A tutorial on the use of temporal principal component analysis in developmental ERP research – Opportunities and challenges

Florian Scharf [a],[*],[1], Andreas Widmann [b],[c],[2], Carolina Bonmassar [b],[3], Nicole Wetzel [b],[d],[e],[4]

[a] *University of Münster, Germany*
[b] *Leibniz Institute for Neurobiology, Magdeburg, Germany*
[c] *Leipzig University, Germany*
[d] *Center for Behavioral Brain Sciences, Magdeburg, Germany*
[e] *University of Applied Sciences Magdeburg-Stendal, Germany*

A B S T R A C T

Developmental researchers are often interested in event-related potentials (ERPs). Data-analytic approaches based on the observed ERP suffer from major problems such as arbitrary definition of analysis time windows and regions of interest and the observed ERP being a mixture of latent underlying components. Temporal principal component analysis (PCA) can reduce these problems. However, its application in developmental research comes with the unique challenge that the component structure differs between age groups (so-called measurement non-invariance). Separate PCAs for the groups can cope with this challenge. We demonstrate how to make results from separate PCAs accessible for inferential statistics by re-scaling to original units. This tutorial enables readers with a focus on developmental research to conduct a PCA-based ERP analysis of amplitude differences. We explain the benefits of a PCA-based approach, introduce the PCA model and demonstrate its application to a developmental research question using real-data from a child and an adult group (code and data openly available). Finally, we discuss how to cope with typical challenges during the analysis and name potential limitations such as suboptimal decomposition results, data-driven analysis decisions and latency shifts.

Electroencephalography (EEG) is a widely used time sensitive and noninvasive method to measure brain activity. More specifically, to enhance the understanding of neuronal and cognitive development throughout the lifetime, developmental researchers are often interested in event-related potentials (ERPs) as a measure of brain activity occurring time-locked to an event of interest (e.g., the presentation of a stimulus). ERPs undergo considerable developmental changes throughout childhood; for instance, there are substantial differences in amplitudes, polarities, latencies, and topographies of ERPs to auditory stimuli between newborns, toddlers, children, and adults (for a review, see Wunderlich et al., 2006). Fig. 1 depicts the ERPs to an auditory stimulus in two groups of children and an adult group. The children's ERPs are characterized by a dominant positive peak which is followed by a negative peak whereas the adults' ERP shows a prototypical series of peaks including two positive and two negative peaks (i.e.,

P1-N1-P2-N2). With increasing age, children's ERPs develop more and more towards the prototypical mature adult ERP. As different cognitive functions are associated with different components in the ERP (e.g., Luck and Kappenman, 2012), ERPs allow for profound insights into the development of cognitive functions and their underlying neuronal networks.

A typical ERP analysis proceeds in the following steps (e.g., Luck, 2014): After recording a continuous EEG from multiple electrode sites during an experiment, the continuous EEG signal is filtered and cleaned from frequent artifacts such as eye-movement and blink activity (e.g., Chaumon et al., 2015) to reduce the amount of non-neural activity in the EEG signal. The EEG signal is then cut into epochs around the events of interest and averaged across all repetitions of each event separately for each event type (e.g., different stimuli) and electrode site per participant. The resulting *ERP dataset* containing data from each
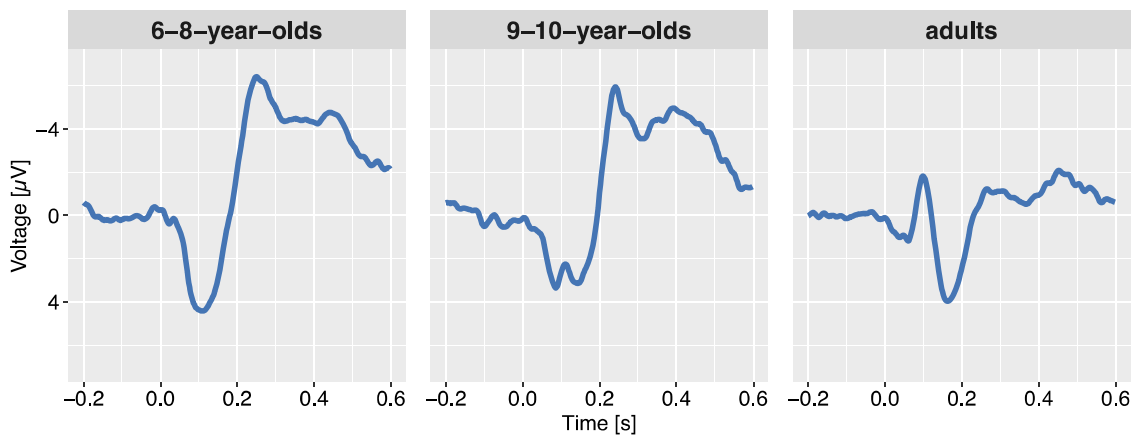
Fig. 1. Event-related potentials to a to-be-ignored auditory stimulus for two groups of children (left panel: 6–8-year-olds, middle-panel: 9–10-year-olds) and a group of adults (right panel). Negative amplitudes are displayed upwards. In the youngest children a broad positive deflection is followed by a large negative deflection. This pattern changes typically with age and results in an alternating sequence of positive and negative deflections.

sampling point × participant × electrode × event type combination is then subjected to subsequent statistical analyses to identify amplitude, latency, polarity, or topographic differences between events. When this ERP dataset is further averaged across participants (especially for data visualization purposes), the resulting sampling point × electrode × event type dataset is referred to as the *grand-average ERP (dataset)*. Before statistical inference is conducted, the ERP dataset is typically further aggregated to reduce noise and complexity. For instance, ERP amplitudes are commonly averaged within specific time windows and across specific electrode sites (i.e., *region of interest*, ROI) to quantify the activity of the ERP component of interest. This aggregated amplitude measure is then used as dependent variable for statistical inference (e.g., via ANOVAs).

This "traditional" data-analytic approach for ERPs suffers from two major, partially related problems: (1) The definition of analysis time windows as well as ROIs often proceeds in a relatively arbitrary manner (please see Luck and Gaspelin, 2016; for a thorough discussion) or is potentially afflicted by double dipping (Kriegeskorte et al., 2009). (2) The observed signal at the scalp is a mixture of underlying signals generated in the brain. The underlying signals overlap to an unknown extent both in the temporal domain (i.e., multiple sources contribute to the voltage at a specific sampling point) and in the spatial domain (i.e., multiple sources contribute to the voltage at a specific electrode site) – making the peaks of the observed signal a poor indicator of the underlying signals. Together, these problems can considerably bias estimates of amplitude, latency, and topography, especially when these estimates are based on the peaks of the observed signal. Therefore, researchers interested in ERPs can benefit a lot from using multivariate *decomposition* methods. These methods aim to describe the observed ERP data as a function of a set of underlying signals and provide objective characterizations of the time courses of these underlying signals. Here, we focus on temporal principal component analysis (PCA) which belongs to a class of procedures decomposing the data based on statistical properties such as the covariance between the voltages at the sampling points (Dien and Frishkoff, 2005; Donchin and Heffley, 1978). Temporal PCA for ERP data aims (a) to identify and disentangle the constituent components of the ERP and (b) to provide dependent measures for statistical analyses which are derived in an objective and transparent way. PCA is particularly suited for the investigation of ERPs in developmental populations reducing problems due to the enhanced noise level (Dien, 2012).

The overarching goal of this article is to enable the reader to conduct a PCA-based ERP analysis of amplitude differences between two age groups on his or her own and to cope with specific challenges when comparisons between children's and adults' ERPs are of interest for the research question. To this end, (1) we will introduce the benefits of a PCA-based approach using a simplistic simulated ERP for illustrative purposes, (2) we will demonstrate a real-data PCA analysis in a step-by-step manner using open source software (including example code and an open dataset), (3) we will explain the challenges which arise from differential ERP structures between age groups and provide recommendations how to cope with them, and lastly, (4) we will briefly compare PCA to other decomposition approaches and discuss potential limitations. In the following, we assume that the reader is familiar with essentials of the ERP technique such as its neural origin, typical research objectives, and pre-processing pipelines (for more basic introductions, see, e.g., Bradley and Keil, 2012; Luck, 2014). In addition, to work with the accompanying *R* code, we recommend that the reader is proficient in at least one high-level programming language (e.g., MATLAB, *R*, or Python).

1. Benefits of PCA-based ERP analyses

1.1. ERPs versus underlying components

The observed ERP signal at the scalp reflects simultaneous activity – mostly post-synaptic potentials – from a large number of neurons in the brain (Nunez and Srinivasan, 2006). Much like a movie is a time series of 2-dimensional images of a 3-dimensional world, an ERP signal can be construed as a time series of the electric potential across the electrode grid on the scalp surface – which is only a projection of the 3-dimensional signal in the brain. Because multiple sources can be active at a certain point in time and can project to the same electrode sites, the signal measured at the scalp suffers from considerable temporal and spatial overlap. Furthermore, the source signal is transmitted to the sensors via volume conduction through several layers with differential resistive properties resulting in the signal at the scalp being "smeared" spatially but also temporally because the spatial overlap of source activity at the sensors also limits the temporal precision which can be achieved in practice (see Burle et al., 2015; for more detailed discussion and possible remedies). In addition, ERPs are typically computed by averaging over many repetitions of the event of interest assuming strict temporal coherence with the event of interest (i.e., no latency jitter). Because brain maturation is not necessarily on the same level in children of the same age, this can have an even greater impact in developmental
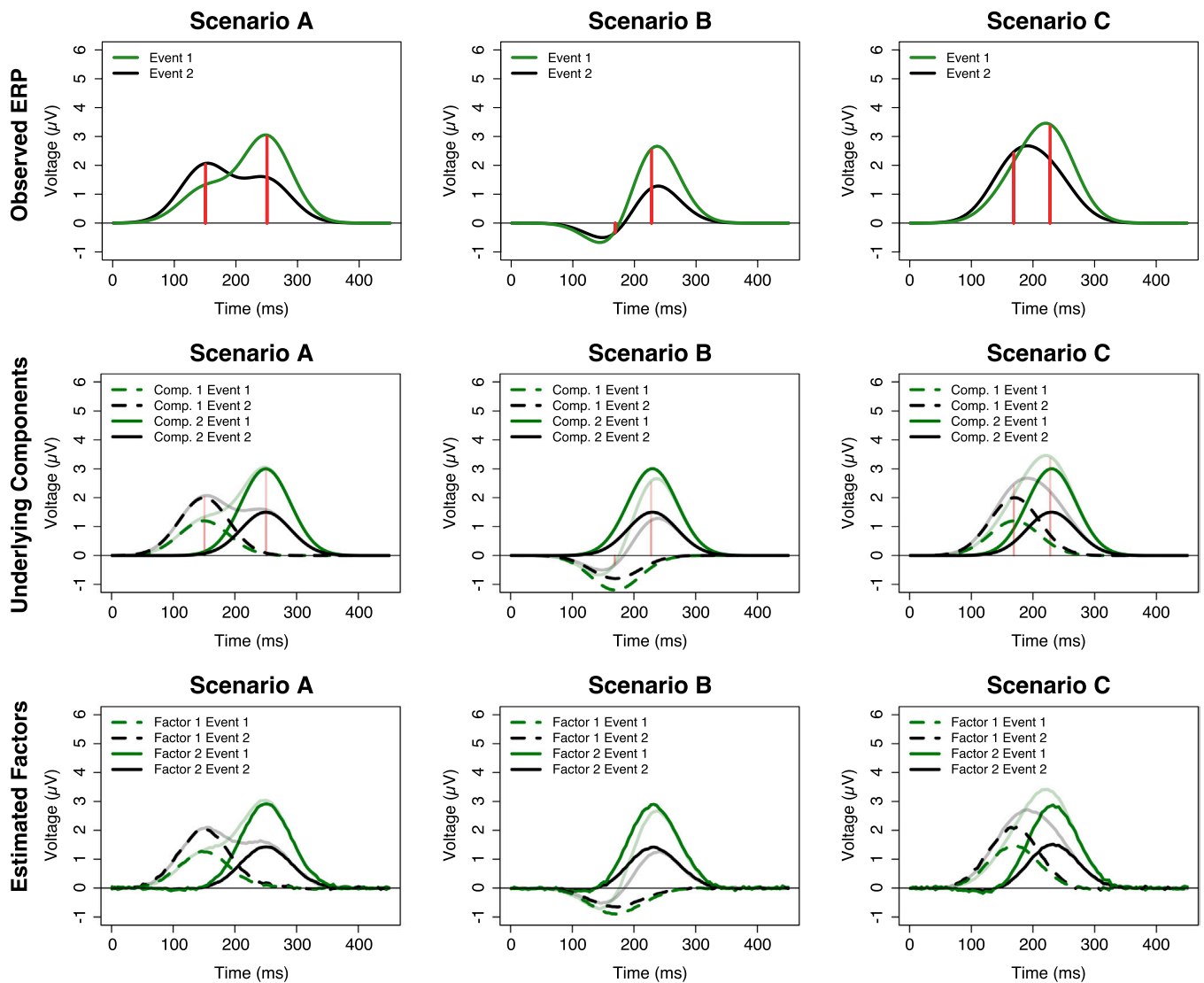
Fig. 2. Illustration of three simplistic ERPs with mild (A) and severe (B & C) temporal overlap of the underlying components. The upper panel depicts the simulated ERPs "elicited" by two events (black and green lines); the vertical red lines mark the true peak of the underlying components shown in the middle panel. The lower panel depicts the factors estimated by the temporal PCA approach explained in the upcoming sections. The estimated factors (lower panel) resemble the underlying components (middle panel) well. Please note that results may be less optimal under more challenging circumstances (e.g., more temporal overlap).

studies. Therefore, in sum, ERP data are a spatially and temporally imprecise representation of the source activity in the brain (Burle et al., 2015). The inherent loss of information regarding time courses, 3-dimensional position and direction of the source signals makes an exact reconstruction of the source signal impossible (that is, without further information and assumptions – this is typically referred to as the *inverse problem* of M/EEG). To appreciate this state of facts and avoid conceptual confusion, we will refer to the ground truth signals at the scalp electrodes that would be recovered by a decomposition technique under *ideal* circumstances as ERP *components*. The term ERP component has been used with many different connotations. In a broad sense, we define a component as coherent voltage deflection which contributes to the observed ERP at the scalp in a replicable way across experiments. A component can be characterized based on: the time of its occurrence within the epoch, its typical amplitude, its sources in the brain and/or its functional interpretation (i.e., its typical behavior under experimental manipulations).

Without further detailed knowledge regarding potential sources in the brain and how the signal spreads through brain, skull, and scalp, the best that researchers can hope for is to gain knowledge regarding

functional interpretations of these components. Therefore, a certain functional independence, which can, for instance, be established by unique sensitivity towards certain experimental manipulations, is often considered a defining property of a component (e.g., Luck, 2014b). The spatial and temporal superposition of underlying components poses a great challenge for researchers who want to interpret the results from ERP experiments because it conceals how many components constitute the observed ERP and how much each individual component contributes to the electric potential at a specific time point and electrode site. That is, temporal and spatial overlap biases the functional interpretation of the ERP. Consequently, inferences regarding differences in component activity between events can be substantially biased when they are based on the observed ERP (Luck, 2005a). In the worst case, this might even lead to erroneous conclusions regarding functional interpretations and underlying processes.

Fig. 2 illustrates three scenarios which mimic typical ERPs in a simplistic manner. These ERPs (upper panel) were simulated so that it is known that there are two underlying components (middle panel). Suppose there are two events for which we wish to investigate differences in ERPs. Due to the simulated nature of the example, we know that both
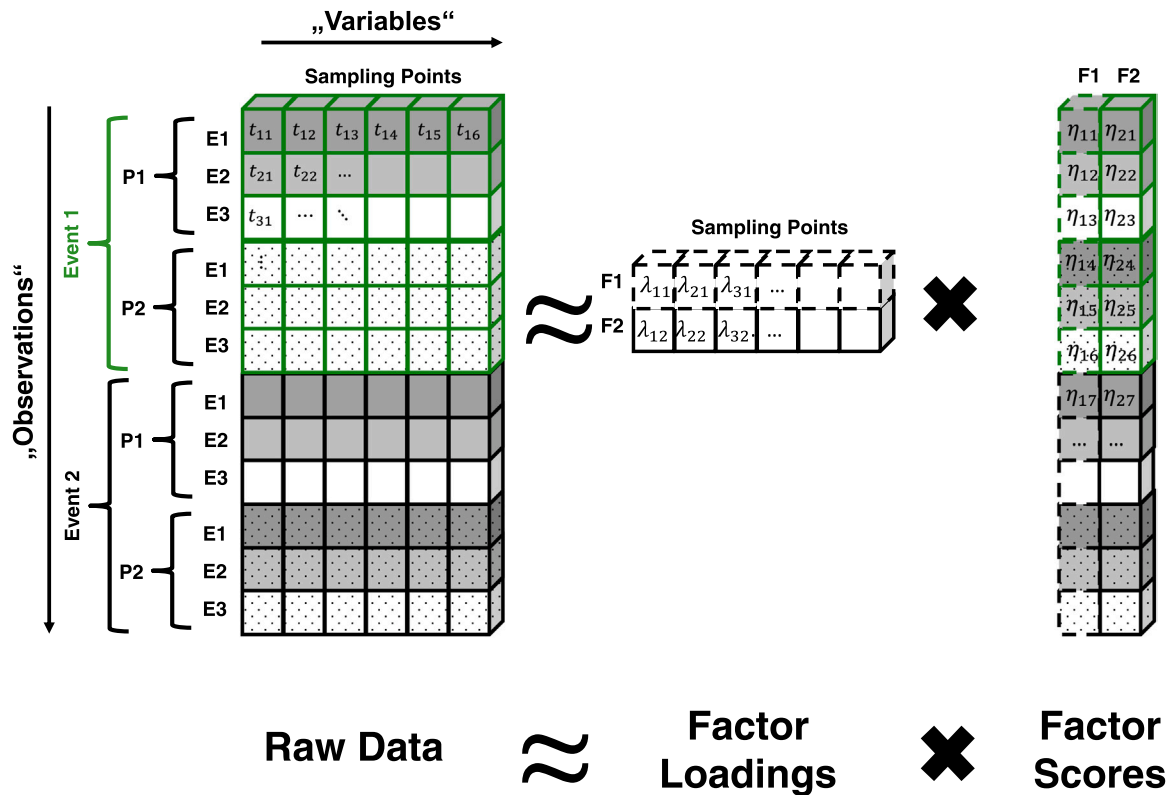
Fig. 3. Prototypical data structure and decomposition into factor loadings and factor scores. P = Participant, E = Electrode, t = voltage at a specific sampling point for a specific observation, F = Factor, λ = factor loading (representing the time course of the activity for each factor), η = factor score (representing the level of activity of each factor for a specific observation). The residuals were left out for the sake of comprehensibility; but typically loadings and scores reconstruct the participant average data only approximately.

components have reverse response patterns to the events: Component 1 has a more positive response to Event 2 compared to Event 1 whereas Component 2 has a more negative response to Event 2 compared to Event 1. When the overlap between components is mild (Scenario A), the observed ERP reflects the underlying components sufficiently because the two peaks of the signal roughly resemble the two peaks of the components and the condition effects resemble the reversed pattern of the underlying components. However, when the overlap between components is large (Scenarios B & C), this is not the case. Although the presence of a positive and a negative peak in Scenario B could be interpreted as a hint towards more than one underlying component, the peaks of the signal are poor indicators of the peaks of the underlying components (marked by the vertical red lines). Consequently, any analysis of amplitude and latency differences based on the peaks would suffer from severe biases. This is especially apparent for Component 1 for which both its true amplitude difference between Events and its latency deviates considerably from the peak-based measure. Scenario C illustrates a case in which peak-based measures are even more fallible because the ERP consists of one broad peak instead of two. In this case, neither the number of underlying components nor the nature of the event effect are as obvious. Solely based on the shape of the ERP, one might easily conclude that there is only one component and that its latency rather than its amplitude differs between events. This conclusion would severely misrepresent the ground truth and would probably result in inappropriate conclusions regarding the functional interpretation of the effect.

