Investigating paired comparisons after principal component analysis in data sets with special structures

J. C. Castura[1*], P. Varela[2], T. Næs[2]

[1]Compusense Inc., 255 Speedvale Av. W., Guelph, Ontario, N1H 1C5, Canada

[2]Nofima AS, Osloveien 1, P.O. Box 210, N-1431 Ås, Norway

* jcastura@compusense.com

Abstract

Principal component analysis (PCA) is a popular technique for summarizing and exploring multivariate data sets, including sensory evaluation data sets. We propose how to conduct PCA of results matrices with a special structure in which only a subset of the product paired comparisons are of interest. We illustrate the proposed approach with two data sets, both from trained sensory panels. In the first example, assessors evaluated the intensities of multiple sensory attributes in a control smoothie and nine test smoothie formulations. In this example, the control-test paired comparisons are of primary interest, not the test-test pair comparisons. In the second example, assessors characterized several yogurt formulations continuously over time during consumption using a method for temporal sensory profiling. In this example, we considered the within-timepoint paired comparisons to be of primary interest. It is possible to conduct PCA conventionally based on each panel's results. Doing so will extract variance from the matrix columns maximally, yielding the optimal space for investigating the variance in all paired comparisons. But this solution does not extract variance maximally from only the relevant subset of paired comparisons, indicating that the PCA conducted conventionally does not yield the optimal space for investigating the variance from only these relevant pairs. In this manuscript, we find this optimal space by submitting to PCA a results matrix containing only the paired comparisons that are of primary interest. The PCA solution extracts a larger proportion of the sum of squares from the relevant paired comparisons and better separates the relevant pairs than a PCA conducted conventionally. We show visually and numerically the advantages of the proposed approach. The methods proposed in this paper can be adapted to investigate data sets that have other special structures in sensory evaluation and in other domains.

Keywords: principal component analysis (PCA); multivariate analysis; bootstrap; paired comparisons; sensory evaluation; trained sensory panel.

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).

Highlights

- In some data sets, only a subset of paired comparisons are of primary interest
- Demonstration of how to conduct PCA focusing on a subset of paired comparisons
- Comparison of PCA conducted conventionally vs PCA of a subset of paired comparisons
- Data sets can be analyzed using both approaches
- Gain from PCA of relevant paired comparisons can be substantial

Abbreviations

PCA principal component analysis

PC principal component

$\mathbf{X}$ a column-centered (J×M) matrix of results to be analyzed

$\mathbf{X}\ominus\mathbf{X}$ a (J^2xM) matrix of all paired differences obtained by crossdiff-unfolding $\mathbf{X}$ (subtracting every row in $\mathbf{X}$ from each row in $\mathbf{X}$; see Castura, Varela & Næs, 2023)

$\mathbf{\Delta}^*$ a ($2C$xM) matrix containing only the $2C$ twinned paired difference rows in $\mathbf{X}\ominus\mathbf{X}$ for the C relevant paired comparisons

$\mathbf{T}_A$ score matrix obtained from PCA of $\mathbf{X}$ retaining the first A PCs

$\mathbf{P}_A$ loading matrix obtained from PCA of either $\mathbf{X}$ or $\mathbf{X}\ominus\mathbf{X}$ retaining the first A PCs

$\mathbf{T}_A$* score matrix obtained from PCA of $\mathbf{\Delta}^*$ retaining the first A PCs

$\mathbf{P}_A$* loading matrix obtained from PCA of $\mathbf{\Delta}^*$ retaining the first A PCs

