On Optimal Allocation of Treatment/Condition Variance in Principal Component Analysis

André Beauducel[1] & Norbert Hilger[1]

[1] Department of Psychology, University of Bonn, Germany

Correspondence: André Beauducel, Department of Psychology, University of Bonn, Kaiser-Karl-Ring, 9, 53111 Bonn, Germany. E-mail: beauducel@uni-bonn.de

Abstract

The allocation of a (treatment) condition-effect on the wrong principal component (misallocation of variance) in principal component analysis (PCA) has been addressed in research on event-related potentials of the electroencephalogram. However, the correct allocation of condition-effects on PCA components might be relevant in several domains of research. The present paper investigates whether different loading patterns at each condition-level are a basis for an optimal allocation of between-condition variance on principal components. It turns out that a similar loading shape at each condition-level is a necessary condition for an optimal allocation of between-condition variance, whereas a similar loading magnitude is not necessary.

Keywords: Principal component analysis, misallocation of variance, within- and between condition effects

1. Introduction

1.1 Condition Effects in Principal Component Analysis

Principal component analysis (PCA) has regularly been performed for the analysis of event-related potentials of the electroencephalogram (Dien, Khoe & Mangun, 2007; Dien, 2010; Kayser & Tenke, 2003, 2005). In the context of event-related potentials, PCA is often performed for observed variables representing k levels of at least one (experimental) condition factor, so that the components represent a mixture of the between- and within-condition variance. However, (experimental) condition factors occur in several areas of research and PCA is performed in several areas of research and has been adapted to several different methodological contexts (Jolliffe & Cadima, 2016). It is therefore interesting to know how experimental condition effects are optimally allocated on principal components.

1.2 Misallocation of Between-condition Variance

Since Wood and McCarthy (1984) it has been regarded as an optimum when a single PCA component combines the complete between-condition variance of a single condition factor with some within-condition variance. Although the allocation of the variance of a single condition factor on a single principal component combining within- and between-condition variance was described as an optimum in research on event-related potentials, this form of variance allocation might also be useful in other contexts of research as it allows for a parsimonious data description. Wood and McCarthy (1984) used the term 'misallocation of variance' in order to denote that the between-condition variance is allocated on more than one principal component. Misallocation of variance has been investigated in simulation studies on methods of PCA component rotation (e.g., Scharf & Nestler, in press; Dien, 2010; Beauducel & Debener, 2003; Wood & McCarthy, 1984) and new methods of component rotation have been proposed that may reduce misallocation of variance (Beauducel, 2018; Beauducel & Leue, 2015). Moreover, Scharf and Nestler (in press) have demonstrated that the covariation of several condition factors may induce misallocation of variance.

It has also been proposed to perform a separate PCA for each group representing a level of a single condition factor because the loading shapes in each condition can be different (Barry, De Blasio, Fogarty & Karamacoska, 2016). Although it might be reasonable to identify condition-specific loading patterns by means of separate PCAs at each level of a condition factor, the effect of this form of analysis on misallocation of variance remains unknown.

1.3 Aims of the Present Paper

The present paper therefore investigates the effects of separate PCAs at each level of a single condition factor on the allocation of between-condition variance on PCA components. First, some definitions for separate PCAs at each level of a single condition factor and for a PCA of the between-condition variance of the condition factor are presented. Second, it is shown that misallocation of condition variance as it has been demonstrated and discussed since Wood and McCarthy (1984) follows necessarily from rotation of components that perfectly represent a single condition effect. Third, it is shown that different condition-specific loading shapes do not allow for an unambiguous allocation of between-condition variance on a single component representing within- and between-condition variance. Finally, it is shown that different condition-specific loading patterns are compatible with an unambiguous allocation of between-condition variance on a single component, when the between-condition differences of the loadings on each component can be accounted for by a scalar.

2. Definitions: PCA for within- and Between-condition Variance

Consider that p random variables have been observed in k levels of a condition factor, so that

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_i & \cdots & \mathbf{x}_k \end{bmatrix}, \ for \ i = 1, 2, \ldots, k. \tag{1}$$