The three scenarios in Fig. 2 demonstrate the central limitation of a data-analytic approach that is based on the observed peaks of the signal

(e.g., averaging in a time window of interest): It assumes that the observed peaks are valid indicators of number, time course, and topography of the underlying components. This is only true when we can assume sufficient temporal and/or spatial separation of components. This limitation is well-known and severe (Luck, 2014a) because it is easily conceivable that many ERP components are never elicited in sufficient isolation to investigate them appropriately. In absence of sufficient separation, the choice of analysis time windows is arbitrary and prone to capitalizing on chance (Luck and Gaspelin, 2016), resulting in reduced reliability (and therefore, reduced statistical power) and potentially in erroneous conclusions regarding experimental effects of interest (Beauducel et al., 2000; Beauducel and Debener, 2003).

1.2. Theoretical introduction to temporal PCA

1.2.1. The temporal PCA model

These outlined limitations of peak-based approaches motivated a search for better suited analytic approaches for ERPs which explicitly acknowledge the mixture of underlying signals. Several *multivariate* analysis methods have been proposed with the purpose of estimating the unobserved underlying components based on *combined* statistical properties of ERP datasets. Among these methods are Principal Component Analysis (PCA), Independent Component Analysis (ICA), Wavelet-based decompositions, multimode PCAs, and source reconstruction methods (Achim and Bouchard, 1997; Chapman and McCrary, 1995; Dien, 2012; Donchin, 1966; Donchin and Heffley, 1978; Groppe et al., 2008; Möcks, 1988b; Mørup et al., 2006a; Slotnick, 2005). Here, we focus on temporal PCA which decomposes the observed ERPs based solely on the statistical

associations between the sampling points. We will briefly discuss the relationship between temporal PCA and other proposed approaches in the last section of this article.

The aim of temporal PCA is to decompose the observed signal into a set of underlying *factors* summarizing sampling points with a similar/coherent activity pattern across participants, electrodes, and conditions. Ideally, one would like to view these *factors* as *estimates* of the true underlying components. However, there is no guarantee that the decomposition is always successful in disentangling *all* underlying components. To avoid any confusion about this fact, we refer to the estimated underlying signals as *factors* and to the ideally recovered signals as *components* throughout this article (deviating from the typical use of the term "component" in PCA and ICA).

The lowest panel in Fig. 2 illustrates a 2-factor temporal PCA solution for each scenario. The transparent lines show the grand-average ERP for both events and the opaque lines are the factor-wise reconstruction of the ERP – summed together, both factors reconstruct the observed ERP. When comparing the estimated factors (lowest panels) with the underlying components from which the data were simulated (middle panels), one can clearly see that temporal PCA was quite successful in recovering the underlying components because the factors resemble the components very well even under severe temporal overlap (Scenarios B & C). Although the decomposition is not always perfect, especially in the presence of strong temporal and spatial overlap (Beauducel and Debener, 2003; Dien, 1998; Dien et al., 2005; Kayser and Tenke, 2003; Möcks and Verleger, 1986; Scharf and Nestler, 2018), one should keep in mind that a simple peak-based approach would have obscured the nature of the underlying amplitude effects, especially in Scenario C. We hope to convince the reader that this data-driven and therefore much more objective characterization of the underlying signals makes temporal PCA a valuable tool for ERP researchers. Furthermore, previous research indicates that temporal PCA can provide more reliable quantifications of ERPs and more accurate estimates of experimental effects (Arruda et al., 1996; Beauducel et al., 2000; Beauducel and Debener, 2003). How does temporal PCA achieve this decomposition? A fully-fledged technical introduction into the mathematical foundations of PCA is beyond the scope of this article but bottom-up introductions including all necessary mathematical basics are available elsewhere (e. g., Gorsuch, 1983; Mulaik, 2010). In the following, we want to provide the reader with an intuitive understanding of the PCA parameters, their estimation and the central assumptions underlying temporal PCA. Temporal PCA typically operates on an individual average dataset (i.e., averaged across trials within participants during pre-processing) in which each row represents a specific electrode site from an individual participant and from a specific experimental condition, and the columns contain the sampling points. Each cell defined by row and column reflects the mean voltage averaged across trials for a specific participant, electrode, condition, and latency. That is, the sampling points are treated as *variables* of interest and electrode sites, conditions, and participants are stacked as *observations* in the rows. Note that this is a variation of how PCA is taught in typical statistics 101 courses in which variables are answers to items in questionnaires and the observations are participants. Fig. 3 illustrates the data structure which enters the computation of a temporal PCA.

The temporal PCA model decomposes the participant average data into two sets of coefficients called *factor loadings* and *factor scores,* respectively (see Fig. 3). Roughly, the factor loadings describe the time course of a factor and are fixed across all electrodes, participants, and conditions. The factor scores describe the contribution (i.e., level of activity) of the factor to the voltage for each observation. Mathematically, this can be summarized in the fundamental equation of PCA in which the voltage at each sampling point is denoted as t_{ij} where the indices represent any specific observation i and any specific sampling point j. The voltage at each sampling point is approximated by a weighted linear combination of the factor *scores* η_{ki} of a fixed number of m factors (with k indicating any specific factor). That is, the factor scores

are multiplied with the respective factor loading of a sampling point and summed up:

$$t_{ij} \approx \lambda_{j1} \cdot \eta_{1i} + \lambda_{j2} \cdot \eta_{2i} + \ldots + \lambda_{jm} \cdot \eta_{mi} \tag{1}$$

The factor loadings λ_{jk} are weights which describe how much and with which sign a specific factor contributes to the voltage at sampling point j ("variable" $t_{.j}$): high loadings indicate a strong influence, and zero loadings indicate no influence of the factor. The factor scores reflect how much a specific factor contributes to the voltage at a specific electrode site for a specific participant in a specific experimental condition and are usually normalized to have unit variance. The unstandardized factor loadings can be interpreted as regression weights indicating by how many microvolts the predicted voltage at a sampling point changes when the score of the respective factor is increased by 1 standard deviation (while keeping all other factor scores constant). The factor scores can be directly used as quantifications of a factor's contribution to the observed ERP. That is, they can be subjected to statistical tests, or they can be averaged across participants to compute a "grand-average" factor score for each electrode and condition, for instance, to illustrate the topography of a factor.

The signs and scales of factor loadings and factor scores are arbitrary and some identifying restrictions are necessary (Mulaik, 2010). In the context of ERP analyses, some specific conventions regarding the factor loadings and factor scores have been established (see also Dien and Frishkoff, 2005) in order to have an interpretable relationship between the signs of the factor scores and the polarity of the ERP signal: (1) the majority of the factor loadings of each factor should be positive, (2) the factor loadings are reported in the original unit µV (i.e., *unstandardized*[5]), (3) the factor scores are normalized[6] (i.e., have unit variance) but they are not centered (i.e., they have a non-zero mean), and (4) the mean (or intercept) of the voltage at a sampling point is zero when no factor is active (i.e., when all loadings and/or scores are zero). This identification strategy maintains an interpretable relationship between factor scores and raw voltage (see, e.g., Kline, 2016, for a thorough explanation in the context of structural equation modeling): That is, for sampling points with a positive loading, a positive/negative factor score reflects positive/negative voltage of the factor at the respective electrode site in a specific participant and condition, and a factor score of zero indicates that a factor is inactive for a certain observation.

A very important implication of the model described above is that factor loadings are assumed to be constant across participants, electrodes, and conditions whereas the factor scores may vary. This assumption is typically referred to as *measurement invariance* in the factor analysis literature and its violation can have severe consequences for the performance of temporal PCA (Beauducel and Hilger, 2018; Meredith, 1993; Möcks, 1986). In general, equal factor loadings across electrode sites can be assumed due to the electrophysiological nature of the ERP signal and it is physiologically plausible to describe the ERP signal as weighted linear combination of the factors (Nunez and Srinivasan, 2006). However, systematic latency differences, latency jitter, or even different component structures can make it implausible to assume equal time courses across participants, conditions, or groups (see Molenaar and Campbell, 2009; for a general discussion).

Whenever measurement invariance is justifiable, it is preferable to analyze data in a *combined* PCA across all observation, because this

---

[5] Please note that different terminologies have been used to differentiate scaling choices in PCA for ERP data (Dien, 2006, 2012; Dien et al., 2005; Kayser and Tenke, 2003, 2005, 2006). For instance, what we call „unstandardized" scale has sometimes been referred to as "microvolt" or "covariance-based" scale.

[6] *Normalization* can have different meanings. Throughout this paper, we refer to a variable as normalized when it was linearly transformed to have a variance of 1 but was not centered to have a mean of 0 to distinguish it from z-standardization which results in a mean of 0 and a variance of 1.

naturally allows for mean comparisons of the factor scores (Beauducel and Hilger, 2018; Mellenbergh, 1989; Meredith, 1993; Nesselroade, 2007; Van De Schoot et al., 2015). When substantially different component structures are conceivable, separate PCAs can help to evaluate in how far the factor structures differ. The analyst should proceed with separate PCAs in case of *major* differences between component structures (e.g., different number of factors, some factors are only present in one PCA, strong latency shifts, etc.). Specific examples in which conducting separate PCAs might be beneficial include comparisons across different modalities (e.g., auditory vs. visual ERPs; see Dien, 2012; for a similar argument), comparisons across experimental conditions with very different factor structures (e.g., Barry et al., 2016), and comparisons between groups with expectably different ERP component structures such as children versus adults or patients versus healthy controls (e.g., Barry et al., 2014).

Separate PCAs violate the assumption of measurement invariance making it difficult to decide which mean comparisons are meaningful. In principle, there are strong theoretical reasons to believe that at least some factors from separate PCAs actually represent the same underlying component (see Nesselroade and Molenaar, 2016b, 2016a; for a detailed discussion). For instance, both adults and children can be expected to have a P2 component in their ERPs and a factor representing the P2 component should occur in the adult as well as in the child PCA. If such matching factors can be identified based on theoretical considerations, quantitative comparisons between both groups can be conducted. Separate PCAs can provide some insights into differential factor structures (e.g., due to latency shifts), but they must be interpreted with caution due to the rotational indeterminacy. Different factor structures could either be caused by different component structures or be an artifact of differential rotated solutions. For instance, the P2 component could be conflated with a P3 component in one group but not in the other (cf. section *Challenges*). Despite these complications, separate PCAs can provide unique analysis opportunities when measurement invariance is unlikely fulfilled. This is especially relevant in developmental populations where ERP structures can vary tremendously between age groups.

1.2.2. Estimation of the temporal PCA model

A large and still growing variety of methods exists to estimate the factor loadings and factor scores (Conti et al., 2014; Mulaik, 2010; Trendafilov, 2014) which differ in many details of their implementation. However, some basic tasks for an estimation procedure can be identified: (1) determination of the number of factors, (2) estimation of the factor loadings, and (3) estimation of the factor scores. In the following, we will describe how these tasks have been solved traditionally in a stepwise manner.

Determination of the Number of Factors. The most fundamental question when estimating PCA is how many factors should be extracted. Ideally, the number of factors would coincide with the number of underlying components so that each factor is an estimate of one underlying component. Because the number of components is typically unknown to ERP researchers, *extraction criteria* can be used to decide how many factors should be extracted (Mulaik, 2010). Most of these criteria are based on the overall fit of the model, for instance, the amount of variance explained by each individual factor or all factors together. Put simply, a factor model is estimated with as many factors as sampling points in which the factors are sorted in descending order of explained variance. The number of factors is then chosen so that each individual factor explains a substantial amount of variance. A popular criterion to identify substantial factors is the so-called *Parallel Analysis* (Horn, 1965). This procedure compares the variance explained by each factor with the variance explained by the corresponding factor from a simulated dataset

of uncorrelated variables. The rationale behind this is that a *substantial* factor should explain more variance than a factor that was extracted from random (white) noise with no underlying factors.

Despite its simplicity and rather heuristic nature, Parallel Analysis has been shown to perform remarkably well under certain conditions and is therefore an extremely popular extraction criterion (Auerswald and Moshagen, 2019; Buja and Eyuboglu, 1992; Crawford et al., 2010). With respect to applications of PCA to ERP data, two opposing views seem to exist in the literature: On the one hand, it has been suggested to apply a *restricted* solution with as many factors as suggested by Parallel Analysis (Dien, 1998, 2006). On the other hand, it has been suggested to use an *unrestricted* solution, that is, to extract as many factors as sampling points (Kayser and Tenke, 2003, 2006). Both recommendations have influenced researchers applying PCA and both practices can be found in published research (e.g., Fogarty et al., 2018; Kieffaber and Hetrick, 2005). Restricted solutions have the advantage of parsimony because they reduce the number of factors to be considered for interpretation (and statistical tests). Extracting too many factors can result in large factors such as slow wave components being split up artificially (Wood et al., 1996). However, unrestricted solutions could be preferable because effect size estimates converge towards a stable value with increasing number of factors (Kayser and Tenke, 2003, 2006). In addition, the amount of variance explained is of limited relevance in temporal PCA because factors which are active for a shorter amount of time always explain less variance (because they load on fewer sampling points) – but this does not automatically mean they are also negligible from a substantive perspective.

These seemingly contradictory recommendations can be reconciled in the light of recent research on factor extraction criteria. Indeed, extracting too few factors (*underextraction*) goes along with much stronger biases than extracting too many factors (*overextraction*; De Winter and Dodou, 2012; Fava and Velicer, 1992; Wood et al., 1996). Put simply, when the number of factors is too small, PCA represents separate components in conflated factors, but when too many factors are extracted, PCA can represent each component correctly as one factor. When extracting too many factors, PCA often results in additional small factors connecting a few sampling points which are negligible for substantive purposes (based on their weak loadings and topography). The findings of Kayser and Tenke (2003) make a strong case for avoiding underextraction: their estimated effects varied dramatically between solutions when only a few (e.g., 1–8) factors were extracted. When extracting more factors, results stabilized – even when extracting extremely many factors (i.e., severe overextraction). Nevertheless, an unrestricted solution is both computationally inefficient and makes the solution harder to interpret. Furthermore, the use of an unrestricted solution poses a massive multiple comparison problem because as many factors as sampling points need to be analyzed. A restricted solution with fewer factors suffers from this problem to much less extent. Considering this, the call for unrestricted solutions could be rephrased to read: *Use a restricted solution, but when in doubt, prefer solutions with too many factors over solutions with too few factors.*

Unfortunately, Parallel Analysis tends to extract too few factors in cases of highly correlated factors (Beauducel, 2001; Green et al., 2012; Lim and Jahng, 2019). Correlated factors are very common in temporal PCA because factors with similar topography are necessarily highly correlated (Dien, 2010a; Scharf and Nestler, 2018) – validating concerns regarding underextraction expressed in previous research (Kayser and Tenke, 2003). The reason for this is that Parallel Analysis is not a formally valid test but rather a heuristic for factors beyond the first factor (Saccenti and Timmerman, 2017). In the initial solution used to determine the number of factors, the variance explained by later factors can easily fall below that of random white noise factors when the factors
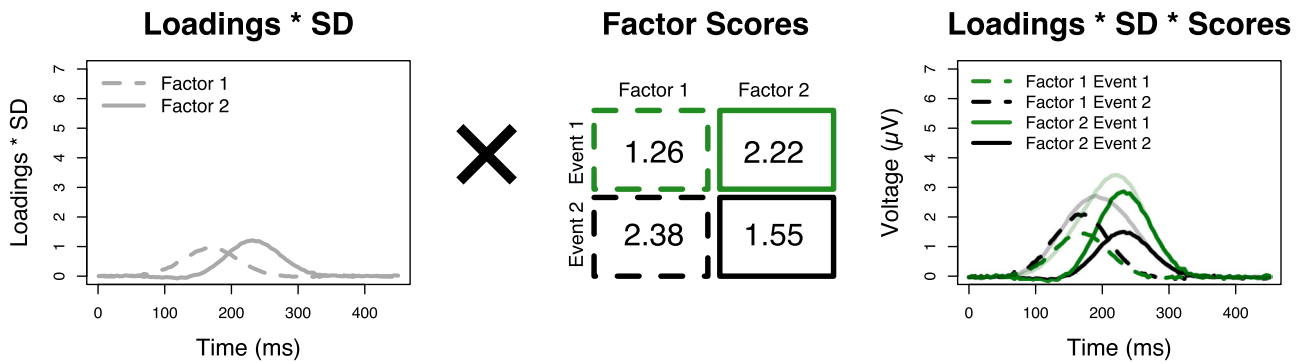
Loadings * SD   Factor Scores   Loadings * SD * Scores

Fig. 4. Factor-wise reconstruction of the observed ERP in Scenario C. The unstandardized factor loadings (left-most panel) represent the time courses of the factors' activities and are fixed across all participants, electrodes, and conditions. The factor scores (middle panel) vary across participants, electrodes and conditions and represent how much and with which sign a factor contributes to the voltage for a specific combination of participant, electrode, and condition. The factor scores can be used to analyze condition effects and to illustrate a factors' topography – by averaging these scores across participants separately for each condition and electrode site. By multiplying the factor loadings and scores, the original data can be reconstructed in a factor-wise manner resulting in an estimate of the underlying components (cf. right-most panel and right-most low panel in Fig. 2).

are substantially correlated (Beaduceul, 2001). Alternative extraction criteria such as the Empirical Kaiser Criterion perform better under these circumstances (EKC; Braeken and Van Assen, 2017; Li et al., 2020). Therefore, we tend to recommend against relying on Parallel Analysis (alone) to determine the number of factors and will consider the EKC in our step-by-step guide.