1. Introduction

Sensory evaluation often produces multivariate data sets that can be investigated using principal component analysis (PCA; Mardia, Bibby & Kent, 1979). PCA compresses most of the variance from the original correlated sensory attributes (variables) into only a few principal components (PCs). The results in these PCs can be plotted to provide a visual summary. Coefficients called *loadings* define the linear combination of variables that comprise each PC. Product coordinates in each PC are called *scores*. Scores and loadings can be visualized in score plots, or together with loadings in biplots. Scores are usually represented as points. Some authors use a bootstrap procedure (Efron & Tibshirani, 1994) to investigate the uncertainty of these points (e.g. Cadoret & Husson, 2013; Courcoux, Qannari, Taylor, Buck & Greenhoff, 2012; Babamoradi, van den Berg & Rinnan, 2013; Lebart, 2007; Kiers & Groenen, 2006; Husson, Lê & Pagès, 2005). The uncertainty of paired difference scores can also be used to determine which pairs of products the panel discriminates (Castura, Varela & Næs, 2023a; Castura, Varela & Næs, 2023b; Castura, Rutledge, Ross & Næs, 2022). In the present manuscript, we consider how to conduct

PCA on data sets that have a special structure. We give two examples of data sets with special structures. For each data set, we discuss how PCA is applied conventionally. Then we discuss why the special structure could lead us to apply PCA to a modified results matrix that leads to a different, more relevant exploratory data analysis.

One type of special structure occurs when there is a control product to be compared with many test products. This structure is exemplified in this paper using a quantitative descriptive analysis of 10 smoothie formulations from a trained sensory panel, presented previously by Galler, Næs, Almli, and Varela (2020); this data set is described further in Section 2.1. Conventionally, the results are summarized in a products-by-attributes matrix of panel mean values. However, we might not be interested in all paired comparisons because this data set has a special structure: one smoothie is a “control” formulation against which the other nine “test” formulations will be compared. In the case, the control-test paired comparisons are what is of primary interest, not the test-test paired comparisons. In Section 2, we show how we analyze these results to focus on these control-test comparisons.

Another type of special structure occurs when there are multiple products, each of which are evaluated continuously over time or at specific time points during consumption by a procedure for dynamic or temporal sensory profiling (Hort, Kemp & Hollowood, 2017). An example of such a data set comes from a study by Nguyen, Næs, and Varela (2018) in which a trained panel evaluated eight yogurt formulations using the temporal check-all-that-apply (TCATA; Castura, Antúnez, Giménez & Ares, 2016a) method. The data set is described in Section 2.2. Conventionally, panel citation proportions are summarized in a matrix with combinations of formulations and timepoints in rows and sensory attributes in columns. PCA is conducted after column-centering the TCATA citation rates matrix (Gonzalez-Estanol et al., 2022; Nguyen & Wismer, 2022; Castura et al., 2022; Berget et al., 2020; Sharma & Duizer, 2019; Poveromo & Hopfer, 2019; Schumaker et al., 2019; Castura, 2018; Esmirino et al., 2017; Reyes, Castura & Hayes, 2017; McMahon et al., 2017; Castura, Baker & Ross, 2016b). After conducting this analysis, Castura et al. (2016b) reported that their PC1 analysis extracted nearly 85% of the total variance and mainly contrasted zero or near-zero citation rates at the start and end of the evaluation with the peak citation rates in the early- to mid-evaluation periods. Consequently, they focused their interpretations mainly on PC2 and PC3, which extracted a far smaller proportion of the total variance, but which they found more discriminating and relevant. This result is typical because most of the variability in a temporal sensory results matrix tends to exist *across* rather than *within* timepoints, which is why the direction of maximum variability (PC1) extracts mostly variability across timepoints. Variability within timepoints, which can be often of greater interest, tends to be extracted in subsequent PCs and usually accounts for only a small proportion of the total variance. In the current manuscript, we will show how to use the special structure of temporal sensory data sets to investigate the relevant within-timepoint differences.

After describing these two data sets (Section 2), we provide background on PCA (Section 3.1). We discuss how all paired comparisons can be investigated after PCA (Section 3.2.1), then give our proposal for investigating only relevant paired comparisons (Section 3.2.2). Next, we describe methods for constructing results sets with only relevant paired comparisons (Section 3.3) and for analyzing these results (Section 3.4). For each of the example data sets, we present the conventional analysis based on all paired comparisons and the approach for investigating only selected paired comparisons (Section 4). Discussion and conclusions follow.