Although the expectation of $\mathbf{x}$ is zero $(\mathrm{E}[\mathbf{x}] = 0)$, the conditions imply $\mathrm{E}[\mathbf{x}_i] \neq 0$. However, when a within- condition PCA is performed separately for the correlations or covariances at each level of the condition factor, $\mathbf{x}_i^v = \mathbf{x}_i - \mathrm{E}[\mathbf{x}_i]$, the mean centered part of $\mathbf{x}_i$, is analyzed, since $\mathrm{Cov}[\mathbf{x}_i, \mathbf{x}_i] = \mathrm{Cov}[\mathbf{x}_i^v, \mathbf{x}_i^v] = \boldsymbol{\Sigma}_i$, so that

$$\mathbf{x}_i^v = \mathbf{A}_i^v \mathbf{c}_i^v, \ for \ i = 1, 2, \ldots k, \tag{2}$$

superscript "v" denotes the within-condition variance and where $\mathbf{A}_i^v$ is a $p \times p$ matrix of component loadings and $\mathbf{A}_i^v{}'\mathbf{A}_i^v$ contains the eigenvalues in decreasing order. The components $\mathbf{c}_i^v$ are assumed to have an expectation zero, $\mathrm{E}[\mathbf{c}_i^v] = 0$. PCA initially yields orthogonal components $(\mathrm{E}[\mathbf{c}_i^v \mathbf{c}_i^v{}'] = \mathbf{I})$, so that each covariance matrix of observed variables can be decomposed into

$$\sum{}_i^v = \mathbf{A}_i^v \mathbf{A}_i^v{}', \ for \ i = 1, 2, \ldots k, \tag{3}$$

Typically, components $\mathbf{c}_i$ are divided into a subset of q wanted components $\mathbf{w}_i$ and $p - q$ unwanted components $\mathbf{u}_i$ ($\mathbf{c}_i = [\mathbf{w}_i, \mathbf{u}_i]$, $\mathbf{A}_i = [\mathbf{M}_i, \mathbf{N}_i]$). Orthogonal and oblique rotations of $\mathbf{M}_i$ and $\mathbf{w}_i$ have been proposed, so that non-zero component inter-correlations are possible $(\mathrm{E}[\mathbf{c}_i \mathbf{c}_i{}'] = \mathbf{Q}_i)$. The covariances of observed variables are then decomposed by

$$\boldsymbol{\Sigma}_i = \mathbf{M}_i \mathbf{Q}_i \mathbf{M}_i' + \mathbf{N}_i \mathbf{N}_i', \ for \ i = 1, 2, \ldots k. \tag{4}$$

It is possible to write the complete data comprising condition variance and within-group variance as

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} \mathbf{x}_i^v + diag(\mathrm{E}[\mathbf{x}_i])\mathbf{1}_i, & \cdots & , \mathbf{x}_k^v + diag(\mathrm{E}[\mathbf{x}_k])\mathbf{1}_k \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}_i^v & \cdots & \mathbf{x}_k^v \end{bmatrix} + \begin{bmatrix} diag(\mathrm{E}[\mathbf{x}_i])\mathbf{1}_i, & \cdots & , diag(\mathrm{E}[\mathbf{x}_k])\mathbf{1}_k \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}_i^v & \cdots & \mathbf{x}_k^v \end{bmatrix} + \begin{bmatrix} \mathbf{x}_i^b & \cdots & \mathbf{x}_k^b \end{bmatrix}, \end{aligned} \tag{5}$$

where $\mathbf{1}_i$ has the dimensions of $\mathbf{x}_i$ and $\mathbf{1}_k$ has the dimensions of $\mathbf{x}_k$. The related within- and between-conditions PCAs yield

$$\mathbf{x} = \begin{bmatrix} \mathbf{A}_i^v \mathbf{c}_i^v, & \cdots & , \mathbf{A}_k^v \mathbf{c}_k^v \end{bmatrix} + \mathbf{A}^b \mathbf{c}^b. \tag{6}$$

Usually q_v wanted within-condition components $\mathbf{w}_i^v$ are separated from $p - q_v$ unwanted within-condition components $\mathbf{u}_i^v$ and q_b wanted between-condition components $\mathbf{w}^b$ from $p - q_b$ unwanted between-condition components $\mathbf{u}^b$. This yields

$$\mathbf{x} = \begin{bmatrix} \mathbf{M}_i^v \mathbf{w}_i^v + \mathbf{N}_i^v \mathbf{u}_i^v, & \cdots & , \mathbf{M}_k^v \mathbf{w}_k^v + \mathbf{N}_k^v \mathbf{u}_k^v \end{bmatrix} + \mathbf{M}^b \mathbf{w}^b + \mathbf{N}^b \mathbf{u}^b, \tag{7}$$

and

$$\mathbf{x}_w = \begin{bmatrix} \mathbf{M}_i^v \mathbf{w}_i^v, & \cdots & , \mathbf{M}_k^v \mathbf{w}_k^v \end{bmatrix} + \mathbf{M}^b \mathbf{w}^b, \tag{8}$$

for the wanted components.