Estimation of the Factor Loadings. When a decision was made regarding the number of factors to be extracted, the final estimates of the factor loadings (and scores) need to be obtained. The factor loadings are typically estimated from the covariance matrix of the sampling points. Eq. (1) implies that these covariances are a function of the factor loadings and factor correlations (Gorsuch, 1983; Mulaik, 2010; Muthén, 2004):

$$\Sigma \approx \Lambda \Phi \Lambda^{'} \qquad (2)$$

Σ is called *model-implied* covariance matrix of the sampling points, it contains the variances and covariances of the sampling points that would be predicted based on the factor model. It can be computed from the factor loading matrix Λ and the factor correlation matrix Φ. Put simply, the PCA model implies that two sampling points should be more correlated the more the same factors contribute to their voltage. This fact can be utilized to estimate the factor loadings, factor correlations and residual variances by minimizing the difference between the model-implied covariance matrix and the observed covariance matrix of the dataset. Alternatively, the standardized correlation matrix could be used but research on PCA for ERPs suggests that the unstandardized covariance matrix is preferable (Dien, 1998; Dien et al., 2005; Kayser and Tenke, 2003) because all sampling points are measured in the same meaningful unit (i.e., microvolts).

An important property of the outlined model is its *rotational indeterminacy*. That is, any specific ERP can be reconstructed by an infinite set of different factor loadings and factor scores. All these solutions fit the data equally well (see, e.g., Fig. 1 in Dien, 1998, for an illustration) and can be transformed into another by an operation called *rotation* (Asparouhov and Muthén, 2009; Gorsuch, 1983; Mulaik, 2010; Scharf and Nestler, 2019a). Rotational indeterminacy can only be overcome by additional assumptions. Typically, an initial or *unrotated* solution is estimated in which a set of restrictions with desirable computational properties is applied (e.g., uncorrelated factors which are extracted in descending order of explained variance). The unrotated solution is then subjected to a factor rotation procedure to obtain a *rotated* solution which is used for substantive interpretations. In the box *Some details about factor rotation*, we provide further details regarding this mathematical operation.

```
##### Factor eP3a #####

# Descriptive Statistics
describeBy(eP3a ~ group + cond,
          data   =   scores[scores$chan   ==
"Fz",])

# Frequentist ANOVA
aov_ez(data = scores[scores$chan == "Fz",],
       dv = "eP3a", id = "subj",
       between = "group", within = "cond",
       type = 3)

# Frequentist Pairwise comparisons:
#  nov vs. sta
#  for Adults
t.test(scores[scores$group == "ad" &
              scores$chan == "Fz" &
              scores$cond == "nov",]$eP3a,
       scores[scores$group == "ad" &
              scores$chan == "Fz" &
              scores$cond == "sta",]$eP3a,
       paired = TRUE)

#  for Children
t.test(scores[scores$group == "ch" &
              scores$chan == "Fz" &
              scores$cond == "nov",]$eP3a,
       scores[scores$group == "ch" &
              scores$chan == "Fz" &
              scores$cond == "sta",]$eP3a,
       paired = TRUE)

# Please see full script for additional
# Bayes analyses.
```

Factor rotation aims at transforming the initial unrotated solution into an alternative rotated solution which optimizes a certain *rotation criterion*. Modern rotation algorithms allow the use of virtually any criterion that can be computed from the factor loadings (Jennrich, 2004b). In the context of ERP research, a *simple structure* rotation criterion is typically applied. Simple structure rotation aims to rotate the factor loadings so that the voltage at each sampling point is attributed to as few factors as possible and so that each factor contributes to a set of sampling

points as uniquely as possible (Dien, 2010a; Dien et al., 2007; Kayser and Tenke, 2003; Thurstone, 1935). Although simple structure rotation has difficulty in disentangling factors when temporal *and* spatial overlap are high (Möcks and Verleger, 1986; Wood and McCarthy, 1984), there is general consensus that the underlying components can be conceived as transient and temporally distinct enough for simple structure rotated results to be useful (Dien, 2010a; Dien and Frishkoff, 2005). For the sake of completeness, we would like to make the reader aware of two relevant alternatives: First, it is possible to rotate towards a user-defined target structure in a more confirmatory manner when sufficient previous data is available (Arruda et al., 1996; Zhang et al., 2018). Second, there are promising recent efforts to apply ERP-specific rotation criteria instead of general simple structure rotation (Beauducel, 2018a; Haumann et al., 2020).

In the following, we will assume that the user wants to apply a simple structure rotation. A variety of rotation techniques exists which differ in the technical criterion of simplicity being optimized (Browne, 2001) and can be roughly divided into *orthogonal* rotation methods and *oblique* rotation techniques. Orthogonal rotation assumes that the factors are strictly uncorrelated whereas oblique rotation allows the factors to be correlated. The factor correlation measures how much the factor *scores* covary across observations (i.e., across participants, electrode sites, and experimental conditions). Due to the specific data structure (cf. Fig. 3), the factor correlation is a hardly interpretable mix of contributions from participants, electrodes and experimental conditions (Scharf and Nestler, 2018) and the factors in temporal PCA are necessarily substantially correlated as a consequence of functional and topographic overlap (Dien, 1998, 2010a; Dien et al., 2005; Scharf and Nestler, 2018). Therefore, we *strongly* recommend applying oblique rotation methods in PCA for ERP data. Among the oblique rotation techniques, *Promax* (Hendrickson and White, 1964) has long been the gold standard based on its good performance in simulation studies (Dien, 2010a). Recent research revealed that *Geomin* (Yates, 1987) and *Componentloss* rotations (Jennrich, 2004a, 2006) may be able to disentangle highly overlapping factors that would be conflated by Promax (Scharf and Nestler, 2018, 2019a).

To sum up, when estimating the factor loadings, temporal PCA must cope with the rotational indeterminacy of the factor model implying that an infinite set of factors can fit the data equally well. To resolve this indeterminacy, rotation is applied which estimates the factor loadings based on the assumption that the components are transient and at least to some extent temporally separable. Since solutions from different rotation techniques cannot be evaluated by model-fit criteria, substantive/physiological considerations should be made to choose among rotated solutions. Recommendations how to detect and improve problematic results of the rotation step will be discussed later in this article (cf. section *Challenges*).

Estimation of the Factor Scores. After the factor loadings are estimated, the factor scores can be estimated. Several factor scoring methods have been proposed in the context of PCA for psychometric applications (Hershberger, 2014; Mulaik, 2010).[7] We compute scores using the Regression method (Thomson, 1938; Thurstone, 1935) following the formula provided by DiStefano et al. (2009; Appendix 2):

$$\widehat{\eta} = T \quad S^{-1} \Lambda \Phi \tag{3}$$

Here, T is the raw data matrix, S is the observed covariance matrix of the sampling points (and S^{-1} is its inverse), and Λ and Φ are the factor loading and factor correlation matrix, respectively, as estimated in the previous step. In other research contexts, the *standardized* raw data

along the correlation matrix instead of the covariance matrix are usually used in Eq. 3 which results in standardized factors scores with zero mean and a variance of 1. Deviating from this procedure, we use the unstandardized (and non-centered) raw data along the covariance matrix. This results in factors scores which obey the restrictions outlined in the section on the temporal PCA model, that is, the factor scores are normalized with a variance of 1 but have a non-zero mean.

The factor scores provide a summary statistic for the activity of the factors in each observation (i.e., per electrode, condition, and participant; cf. Fig. 3). The factor scores can be used to reconstruct the data in a factor-wise manner, or they can be directly used as dependent measures, for instance, they can be subjected to ANOVAs to test for condition effects. In addition, average factor scores at each electrode site can be used to plot the topography of a factor. As a simple intuition, one can imagine factor scores as *weighted* averages of the voltage across the whole epoch in which the highest weight is given to the sampling points with high factor loadings (or vice versa: peak-based measures can be seen as extremely simplistic factor scores, see Beauducel and Debener, 2003; DiStefano et al., 2009; Donchin, 1966).

1.2.3. Factor-wise reconstruction of the ERP

The factor loadings and scores can be used to provide a factor-wise decomposition of the average ERPs. Fig. 4 illustrates how we computed the factor-wise reconstruction of the ERP in Scenario C (rightmost column, lowest-panel in Fig. 2). The left-most panel illustrates the *unstandardized* factor loadings reflecting the time course of the activity of the factors. By convention, many software packages return *standardized* factor loadings, referring to both the voltages at the sampling points and the factor scores in z-standardized units much like standardized regression weights. Unstandardized factor loadings are scaled in μV units.[8] For ERP applications, it is preferable to report unstandardized loadings because the original unit of the voltage (i.e., μV) is meaningful and interpretable for researchers and unstandardized factor loadings thus provide a better link to the original ERP (Dien, 1998, 2012; Dien et al., 2005). To compute the unstandardized factor loadings from the standardized loadings, the standardized factor loading at a sampling point needs to be multiplied with the standard deviation of the voltage at the respective time point (Dien, 1998; Dien et al., 1997; Muthén and Muthén, 2004).[9]

To reconstruct the ERP in a factor-wise manner, the unstandardized loadings need to be multiplied with the factor scores (see middle panel in Fig. 4). The factor scores in the illustrative example were computed by averaging the factor scores across all participants separately for each event, that is, these are the grand-average factor scores. Multiplying these with the factor loadings results in the factor-wise reconstruction of the ERP in the right-most panel of Fig. 4. Similarly, the averaged factor scores across participants (per condition) at a specific electrode site can be used to reconstruct the ERP at a specific electrode site. The factor-wise reconstructions of the ERP provide a dependent variable in the original unit μV which is comparable across different PCA solutions, and which can be subjected to any further analysis.

2. Step-by-step analysis of an ERP dataset

In this section, we proceed with a step-by-step guide through a PCA analysis of real ERP data (see Fig. 5 for an overview). Extending previous treatments (Boxtel, 1998; Dien, 2012; Dien and Frishkoff, 2005;

---

[7] In general, the choice of scoring method matters when the factor scores are subjected to further statistical analyses (Skrondal and Laake, 2001). However, for ERP data, the factors typically explain most of the variance (Dien et al., 2005) and the choice of the scoring method is therefore of minor importance (Scharf and Nestler, 2019b).

[8] More precisely, an unstandardized factor loading of 1 implies that a change of the factor score by 1 *standard deviation* predicts a 1 μV increase of the voltage at that sampling point.

[9] As in multiple regressions the relationship between the standardized and unstandardized loadings is simply $\lambda_{stand} = \frac{s_\eta}{s_t} \lambda_{unstand}$ (indices left out for readability). As the factor scores are typically normalized (i.e., $s_\eta = 1$; see below), it follows that $\lambda_{unstand} = s_t \cdot \lambda_{stand}$.

In case of suboptimal solution

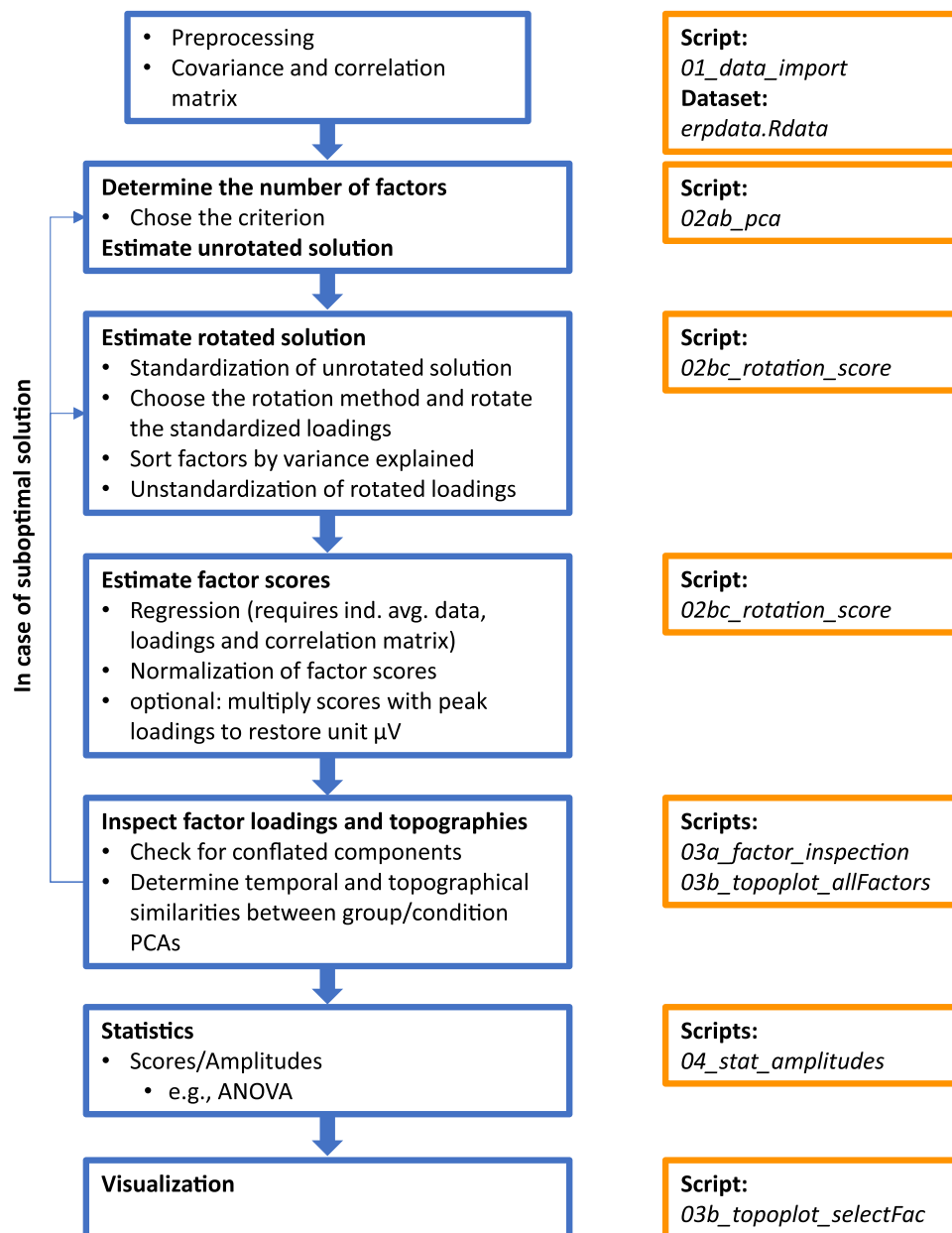| Process Step	Script/Dataset
• Preprocessing • Covariance and correlation matrix	**Script:** *01_data_import* **Dataset:** *erpdata.Rdata*
Determine the number of factors • Chose the criterion **Estimate unrotated solution**	**Script:** *02ab_pca*
Estimate rotated solution • Standardization of unrotated solution • Choose the rotation method and rotate the standardized loadings • Sort factors by variance explained • Unstandardization of rotated loadings	**Script:** *02bc_rotation_score*
Estimate factor scores • Regression (requires ind. avg. data, loadings and correlation matrix) • Normalization of factor scores • optional: multiply scores with peak loadings to restore unit μV	**Script:** *02bc_rotation_score*
Inspect factor loadings and topographies • Check for conflated components • Determine temporal and topographical similarities between group/condition PCAs	**Scripts:** *03a_factor_inspection* *03b_topoplot_allFactors*
Statistics • Scores/Amplitudes • e.g., ANOVA	**Scripts:** *04_stat_amplitudes*
Visualization	**Script:** *03b_topoplot_selectFac*

Fig. 5. Overview of all processing steps and accompanying scripts in our step-by-step guide.

Donchin, 1978), we provide openly available *R* scripts (R Core Team, 2020) and an example dataset which can be adapted for further use by the reader. Some important code sections are also highlighted in code boxes to provide a link between the code and the corresponding explanations. We would like to make the reader aware of the *ERP PCA Toolkit* (Dien, 2010b) which is a MATLAB-based software specifically for PCA analyses with a graphical user interface. We did not want the user to depend on any commercial software to apply the methods described here. Therefore, our choice should rather be interpreted as a decision in favor of maximal accessibility rather than against the ERP PCA Toolkit. All codes and data are available from GitHub (https://github.com/FlorianScharf/PCA_Tutorial/).

2.1. Prerequisites

We assume that the EEG dataset has already undergone basic pre-

processing such as filtering, artifact cleaning, epoching, and averaging across trials within participants. That is, we apply temporal PCA to a dataset with an average ERP per electrode site, participant, and condition in each row. As for any ERP analysis, pre-processing might have beneficial or adverse effects depending on the extent to which the signal is isolated from noise and/or confounding influences. Systematic investigations of pre-processing choices in the context of temporal PCA are scarce but it is known that re-referencing or centering of the data has profound consequences because it can remove variance sources from the data. For instance, some variability across participants can be removed by applying an average reference within participants (Dien, 2012; Dien et al., 2005; Möcks and Verleger, 1985). In addition, one should be aware that slow-wave potentials can substantially confound the solutions because it is hard for temporal PCA to disentangle factors which are perfectly overlapping (Beauducel, 2018a; Verleger and Möcks, 1987; Wood and McCarthy, 1984). Similarly, filtering could affect the PCA
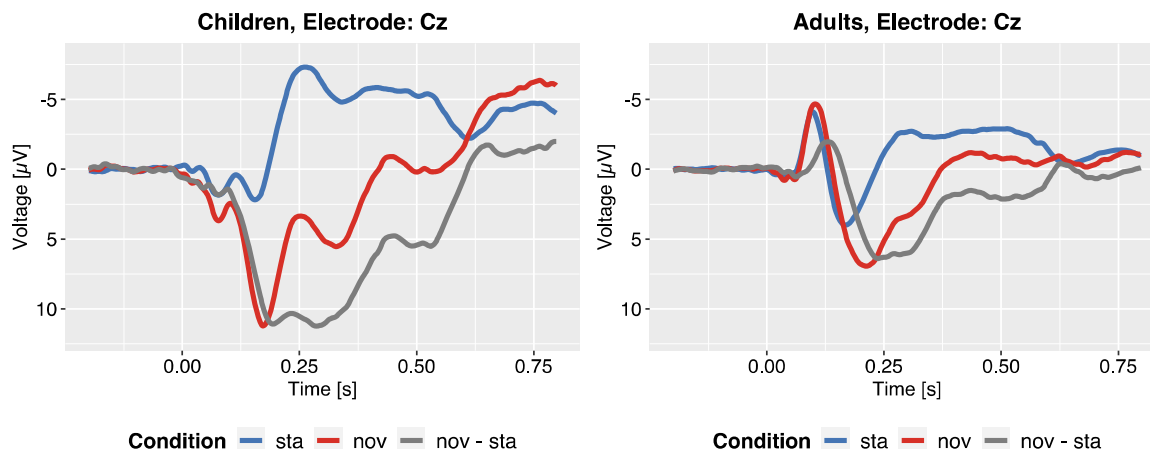
Fig. 6. Grand-average ERPs to standards and novels at Cz for children and adults. Very similar to the simplistic illustrations before, the observed ERPs are characterized by a series of peaks with different amplitudes and polarities. Nevertheless, it is hard to tell from the observed ERP which sampling points reflect specific underlying components. Note that all novel sounds were analyzed together irrespective of their emotional content.

solution because high-pass and low-pass filters are known to reduce or increase, respectively, the auto-correlation of the signal (i.e., the correlation of the voltage between adjacent sampling points entering PCA; Piai et al., 2015). Whether this is beneficial or harmful should depend on the extent to which the signal-to-noise ratio is improved (Widmann et al., 2015), but we are not aware of systematic investigations in the context of the PCA model.