2. Materials & Methods

2.1. Smoothie data set

A trained panel conducted a quantitative descriptive analysis of 10 berry-banana smoothies as part of a study reported by Galler et al. (2020). Smoothie formulations are given in Table 1. Each formulation was evaluated in duplicate by 9 trained assessors on 18 sensory attributes using 10-cm continuous line scales. Results from univariate two-way analyses of variance with factors smoothie formulation and assessor showed that the panel discriminated the smoothie formulations with 95% confidence on all but two attributes (*sour odour* and *metallic taste*), which were dropped so that they did not influence the PCA solutions (Næs, Tomic, Endrizzi & Varela, 2021). The 16 retained attributes [attribute abbreviations used when plotting results] included odour attributes (*odour intensity* [i], *fruit/berry odour* [b], *artificial odour* [r]), appearance attributes (*colour strength* [c], *whiteness* [w]), and taste, flavour and mouthfeel attributes (*taste intensity* [I], *acidity* [A], *sweetness* [E], *sourness* [S], *bitterness* [T], *fruit/berry flavour* [B], *artificial flavour* [R], *fullness* [F], *viscosity* [V], *astringency* [Y], and *pungency* [P]). See Galler et al. (2020) for further details on this study.

<<TABLE 1 APPROXIMATELY HERE>>

Table 1. Smoothie formulations from the study by Galler et al. (2020). The Control formulation was a raspberry-strawberry-blueberry-banana smoothie. Test products were formulated by adding one or more ingredients, where + indicates the addition of an ingredient and ++ indicates the addition of extra quantities of the ingredient.

Code	Formulation	Xanthan gum	Beetroot powder	Lemon juice	Expected sensory change relative to Control
C	Control				-
T1	Test 1	+			thicker
T2	Test 2		+		redder
T3	Test 3	+	+		thicker, redder
T4	Test 4			+	more sour
T5	Test 5	+		+	thicker, more sour
T6	Test 6	+		+	redder, more sour
T7	Test 7	+	+	+	thicker, redder, more sour
T8	Test 8		++	+	much redder, more sour
T9	Test 9		++		much redder

2.2. Yogurt data set

Nguyen et al. (2018) describe a study in which yogurts were formulated with the same ingredients but processed differently to deliver different textural properties. Formulations were obtained from a 2^3

factorial design with factors viscosity (levels: thick, thin), particle size (levels: flakes, flour) and flavour intensity (levels: optimal, low) (Table 2). As part of a larger study, each formulation was then evaluated in triplicate by eight trained panelists using the TCATA method (Castura et al., 2016a) on 10 taste, flavour and mouthfeel attributes: *acidic* [A], *bitter* [B], *cloying* [C], *dry* [D], *gritty* [G], *sandy* [S], *sweet* [W], *thick* [K], *thin* [N], and *vanilla* [V]. [Abbreviations will be used when plotting results.]

Since there were excessive delays before the first attribute was selected in some evaluations, which suggested that some assessors pressed Start and only later put the sample into the mouth to be evaluated, we left-trimmed each evaluation to begin when the first attribute was selected, as advocated by Castura (2020). Time units were kept on the original scale (seconds), not standardized, to avoid data warping (see Castura, 2020, 2018). Analysis focused on the results recorded at 1-s increments between 0 s and 55 s. This data set has also been analyzed by Nguyen and Varela (2021), Nguyen et al. (2020a), Meyners (2020), Berget et al. (2020), and Castura (2020). These eight yogurt formulations have also been investigated in other sensory tests (Asioli, Nguyen, Varela, & Næs, 2022; Nguyen, Næs, Almøy, & Varela, 2020b). Readers are referred Nguyen et al. (2018) for further details on the yogurt formulations and data collection methods.

