde, A. K., van der Werf, M., Bijlsma, S., van der Werff-van der Vat, B., & Jellema, R. (2005). Fusion of mass spectrometry-based metabolomics data. *Analytical Chemistry*, *77*, 6729–6736.

Smilde, A. K., Westerhuis, J. A., & Boqué, R. (2000). Multiway multiblock component and covariates regression models. *Journal of Chemometrics*, *14*, 301–331.

Smilde, A. K., Westerhuis, J. A., & de Jong, S. (2003). A framework for sequential multiblock component methods. *Journal of Chemometrics*, *17*, 323–337.

Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

ten Berge, J. M. F. (1986). Rotation to perfect congruence and the cross-validation of component weights across populations. *Multivariate Behavioral Research*, *21*, 41–64.

ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis*. DSWO Press.

Thurstone, L. L. (1947). *Multiple-factor analysis*. The University of Chicago Press.

Timmerman, M. E., & Kiers, H. A. L. (2000). Three-mode principal component analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology*, *53*, 1–16.

Timmerman, M. E., & Kiers, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, *68*, 105–121.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies*. Personnel research section rapport # 984. Department of the Army.

Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 122–137). University of Wisconsin Press.

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 109–127). Holt, Rinehart and Winston.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.

Ulungu, E., & Teghem, J. (1994). The two-phase method: An efficient procedure to solve biobjective combinatorial optimization problems. *Foundations of Computing and Decision Sciences*, *20*, 149–165.

Van Coillie, H., Van Mechelen, I., & Ceulemans, E. (2006). Multidimensional individual differences in anger-related behaviors. *Personality and Individual Differences*, *41*, 27–38.

van den Berg, R. A., Van Mechelen, I., Wilderjans, T. F., Van Deun, K., Kiers, H. A. L., & Smilde, A. K. (2009). Integrating functional genomics data using maximum likelihood based simultaneous component analysis. *BMC Bioinformatics*, *10*, 340.

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Van der Ploeg, H. (2000). *Handleiding bij de Zelf-Beoordelings Vragenlijst. Een Nederlandstalige bewerking van de Spielberger State-Trait Anxiety Inventory STAI-DY*. Swets & Zeitlinger.
Van Deun, K., Smilde, A. K., Thorrez, L., Kiers, H. A. L., & Van Mechelen, I. (2013). Identifying common and distinctive processes underlying multiset data. *Chemometrics and Intelligent Laboratory Systems*, *129*, 40–51.
Van Mechelen, I., & Kiers, H. A. L. (1999). Individual differences in anxiety responses to stressful situations: A three-mode component analysis model. *European Journal of Personality*, *13*, 409–428.
Van Mechelen, I., & Schepers, J. (2007). A unifying model involving a categorical and/or dimensional reduction for multimode data. *Computational Statistics and Data Analysis*, *52*, 537–549.
Van Mechelen, I., & Smilde, A. K. (2010). A generic linked-mode decomposition model for data fusion. *Chemometrics and Intelligent Laboratory Systems*, *104*, 83–94.
Vervloet, M., Kiers, H. A. L., Van den Noortgate, W., & Ceulemans, E. (2015). PCovR: An R package for principal covariates regression. *Journal of Statistical Software*, *65*, 1–14.
Vervloet, M., Van Deun, K., Van den Noortgate, W., & Ceulemans, E. (2013). On the selection of the weighting parameter value in principal covariates regression. *Chemometrics and Intelligent Laboratory Systems*, *123*, 36–43.
Vervloet, M., Van Deun, K., Van den Noortgate, W., & Ceulemans, E. (2016). Model selection in principal covariates regression. *Chemometrics and Intelligent Laboratory Systems*, *151*, 26–33.
Wansbeek, T., & Verhees, J. (1989). Models for multidimensional matrices in econometrics and psychometrics. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 543–552). Elsevier.
Wilderjans, T. F., & Ceulemans, E. (2013). Clusterwise Parafac to identify heterogeneity in three-way data. *Chemometrics and Intelligent Laboratory Systems*, *129*, 87–97.
Wilderjans, T. F., Ceulemans, E., Kiers, H. A. L., & Meers, K. (2009). The LMPCA program: A graphical user interface for fitting the linked-mode PARAFAC-PCA model to coupled real-valued data. *Behavior Research Methods*, *41*, 1073–1082.
Wilderjans, T. F., Ceulemans, E., & Kuppens, P. (2012). Clusterwise HICLAS: A generic modeling strategy to trace similarities and differences in multi-block binary data. *Behavior Research Methods*, *44*, 532–545.
Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull-based model selection method. *Behavior Research Methods*, *45*, 1–15.
Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (2008). The CHIC model: A global model for coupled binary data. *Psychometrika*, *73*, 729–751.
Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (2009). Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics and Data Analysis*, *53*, 1086–1098.
Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (2012). The SIMCLAS model: Simultaneous analysis of coupled binary data matrices with noise heterogeneity between and within data blocks. *Psychometrika*, *77*, 724–740.
Wilderjans, T. F., Ceulemans, E., Van Mechelen, I., & van den Berg, R. A. (2011). Simultaneous analysis of coupled data matrices subject to different amounts of noise. *British Journal of Mathematical and Statistical Psychology*, *64*, 277–290.
Zehfuss, G. (1858). Ueber eine gewisse determinante. *Zeitschrift für Mathematik und Physik*, *3*, 298–301.
Zheng, V., Cao, B., Zheng, Y., & Xie, Q. (2010). Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI'10: Proceedings of the 24th Conference on Artificial Intelligence* (pp. 236–241).