In the following, we re-analyze an ERP dataset from a passive auditory oddball task that was previously reported by Bonmassar et al. (2020). A child group (*N* = 32, 7–10-years, mean 8 years and 10 months) and an adult group (*N* = 32, 18–36-years, mean 26 years and 6 months) watched a silenced video while a sequence of sounds was being presented to them. The sound sequence consisted of 80% standard sounds and 20% novel sounds (i.e., environmental novel sounds which were presented only four times during the experiment). Half of the novel sounds was categorized as "emotional" whereas the other half was categorized as "neutral". Participants were instructed to ignore the stream of sounds. The authors investigated amplitude and latency differences between adults and children regarding multiple ERP components. For the sake of comprehensibility, we will focus on the early P3a component and ignore the emotional content of the sounds when we illustrate each specific analysis step in the application of a temporal PCA. Fig. 6 depicts the grand averages across all participants in both groups.

We import the pre-processed ERP data from the file *erpdata.Rdata* (also available as a human-readable file in *erpdata.csv*). One could also import the data from toolboxes such as *ERPLAB* (Lopez-Calderon and Luck, 2014), *EEGLAB* (Delorme and Makeig, 2004), or *FieldTrip* (Oostenveld et al., 2011) – or consider conducting the pre-processing in *R* (see the R package *eegUtils* by Craddock, 2021).

```
## Load pre-processed data
load("results/01_data_import/erpdata.Rdata")
```

2.2. Combined or separate analysis of groups?

First, one should consider whether it is justified to conduct a combined PCA over all available data, that is, measurement invariance can be assumed. In the following, we will analyze data from different stimulus types in combined PCAs and data from children and adults in separate PCAs, demonstrating how the analysis is conducted in any of the two cases. Therefore, most of the analysis scripts contain a variable to indicate for which group the specific analysis step should be conducted.

```
## Select group to be analyzed
# ad = adults; ch = children
group = "ad"
## The variable age_group_data contains the
# data from the selected
# group – but only the sampling points
#  (i.e., no label variables for electrode
site, # etc.).

## Compute correlation matrix of
# the sampling points
R = cor(age_group_data)

## Determine number of factors using EKC
res_ekc = EKC(R, N = nrow(age_group_data))

## Save the number of factors for use
# in later analysis steps
nFac = res_ekc$n_factors
```

2.3. Estimation of the PCA model

In the next step, the parameters of the factor model need to be estimated. That is, the number of factors needs to be determined, the factor loadings and the factor scores need to be estimated. In this section, we go through these steps separately for the adult and child data. The presented code boxes are taken from the accompanying scripts *02ab_pca.R* and *02bc_rotation_score.R*.

2.3.1. Determination of the number of factors

We applied the EKC to determine the number of factors (Braeken and Van Assen, 2017). A user-friendly implementation of the EKC is available from the package *EFAtools* (Steiner and Grieder, 2020). The correlation matrices of the sampling points need to be computed for adults and children separately and are then used to compute the EKC.

```
## Estimate an unstandardized and unrotated
PCA
PCAFit = fa simplified(
                       # Dataset for PCA
                       r = age_group_data,
                       # Number of factors
                       nfactors = nFac,
                       # Unrotated Solution
                       rotate = "none",
                       #   Use   covariance
matrix
                       covar = TRUE)
```

Fig. 7 depicts the variance explained by each unrotated factor in descending order and the value of the EKC with which it is compared. Remember that the number of factors in the initial unrestricted solution is equal to the number of sampling points. That is, there are 500 factors in our example. The number of factors *m* is the number of factors (red dots) lying above the EKC reference line. In the child PCA, 21 factors were extracted compared to 23 factors in the adult PCA. The number of factors is typically high for ERP data (e.g., > 10), which may still be challenging to interpret. We recommend reflecting on components of interest and their potential time range before the analysis. For instance, if the main research question focuses on an N100 component, one would focus on factors in a time range around 100 ms. The plausibility of the solution cannot be judged at this point. We will discuss how to detect potential under- and overextraction when we address common challenges.

2.3.2. Factor loading estimation

We proceed with the estimation of the factor loadings. We first compute an unrotated solution and perform the factor rotation in a separate step. We estimate the unrotated solution using a custom[10] version of a function[11] from the package *psych* (Revelle, 2021). We note that the estimation of the unrotated model can take some time depending on the size of the dataset. To improve computation speed, the user can consider reducing the sampling rate. The correlation between adjacent sampling points is typically extremely high (e.g., between .97 and 1 in our dataset) so that reducing the sampling rate would not

remove much information. However, one should make sure that the sampling rate is high enough to represent each theoretically expected component with multiple sampling points.

```
## Standardize factor loadings before
rotation
PCAFit$loadings = PCAFit$loadings /
                  PCAFit$varSD
```

Before we subject the unrotated loadings to the rotation algorithm, we *standardize* them, that is, we divide the factor loadings by the standard deviations of the sampling points. It is common to standardize the factor loadings before rotation to prevent large factors from dominating the results of the rotation step (Asparouhov and Muthén, 2009; Appendix B; Dien et al., 2005; Grieder and Steiner, 2020; Mulaik, 2010). There is evidence that this can lead to better rotation results (Cureton, 1976; Mulaik, 2010) and that rotating unstandardized loadings is suboptimal in the context of ERP data (Dien et al., 2005). The standardized unrotated factor loadings are then subjected to factor rotation. We applied a *Geomin* rotation (Yates, 1987) with the rotation parameter ϵ set to 0.01.

```
## Use Geomin rotation with 30
# random start values.
# Please be aware that this function can take
```

The estimation of the rotated solution is done with *Gradient Projection* as implemented in the *R* package *GPArotation* (Bernaards and Jennrich, 2005; Jennrich, 2004b). Gradient projection can be used to rotate the factor loadings optimizing any mathematical criterion calculated from the factor loadings – enabling researchers to choose from a variety of rotation criteria. However, the user must be aware of the occurrence of *local optima* (Hattori et al., 2017; Weide and Beauducel, 2019; Nguyen and Waller, 2022). In simplified terms, Gradient projection proceeds as follows: (1) Choose a random rotation matrix for a start; (2) Compute the rotated factor loadings for this starting rotation matrix; (3) Evaluate the rotation criterion for these intermediate rotated factor loadings; (4) Compute a new rotation matrix to achieve a more optimal value of the rotation criterion; (5) Re-iterate through steps (2) to (4) until the value of the rotation criterion does not change substantially any more between iterations; (6) The rotated factor loadings from the last iteration are the final rotated solution. This estimation procedure is "short-sighted", that is, it tends to pick a solution optimizing the rotation criterion in close proximity to the starting value (local) while overlooking (globally) more optimal solutions "further away". One might imagine a soapbox race car on a street which – on average – descends a hill but every now and then there is a minor ascent. If the soapbox car is too slow, it may get stuck due to one of these ascents and never reach the real valley. This leads to a strong dependence of the solution on the random starting value, especially for Geomin rotation (Hattori et al., 2017). When random starting values are used, it could even be that re-running the rotation procedure with the same data yields a different result every time. That is, our soapbox car might end up in different places depending on our starting point along the hill.

To increase the probability of finding the global optimum, the

---

[10] We removed the computation of some fit indices to improve computational efficiency for large datasets.

[11] Technically, this function estimates an exploratory factor analysis (EFA). In general, the differences between PCA and EFA estimates are negligible for ERP data due to the high correlations between sampling points and due to the high number of variables (Dien and Frishkoff, 2005; Scharf and Nestler, 2019b; Widaman, 1990, 2007).

procedure outlined above should be repeated from many (e.g., 30) different random starting points and the globally optimal solution among all random starts should be used. With respect to the soapbox car, we just restart from different places on the hill and declare the deepest point that we have reached in all runs the "global valley". Even relatively few random starts (usually between 30 and 100) can suffice to find the global optimum (Hattori et al., 2017). When in doubt, the user can check for global optimality by increasing the number of random starts. If the solution does not change after considering (many) more random starts, it is very likely globally optimal. We provide a custom implementation of the Geomin rotation with multiple starts based on the package *GPArotation*. We set the number of random starts to 30 for computational efficiency but we made sure that more random starts (we tested as many as 500) do not change the results. The validity of the rotated solution can only be judged by careful inspection of the solution and will be discussed later.

```
# a while.
##
# Note: We decided to name all parameters
# consistently with the GPArotation package
# despite its deviation from the conventional
# naming epsilon.
##

rotFit <- geominQ_multstart(
                        #        unrotated
loadings
                        A             =
PCAFit$loadings,
                        #        rotation
parameter
                        # (geomin epsilon)
                        delta = 0.01,
                        # No additional
                        # standardization
                        normalize = F,
                        # Use multiple
                        # random starts
                        rand.start = T,
                        # Number of random
                        # starts
                        start.values = 30,
                        ...)
```

As mentioned before, ordering and signs or the factors are generally arbitrary. That is, after rotation, the factors can appear in an arbitrary order and can have mainly negative loadings. As the ERP PCA Toolkit (Dien, 2010b), we reorder the factors in descending order of variance explained. This makes visual inspection easier because large factors (e.g., P3 components) tend to appear among the first factors. We also reverse the signs of factors with mainly negative loadings so that a positive factor score indicates positive polarity at least for sampling points with positive loadings.

2.3.3. Factor score estimation

Using the original data (and their covariance matrix), the *unstandardized* rotated factor loadings, and the estimated factor correlations, we estimate the factor scores using the regression method (Thomson, 1935; Thurstone, 1935). We subject the *unstandardized* original data to

the factor scoring step (Dien and Frishkoff, 2005). This results in factor scores with a non-zero mean and unit-variance. For instance, the first factor from the adult PCA has a mean of -0.20 and a standard deviation of 1.

```
## Estimate factor scores
# The formula is taken from
# DiStefano et al. (2009; Appendix 2).
# Note that "age group data" contains the
# unstandardized ERP individual
# averages of the respective group.
# Therefore, the factor scores
# are not centered.
FacScr = age_group_data %*% PCAFit$Sinv %*%
        (rotFit$loadings * rotFit$varSD) %*%
        rotFit$Phi

# PCAFit$Sinv ... generalized inverse of the
# covariance matrix of
# the sampling points

# rotFit$loadings * rotFit$varSD ...
# unstandardized factor loadings

# rotFit$Phi ... factor correlation matrix

## We save all estimated objects for later
steps.
save(PCAFit, rotFit, scores,
     # file name
     file = paste0("rotfit_",
                PCAFit$group,
                PCAFit$factors,
                "_geomin0.01.Rdata")
                )
```

2.4. Visual inspection of time courses and topography

Before conducting further analyses of condition effects or group differences, it is important to inspect the rotated solution, judge its substantive plausibility and interpret the factors based on their time courses and topography. We recommend looking at the time courses (i. e., the unstandardized factor loadings) first and inspect the topographies afterwards. We generated a series of plots (Figs. 8–10) that can be used to conduct these steps and provide the code to create them in the scripts. We also provide a script *03b_topoplot_selectFactor.R* which generates a combined plot of time course and topography of a *selected* factor as a service to the reader. All topography plots rely on the package *eegUtils* (Craddock, 2021).

Fig. 8 depicts the unstandardized rotated factor loadings for the child and adult PCA, respectively. As expected, the factor solutions between adults and children differ profoundly. The loadings in the child PCA were generally higher – reflecting the typically larger ERPs and larger variability in children compared to adults. Although major factors were extracted in similar time ranges, the *relative* contributions of these factors differ considerably between both groups. For instance, Factor 1 has much lower peak loadings (e.g., relative to Factor 2) in the adult group than in the child group. Finally, a close look reveals that there could also be some latency differences (e.g., for Factor 2) between both groups.

The depicted solutions are quite typical of a PCA solution for ERP data. There are often about 8–12 major factors with large peak loadings

and many minor factors which explain only a small proportion of the variation in the data. In our experience, meaningful ERP components are most often represented among the major factors (see also Dien, 2010b, 2012). The interpretation of minor factors is difficult and often unnecessary regarding the actual research questions. These factors could represent remaining autocorrelations in the noise, minor latency jitter between participants or idiosyncrasies in the ERPs of certain individuals among the participants (Dien, 2012, 2018; Möcks, 1986). Unless there are specific substantive hypotheses regarding a small but established component which must be identified among the minor factors, it is not essential to specifically explain the occurrence of every factor in the solution. Nevertheless, we want to emphasize again that the extraction of minor factors can improve the overall solution and that over-extraction is typically preferable to underextraction. Even if only a subset of the major factors is relevant for further analyses, we recommend to always report the rotated loadings of *all* extracted factors. This ensures that readers (and reviewers) can evaluate the adequacy of the conclusions and that potential problems with the rotated solutions can be detected (cf. section *Challenges*).

To derive substantive interpretations, the topography of the factors needs to be considered in addition to the loadings. As in the ERP PCA Toolkit (Dien, 2010b), we computed the average factor scores across participants at each electrode site, for each stimulus type, and for each group and multiplied these scores with the factor loadings. The resulting factor-wise reconstructions of the grand-average ERP served as the basis for the topography plots. Specifically, we plotted the time courses at the electrode site with the maximal factor score and the topography at the sampling point with the maximal factor loading. Please note that this heuristic is not always optimal (e.g., for bipolar factors) and that the automatically chosen electrode site may not always be theoretically reasonable. In the following, we focus on the identification of the early P3a among the factors in both groups for the sake of comprehensibility. In a complete analysis, the topographies of all factors should be inspected, and we encourage authors to always make them available in the spirit of transparency. The full topography plots for our example dataset are available from the online supplement.

There are several major factors in a typical P300 time range between 200 and 400 ms in each PCA (highlighted in Fig. 8). Based on the loadings alone, these factors are candidates to represent the early P3a. Figs. 9 and 10 depict the topographies and reconstructed grand-average ERPs for the child and adult PCA, respectively. For both groups, there are two factors in the time range between 200 ms and 400 ms with a central, positive topography for novel sounds (child PCA: F5 and F6; adult PCA: F3 and F5). All remaining candidates are characterized by a negative topography which excludes them as early P3a factors (Escera et al., 2000). Based on their peak latencies, we labeled the adult factor F5 (peak loading at 222 ms) and the child factor F6 (peak loading at 296 ms) as "early P3a" and the respective other factor as "late P3a". Despite their considerably different latency, we think that comparing these factors between groups is substantively justified because it is known that the respective components occur later in children compared to adults (Riggins and Scott, 2020). While the considerable latency differences support the appropriateness of separate PCAs in this case, they also demonstrate that detailed substantive reasoning is mandatory when matching factors from separate PCAs (Nesselroade, 2007; Nesselroade and Molenaar, 2016b).

In general, researchers must face several challenges when matching factors from separate PCAs: (1) the order in which the factors were extracted can vary, (2) systematic latency shifts of factors can occur, (3) matching factors solely based on visual inspection can be very subjective, and (4) the number of factors to compare is typically very large. Quantitative similarity measures for the factor loadings and topographies such as Pearson's correlation or Tucker's congruency coefficient can support the search for matching factors (Lorenzo-Seva and ten Berge, 2006). These measures are regularly applied in simulation studies to quantify the similarity of loading patterns (De Winter and Dodou, 2016; Dien, 1998; Scharf and Nestler, 2018, 2019a, 2019b). Therefore, we think that they can be helpful to researchers applying PCA as well. Specifically, we recommend matching the factors by temporal similarity first followed by topographic similarity because factors tend to be more distinct with respect to their time courses than to their topographies (Dien, 2010a). In our experience, this procedure allows to identify clearly corresponding factors very easily. However, we note that factors with systematic latency shifts *cannot* be matched based on temporal similarity. In the accompanying scripts, we provide a detailed demonstration how Pearson's correlation can be used to match the corresponding factors from the child and adult PCA (see script *03a_factor_inspection.R*).

A closer inspection of the contributions of early and late P3a to the grand average ERPs (Figs. 6, 9 and 10) underlines the usefulness of the temporal PCA approach. Both factors overlap considerably in time and space with each other and with the P2 factor and are hardly distinguishable based on the grand average ERP alone. That is, there is no obvious time-window (or baseline-to-peak relation) that could be utilized to disentangle these factors. Given that these separate factors could not even be identified without statistical decomposition, any analysis based on the observed waveform would suffer from a considerable risk of false conclusions due to the conflation of separable factors. To be clear, we do not claim that all such analyses are wrong – to what extent a conflation of factors biases the results is highly specific for a given dataset – but the risk of missing important features of an ERP is much higher without decomposition approaches.