<<TABLE 2 APPROXIMATELY HERE>>

Table 2. Yogurt formulations from the study by Nguyen et al. (2018).

Code	Viscosity	Particle size	Flavour Intensity
tFl	thin	flakes	low
TFl	thick	flakes	low
Tfl	thin	flour	low
Tfl	thick	flour	low
tFo	thin	flakes	optimal
TFo	thick	flakes	optimal
Tfo	thin	flour	optimal
Tfo	thick	flour	optimal

3. Theory and calculations

3.1. Statistical methods: PCA, uncertainty, and making paired comparisons

In this section, we provide details of new and existing methods for investigating paired comparisons in PCA results. In Section 3.1, we give an overview of PCA. In Section 3.2, we discuss the goal of finding an optimal space for investigating variance in only the relevant paired differences after conducting PCA. We propose a new approach for finding an optimal space for investigating the variance in selected paired comparisons. The approach is applied to two types of data sets with special structures (Section 3.3). In Section 3.4, we describe how we will investigate uncertainty and whether paired differences are discriminated.

3.1. Overview of principal component analysis (PCA)

Submitting a column-centered (J×M) matrix **X** with rank R to singular value decomposition (SVD; Mardia, Bibby & Kent, 1979) yields

$$\mathbf{X} = \mathbf{UDP}^{\mathrm{T}} \tag{1}$$

where columns of the (J×R) matrix **U** are left singular vectors, diagonal elements of the (R×R) diagonal matrix **D** are singular values, and columns of the (M×R) matrix **P** are right singular vectors. Singular vectors are orthonormal, so $\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}_R$ and $\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_R$. Standardizing the columns in **X** before SVD allows variables that are collected on different scales to participate equally in the analysis.

The sum of squared singular values, *trace*($\mathbf{D}^2$) equals the sum of squared elements of **X**, which is sometimes called the total inertia (Abdi & Williams, 2010). Dividing the squared singular value for component a by the sum of squares of **X** gives the proportion of the total sum of squares that is extracted by PC a; expressed as a percentage, it is equivalent to the percentage of variance accounted for (%VAF) by PC a.

SVD is related to the eigendecomposition (Mardia et al., 1979). Eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{XX}^{\mathrm{T}}$ are identical to **P** and **U**, respectively. The eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ are identical to diagonal elements of $\mathbf{D}^2$ (Mardia et al., 1979).

SVD is also related to PCA, which reduces (1) to two matrices. In PCA of sensory evaluation results, it is conventional to multiply **U** and **D** to obtain the score matrix **T**, where

$$\mathbf{X} = \mathbf{TP}^{\mathrm{T}} \tag{2}$$

The first PC extracts variance maximally from **X**. Each subsequent PC extracts variance maximally from the residual matrix. Dimension reduction to A PCs reduces (2) to

$$\mathbf{X} = \mathbf{T}_A\mathbf{P}_A^{\mathrm{T}} + \mathbf{E} \tag{3}$$

where the (J×M) matrix $\mathbf{T}_A\mathbf{P}_A$ contains most of the variance and is considered to be "signal", and the (J×M) residual matrix **E** is considered to be "noise". Interpretation focuses on the truncated (J×A) score matrix $\mathbf{T}_A$ and the truncated (M×A) loading matrix $\mathbf{P}_A$.