Typically, the wanted components are rotated in order to improve the interpretation (Dien, 2010; Kayser & Tenke, 2003). If there is an additional condition factor, there can be additional groupings of PCAs for each level of the condition factor and an additional PCA across the levels of the condition factor. If the sample size is sufficiently large, it is also possible to perform a PCA for each of the combinations of condition levels and across all combinations of conditions of the two condition factors.

3. Misallocation of Variance

3.1 Misallocation of Variance and Component Rotation

When there are only a few condition factors the number of wanted within-condition components is probably larger than the number of wanted between-condition components. For example, when there is only one condition factor with two levels, PCA of the between-condition variance without subsequent component rotation will result in only one between-condition component. When $q^v > q^b = 1$ it is possible to write Equation 8 as

$$\mathbf{x}_w = \left[\left[\mathbf{m}_{ji}^v, \quad \cdots \quad , \mathbf{m}_{qi}^v\right] \begin{bmatrix} \mathbf{w}_{ji}^v \\ \vdots \\ \mathbf{w}_{qi}^v \end{bmatrix}, \quad \cdots \quad , \left[\mathbf{m}_{jk}^v, \quad \cdots \quad , \mathbf{m}_{qk}^v\right] \begin{bmatrix} \mathbf{w}_{jk}^v \\ \vdots \\ \mathbf{w}_{qk}^v \end{bmatrix} \right] + \mathbf{m}^b \mathbf{w}^b, \tag{9}$$

where j denotes the number of the respective within-condition component. For $q^b = 1$ and $\mathbf{m}_{1i}^v = \mathbf{m}^b, \quad \cdots \quad , \mathbf{m}_{1k}^v = \mathbf{m}^b$ Equation 9 can be written as

$$\mathbf{x}_w = \mathbf{M} \left[\begin{bmatrix} \mathbf{w}_{ji} \\ \vdots \\ \mathbf{w}_{qi} \end{bmatrix}, \quad \cdots \quad , \begin{bmatrix} \mathbf{w}_{jk} \\ \vdots \\ \mathbf{w}_{qk} \end{bmatrix} \right], \tag{10}$$

$$with \ \mathbf{M} = \left[\mathbf{m}^b, \mathbf{m}_{2i}^v, \quad \cdots \quad , \mathbf{m}_{qi}^v\right] \ and \ \left[\mathbf{w}_{1i}, \cdots, \mathbf{w}_{1k}\right] = \left[\mathbf{w}_{1i}^v + \mathbf{w}_{1i}^b, \cdots, \mathbf{w}_{1k}^v + \mathbf{w}_{1k}^b\right],$$

where $\mathbf{w}_{1i}^v, \cdots, \mathbf{w}_{1k}^v$ denotes the scores on the first wanted component (j=1) at each level i of the condition factor, and $\mathbf{w}_{1i}^b, \cdots, \mathbf{w}_{1k}^b$ denotes the expectancy of the first wanted component on each level i of the condition factor, which corresponds to the expectancy of the observed scores on condition level i, with $[\mathbf{w}_{1i}^b, \cdots, \mathbf{w}_{1k}^b] = [\mathrm{E}(\mathbf{w}_{1i}), \cdots, \mathrm{E}(\mathbf{w}_{1k})] = [\mathrm{E}(\mathbf{x}_i), \cdots, \mathrm{E}(\mathbf{x}_k)]$.

Equation 10 describes what is typically regarded as an optimal allocation of variance, namely, that a condition effect occurs on a single component that combines within- and between-condition variance. The simulation studies on this issue were based on a single condition effect that was introduced exclusively on a single component when the data were generated (Wood & McCarthy, 1984; Dien, 2010; Beauducel & Debener, 2003; Beauducel & Leue, 2015) and that occurred on more than one component after PCA followed by component rotation.