2.5. Analysis of amplitude effects

Either the factor scores or a factor-wise reconstruction of the ERPs could be used for inferential statistics. In a *combined* PCA, the choice does not matter – at least not for the results of a general linear model (i. e., *t*-tests, ANOVAs, regressions) – because common amplitude measures (e.g., peak amplitude or time window averages) are only rescaled factor scores. This can be shown very easily: The peak amplitudes of the reconstructed ERPs are just the factor scores of the respective factor multiplied with the peak factor loading: $\widehat{t}_{peak} = \lambda_k^{(peak)} \cdot \eta_k$. Similarly, one can show that the average reconstructed voltage in a time window is simply the average factor loading in that time window multiplied by the factor scores:

$$\bar{\widehat{t}}_{timewindow} = \frac{\sum\limits_{i \in timewindow} \lambda_{ik} \cdot \eta_k}{length(timewindow)} = \eta_k \frac{\sum\limits_{i \in timewindow} \lambda_{ik}}{length(timewindow)}$$

$$= \eta_k \cdot \bar{\lambda}_{timewindow,k}$$

These considerations underline that both approaches are equally valid and will eventually lead to the same conclusions in a combined PCA. However, in *separate* PCAs only the analysis of reconstructed ERPs ensures a shared scale (i.e., μV) – directly comparing the factor scores between children and adults would be problematic because both scores were normalized relative to the variability *within* the respective dataset. A natural way to bring them back to a shared scale is the factor-wise reconstruction of the ERP. After this rescaling, again, it does not matter whether the reconstructed peak amplitudes are used or a measure of the area under the curve. In the following, we will use the peak of the reconstructed ERP $\widehat{t}_{peak}$ as dependent variable in our analyses. We generally recommend this approach because it reinstates the original unit μV which ERP researchers are most familiar with. We compute the peaks of the reconstructed ERPs in the script *04_stat_amplitudes.R*:

```
## Adult PCA
# Compute unstandardized factor loadings
adPCA$rotFit$loadings_unstd <-
    adPCA$rotFit$loadings                               *
adPCA$rotFit$varSD
# Compute the peak loading for each factor
adPCA$peak_loadings <-
    apply(adPCA$rotFit$loadings_unstd,
        MARGIN = 2, max)
```

For the sake of simplicity, we conduct further amplitude analyses on a single electrode site for each factor, but it would also be possible to average across a spatial region of interest instead. It should be noted that the electrode site should be chosen a-priori to avoid capitalization on chance or "double dipping" (Kriegeskorte et al., 2009; Luck and Gaspelin, 2016). For each of the four selected factors, we conducted the same statistical analyses: (1) Compute descriptive statistics using the package *psych* (Revelle, 2021), (2) conduct an analysis of variance with the between factor "group" and the within factor "condition", and (3) conduct some selected pairwise comparisons of interest. This should be read as a basic suggestion; the reader is free to use the (rescaled) factor scores for any analysis for which peak-based ERP scores could have been used. We demonstrate frequentist inferential statistics using the package *afex* (Singmann et al., 2021) as well as Bayesian inferential statistics as implemented in the packages *BayesFactor* (Morey and Rouder, 2018) and *bayestestR* (Makowski et al., 2019). We report generalized η^2 (η^2_{gen}) as effect size measure (Bakeman, 2005).

```
# Multiply peak loadings with factor scores
adPCA$scores[,-c(1:4)]   <-
    as.matrix(adPCA$scores[,-c(1:4)]) %*%

diag(adPCA$peak_loadings)

## Child PCA
[...]
```

In the following, we briefly go through the results of these steps for the early P3a factor. As illustrated in Figs. 9 and 10 (child PCA: F6; adult PCA: F5), the voltage of the early P3a factor at electrode site Cz is negative for standard sounds in both groups (Children: $M = -3.58$ μV, $SD = 2.35$ μV; Adults: $M = -0.064$ μV, $SD = 1.60$ μV). For novel sounds, the voltage was more positive compared to standard sounds (Children: $M = 3.56$ μV, $SD = 3.82$ μV; Adults: $M = 4.52$ μV, $SD = 2.90$ μV). The ANOVA revealed a significant main effect of condition, indicating generally more positive voltage of the factor for novels, $F(1,62) = 235.97$, $p < .001$, $\eta^2_{gen} = 0.533$. There was also a significant main effect of group, $F(1,62) = 14.68$, $p < .001$, $\eta^2_{gen} = 0.142$, and a significant group × condition interaction, $F(1,62) = 11.26$, $p = .001$, $\eta^2_{gen} = 0.052$. Both effects were most likely driven by the more negative voltage of the factor for standards in children – but we did not follow up on them to keep this demonstration simple. Most importantly, both in adults, $t(31) = 10.15$, $p < .001$, and in children, $t(31) = 11.61$, $p < .001$, paired t-tests revealed significant differences in the voltage between standards

and novels.

To sum up, amplitude differences in the voltage of a specific factor can be analyzed either using the factor scores directly or based on the reconstructed participant average data. Both yield equivalent test results in combined PCAs but the reconstructed ERP approach offers the advantage of restoring the original unit μV and enabling comparisons across separate PCAs. The factor score based measures can be used for any further analysis as a quantification of a factor's contribution.

3. Challenges in developmental studies

In the previous sections, we have introduced the PCA model and its potential merits for ERP researchers and we have provided a step-by-step guide through a PCA-based analytic approach demonstrating how amplitude effects can be analyzed. In this section, we want to address several common challenges and discuss how the data-analyst can cope with them: (1) How can the visual inspection be used to reveal suboptimal decomposition results and what can be done about it? (2) How can "double dipping" be avoided? (3) Further considerations on measurement non-invariance: What are possible reasons for different results from separate PCAs? What should one do when quantitative comparisons seem unreasonable? (4) How can PCA be used to analyze longitudinal data with multiple measurement occasions for each participant? (5) How can separate PCAs be used to investigate latency differences?

3.1. How to detect and handle suboptimal PCA results

Careful visual inspection of the PCA solution is essential to make the *substantive* decision whether a specific PCA solution is "reasonable". The most likely reasons for suboptimal solutions are under- or overextraction of factors and difficulty of the rotation method in disentangling highly overlapping factors. A very common symptom are *conflated* factors (Scharf and Nestler, 2019a), that is, factors occur in the solution which seem to be a mixture of established (sub-) components. The likelihood of this happening increases with temporal and spatial overlap. To illustrate this, Fig. 11 contrasts the loadings from the adult PCA presented before (23 factors) with a suboptimal solution with only 8 factors. The extraction of too few factors forces PCA to conflate separable components into a single factor. In this case, the early and late P3a factors from the left panel (Factors 3 and 5) are conflated into a single factor in the right panel (Factor 1). The reason for this is that both factors overlap in time and have very similar topographies. Such a conflation of prominent separate components is a clear warning sign that the solution is not optimal and is often also recognizable in the topographies which are also a mixture of the conflated factors' typical topographies.

In general, temporal PCA tends to err in the direction of conflating factors which should be separated rather than splitting factors artificially (but see Wood et al., 1996). The spatial and temporal overlap which make it hard to identify these factors by visual inspection of the original ERP also make it hard to separate them based on statistical properties of the data. Although conflated factors cannot always be treated successfully, there are some systematic steps researchers can try to improve suboptimal solutions: (1) Try to increase the numbers of random starts and iterations for the rotation algorithm in case the suboptimal solution was due to a local optimum of the rotation criterion. (2) Investigate whether the solution improves when the number of factors is increased to rule out underextraction. (3) Adapt the rotation procedure. Promax has been the gold standard for a long time, but alternative rotation techniques such as Geomin and Componentloss rotation can sometimes achieve better results (Scharf and Nestler, 2019a). Alternatively, some rotation techniques have a tuning parameter that can be changed. For instance, increasing the Geomin ε improves the ability of
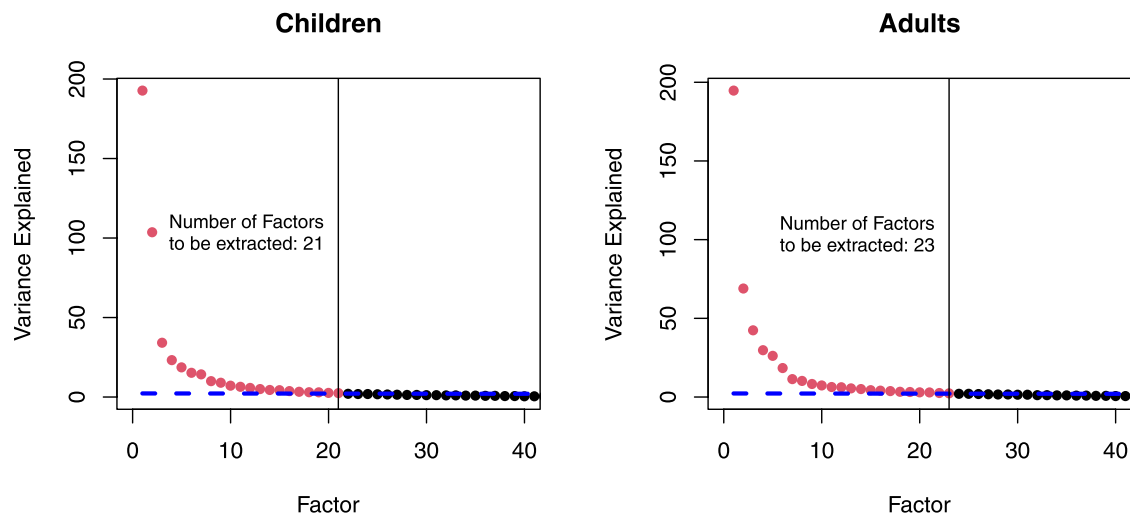
Children

Adults

Fig. 7. Variance explained by each factor in the initial solution for adults and children. Note. Remember that the factors are extracted in descending order of variance explained in the initial solution. The dots in the figures illustrate the variance explained by each factor in the initial solution for all factors (so-called Scree-Plot). There are 500 factors in the initial solution (as many as sampling points in the epoch), but we cut the plot at 40 factors to focus on the comparison with the reference line (blue line) for the EKC. The number of factors is marked by the last factor lying above the reference line (additionally marked by the vertical line). The red dots are above the reference line, the black dots are below the reference line. For the child PCA, the EKC indicates that 21 factors should be extracted, compared to 23 factors in the adult PCA.

the rotation to separate overlapping factors (but at the risk of splitting other factors artificially). (4) When none of the above options is viable, the assumptions of PCA may simply not fit the present purpose and alternative decomposition methods could help (see below).[12]

3.2. How can "double dipping" be avoided?

Although such data-driven adaptations of the analysis procedure can improve the physiological plausibility of the decomposition, one must be aware of the considerable risk of "double dipping" (Kriegeskorte et al., 2009). Technically, data-driven adaptations of the analyses present a case of selective inference in which statistical tests are applied after the statistical "model" has been chosen based on the same dataset (Benjamini, 2010; Kriegeskorte et al., 2009; Luck and Gaspelin, 2016). Strictly speaking, any post-hoc model modification is potentially problematic and may inflate the type I error rate – be it a change of the number of factors or a change of the rotation method. A simple and effective strategy to avoid this is a strict separation of the model specification step from the statistical inference step. Preregistrations can be used to decide on the parameters for PCA *before* beginning the analysis and to make all deviations from the original analysis scheme transparent. A preregistration of a PCA-based analysis should *at least* contain the following aspects:

(1) How are the data pre-processed and cleaned?
(2) Which subsets of the data are submitted to combined versus separate PCAs?
(3) How is the number of factors determined?
(4) Is the analysis based on standardized data (i.e., a correlation matrix) or unstandardized data (i.e., a covariance matrix)?

(5) Which rotation technique is applied? (if applicable, including the values of rotation parameters and standardization scheme of loadings during rotation)
(6) How will a substantive interpretation of the factors be derived?
(7) How will factors from separate PCAs be matched?
(8) How will statistical inference be conducted? Which electrode sites/ROIs will be included in the analysis for each presumed factor?

To increase the objectivity of interpretations, we also encourage groups of cooperating researchers to interpret the results independently in a first step and to discuss deviating interpretations in a second step.

Even for very experienced researchers, it is hardly possible to anticipate all characteristics of ERP data with sufficient accuracy to always make ideal a-priori choices. Instead of avoiding modifications at all costs, we think that it is important to make transparent which decisions were made post-hoc. Ultimately, each exploratory modeling decision increases the need for confirmatory replication studies to establish the robustness of the findings. If the sample size is sufficient, researchers may consider data splitting (i.e., tuning analysis parameters in one half of the data and conducting statistical inference on the other half). Data splitting is an effective, easily implemented and non-parametric method to assure the validity of statistical inference after model selection (Rinaldo et al., 2019; Wasserman and Roeder, 2009), although we acknowledge that sufficiently large samples are hard to collect in a developmental context. Nevertheless, replication studies and studies trying to manipulate certain factors in isolation are essential next steps to validate the decomposition results (Widmann et al., 2018). Alternatively, when many analyses have been tried out, it may even be best to make their results transparent and to show which of the results is sensitive towards specific data-analytic choices (Steegen et al., 2016). In sum, none of these threads for replicability is specific to temporal PCA; peak-based analyses suffer from the very same issues, and the same strategies can (and should) be applied to maximize the chance of replicable findings.

3.3. Further considerations on measurement invariance

As we have seen, equality of the factor loadings across all observations is arguably one of the most questionable assumptions of the

---

[12] Some readers may wonder whether a peak-based measure could be used alternatively. In theory, this could work when the temporal overlap is low and the prior knowledge on the component structure is very precise. However, we doubt that such detailed knowledge is commonly available. Therefore, we rather recommend the use of decomposition methods with different assumptions or incorporating prior knowledge in a suitable decomposition framework (e.g., Huang, 2020; Huang et al., 2017). For instance, there have even been attempts to utilize prior knowledge on condition effects during factor rotation (Beauducel and Leue, 2015).

temporal PCA model. Apart from differences between groups, systematic latency shifts may also occur between experimental conditions (Barry et al., 2016) or between participants resulting in additional factors (Möcks, 1986). We recommend the use of separate PCAs whenever measurement invariance is questionable. Specifically, developmental researchers should conduct separate PCAs for different age groups because the evidence for fundamentally different component structures between different age groups is strong (Wunderlich et al., 2006). With respect to experimental conditions, separate PCAs should be conducted when substantial latency differences are expected. Otherwise, separate PCAs can be conducted to check for unanticipated latency shifts. In the presence of very strong latency jitter, PCA might not be an appropriate analysis model and alternative decomposition methods might be considered (see below).

When separate PCAs reveal different factor patterns, this may either be an artifact of the rotation step or indicate genuinely different component structures. That is, measurement non-invariance does not inevitably imply different component structures (see Welzel et al., 2021; for a similar discussion in the context of psychometric data). For instance, PCA may conflate certain factors in one group but not in the other. This possibility can hardly be ruled out using a single dataset, although isolated cases of conflated factors could be identified as outlined in the previous section. Nevertheless, measurement non-invariance can be a hint towards qualitative differences between the groups and there is no guarantee that separate PCAs map as well as in our example dataset – and it is therefore not a failure of the analyst if a specific factor from one group cannot be matched with any factor from the other group. For instance, it is well-known that babies have very different ERPs from school children and adults (Wunderlich et al., 2006) and it would not be surprising if separate PCAs in these groups yielded very different results. Quantitative comparisons may be unreasonable in such situations, and we suggest that potential explanations for the qualitative differences should be discussed instead of enforcing a quantitative comparison of fundamentally different entities.

3.4. Longitudinal designs

In the present article, we focused on an example of a cross-sectional temporal PCA in which independent age groups were compared. However, developmental researchers could also use longitudinal designs with repeated measurement occasions for the same group of participants. Two specific challenges occur in the context of longitudinal designs: First, researchers should reflect on the plausibility of measurement invariance across measurement occasions, and second, they need to account for the dependencies between measurements occasions from the same participant. With respect to the first challenge, temporal proximity between measurement occasions is crucial. Assuming that the main driver of different factor structures is cognitive and neuronal development, the plausibility of measurement invariance largely depends on the time lag between measurements. For instance, when ERPs are recorded from a group of children twice within one week, it is unlikely that fundamental development has taken place in these areas, and one would expect to observe the same factor structure across both measurement occasions. However, if the time lag between measurements would be a year or longer, substantial developmental changes could have occurred and all considerations on measurement invariance from the previous sections apply.

When measurement invariance can be assumed, researchers can conduct a combined PCA across all measurement occasions. Importantly, for statistical inference, the measurement occasions need to be considered a within-person factor, essentially in the same way as we did it with *stimulus type* in our example. When measurement invariance cannot be assumed, researchers could combine separate PCAs with so-called target rotations (for an overview, see Browne, 2001). For instance, a separate PCA with simple structure rotation could be applied for the first measurement occasion. For the second measurement

occasion, this rotated loading matrix could serve as a target for the rotation – aiming for as similar factor structures as possible. We think this approach is most reasonable when the time lag is substantial but not too large. In this case, factor structures may still be relatively similar with some specific changes (e.g., latency shifts of certain factors). In such cases, target rotation could help to uncover the nature of the developmental changes. More recent approaches to longitudinal analyses such as latent growth models can be applied in the context of exploratory structural equation models (Asparouhov and Muthén, 2009; Marsh et al., 2009, 2013, 2014; Morin et al., 2013; Scharf and Nestler, 2019b) including a more general approach to rotate towards a similar factor structure between multiple measurement occasions (Marsh et al., 2017). Please note that further methodological evaluations of these approaches are necessary because we are not aware of any previous applications of these recently developed methods to ERP data.

3.5. Analysis of latency effects

Apart from the analysis of amplitude differences between conditions, ERP researchers are often interested in latency differences between groups or conditions. Some methods have been proposed which are able to provide individual factor-wise latency estimates for participants (e.g., Möcks et al., 1988; Tuan et al., 1987) but these methods have not been widely adopted to our knowledge. In a combined PCA, latency differences are mostly concealed due to the strict measurement invariance assumption, but extreme latency differences may result in split-up factors representing different latencies. Separate PCAs can be used to detect latency differences assuming that none of the solutions is affected by rotation artifacts such as conflated factors – but they do not quantify the uncertainty of the estimated latencies. In the original analysis, we adopted an established Jackknife approach to test for latency effects (Kiesel et al., 2008; Miller et al., 1998; Ulrich and Miller, 2001). Instead of applying this resampling approach to the observed ERP, we applied it to the factor loadings re-estimating the rotated loading for each re-sampled dataset (Bonmassar et al., 2020). Systematic evaluations of the utility of this approach are not yet available but we provide an implementation and documentation of this approach as a supplement in the GitHub repository for any interested reader. In sum, despite the aforementioned and other methodological efforts to lift this limitation (De Munck et al., 2004; Ouyang et al., 2017; Truccolo et al., 2003), using temporal PCAs to analyze latency effects remains challenging and no validated wide-spread solution is available so far.