How to cite this article: Frutos-Bernal, E., Ceulemans, E., Galindo-Villardón, P., & Wilderjans, T. F. (2025). Data fusion by T3–PCA: A global model for the simultaneous analysis of coupled three-way and two-way real-valued data. *British Journal of Mathematical and Statistical Psychology*, *78*, 672–709. https://doi.org/10.1111/bmsp.12372

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX 1

ALGORITHM

Algorithm 1 (T3-PCA-ALS)

Require: $\underline{\mathbf{D}}^1$, $\mathbf{D}^2$, MaxIter and τ;
Ensure: $\mathbf{A}$, $\mathbf{B}_1$, $\mathbf{C}$, $\mathbf{B}_2$ and $\underline{\mathbf{G}}$
Obtain random or rational starting values for $\mathbf{A}$, $\mathbf{B}_1$ and $\mathbf{C}$;
Compute $\underline{\mathbf{G}}$ as $\hat{\mathbf{G}}_a = (\mathbf{A}'\mathbf{A})^{-1}\,\mathbf{A}'\,\mathbf{D}^1_a\,(\mathbf{C}\otimes\mathbf{B})\,[(\mathbf{C}'\mathbf{C})^{-1}\otimes(\mathbf{B}_1'\mathbf{B}_1)^{-1}]$;
Compute $\mathbf{B}_2$ as $\hat{\mathbf{B}}_2 = (\mathbf{D}^2)'\,\mathbf{A}\,(\mathbf{A}'\mathbf{A})^{-1}$;
Compute current loss $fold$ with current values of $\mathbf{A}$, $\mathbf{B}_1$, $\mathbf{C}$, $\mathbf{B}_2$ and $\underline{\mathbf{G}}$;
Set Iter := 0;
Set LossIncrease := TRUE;
while (Iter < MaxIter) and LossIncrease **do**
Set Iter = Iter + 1;
Update $\mathbf{A}$ as $\hat{\mathbf{A}} = \mathbf{U}\mathbf{V}'$ where $\mathbf{U}$ and $\mathbf{V}$ are the P left and right singular vectors, respectively, associated with the P largest singular values of $\mathbf{Q} = \mathbf{Y}\mathbf{Z}'$, with $\mathbf{Y} = [\sqrt{\alpha}\times\mathbf{D}^1_a \mid \sqrt{(1-\alpha)}\times\mathbf{D}^2]$ and $\mathbf{Z} = [\sqrt{\alpha}\times\mathbf{G}_a\,(\mathbf{C}\otimes\mathbf{B}_1)' \mid \sqrt{1-\alpha}\times\mathbf{B}_2']$;
Update $\mathbf{B}_1$ as $\hat{\mathbf{B}}_1 = \mathbf{U}\mathbf{V}'$ where $\mathbf{U}$ and $\mathbf{V}$ are the P left and right singular vectors, respectively, associated with the P largest singular values of $\mathbf{Q} = \mathbf{Y}\mathbf{Z}'$, with $\mathbf{Y} = \mathbf{D}^1_b$ and $\mathbf{Z} = \mathbf{G}_b\,(\mathbf{A}\otimes\mathbf{C})'$;
Update $\mathbf{C}$ as $\hat{\mathbf{C}} = \mathbf{U}\mathbf{V}'$ where $\mathbf{U}$ and $\mathbf{V}$ are the P left and right singular vectors, respectively, associated with the P largest singular values of $\mathbf{Q} = \mathbf{Y}\mathbf{Z}'$, with $\mathbf{Y} = \mathbf{D}^1_c$ and $\mathbf{Z} = \mathbf{G}_c\,(\mathbf{B}_1\otimes\mathbf{A})'$;
Update $\underline{\mathbf{G}}$ as $\hat{\mathbf{G}}_a = (\mathbf{A}'\mathbf{A})^{-1}\,\mathbf{A}'\,\mathbf{D}^1_a\,(\mathbf{C}\otimes\mathbf{B})\,[(\mathbf{C}'\mathbf{C})^{-1}\otimes(\mathbf{B}_1'\mathbf{B}_1)^{-1}]$;
Update $\mathbf{B}_2$ as $\hat{\mathbf{B}}_2 = (\mathbf{D}^2)'\,\mathbf{A}\,(\mathbf{A}'\mathbf{A})^{-1}$;
Compute new loss f with the updated values of $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}_1$, $\hat{\mathbf{C}}$, $\hat{\mathbf{B}}_2$ and $\hat{\underline{\mathbf{G}}}$;
Set LossIncrease = ($fold$ - f) > ($fold$ * τ);
Set $fold = f$;
end while