Component rotation means that the $\mathbf{M}$ is rotated by means of postmultiplication by a $q^v \times q^v$ transformation matrix $\mathbf{T}$ (Harman, 1976) and that the component scores are counter-rotated by means of premultiplication with $\mathbf{T}^{-1}$, so that

$$\mathbf{x}_w = \mathbf{MT} \left[\mathbf{T}^{-1} \begin{bmatrix} \mathbf{w}_{ji} \\ \vdots \\ \mathbf{w}_{qi} \end{bmatrix}, \quad \cdots \quad , \mathbf{T}^{-1} \begin{bmatrix} \mathbf{w}_{jk} \\ \vdots \\ \mathbf{w}_{qk} \end{bmatrix} \right], \ with \ \left[\mathbf{w}_{1i}, \cdots, \mathbf{w}_{1k}\right] = \left[\mathbf{w}_{1i}^v + \mathbf{w}_{1i}^b, \cdots, \mathbf{w}_{1k}^v + \mathbf{w}_{1k}^b\right]. \tag{11}$$

For a single condition i the rotation of the infinite matrices containing the population of individual component scores l can be written as

$$\mathbf{T}^{-1} \begin{bmatrix} \mathbf{w}_{ji} \\ \vdots \\ \mathbf{w}_{qi} \end{bmatrix} = \mathbf{T}^* \begin{bmatrix} \mathbf{w}_{jil}, \cdots \\ \vdots \ \ \ddots \\ \mathbf{w}_{qil}, \cdots \end{bmatrix} = \begin{bmatrix} \left(\mathbf{t}^*\mathbf{w}\right)_{jil}, \cdots \\ \vdots \ \ \ddots \\ \left(\mathbf{t}^*\mathbf{w}\right)_{qil}, \cdots \end{bmatrix}, \tag{12}$$

$$for \ l = 1, \ldots, \infty, \ with \ \left[\mathbf{w}_{1il}, \quad \cdots\right] = \left[\mathbf{w}_{1il}^v + \mathrm{E}\left(\mathbf{w}_{1i}\right), \quad \cdots\right] \ and \ \mathbf{T}^{-1} = \mathbf{T}^*.$$

Theorem 1 describes that a non-zero expectation that is initially only on the first component leads to a non-zero expectation on others than the first component after component rotation.

Theorem 1. *If* $\mathrm{E}\left(\mathbf{w}_{ji}\right) = \begin{cases} \mathrm{E}\left(\mathbf{w}_{ji}\right) \neq 0 \ for \ j = 1 \\ \mathrm{E}\left(\mathbf{w}_{ji}\right) = 0 \ for \ j = 2, \ldots, q \end{cases}$ *and* $\mathbf{t}_{jh}^* \neq 0, for \ j = 1, \ldots, q, h = 1, \ldots, q,$

then $\mathrm{E}\left(\mathbf{t}^*\mathbf{w}\right)_{ji} \neq 0, \ for \ j > 1, \ldots, q.$

Proof. A single element for condition *i* of the matrix resulting from Equation 12 is given by

$$\left(\mathbf{t}^*\mathbf{w}\right)_{jil} = \sum_{h=1}^{q} \mathbf{t}^*_{jh}\mathbf{w}_{hil}, \; with \; \mathbf{w}_{1il} = \mathbf{w}^v_{1il} + \mathrm{E}\left(\mathbf{w}_{1i}\right). \tag{13}$$

Equation 13 can be written as

$$\left(\mathbf{t}^*\mathbf{w}\right)_{jil} = \mathbf{t}^*_{j1}\left(\mathbf{w}^v_{1il} + \mathrm{E}\left(\mathbf{w}_{1i}\right)\right) + ,\ldots, + \mathbf{t}^*_{jq}\mathbf{w}_{qil}. \tag{14}$$

Equation 14 implies that the expectation for the population of scores even for $j > 1$ is $\mathrm{E}\left(\mathbf{t}^*\mathbf{w}\right)_{ji} = \mathbf{t}^*_{j1}\mathrm{E}\left(\mathbf{w}_{1i}\right)$.

This completes the proof. □

Theorem 1 implies that a condition effect that occurs only on the first component before rotation, also occurs on other components after rotation. Thus, Theorem 1 shows that misallocation of variance as it has typically been investigated in simulation studies since Wood and McCarthy (1984) is a necessary consequence of any rotation of an initial set of components combining unambiguously within- and between-condition effects. Therefore, the attempts to reduce misallocation of variance are attempts to recover the initial combination of within- and between-condition components (Dien, 2010; Beauducel & Leue, 2015; Beauducel, 2018) so that the matrix **T**, transforming the original components to the given components becomes **I**. This implies $\mathbf{T}^* = \mathbf{I}$ and $\mathbf{t}^*_{jh} = 0, \; for \; j \neq h$ so that Theorem 1 does not hold. Eliminating variance misallocation by means of component rotation precludes that there exists a PCA solution for the data at hand where each between-condition effect can be allocated on a separate single component. This is, however, not necessarily the case for any data set.