3.2. Investigated paired comparisons in PCA

3.2.1. Investigating all paired comparisons

A row vector $\mathbf{t}_1$ in $\mathbf{T}_A$ can be obtained by multiplying the row vector $\mathbf{x}_1$ in **X** by the loadings. A paired difference between $\mathbf{t}_1$ and $\mathbf{t}_2$ in $\mathbf{T}_A$ can also be obtained by multiplying the difference between the row vectors by the loadings, since

$$(\mathbf{t}_1 - \mathbf{t}_2)^{\mathrm{T}} = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}\mathbf{P}_A \tag{4}$$

The covariance matrix of **X** and the covariance matrix of its "crossdiff-unfolded" matrix **X**⊖**X**, which is the (J^2xM) matrix of all paired differences that is obtained by subtracting every row in **X** from each row in **X**, differ only by a scalar that depends on the number of rows in **X**, not on the data (Castura et al., 2023b). Consequently, PCA of **X**⊖**X** yields

$$\mathbf{X} \ominus \mathbf{X} = (\mathbf{T} \ominus \mathbf{T})\mathbf{P}^{\mathrm{T}} \quad (5)$$

where $\mathbf{X}\ominus\mathbf{X}$ and $\mathbf{T}\ominus\mathbf{T}$ are the crossdiff-unfolded **X** and **T** matrices from (2) and the loading matrix **P** in (2) and in (5) are identical. In other words, the columns of **P** span the optimal space for investigating variance in both the original row objects and in all paired comparisons of row objects. Truncating (5) to *A* PCs yields

$$\mathbf{X} \ominus \mathbf{X} = (\mathbf{T}_A \ominus \mathbf{T}_A)\mathbf{P}_A^{\mathrm{T}} + (\mathbf{E} \ominus \mathbf{E}) \quad (6)$$

which can be obtained from (3) directly (Castura et al., 2023b). These findings were our starting point for investigating only a relevant subset of paired comparisons.

3.2.2. Investigating a subset of paired comparisons

As just discussed, **P** is the optimal space for investigating variance in the original row objects and in all paired comparisons. Appendix A.1 shows that **P** does not extract variance maximally from a subset of paired comparisons.

To find the optimal space for exploring variance in only *C* relevant paired comparisons, we construct a ($2C$xM) matrix $\mathbf{\Delta}^*$. Each of the *C* paired comparisons is represented by its twinned paired differences; e.g., the paired comparison $\mathbf{x}_1$ vs $\mathbf{x}_2$ is represented by its twinned paired differences $\mathbf{x}_1$-$\mathbf{x}_2$ and $\mathbf{x}_2$-$\mathbf{x}_1$, such that the analysis applies equally to both paired differences. The ($2C\times M$) matrix $\mathbf{\Delta}^*$ is identical to the ($2C\times M$) submatrix of relevant rows in $\mathbf{X}\ominus\mathbf{X}$.

This matrix of relevant paired comparisons ($\mathbf{\Delta}^*$) is new. Its construction is determined by which paired comparisons are of interest to the researcher. PCA of $\mathbf{\Delta}^*$ yields

$$\mathbf{\Delta}^* = \mathbf{T}^*(\mathbf{P}^*)^{\mathrm{T}} \quad (7)$$

Dimension reduction to *A* PCs yields

$$\mathbf{\Delta}^* = \mathbf{T}_A^*(\mathbf{P}_A^*)^{\mathrm{T}} + \mathbf{E}^* \quad (8)$$

The ($J^2\times M$) matrix $\mathbf{X}\ominus\mathbf{X}$ was defined previously as containing the paired differences between rows in the ($J\times M$) matrix **X** (Castura et al., 2023b). Now, we define the matrix $\mathbf{\Delta}^*$ to include only the relevant paired comparisons. So, if all paired comparisons are relevant, then $\mathbf{\Delta}^*$ contains only (J^2-J) rows because this matrix excludes the *J* rows of only zeros that occur when a row in **X** is subtracted from itself. Although $\mathbf{\Delta}^*$ and $\mathbf{X}\ominus\mathbf{X}$ are not identical when all paired comparisons are relevant, these matrices are related by an important property: their respective covariance matrices differ only by the scalar (J^2-J-1)/(J^2-1), which occurs only the covariances in $\mathbf{X}\ominus\mathbf{X}$ are each calculated based on having *J* more zeros than covariances in $\mathbf{\Delta}^*$. For this reason, PCA of $\mathbf{X}\ominus\mathbf{X}$ and PCA of $\mathbf{\Delta}^*$ each yield the same loading matrix, **P**. As noted previously, this loading matrix is also identical to **P** from PCA of **X** (Castura et al., 2023b), so when all paired comparisons are of interest, $\mathbf{P}^*$ in (7) is identical to **P** in (1), (2), and (5). In this case, the corresponding rows of $\mathbf{T}\ominus\mathbf{T}$ in (6) and $\mathbf{T}^*$ in (8), as is evident from (4).