4. Comparison with alternative decomposition methods

The challenges described above demonstrate that PCA is a data-analytic approach with strengths and weaknesses as any other approach (Delorme et al., 2012; Dien et al., 2007). Since the original proposal of PCA decompositions for ERP data (Donchin, 1966; Donchin and Heffley, 1978), its usefulness has been demonstrated many times (Chapman and McCrary, 1995; Dien, 2012; Dien et al., 2003; Dien and Frishkoff, 2005). But there have also been lively debates regarding its weaknesses (Achim and Marcantoni, 1997; Beauducel and Debener, 2003; Collet, 1989; Dien, 1998; Dien et al., 2005; Donchin, 1989; Kayser and Tenke, 2003, 2005; Möcks, 1989; Möcks and Verleger, 1986; Wood and McCarthy, 1984). The term *variance misallocation* has been coined by Wood and McCarthy (1984) to describe the fact that condition effect estimates based on PCA can be biased. The reason for this is that the loading estimates can be biased which is a result of a partial mismatch between the goal of the rotation technique and the ground truth (Möcks and Verleger, 1986). Specifically, orthogonal rotations must be avoided because they *always* estimate uncorrelated factors which is not appropriate (Dien, 1998, 2010a; Dien et al., 2005; Scharf and Nestler, 2018). In addition, when the ground truth is characterized by components with high spatial and temporal overlap, especially in the presence of slow-wave components, simple structure rotation can by definition not
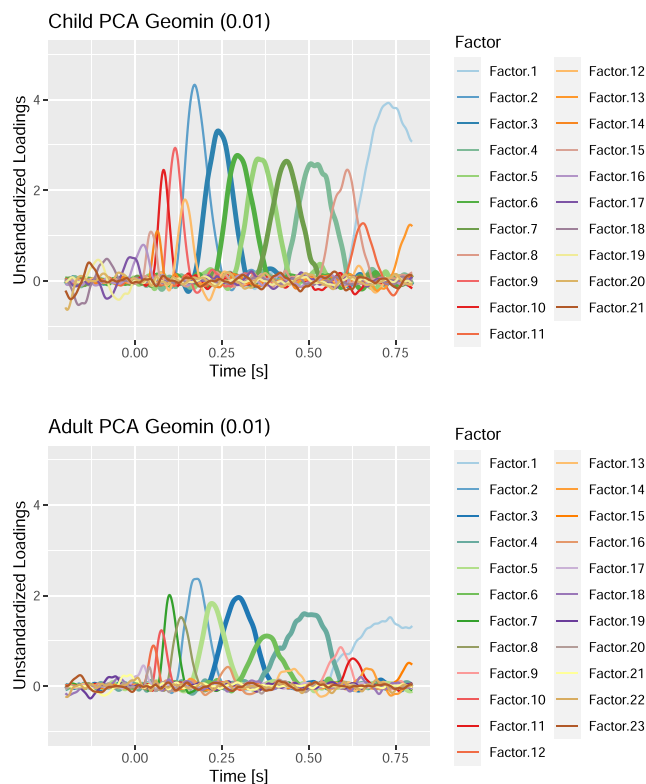
Fig. 8. Unstandardized Factor Loadings after Geomin (0.01) Rotation in both PCAs. Each colored line represents the factor loadings of a factor. Higher factor loadings implicate that the factor contributes more to the voltage at a sampling point. Thus, one can conceive the loadings as a representation of the time course of a factor's activity. Please note that the factors are numbered by the amount of variance they explain – not by their temporal order. Thicker lines were used to highlight factors in the typical P300 time range. This plot can be created using the script *03a_factor_inspection.R.*

achieve perfect separation but will conflate, for instance, slow-wave components with other components (Scharf and Nestler, 2018, 2019a; Verleger and Möcks, 1987). ERP-specific rotations and estimation algorithms could further reduce this problem (Beauducel, 2018b; Haumann et al., 2020) but more experience is needed with applications of these methods for further evaluation.

Despite this methodological progress and the wide-spread use of decomposition methods such as ICA for artifact removal (e.g., Chaumon et al., 2015), the use of decomposition techniques during the actual quantitative analysis step is sometimes met with skepticism. Some of these concerns have been expressed in popular introductory treatments for ERP analysis (Luck, 2005b).[13] Therefore, we would like to put some arguments made there into perspective:

The PCA technique, in particular, is problematic because it does not yield a single, unique set of underlying components without additional assumptions (see, e.g., Rösler and Manzey, 1981). That is, PCA really just provides a means of determining the possible set of latent component waveshapes, but additional assumptions are necessary to decide on one set of component waveshapes (and there is typically no way to verify that the assumptions are correct). (Luck, 2005b, p. 22).

---

[13] We would like to emphasize that we merely take these quotes as an example to illustrate the presence of such concerns. Although we tend to disagree in this specific matter, this should not be taken as a criticism of the remainder of this chapter which contains many valid and important recommendations for ERP research.

It is true that the PCA model needs restrictions to identify the model parameters. However, this is *not* a property of the method itself but rather of the inverse problem in ERP research. We try to make inferences regarding the functional significance of components (or even the sources in source localization analyses) from the limited information provided by the ERP observed at the scalp. From a mathematical perspective, this problem is *ill-posed*. That is, the ERP signal at the scalp does not contain sufficient information to recover the ground truth uniquely. Consequently, it is impossible to infer the components without additional assumptions or additional knowledge (e.g., co-registered fMRI). Importantly, mean amplitudes in time-windows, peak- or area-based measures make even more restrictive assumptions: All these measures can be conceived as extremely simplistic factor scores (e.g., with rectangular factor loadings in case of mean amplitude time-windows) *assuming* that the voltage in the analyzed time-window is uniquely due to one factor, that is, that there is zero temporal or spatial overlap with other components (DiStefano et al., 2009; Donchin and Heffley, 1978). Because these methods are simply naïve scoring methods, it is not surprising that they suffer from even worse variance misallocation than PCA (Beauducel and Debener, 2003). The appropriateness of simple structure rotations is certainly up to discussion in many contexts but we see it as an advantage of PCA that we can explicitly name and discuss this assumption rather than simply making it implicitly by using naïve scoring methods.

Another common concern is that decomposition techniques fail in the presence of latency shifts and latency jitter:

[A]ny correlation-based method will have significant limitations. One limitation is that when two separate cognitive processes covary, they may be captured as part of a single component even if they occur in very different brain areas and represent different computational functions. […] Another very important limitation is that, when a component varies in latency across conditions, both PCA and ICA will treat this single component as multiple components (Luck, 2005b, p. 22).

It is true that *temporal* PCA tends to conflate components which overlap temporally, spatially and/or functionally (i.e., when they have the similar condition effects; but see below for potential solutions). It is also true that latency jitter can result in artificial additional factors (Möcks, 1986) and that a combined PCA cannot represent latency shifts but, as we have demonstrated, separate PCAs can be used to improve this situation to some extent. In the context of exploratory structural equation models, measurement invariance can even be tested formally (Asparouhov and Muthén, 2009; Putnick and Bornstein, 2017; Scharf and Nestler, 2019b). Furthermore, for measures based on the observed ERP it is by no means clear whether ERP differences are due to amplitude differences or latency shifts as is apparent even in a simplistic example such as Scenario C in Fig. 2. PCA offers a much more objective way to disentangle these scenarios than visual inspection of observed ERPs. An additional advantage of PCA-based analyses is that we know the direction in which PCA tends to err and we know the symptoms: Conflated factors and artificial "jitter-factors" are well-documented (Möcks, 1986; Verleger and Möcks, 1987) and it is possible to inspect the PCA solution for these signs of suboptimal solutions – offering much deeper insight into the potential component structure.

Ultimately, researchers must decide – based on their knowledge on the subject matter – which assumptions can be made and, hence, which analytic method is appropriate. In that sense, we agree with (Luck's 2005b, p.22) conclusion that PCA "[does] not provide a magic bullet for determining which components an experimental manipulation influences" but neither do the "traditional" approaches which are based on the observed ERP. Hence, we strongly advocate that researchers carefully reflect on the assumptions and make them explicit. To support these reflections, we briefly want to make the reader aware of some further decomposition methods and when they could be helpful. A
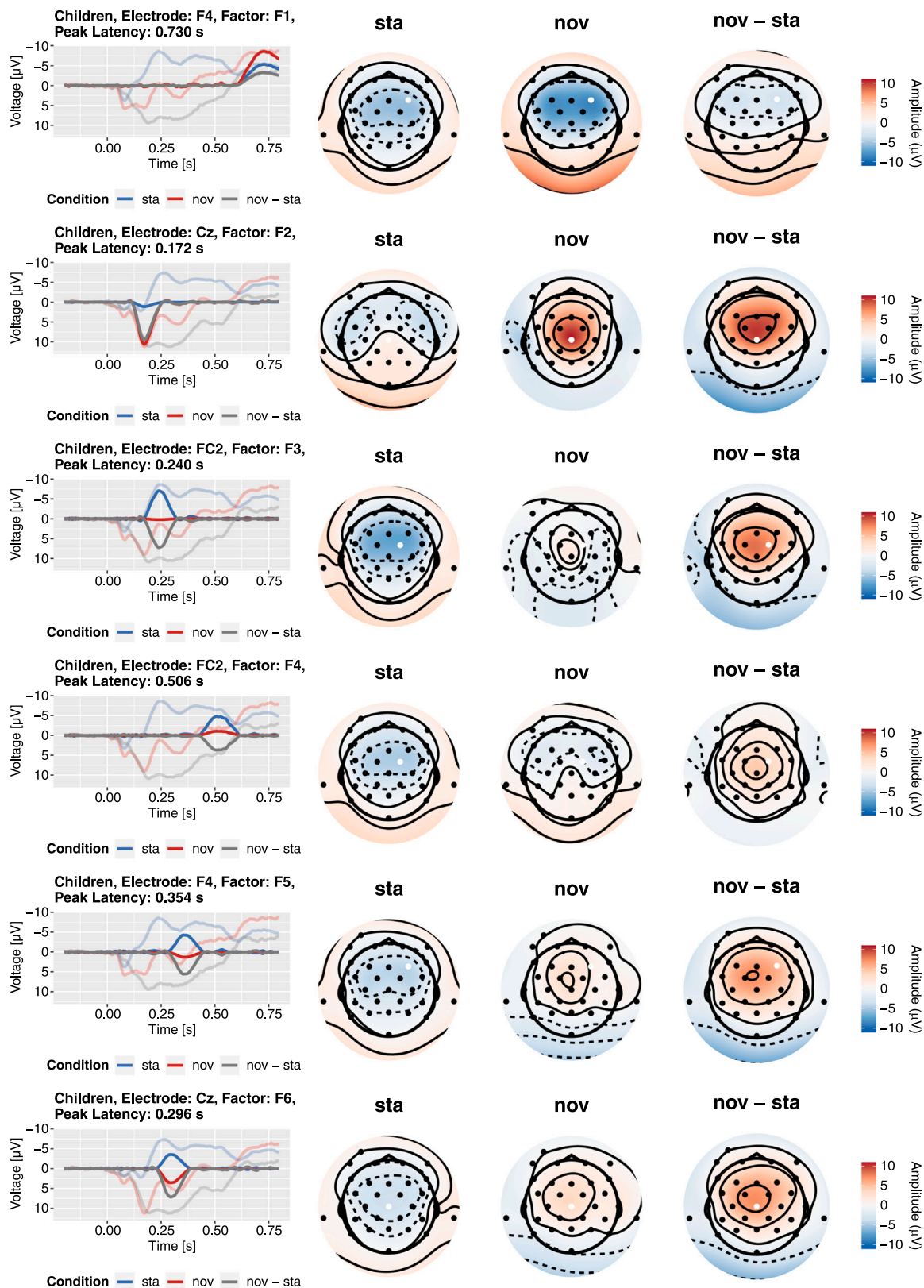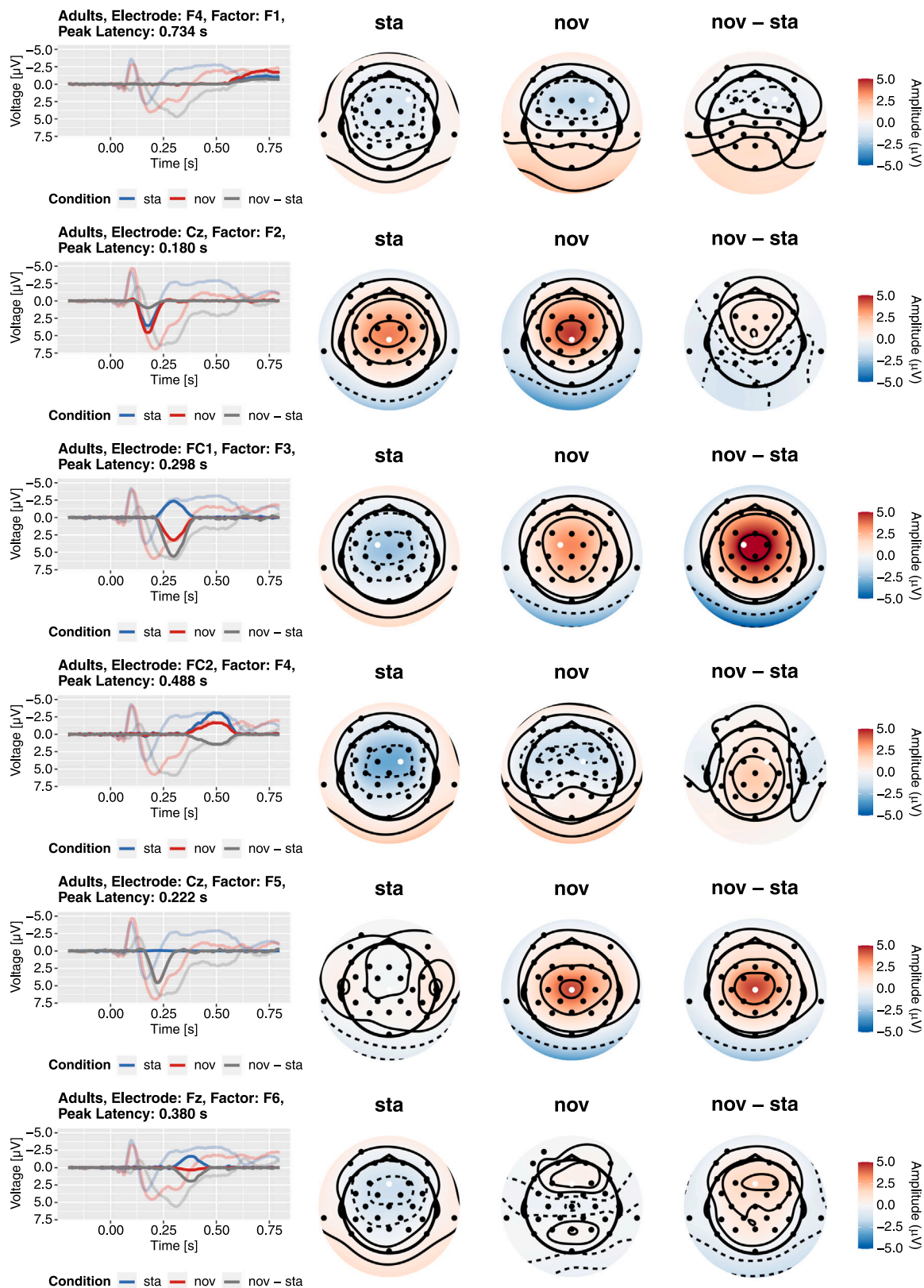
Fig. 9. *T*opographies and reconstructed ERPs for the first six factors from the child PCA. Each row represents a factor in the rotated solution. The left-most column depicts the contribution of the factor to the grand-average ERP (transparent lines) at the maximum electrode (marked by white dot). The other columns present the factors' topographies separately for all conditions. Please note that the extrapolated topography becomes unreliable outside the area covered by electrodes and tends towards extreme values due to the simplistic biharmonic interpolation. This plot is created in the script *03b_topoplot_allFactors.R*.

Fig. 10. Topographies and reconstructed ERPs for the first six factors from the adult PCA. Each row represents a factor in the rotated solution. The left-most column depicts the contribution of the factor to the grand-average ERP (transparent lines) at the maximum electrode (marked by white dot). The other columns present the factors' topographies separately for all conditions. Please note that the extrapolated topography becomes unreliable outside the area covered by electrodes and tends towards extreme values due to the simplistic biharmonic interpolation. This plot is created in the script *03b_topoplot_allFactors.R*.
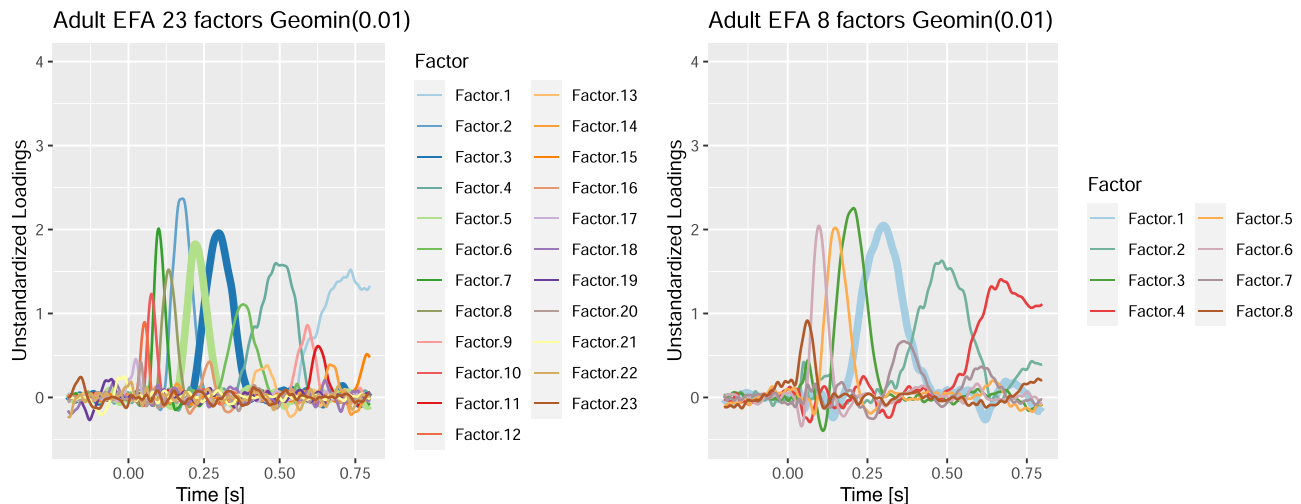
Fig. 11. Comparison of the reported solution with two suboptimal solutions. The left-hand panel contains our reported solution. The right-hand panel contains a suboptimal solution. The thicker lines highlight the factors mentioned in the text. In the right-hand panel, early and late P3a (Factors 3 and 5 in left panel) are conflated into a single factor.

detailed treatment of these methods is beyond the scope of this introductory tutorial, but we strongly encourage the reader to explore them following the familiarization with the temporal PCA approach.