3.2 Misallocation of Variance in Combined within- and Between-Condition Components

Theorem 1 describes misallocation of variance as it can occur when PCA is performed for the total sample, i.e., across the levels of a between-condition factor. When separate within-condition components $\mathbf{c}^v_i, \ldots, \mathbf{c}^v_k$ are computed, the within-condition components $\mathbf{c}^v_i, \ldots, \mathbf{c}^v_k$ are completely unrelated to $\mathbf{c}^b$ so that within- and between-condition variance is completely disentangled. This yields the question under which constraints within- and between-condition components can be combined into a single component representing within- and between-condition variance unambiguously. Theorem 2 describes a constraint for the component loadings that implies $\mathbf{c} = [\mathbf{c}^v_i + \mathbf{c}^b_i, \quad \cdots \quad, \mathbf{c}^v_k + \mathbf{c}^b_k]$, i.e., that each component in **c** can be decomposed into a separate within- and between-condition component. This implies that no misallocation of variance occurs because each between-condition component is uniquely combined with another within-condition component.

Theorem 2. *If* $\mathbf{A}^v_i = \mathbf{A}^b, \cdots, \mathbf{A}^v_k = \mathbf{A}^b$, *then* $\mathbf{c} = \left[\mathbf{c}^v_i, \quad \cdots \quad, \mathbf{c}^v_k\right] + \left[\mathbf{c}^b_i, \quad \cdots \quad, \mathbf{c}^b_k\right]$.

Proof. Since $\mathbf{c}^b = [\mathbf{c}^b_i, \ldots, \mathbf{c}^b_k]$ Equation 6 can be written as

$$\mathbf{x} = \left[\mathbf{A}^v_i\mathbf{c}^v_i, \quad \cdots \quad, \mathbf{A}^v_k\mathbf{c}^v_k\right] + \mathbf{A}^b\left[\mathbf{c}^b_i, \quad \cdots \quad, \mathbf{c}^b_k\right], \tag{15}$$

inserting $\mathbf{A}^b$ for $\mathbf{A}^v_i, \cdots, \mathbf{A}^v_k$ into Equation 15 yields

$$\begin{aligned} \mathbf{x} &= \mathbf{A}^b\left[\mathbf{c}^v_i, \quad \cdots \quad, \mathbf{c}^v_k\right] + \mathbf{A}^b\left[\mathbf{c}^b_i, \quad \cdots \quad, \mathbf{c}^b_k\right] \\ &= \mathbf{A}^b\left[\mathbf{c}^v_i + \mathbf{c}^b_i, \quad \cdots \quad, \mathbf{c}^v_k + \mathbf{c}^b_k\right] = \mathbf{A}^b\mathbf{c}. \end{aligned} \tag{16}$$

This completes the proof. □

Thus, when the within-condition loading matrices at each condition level are identical to the between-condition loading matrix, this implies a component model where all components combine their respective within- and between-condition variance. Theorem 2 implies that no misallocation of variance occurs when each condition-specific loading pattern is identical to the between-condition loading pattern. When Theorem 2 holds, it would be possible to find a solution without variance misallocation by means of component rotation. Writing loading vectors $\mathbf{A}^v_i = [\mathbf{a}^v_{si}, \cdots, \mathbf{a}^v_{pi}], \mathbf{A}^b = [\mathbf{a}^b_s, \cdots, \mathbf{a}^b_p]$ and component score

vectors $\mathbf{c}^v_i = \begin{bmatrix} \mathbf{c}^v_{si} \\ \vdots \\ \mathbf{c}^v_{pi} \end{bmatrix}, \mathbf{c}^b_i = \begin{bmatrix} \mathbf{c}^b_{si} \\ \vdots \\ \mathbf{c}^b_{pi} \end{bmatrix}$ for the *s* to *p* components in Equation 16 yields

$$\mathbf{x} = \left[[\mathbf{a}_{si}^v, \cdots, \mathbf{a}_{pi}^v] \begin{bmatrix} \mathbf{c}_{si}^v \\ \vdots \\ \mathbf{c}_{pi}^v \end{bmatrix}, \quad \cdots \quad , [\mathbf{a}_{sk}^v, \cdots, \mathbf{a}_{pk}^v] \begin{bmatrix} \mathbf{c}_{sk}^v \\ \vdots \\ \mathbf{c}_{pk}^v \end{bmatrix} \right] + [\mathbf{a}_s^b, \cdots, \mathbf{a}_p^b] \left[\begin{bmatrix} \mathbf{c}_{si}^b \\ \vdots \\ \mathbf{c}_{pi}^b \end{bmatrix}, \quad \cdots \quad , \begin{bmatrix} \mathbf{c}_{sk}^b \\ \vdots \\ \mathbf{c}_{pk}^b \end{bmatrix} \right].$$