The paired comparison $\mathbf{x}_1$ vs $\mathbf{x}_2$ is represented in $\mathbf{\Delta}^*$ by the twinned paired differences $\mathbf{x}_1$-$\mathbf{x}_2$ and $\mathbf{x}_2$-$\mathbf{x}_1$. Had a researcher represented this and other paired comparisons by choosing only one of the twinned

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).

paired differences, then different researchers might obtain different matrices of paired differences, which would yield different results. Such idiosyncratic solutions are avoided by always investigating C paired comparisons using all $2C$ twinned paired differences. Since columns of $\mathbf{X}$ are centered, the twinned paired differences always sum to zero in every attribute so the columns of $\mathbf{\Delta}^*$ center naturally. After PCA of $\mathbf{\Delta}^*$, interpretation of any paired difference will mirror the interpretation of its twinned paired difference exactly. These advantages—centered columns, identical interpretations of the twinned paired differences—hold whether $\mathbf{\Delta}^*$ contains twinned paired differences of all or only a subset of the paired comparisons.

PCA extracts variance maximally from $\mathbf{\Delta}^*$ whether it contains all or only a relevant subset of the paired comparisons. Since the truncated loading matrices $\mathbf{P}_A^*$ in (8) and $\mathbf{P}_A$ from (6) are nearly always different, we must consider which of these PCA solutions is superior.

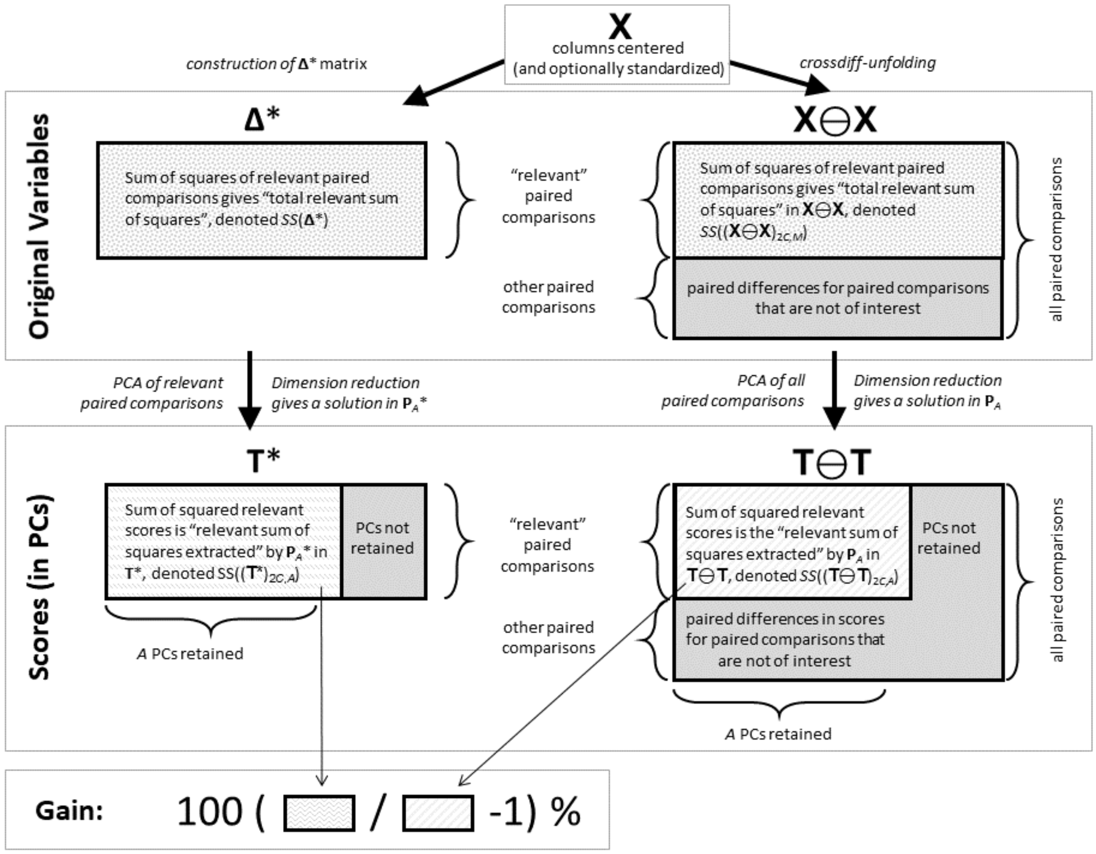
3.2.3. Gain from PCA focusing on relevant paired comparisons