Note: $\mathbf{D}^1_a$ ($\mathbf{D}^1_b$, $\mathbf{D}^1_c$) is the $I\times JK$ ($J\times KI$, $K\times IJ$) matricized version of $\mathbf{D}^1$. Similarly, $\mathbf{G}_a$ ($\mathbf{G}_b$, $\mathbf{G}_c$) is the $P\times QR$ ($Q\times RP$, $R\times PQ$) matricized version of $\mathbf{G}$. $\mathbf{G}$ can also be updated as $\widehat{\mathbf{G}_b} = (\mathbf{B}_1'\mathbf{B}_1)^{-1}\,\mathbf{B}_1'\,\mathbf{D}^1_b\,(\mathbf{A}\otimes\mathbf{C})\,[(\mathbf{A}'\mathbf{A})^{-1}\otimes(\mathbf{C}'\mathbf{C})^{-1}]$ or as $\widehat{\mathbf{G}_c} = (\mathbf{C}'\mathbf{C})^{-1}\bar{\mathbf{C}}'\,\bar{\mathbf{D}}^1_c\,(\mathbf{B}_1\otimes\mathbf{A})\,[(\mathbf{B}_1'\mathbf{B}_1)^{-1}\otimes(\mathbf{A}'\mathbf{A})^{-1}]$.

APPENDIX 2

PERCENTAGE OF SIMULATED DATASETS IN WHICH T3–PCA OUTPERFORMS SEQUENTIAL AND LMPCA STRATEGIES

TABLE A1 Percentage of simulated datasets for which simultaneous T3–PCA strategy outperforms sequential strategy in terms of recovery of component matrices ($GOR^{diff} > 0$) and core array ($BOR^{diff} < 0$), computed across all datasets and analyses and for each level of α separately.

α Level	GOR^{diff}_{A}	$GOR^{diff}_{B_1}$	GOR^{diff}_{C}	$GOR^{diff}_{B_2}$	$BOR^{diff}_{\bar{G}}$
.050	67.81	25.10	22.50	58.65	72.29
.100	68.12	25.73	23.02	59.17	72.71
.300	90.73	31.67	30.63	75.42	95.10
.500	99.90	39.79	38.96	96.56	99.27
.600	100.00	41.98	41.77	96.98	99.06
.700	99.48	43.75	43.54	97.19	96.67
.800	99.17	45.52	45.42	97.29	94.69
.900	98.75	46.77	47.29	96.88	92.60
.950	98.65	48.02	48.12	96.77	92.08
.999	97.19	47.60	49.17	94.79	89.90
Overall	91.98	39.59	39.04	86.97	90.44

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A2 Percentage of simulated datasets for which simultaneous T3–PCA strategy outperforms simultaneous LMPCA strategy in terms of recovery of component matrices ($GOR^{diff} > 0$), computed overall across datasets with rank ($P = 2$, $Q = 2$, $R = 2$) and for each level of α separately.

α Level	GOR_A^{diff}	$GOR_{B_1}^{diff}$	GOR_C^{diff}	$GOR_{B_2}^{diff}$
.050	87.29	100.00	100.00	100.00
.100	83.33	99.79	100.00	100.00
.300	93.75	100.00	99.58	100.00
.500	99.79	100.00	100.00	99.79
.600	99.58	100.00	100.00	99.79
.700	98.54	100.00	100.00	99.17
.800	97.50	100.00	100.00	99.17
.900	94.79	100.00	100.00	98.96
.950	93.96	100.00	100.00	98.75
.999	93.12	100.00	100.00	98.54
Overall	94.17	99.98	99.96	99.42

20448317, 2025, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bmsp.12372, Wiley Online Library on [05/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License