$$(17)$$

Note that the scores $\mathbf{c}_{si}^b$ are equal for each between-condition component s at each condition-level i. For convenience, the raw data reproduced from the first component are considered. This yields

$$\mathbf{x}_1^* = \left[\mathbf{a}_{1i}^v \mathbf{c}_{1i}^v, \quad \cdots \quad , \mathbf{a}_{1k}^v \mathbf{c}_{1k}^v \right] + \mathbf{a}_1^b \left[\mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^b \right]. \tag{18}$$

It follows from $\mathbf{a}_{1i}^v = \mathbf{a}_1^b, \cdots, \mathbf{a}_{1k}^v = \mathbf{a}_1^b$ that $\mathbf{c}_1 = [\mathbf{c}_{1i}^v + \mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^v + \mathbf{c}_{1k}^b]$ and that $\mathbf{x}_1^* = \mathbf{a}_1^b \mathbf{c}_1$. Thus, it is possible that only a subset of the within-condition loading matrices and between-condition loading matrices is identical and that this subset of components combines within- and between-condition variance. When there is only one between-condition component, i.e., $q^v = p > q^b = 1$, Equation 17 can be written as

$$\mathbf{x} = \left[[\mathbf{a}_{si}^v, \cdots, \mathbf{a}_{pi}^v] \begin{bmatrix} \mathbf{c}_{si}^v \\ \vdots \\ \mathbf{c}_{pi}^v \end{bmatrix}, \quad \cdots \quad , [\mathbf{a}_{sk}^v, \cdots, \mathbf{a}_{pk}^v] \begin{bmatrix} \mathbf{c}_{sk}^v \\ \vdots \\ \mathbf{c}_{pk}^v \end{bmatrix} \right] + \mathbf{a}_1^b \left[\mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^b \right]. \tag{19}$$

Theorem 3 describes constraints for the loadings that are compatible with a model combining a single between-condition component with the first within-condition component.

Theorem 3. *If* $q^v = p > q^b = 1, and$ $\mathbf{a}_{1i}^v = \mathbf{a}_1^b, \cdots, \mathbf{a}_{1k}^v = \mathbf{a}_1^b$ *and* $\mathbf{a}_{si}^v \neq \mathbf{a}_1^b, \cdots, \mathbf{a}_{sk}^v \neq \mathbf{a}_1^b$,

then $\mathbf{c}_1 = \left[\mathbf{c}_{1i}^v + \mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^v + \mathbf{c}_{1k}^b \right]$ *and* $\mathbf{c}_s \neq \left[\mathbf{c}_{si}^v + \mathbf{c}_{si}^b, \quad \cdots \quad , \mathbf{c}_{sk}^v + \mathbf{c}_{sk}^b \right]$, *for* $s = 2, ..., p$.

Proof. For $\mathbf{a}_{1i}^v = \mathbf{a}_1^b, \cdots, \mathbf{a}_{1k}^v = \mathbf{a}_1^b$ Equation 18 can be written as

$$\begin{aligned} \mathbf{x} &= \left[[\mathbf{a}_1^b, \mathbf{a}_{si}^v, \cdots, \mathbf{a}_{pi}^v] \begin{bmatrix} \mathbf{c}_{1i}^v \\ \mathbf{c}_{si}^v \\ \vdots \\ \mathbf{c}_{pi}^v \end{bmatrix}, \quad \cdots \quad , [\mathbf{a}_1^b, \mathbf{a}_{sk}^v, \cdots, \mathbf{a}_{pk}^v] \begin{bmatrix} \mathbf{c}_{1k}^v \\ \mathbf{c}_{sk}^v \\ \vdots \\ \mathbf{c}_{pk}^v \end{bmatrix} \right] + \mathbf{a}_1^b \left[\mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^b \right] \\ &= \left[[\mathbf{a}_{si}^v, \cdots, \mathbf{a}_{pi}^v] \begin{bmatrix} \mathbf{c}_{si}^v \\ \vdots \\ \mathbf{c}_{pi}^v \end{bmatrix}, \quad \cdots \quad , [\mathbf{a}_{sk}^v, \cdots, \mathbf{a}_{pk}^v] \begin{bmatrix} \mathbf{c}_{sk}^v \\ \vdots \\ \mathbf{c}_{pk}^v \end{bmatrix} \right] + \mathbf{a}_1^b \mathbf{c}_1 , \end{aligned} \tag{20}$$

with $\mathbf{c}_1 = \left[\mathbf{c}_{1i}^v + \mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^v + \mathbf{c}_{1k}^b \right]$, *for* $s = 2, ..., p$.