Instead of the sampling points as in temporal PCA, the electrode sites can be treated as variables in a *spatial* PCA (Dien, 2010a; Dien et al., 2004, 2007; Dien and Frishkoff, 2005). Spatial PCA assumes equal topographic weights across participants, time points, and conditions but allows the time courses to vary. Whereas measurement invariance of the topography is clearly plausible in many situations (Delorme et al., 2012), spatial PCA requires the use of special rotation techniques because factors typically contribute to the voltage at *all* electrode sites. These rotation techniques (e.g., *Infomax*) are closely related to independent component analysis (ICA; Dien et al., 2007), and they aim to extract factors which are as stochastically independent as possible instead of rotating towards a simple structure (for mathematical introductions and tutorial treatments, see, e.g., De Lathauwer et al., 2000; Groppe et al., 2008; Huster and Raud, 2018; Makeig et al., 1996; Stone, 2004). Please note that there is an ongoing debate regarding the use of restricted versus unrestricted solutions for spatial decompositions as well (Artoni et al., 2018).

A further noteworthy analytic approach is to combine spatial and temporal PCA. That is, first, a temporal PCA is applied, followed by a spatial PCA on the temporal PCA results in a second step. This is called temporospatial PCA or spatiotemporal PCA when the PCAs are conducted in reversed order (Dien, 2010a; Dien et al., 2004; Spencer et al., 2001). The main idea is to complement the strengths of both approaches: Temporal PCA benefits from the fact that many of the factors are temporally separable but it fails to separate factors with identical time courses but separable topography. Spatial PCA suffers from the strong spatial overlap and could conflate factors with identical topography but separable time courses. A combination of these approaches could therefore considerably reduce the risk of conflating factors. In addition, spatial PCA offers an aggregated quantification across all electrodes which uses more of the available information compared to factor scores from single electrodes or averages across regions of interest. Alternatively, a trilinear decomposition of the ERP dataset has been proposed as an advancement of the PCA model which acknowledges the role of

electrode sites *and* sampling points as measurement occasions instead of treating one or the other as "observations" (Achim and Bouchard, 1997; Cong et al., 2015; Möcks, 1988b, 1988a; Wang et al., 2000). There have even been proposals how latency shifts can be directly acknowledged within such a model (Harshman et al., 2003; Harshman and Lundy, 1994; Mørup et al., 2008, 2011). Apart from modifications of the PCA model to ERP data, it has been suggested that PCA can also be helpful when conducted on typical transformations of ERP data such as scalp current density (Kayser and Tenke, 2015b, 2015a), time-frequency decompositions (Barry et al., 2019), or wavelet-transformations (Mørup et al., 2006b).

In sum, a variety of methods has been proposed to address certain limitations of the temporal PCA explained in this article or to adapt it to other use cases. All these methods are either direct modifications of the outlined approach or they estimate a very similar mathematical model. In that sense, the reader could see the temporal PCA approach as an entry point to the world of decomposition methods. The limitations of peak-based methods have been acknowledged for a long time. We are convinced that a basic understanding of decomposition methods is crucial for ERP researchers and that it will contribute to further progress in the field of developmental cognitive neuroscience. We hope that this tutorial and the provided code encourage the readers to apply this method to learn more about the development of the brain.

5. Conclusion

We set out from the fact that observed ERP data are a mixture of latent signals and that ignoring this issue can considerably bias substantive conclusions drawn from ERP data. Decomposition methods such as temporal PCA aim to model the latent signals. Although there are common challenges one must be aware of, temporal PCA is an effective tool for ERP researchers, especially in a developmental context where the noise level is typically high and the comparability of the ERP signals from different age groups is always at stake due to maturational changes. Extending previous tutorial treatments, the analytic procedure outlined here explicitly acknowledges that ERPs differ systematically between children and adults. We conclude that the use of decomposition

methods in ERP analysis can substantially improve research on the development of neuronal network activity underlying cognitive functions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Achim, A., Bouchard, S., 1997. Toward a dynamic topographic components model. Electroencephalogr. Clin. Neurophysiol. 103 (3), 381–385. https://doi.org/10.1016/S0013-4694(97)96055-0.

Achim, A., Marcantoni, W., 1997. Principal component analysis of event-related potentials: misallocation of variance revisited. Psychophysiology 34 (5), 597–606. https://doi.org/10.1111/j.1469-8986.1997.tb01746.x.

Arruda, J.E., Weiler, M.D., Valentino, D., Willis, W.G., Rossi, J.S., Stern, R.A., Gold, S.M., Costa, L., 1996. A guide for applying principal-components analysis and confirmatory factor analysis to quantitative electroencephalogram data. Int. J. Psychophysiol. 23 (1–2), 63–81. https://doi.org/10.1016/0167-8760(96)00032-3.

Artoni, F., Delorme, A., Makeig, S., 2018. Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition. NeuroImage 175 (February), 176–187. https://doi.org/10.1016/j.neuroimage.2018.03.016.

Asparouhov, T., Muthén, B., 2009. Exploratory structural equation modeling. Struct. Equ. Model.: A Multidiscip. J. 16 (3), 397–438. https://doi.org/10.1080/10705510903008204.

Auerswald, M., Moshagen, M., 2019. How to determine the number of factors to retain in exploratory factor analysis: a comparison of extraction methods under realistic conditions. Psychol. Methods 24 (4), 468–491. https://doi.org/10.1037/met0000200.

Bakeman, R., 2005. Recommended effect size statistics for repeated measures designs. Behav. Res. Methods 37 (3), 379–384. https://doi.org/10.3758/BF03192707.

Barry, R.J., De Blasio, F.M., Borchard, J.P., 2014. Sequential processing in the equiprobable auditory Go/NoGo task: children vs. adults. Clin. Neurophysiol. 125 (10), 1995–2006. https://doi.org/10.1016/j.clinph.2014.02.018.

Barry, R.J., De Blasio, F.M., Fogarty, J.S., Karamacoska, D., 2016. ERP Go/NoGo condition effects are better detected with separate PCAs. Int. J. Psychophysiol. 106, 50–64. https://doi.org/10.1016/j.ijpsycho.2016.06.003.

Barry, R.J., De Blasio, F.M., Karamacoska, D., 2019. Data-driven derivation of natural EEG frequency components: an optimised example assessing resting EEG in healthy ageing. J. Neurosci. Methods 321, 1–11. https://doi.org/10.1016/j.jneumeth.2019.04.001.

Beauducel, A., 2001. Problems with parallel analysis in data sets with oblique simple structure. MPR-Online 6, 141–157. ⟨http://www.mpr-online.de⟩.

Beauducel, A., 2018a. Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. J. Neurosci. Methods 295, 20–36. https://doi.org/10.1016/j.jneumeth.2017.11.011.

Beauducel, A., 2018b. Recovering Wood and McCarthy's ERP-prototypes by means of ERP-specific procrustes-rotation. J. Neurosci. Methods 295, 20–36. https://doi.org/10.1016/j.jneumeth.2017.11.011.

Beauducel, A., Debener, S., 2003. Misallocation of variance in event-related potentials: simulation studies on the effects of test power, topography, and baseline-to-peak versus principal component quantifications. J. Neurosci. Methods 124 (1), 103–112. https://doi.org/10.1016/S0165-0270(02)00381-3.

Beauducel, A., Debener, S., Brocke, B., Kayser, J., 2000. On the reliability of augmenting/reducing: peak amplitudes and principal component analysis of auditory evoked potentials. J. Psychophysiol. 14 (4), 226–240.

Beauducel, A., Hilger, N., 2018. On optimal allocation of treatment / condition variance in principal component analysis. Int J. Stat. Probab. 7 (4), 50–56. https://doi.org/10.5539/ijsp.v7n4p50.

Beauducel, A., Leue, A., 2015. Controlling for experimental effects in event-related potentials by means of principal component rotation. J. Neurosci. Methods 239, 139–147. https://doi.org/10.1016/j.jneumeth.2014.10.008.

Benjamini, Y., 2010. Simultaneous and selective inference: current successes and future challenges. Biom. J. 52 (6), 708–721. https://doi.org/10.1002/bimj.200900299.

Bernaards, C.A., Jennrich, R.I., 2005. Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. Educ. Psychol. Meas. 65, 676–696.

Bonmassar, C., Widmann, A., Wetzel, N., 2020. The impact of novelty and emotion on attention-related neuronal and pupil responses in children. Dev. Cogn. Neurosci. 42 (February), 100766 https://doi.org/10.1016/j.dcn.2020.100766.

Boxtel, G.J.M., 1998. Computational and statistical methods for analyzing event-related potential data. Behav. Res. Methods, Instrum., Comput. 30 (1), 87–102. https://doi.org/10.3758/BF03209419.

Bradley, M.M., Keil, A., 2012. Event-Related Potentials (ERPs). Encyclopedia of Human Behavior. Elsevier, pp. 79–85. https://doi.org/10.1016/B978-0-12-375000-6.00154-3.

Braeken, J., Van Assen, M.A.L.M., 2017. An empirical Kaiser criterion. Psychol. Methods 22 (3), 450–466. https://doi.org/10.1037/met0000074.

Browne, M.W., 2001. An overview of analytic rotation in Exploratory Factor Analysis. Multivar. Behav. Res. 36 (1), 111–150. https://doi.org/10.1207/S15327906MBR360105.

Buja, A., Eyuboglu, N., 1992. Remarks on parallel analysis. Multivar. Behav. Res. 27 (4), 509–540. https://doi.org/10.1207/s15327906mbr2704_2.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., Vidal, F., 2015. Spatial and temporal resolutions of EEG: is it really black and white? A scalp current density view. Int. J. Psychophysiol. 97 (3), 210–220. https://doi.org/10.1016/j.ijpsycho.2015.05.004.

Chapman, R.M., McCrary, J.W., 1995. EP component identification and measurement by principal components analysis. Brain Cogn. Vol. 27 (Issue 3), 288–310. https://doi.org/10.1006/brcg.1995.1024.

Chaumon, M., Bishop, D.V.M., Busch, N.A., 2015. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J. Neurosci. Methods 250, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025.

Collet, W., 1989. Doubts on the adequacy of the principal component varimax analysis for the identification of event-related brain potential components: a commentary on Glaser and Ruchkin, and Donchin and Heffley. Biol. Psychol. 28 (2), 163–172. ⟨http://www.ncbi.nlm.nih.gov/pubmed/2775805⟩.

Cong, F., Lin, Q.-H., Kuang, L.-D., Gong, X.-F., Astikainen, P., Ristaniemi, T., 2015. Tensor decomposition of EEG signals: a brief review. J. Neurosci. Methods 248, 59–69. https://doi.org/10.1016/j.jneumeth.2015.03.018.

Conti, G., Frühwirth-Schnatter, S., Heckman, J.J., Piatek, R., 2014. Bayesian exploratory factor analysis. J. Econ. 183 (1), 31–57. https://doi.org/10.1016/j.jeconom.2014.06.001.

Craddock, M., 2021. eegUtils: Utilities for Electroencephalographic (EEG) Analysis.

Crawford, A.V., Green, S.B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., Thompson, M.S., 2010. Evaluation of parallel analysis methods for determining the number of factors. Educ. Psychol. Meas. 70, 885–901. https://doi.org/10.1177/0013164410379332.

Cureton, E.E., 1976. Studies of the Promax and Optres rotations. Multivar. Behav. Res.

De Lathauwer, L., De Moor, B., Vandewalle, J., 2000. An introduction to independent component analysis. J. Chemom. 14 (3), 123–149. https://doi.org/10.1002/1099-128X(200005/06)14:3<123::AID-CEM589>3.0.CO;2-1.

De Munck, J.C., Bijma, F., Gaura, P., Sieluzycki, C.A., Branco, M.I., Heethaar, R.M., 2004. A maximum-likelihood estimator for trial-to-trial variations in noisy MEG/EEG data sets. IEEE Trans. Biomed. Eng. 51 (12), 2123–2128. https://doi.org/10.1109/TBME.2004.836515.

De Winter, J.C.F., Dodou, D., 2012. Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. J. Appl. Stat. 39 (4), 695–710. https://doi.org/10.1080/02664763.2011.610445.

De Winter, J.C.F., Dodou, D., 2016. Common factor analysis versus principal component analysis: a comparison of loadings by means of simulations. Commun. Stat. - Simul. Comput. 45 (1), 299–321. https://doi.org/10.1080/03610918.2013.862274.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S., 2012. Independent EEG sources are dipolar. PLoS ONE 7 (2), 30135. https://doi.org/10.1371/journal.pone.0030135.

Dien, J., 1998. Addressing misallocation of variance in principal components analysis of event-related potentials. Brain Topogr. 11 (1), 43–55. https://doi.org/10.1023/A:1022218503558.

Dien, J., 2006. Progressing towards a consensus on PCA of ERPs. Clin. Neurophysiol. 117 (3), 699–702. https://doi.org/10.1016/j.clinph.2005.09.029.

Dien, J., 2010a. Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. Psychophysiology 47 (1), 170–183. https://doi.org/10.1111/j.1469-8986.2009.00885.x.

Dien, J., 2010b. The ERP PCA Toolkit: an open source program for advanced statistical analysis of event-related potential data. J. Neurosci. Methods 187 (1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009.

Dien, J., 2012. Applying principal components analysis to event-related potentials: a tutorial. Dev. Neuropsychol. 37 (6), 497–517. https://doi.org/10.1080/87565641.2012.697503.

Dien, J.2018. ERP PCA Toolkit 2.66 Tutorial. University of Maryland. ⟨http://sourceforge.net/projects/erppcatoolkit⟩.

Dien, J., Beal, D.J., Berg, P., 2005. Optimizing principal components analysis of event-related potentials: matrix type, factor loading weighting, extraction, and rotations. Clin. Neurophysiol. 116 (8), 1808–1825. https://doi.org/10.1016/j.clinph.2004.11.025.

Dien, J., Frishkoff, G.A., 2005. Principal components analysis of event-related potential datasets. In: Handy, T. (Ed.), Event-related Potentials: A Methods Handbook. MIT Press, pp. 189–208.

Dien, J., Khoe, W., Mangun, G.R., 2007. Evaluation of PCA and ICA of simulated ERPs: Promax vs. infomax rotations. Hum. Brain Mapp. 28 (8), 742–763. https://doi.org/10.1002/hbm.20304.

Dien, J., Spencer, K.M., Donchin, E., 2003. Localization of the event-related potential novelty response as defined by principal components analysis. Brain Res. Cogn. Brain Res. 17 (3), 637–650. 〈http://www.ncbi.nlm.nih.gov/pubmed/14561451〉.

Dien, J., Spencer, K.M., Donchin, E., 2004. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology 41 (5), 665–678. https://doi.org/10.1111/j.1469-8986.2004.00193.x.

Dien, J., Tucker, D.M., Potts, G., Hartry-Speiser, A., 1997. Localization of auditory evoked potentials related to selective intermodal attention. J. Cogn. Neurosci. 9 (6), 799–823. https://doi.org/10.1162/jocn.1997.9.6.799.

DiStefano, C., Zhu, M., Mîndrilă, D., 2009. Understanding and using factor scores: considerations for the applied researcher. *Practical Assessment.* Res. Eval. 14 (20), 1–11 https://doi.org/10.1.1.460.8553.

Donchin, E., 1966. A multivariate approach to the analysis of average evoked potentials. IEEE Trans. Biomed. Eng. 13 (3), 131–139. https://doi.org/10.1109/TBME.1966.4502423.

Donchin, E., 1978. Multivariate analysis of event-related potential data: a tutorial review. Multidiscip. Perspect. Event-Relat. Brain Potential Res. 555–572.

Donchin, E., 1989. On why Collet's doubts regarding the PCA are misplaced. Biol. Psychol. 28 (2), 181–186. https://doi.org/10.1016/0301-0511(89)90100-2.

Donchin, E., Heffley, E.F., 1978. Multivariate analysis of ERP data: a tutorial review. In: Otto, D. (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Government Printing Office, pp. 555–572.

Escera, C., Alho, K., Schröger, E., Winkler, I.W., 2000. Involuntary attention and distractibility as evaluated with event-related brain potentials. Audiol. Neurotol. 5 (3–4), 151–166. https://doi.org/10.1159/000013877.

Fava, J.L., Velicer, W.F., 1992. The effects of overextraction on factor and component analysis. Multivar. Behav. Res. 27 (3), 387–415. https://doi.org/10.1207/s15327906mbr2703_5.

Fogarty, J.S., Barry, R.J., De Blasio, F.M., Steiner, G.Z., 2018. ERP components and behavior in the auditory equiprobable go/no-go task: Inhibition in young adults. Psychophysiology., July 2017, e13065. https://doi.org/10.1111/psyp.13065.

Gorsuch, R.L. , 1983. Factor Analysis (2nd ed.).

Green, S.B., Levy, R., Thompson, M.S., Lu, M., Lo, W.J., 2012. A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. Educ. Psychol. Meas. 72 (3), 357–374. https://doi.org/10.1177/0013164411422252.

Grieder, S., Steiner, M.D., 2020. Algorithmic jingle jungle: a comparison of implementations of principal axis factoring and promax rotation in R and SPSS. J. Mater. Process. Technol. Vol. 1 (Issue 1) https://doi.org/10.31234/osf.io/7hwrm.

Groppe, D.M., Makeig, S., Kutas, M., Diego, S., 2008. Identifying reliable independent components via split-half comparisons. Indep. Compon. Anal. Event-Relat. Potentials 45, 1–44.

Harshman, R.A., Hong, S., Lundy, M.E., 2003. Shifted factor analysis – part I: models and properties. J. Chemom. 17 (7), 363–378. https://doi.org/10.1002/cem.808.

Harshman, R.A., Lundy, M.E., 1994. PARAFAC: parallel factor analysis. Comput. Stat. Data Anal. 18 (1), 39–72. https://doi.org/10.1016/0167-9473(94)90132-5.

Hattori, M., Zhang, G., Preacher, K.J., 2017. Multiple local solutions and geomin rotation. Multivar. Behav. Res. 52 (6), 720–731. https://doi.org/10.1080/00273171.2017.1361312.

Haumann, N.T., Hansen, B., Huotilainen, M., Vuust, P., Brattico, E., 2020. Applying stochastic spike train theory for high-accuracy human MEG/EEG. J. Neurosci. Methods 340, 108743. https://doi.org/10.1016/j.jneumeth.2020.108743.

Hendrickson, A.E., White, P.O., 1964. PROMAX: a quick method for rotation to oblique simple structure. Br. J. Stat. Psychol. 17 (1), 65–70. https://doi.org/10.1111/j.2044-8317.1964.tb00244.x.

Hershberger, S.L., 2014. Factor score estimation. Wiley StatsRef: Stat. Ref. Online 2, 636–644. https://doi.org/10.1002/9781118445112.stat06532.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30 (2), 179–185. https://doi.org/10.1007/BF02289447.