This completes the proof. □

The identity of the loading patterns of the first unrotated within- and between-condition components is a sufficient constraint for the allocation of the between- and within-condition variance on a common component. Theorem 4 describes a somewhat relaxed constraint that is based on an identical shape of the loadings of the first within- and between-condition components but allows for a different scale.

Theorem 4. *If* $q^v = p > q^b = 1,$ *and* $\theta_i \mathbf{a}_{1i}^v = \mathbf{a}_1^b, \cdots, \theta_k \mathbf{a}_{1k}^v = \mathbf{a}_1^b$ *and* $\theta_i \mathbf{a}_{si}^v \neq \mathbf{a}_1^b, \cdots, \theta_k \mathbf{a}_{sk}^v \neq \mathbf{a}_1^b,$

then $\mathbf{c}_s \neq \left[\theta_i \mathbf{c}_{si}^v + \mathbf{c}_{si}^b, \quad \cdots \quad , \theta_k \mathbf{c}_{sk}^v + \mathbf{c}_{sk}^b \right]$, *for* $s = 2, ..., p$ *and* $\theta_i > 0, \cdots, \theta_k > 0$. *and* $\mathbf{c}_1 = \left[\theta_i \mathbf{c}_{1i}^v + \mathbf{c}_{1i}^b, \quad \cdots \quad , \theta_k \mathbf{c}_{1k}^v + \mathbf{c}_{1k}^b \right]$

Proof. For $\theta_i \mathbf{a}_{1i}^v = \mathbf{a}_1^b, \cdots, \theta_k \mathbf{a}_{1k}^v = \mathbf{a}_1^b$ Equation 18 can be written as

$$\mathbf{x} = \left[[\theta_i \mathbf{a}_1^b, \mathbf{a}_{si}^v, \cdots, \mathbf{a}_{pi}^v] \begin{bmatrix} \mathbf{c}_{1i}^v \\ \mathbf{c}_{si}^v \\ \vdots \\ \mathbf{c}_{pi}^v \end{bmatrix}, \quad \cdots \quad , [\theta_k \mathbf{a}_1^b, \mathbf{a}_{sk}^v, \cdots, \mathbf{a}_{pk}^v] \begin{bmatrix} \mathbf{c}_{1k}^v \\ \mathbf{c}_{sk}^v \\ \vdots \\ \mathbf{c}_{pk}^v \end{bmatrix} \right] + \mathbf{a}_1^b \left[\mathbf{c}_{1i}^b, \quad \cdots \quad , \mathbf{c}_{1k}^b \right] \tag{21}$$

$$= \left[[\mathbf{a}_{si}^v, \cdots, \mathbf{a}_{pi}^v] \begin{bmatrix} \mathbf{c}_{si}^v \\ \vdots \\ \mathbf{c}_{pi}^v \end{bmatrix}, \quad \cdots \quad , [\mathbf{a}_{sk}^v, \cdots, \mathbf{a}_{pk}^v] \begin{bmatrix} \mathbf{c}_{sk}^v \\ \vdots \\ \mathbf{c}_{pk}^v \end{bmatrix} \right] + \mathbf{a}_1^b \mathbf{c}_1 ,$$

$$with \ \mathbf{c}_1 = \left[\theta_i \mathbf{c}_{1i}^v + \mathbf{c}_{1i}^b, \quad \cdots \quad , \theta_k \mathbf{c}_{1k}^v + \mathbf{c}_{1k}^b \right], \ for \ s = 2, ..., p.$$

This completes the proof. □

Theorem 4 shows that condition-specific loading patterns that have the same shape, but a different scale are compatible with a model where a single between-condition component is unambiguously allocated on a single within-condition component.

4. Discussion

According to Wood and McCarthy (1984) misallocation of variance occurs when a single between-condition effect that can in principle be allocated on a single PCA component is allocated on more than one component in a given PCA solution. The present study describes constraints that are to be imposed on the component loading matrices in order to avoid misallocation of variance. The following conclusions can be drawn: When a single between-condition effect is allocated on a single component of an initial PCA solution, any rotation of these initial components will result in a misallocation of variance (Theorem 1). This is an algebraic demonstration of what has been discussed elsewhere (Scharf & Nestler, in press; Dien, 2010; Beauducel & Leue, 2015; Beauducel, 2018), namely that, at the level of combined within- and between-condition components, the misallocation of variance is directly related to component rotation. However, component rotation can only result in an optimal allocation of between-condition variance when such a rotational solution exists for a given data set.

Since it has been proposed to perform separate PCAs at each level of a condition factor (Barry et al., 2016), the consequences of this procedure for misallocation of variance were explored. When a PCA is calculated at each level of a condition factor and when a PCA is calculated for a single between-condition factor, an unambiguous allocation of the between-condition variance on a single component combining within- and between-condition variance is possible when the within-condition component loadings have the same shape, even when their scale is different (Theorem 3 and 4). Thus, only when the constraints of Theorem 3 and 4 hold for a given data set, it would be possible to find the solution with optimal allocation of between-condition variance by means of component rotation.

Theorem 3 and 4 also imply that separate PCAs at each level of a condition-factor are not necessarily a way to avoid or eliminate misallocation of variance. When different loading shapes occur at each level of a condition factor in separate PCAs, this indicates that misallocation of variance would occur when the separate components are combined into within- and between variance components. In contrast, when the loading shape is similar in the different PCAs with larger or smaller loadings at each level of the condition factor, the components can be combined into within- and between-components without misallocation of variance.

Finally, it follows from Theorem 4 that perfect congruence coefficients (Tucker, 1951; Wrigley & Neuhaus, 1955) of the loadings of respective components at different levels of the condition factor are not a necessary condition for optimal variance allocation because congruence coefficients also refer to the similarity of the loading magnitude. For optimal variance allocation, a perfect Pearson correlation of the loadings of the respective components at different levels of the condition factor would be necessary.

References

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology, 106*, 50-64. http://dx.doi.org /10.1016/j.ijpsycho.2016.06.003

Beauducel, A., & Debener, S. (2003). Misallocation of variance in event-related potentials: simulation studies on the effects of test power, topography, andbaseline-to-peak versus principal component quantifications. *Journal of Neuroscience Methods, 124*, 103–112. http://dx.doi.org/10.1016/S0165-0270(02)00381-3.

Beauducel, A. (2018). Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific Procrustes-rotation. *Journal of Neuroscience Methods, 295*, 20-36. https://doi.org/10.1016 /j.jneumeth.2017.11.011

Beauducel, A., & Leue, A. (2015). Controlling for experimental effects in event-related potentials by means of principal component rotation. *Journal of Neuroscience Methods, 239*, 139–147. http://dx.doi.org /10.1016/j.jneumeth.2014.10.008

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology, 47*, 170–183. https://doi.org/10.1111/j.1469-8986.2009.00885.x

Dien J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax vs. Infomax rotations. *Human Brain Mapping, 28*, 742–763. https://doi.org/10.1002/hbm.20304

Harman, H. H. (1976). *Modern factor analysis (3rd ed.).* Chicago, IL: University of Chicago Press.

Jolliffe, I. T., & Cadima J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions Royal Society A, 374*: 20150202. http://dx.doi.org/10.1098/rsta.2015.0202

Kayser, J., & Tenke, C. E. (2003). Optimizing principal component analysis PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Psychophysiology, 114*, 2307–2325. https://doi.org/10.1016/S1388-2457(03)00241-4

Kayser, J., & Tenke, C. E. (2005). Trusting in or breaking with convention: towards a renaissance of principal component analysis in electrophysiology. *Clinical Neurophysiology, 116*, 1747–1753. https://doi.org/ 10.1016/j.clinph.2005.03.020

Scharf, F., & Nestler, S. (in press). Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: Insights from an inter-factor covariance decomposition. *International Journal of Psychophysiology.* https://doi.org/10.1016/j.ijpsycho.2018.03.019

Tucker, L. R. (1951). A Method for Synthesis of Factor Analysis Studies *(Personnel Research Section Report No. 984).* Department of the Army, Washington, DC.

Wood, C. C., & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and Clinical Neurophysiology, 59*, 249–260. https://doi.org/10.1016/0168-5597(84)90064-9

Wrigley, C. S., & Neuhaus, J. O. (1955). The matching of two sets of factors. *American Psychologist, 10*, 418–419.