Huang, P.-H., 2020. Lslx: Semi-confirmatory structural equation modeling via penalized likelihood. J. Stat. Softw. 93 (7) https://doi.org/10.18637/jss.v093.i07.

Huang, P.-H., Chen, H., Weng, L.-J., 2017. A penalized likelihood method for structural equation modeling. Psychometrika 82 (2), 329–354. https://doi.org/10.1007/s11336-017-9566-9.

Huster, R.J., Raud, L., 2018. A tutorial review on multi-subject decomposition of EEG. Brain Topogr. 31 (1), 3–16. https://doi.org/10.1007/s10548-017-0603-x.

Jennrich, R.I., 2004a. Rotation to simple loadings using component loss functions: the orthogonal case. Psychometrika 69 (2), 257–273. https://doi.org/10.1007/BF02295943.

Jennrich, R.I., 2004b. Derivative free gradient projection algorithms for rotation. Psychometrika 69 (3), 475–480. https://doi.org/10.1007/BF02295647.

Jennrich, R.I., 2006. Rotation to simple loadings using component loss functions: the oblique case. Psychometrika 71 (1), 173–191. https://doi.org/10.1007/s11336-003-1136-B.

Kayser, J., Tenke, C.E., 2003. Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. Clin. Neurophysiol. 114 (12), 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 components analysis in electrophysiology. Clin. Neurophysiol. 116 (8), 1747–1753. https://doi.org/10.1016/j.clinph.2005.03.020.

Kayser, J., Tenke, C.E., 2006. Consensus on PCA for ERP data, and sensibility of unrestricted solutions. Clin. Neurophysiol. 117 (3), 703–707. https://doi.org/10.1016/j.clinph.2005.11.015.

Kayser, J., Tenke, C.E., 2015a. Issues and considerations for using the scalp surface laplacian in EEG/ERP research: a tutorial review. Int. J. Psychophysiol. 97 (3), 189–209. https://doi.org/10.1016/j.ijpsycho.2015.04.012.

Kayser, J., Tenke, C.E., 2015b. On the benefits of using surface Laplacian (current source density) methodology in electrophysiology. Int. J. Psychophysiol. 97 (3), 171–173. https://doi.org/10.1016/j.ijpsycho.2015.06.001.

Kieffaber, P.D., Hetrick, W.P., 2005. Event-related potential correlates of task switching and switch costs. Psychophysiology 42 (1), 56–71. https://doi.org/10.1111/j.1469-8986.2005.00262.x.

Kiesel, A., Miller, J., Jolicœur, P., Brisson, B., 2008. Measurement of ERP latency differences: a comparison of single-participant and jackknife-based scoring methods. Psychophysiology 45 (2), 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x.

Kline, R.B., 2016. In: Little, T.D. (Ed.), Principles and Practice of Structural Equation Modeling, fourth ed.,. The Guilford Press.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540. https://doi.org/10.1038/nn.2303.

Li, Y., Wen, Z., Hau, K.T., Yuan, K.H., Peng, Y., 2020. Effects of Cross-loadings on determining the number of factors to retain. Struct. Equ. Model. 27 (6), 841–863. https://doi.org/10.1080/10705511.2020.1745075.

Lim, S., Jahng, S., 2019. Determining the number of factors using parallel analysis and its recent variants. Psychol. Methods 24 (4), 452–467. https://doi.org/10.1037/met0000230.

Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8 (1 APR), 1–14. https://doi.org/10.3389/fnhum.2014.00213.

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. https://doi.org/10.1027/1614-2241.2.2.57.

Luck, S.J., 2005a. An Introduction to Event-Related Potentials and Their Neural Origins. An Introduction to the Event-Related Potential Technique, 2–50. doi: 10.1007/s10409-008-0217-3.

Luck, S.J., 2005b. Ten simple rules for designing ERP experiments. In: Handy, T.C. (Ed.), Event-Related Potentials: A Methods Handbook. MIT Press, pp. 17–32.

Luck, S.J., 2014. An Introduction to the Event-related Potential Technique, second ed. MIT Press.

Luck, S.J., Gaspelin, N., 2016. How to get statistically significant effects in any ERP experiment (and Why You Shouldn't). Psychophysiology 1 (4424), 1–44. https://doi.org/10.1111/psyp.12639.

Luck, S.J., Kappenman, E.S. (Eds.), 2012. The Oxford Handbook of Event-Related Potential Components. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.001.0001.

Makeig, S., Bell, A., Jung, T.-P., Sejnowski, T., 1996. Independent Component Analysis of Electroencephalographic Data. 8.

Makowski, D., Ben-Shachar, M.S., Lüdecke, D., 2019. bayestestR: describing effects and their uncertainty, existence and significance within the Bayesian Framework. J. Open Source Softw. 4 (40), 1541. https://doi.org/10.21105/joss.01541.

Marsh, H.W., Guo, J., Parker, P.D., Nagengast, B., Asparouhov, T., Muthén, B., Dicke, T., 2017. What to do when scalar invariance fails: the extended alignment method for multi-group factor analysis comparison of latent means across many groups. Psychol. Methods 23, 524–545. https://doi.org/10.1037/met0000113.

Marsh, H.W., Morin, A.J.S., Parker, P.D., Kaur, G., 2014. Exploratory Structural Equation Modeling: An integration of the best features of Exploratory and Confirmatory Factor Analysis. Annu. Rev. Clin. Psychol. 10, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S., Trautwein, U., 2009. Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. Struct. Equ. Model.: A Multidiscip. J. 16 (3), 439–476. https://doi.org/10.1080/10705510903008220.

Marsh, H.W., Nagengast, B., Morin, A.J.S., 2013. Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. Dev. Psychol. 49 (6), 1194–1218. https://doi.org/10.1037/a0026913.

Mellenbergh, G.J., 1989. Item bias and item response theory. Int. J. Educ. Res. 13 (2), 127–143. https://doi.org/10.1016/0883-0355(89)90002-5.

Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58 (4), 525–543. https://doi.org/10.1007/BF02294825.

Miller, J., Patterson, T., Ulrich, R., 1998. Jackknife-based method for measuring LRP onset latency differences. Psychophysiology 35 (1), 99–115. https://doi.org/10.1017/S0048577298000857.

Möcks, J., 1986. The influence of latency jitter in principal component analysis of event-related potentials. Psychophysiology Vol. 23 (Issue 4), 480–484. https://doi.org/10.1111/j.1469-8986.1986.tb00659.x.

Möcks, J., 1988a. Decomposing event-related potentials: a new topographic components model. Biol. Psychol. 26 (1–3), 199–215. https://doi.org/10.1016/0301-0511(88)90020-8.

Möcks, J., 1988b. Topographic components model for event-related potentials and some biophysical considerations. IEEE Trans. Biomed. Eng. 35 (6), 482–484. https://doi.org/10.1109/10.2119.

Möcks, J., 1989. Doubting these doubts—a reply to Collet. Biol. Psychol. 28 (2), 173–180. https://doi.org/10.1016/0301-0511(89)90099-9.

Möcks, J., Köhler, W., Gasser, T., Pham, D.T., 1988. Novel approaches to the problem of latency jitter. Psychophysiology 25 (2), 217–226. https://doi.org/10.1111/j.1469-8986.1988.tb00992.x.

Möcks, J., Verleger, R., 1985. Nuisance sources of variance in principal components analysis of event-related potentials. Psychophysiology 22 (6), 674–688. https://doi.org/10.1111/j.1469-8986.1985.tb01667.x.

Möcks, J., Verleger, R., 1986. Principal component analysis of event-related potentials: a note on misallocation of variance. Electroencephalogr. Clin. Neurophysiol. / Evoked Potentials 65 (5), 393–398. https://doi.org/10.1016/0168-5597(86)90018-3.

Molenaar, P.C.M., Campbell, C.G., 2009. The new person-specific paradigm in psychology. Curr. Dir. Psychol. Sci. 18 (2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x.

Morey, R.D., Rouder, J.N., 2018. BayesFactor: Computation of Bayes Factors for Common Designs. https://cran.r-project.org/package=BayesFactor.

Morin, A.J.S., Marsh, H.W., Nagengast, B., 2013. Exploratory structural equation modeling: an introduction. In: Hancock, G.R., Mueller, R.O. (Eds.), Structural Equation Modeling: A Second Course, second ed. Information Age Publishing, Inc,.

Mørup, M., Hansen, L.K., Arnfred, S.M., Lim, L.H., Madsen, K.H., 2008. Shift-invariant multilinear decomposition of neuroimaging data. NeuroImage 42 (4), 1439–1450. https://doi.org/10.1016/j.neuroimage.2008.05.062.

Mørup, M., Hansen, L.K., Herrmann, C.S., Parnas, J., Arnfred, S.M., 2006a. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. NeuroImage 29 (3), 938–947. https://doi.org/10.1016/j.neuroimage.2005.08.005.

Mørup, M., Hansen, L.K., Herrmann, C.S., Parnas, J., Arnfred, S.M., 2006b. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. NeuroImage 29 (3), 938–947. https://doi.org/10.1016/j.neuroimage.2005.08.005.

Mørup, M., Hansen, L.K., Madsen, K.H.2011. Modeling latency and shape changes in trial based neuroimaging data. Conference Record – Asilomar Conference on Signals, Systems and Computers, 439–443. doi: 10.1109/ACSSC.2011.6190037.

Mulaik, S.A., 2010. Foundations of Factor Analysis, second ed. CRC Press.

Muthén, B.O. , 2004. Mplus Technical Appendices. ⟨https://www.statmodel.com/download/techappen.pdf⟩.

Muthén, B.O., Muthén, L.K., 2004. Technical Appendices. Muthén& Muthén, Los Angeles.

Nesselroade, J.R., 2007. Factoring at the individual level: some matters for the second century of factor analysis. In Factor Analysis at 100: Historical Developments and Future Directions (pp. 249–264).

Nesselroade, J.R., Molenaar, P.C.M., 2016a. A Rejoinder. Multivar. Behav. Res. 51 (2–3), 428–431. https://doi.org/10.1080/00273171.2015.1101368.

Nesselroade, J.R., Molenaar, P.C.M., 2016b. Some behavioral science measurement concerns and proposals. Multivar. Behav. Res. 51 (2–3), 396–412. https://doi.org/10.1080/00273171.2015.1050481.

Nguyen, Hoang V., Waller, Niels G., 2022. Local Minima and Factor Rotations in Exploratory Factor Analysis. Psychological Methods. https://doi.org/10.1037/met0000467. In press.

Nunez, P.L., Srinivasan, R., 2006. Electric Fields of the Brain: The Neurophysics of EEG, second ed. Oxford Univ. Press ⟨https://katalog.ub.uni-leipzig.de/Record/0000771752⟩.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1–9. https://doi.org/10.1155/2011/156869.

Ouyang, G., Hildebrandt, A., Sommer, W., Zhou, C., 2017. Exploiting the intra-subject latency variability from single-trial event-related potentials in the P3 time range: a review and comparative evaluation of methods. Neurosci. Biobehav. Rev. 75, 1–21. https://doi.org/10.1016/j.neubiorev.2017.01.023.

Piai, V., Dahlslätt, K., Maris, E., 2015. Statistically comparing EEG/MEG waveforms through successive significant univariate tests: how bad can it be? Psychophysiology 52 (3), 440–443. https://doi.org/10.1111/psyp.12335.

Putnick, D.L., Bornstein, M.H., 2017. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev. Rev. 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.Measurement.

R. Core Team, 2020. R: A Language and Environment for Statistical Computing (4.0.1). https://www.r-project.org/.

Revelle, W. , 2021. psych: Procedures for Psychological, Psychometric, and Personality Research. ⟨https://cran.r-project.org/package=psych⟩.

Riggins, T., Scott, L.S., 2020. P300 development from infancy to adolescence. Psychophysiology 57 (7), 13346. https://doi.org/10.1111/psyp.13346.

Rinaldo, A., Wasserman, L., G'Sell, M., 2019. Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. Ann. Stat. 47 (6), 3438–3469. https://doi.org/10.1214/18-aos1784.

Rösler, F., Manzey, D., 1981. Principal components and varimax-rotated components in event-related potential research: Some remarks on their interpretation. Biological Psychology 13, 3–26. https://doi.org/10.1016/0301-0511(81)90024-7.

Saccenti, E., Timmerman, M.E., 2017. Considering Horn's Parallel Analysis from a random matrix theory point of view. Psychometrika 82 (1), 186–209. https://doi.org/10.1007/s11336-016-9515-z.

Scharf, F., Nestler, S., 2018. Principles behind variance misallocation in temporal exploratory factor analysis for ERP data: insights from an inter-factor covariance decomposition. Int. J. Psychophysiol. 128, 119–136. https://doi.org/10.1016/j.ijpsycho.2018.03.019.

Scharf, F., Nestler, S., 2019a. A comparison of simple structure rotation criteria in temporal exploratory factor analysis for event-related potential data. Methodology 15 (Supplement 1), 43–60. https://doi.org/10.1027/1614-2241/a000175.

Scharf, F., Nestler, S., 2019b. Exploratory structural equation modeling for event-related potential data—an all-in-one approach? Psychophysiology 56 (3), e13303. https://doi.org/10.1111/psyp.13303.

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M.S. , 2021. afex: Analysis of Factorial Experiments. ⟨https://cran.r-project.org/package=afex⟩.

Skrondal, A., Laake, P., 2001. Regression among factor scores. Psychometrika 66 (4), 563–575. https://doi.org/10.1007/BF02296196.

Slotnick, S.D., 2005. Source localization. In: Handy, T. (Ed.), Event-related Potentials: A Methods Handbook. MIT Press, pp. 149–166.

Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology 38 (2), 343–358. https://doi.org/10.1017/S0048577201000324.

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing transparency through a multiverse analysis. Perspect. Psychol. Sci. 11 (5), 702–712. https://doi.org/10.1177/1745691616658637.

Steiner, M.D., Grieder, S., 2020. EFAtools: an R package with fast and flexible implementations of exploratory factor analysis tools. J. Open Source Softw. 5 (53), 2521. https://doi.org/10.21105/joss.02521.

Stone, J.V., 2004. Enzymatic digestion of proteins in gels for mass spectrometric identification and structural analysis. Indep. Compon. Anal.: a Tutor. Introd. Chapter 11, 11.

Thomson, G.H., 1935. The definition and measurement of "g" (general intelligence). J. Educ. Psychol. 26 (4), 241–262. https://doi.org/10.1037/h0059873.

Thomson, G.H., 1938. Methods of estimating mental factors, 246–246 Nature 141 (3562), 246. https://doi.org/10.1038/141246a0.

Thurstone, L.L., 1935. The Vectors of Mind: Multiple-factor Analysis for the Isolation of Primary Traits. University of Chicago Press. https://doi.org/10.1037/10018-000.

Trendafilov, N.T., 2014. From simple structure to sparse components: a review. Comput. Stat. 29 (3–4), 431–454. https://doi.org/10.1007/s00180-013-0434-5.

Truccolo, W., Knuth, K.H., Shah, A., Bressler, S.L., Schroeder, C.E., Ding, M., 2003. Estimation of single-trial multicomponent ERPs: differentially variable component analysis (dVCA). Biol. Cybern. 89 (6), 426–438. https://doi.org/10.1007/s00422-003-0433-7.

Tuan, P.D., Mocks, J., Kohler, W., Gasser, T., 1987. Variable latencies of noisy signals: Estimation and testing in brain potential data. Biometrika Vol. 74 (Issue 3), 525–533. ⟨https://academic.oup.com/biomet/article-abstract/74/3/525/238523⟩.

Ulrich, R., Miller, J., 2001. Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. Psychophysiology 38 (5), 816–827. https://doi.org/10.1017/S0048577201000610.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., Zondervan-Zwijnenburg, M., 2015. Editorial: measurement invariance. Front. Psychol. 6 (July), 10–13. https://doi.org/10.3389/fpsyg.2015.01064.

Verleger, R., Möcks, J., 1987. Varimax may produce slow-wave-like shapes by merging monotonic trends with other components. J. Psychophysiol. 1 (3), 265–270.

Wang, K., Begleiter, H., Porjesz, B., 2000. Trilinear modeling of event-related potentials. Brain Topogr. 12 (4), 263–271. ⟨http://www.ncbi.nlm.nih.gov/pubmed/10912734⟩.

Wasserman, L., Roeder, K., 2009. High-dimensional variable selection. Ann. Stat. 37 (5A), 2178–2201. https://doi.org/10.1214/08-AOS646.

Weide, A.C., Beauducel, A., 2019. Varimax rotation based on gradient projection is a feasible alternative to SPSS. Front. Psychol. 10. https://doi.org/10.3389/fpsyg.2019.00645.

Welzel, C., Brunkert, L., Kruse, S., Inglehart, R.F., 2021. Non-invariance? An overstated problem with misconceived causes. Sociol. Methods Res., 004912412199552 https://doi.org/10.1177/0049124121995521.

Widaman, K.F., 1990. Bias in pattern loadings represented by common factor analysis and component analysis. Multivar. Behav. Res. 25 (1), 89–95. https://doi.org/10.1207/s15327906mbr2501_11.

Widaman, K.F., 2007. Common factors versus components: principals and principles, errors and misconceptions. In. In: Cudeck, R., MacCallum, R.C. (Eds.), Factor Analysis at 100: Historical Developments and Future Directions. Lawrence Erlbaum Associates, pp. 177–203.

Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data – a practical approach. J. Neurosci. Methods 250, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002.

Widmann, A., Schröger, E., Wetzel, N., 2018. Emotion lies in the eye of the listener_ Emotional arousal to novel sounds is reflected in the sympathetic contribution to the pupil dilation response and the P3. doi: 10.1016/j.biopsycho.2018.01.010.

Wood, C.C., McCarthy, G., 1984. Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. Electroencephalogr. Clin. Neurophysiol. 59 (3), 249–260. ⟨http://www.ncbi.nlm.nih.gov/pubmed/6203715⟩.

Wood, J.M., Tataryn, D.J., Gorsuch, R.L., 1996. Effects of under- and overextraction on principal axis factor analysis with varimax rotation. Psychol. Methods 1 (4), 354–365. https://doi.org/10.1037/1082-989X.1.4.354.

Wunderlich, J.L., Cone-Wesson, B.K., Shepherd, R., 2006. Maturation of the cortical auditory evoked potential in infants and young children. Hear. Res. 212 (1–2), 185–202. https://doi.org/10.1016/j.heares.2005.11.010.

Yates, A., 1987. Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis. State Univ. New York Press.

Zhang, G., Hattori, M., Trichtinger, L.A., Wang, X., 2018. Target rotation with both factor loadings and factor correlations. Psychol. Methods 24 (3), 390–402. https://doi.org/10.1037/met0000198.