Now, we quantify the benefit of investigating the relevant paired comparisons in the directions of $\mathbf{P}_A^*$ instead of the directions of $\mathbf{P}_A$ from PCA based on all paired comparisons. We cannot use %VAF to compare PCA of matrices with different dimensions. As we show in Suppl. Fig. S1 (eComponent), PCA tends to extract a larger proportion of variance from a matrix with fewer rows than from a matrix with more rows. Since $\mathbf{\Delta}^*$ has fewer rows than $\mathbf{X}\ominus\mathbf{X}$, we must compare their PCA solutions using a method other than %VAF.

Instead, we compare these PCA solutions by calculating the relevant sum of squares extracted (see inertia; Section 3.1). The first A PCs of $\mathbf{\Delta}^*$ extract the largest possible sum of squares from $\mathbf{\Delta}^*$. The $\mathbf{X}\ominus\mathbf{X}$ matrix contains the $2C$ rows that are also in $\mathbf{\Delta}^*$, plus rows for other paired comparisons that are not considered to be relevant. The first A PCs of $\mathbf{X}\ominus\mathbf{X}$ extract the sum of squares maximally from $\mathbf{X}\ominus\mathbf{X}$. But only $2C$ rows in the score matrix $\mathbf{T}\ominus\mathbf{T}$ are relevant; other rows pertain to other paired comparisons that are not relevant. For this reason, we use the procedure shown diagrammatically in Fig. 1 to obtain an index that compares the relevant sum of squares from the two PCA solutions.

<<FIG. 1 APPROXIMATELY HERE>>

Fig. 1. Sum-of-squares calculations used to quantify the benefit (Gain) of investigating the relevant paired comparisons using $\mathbf{P}_A$* *instead of all paired comparison using* $\mathbf{P}_A$.



The benefit of investigating the relevant paired comparisons using $\mathbf{P}_A$* from (8) instead of all paired comparison using $\mathbf{P}_A$ from (6) can be quantified as a percentage:

$$\text{Gain} = 100\left(\frac{SS((\mathrm{T}*)_{2C,A})}{SS((\mathrm{T}\ominus\mathrm{T})_{2C,A})} - 1\right)\% \quad (9)$$

where $SS((\mathbf{T}^*)_{2C,A})$ and $SS((\mathbf{T}\ominus\mathbf{T})_{2C,A})$ are the *relevant sum of squares extracted* in the first A PCs of **Δ*** and the first A PCs of **X⊖X**, respectively. A Gain that is larger indicates a greater benefit from focusing on $\mathbf{P}_A$* instead of $\mathbf{P}_A$. Gain cannot be negative. If all paired comparisons are relevant, then Gain is zero. If only a subset of paired comparisons are relevant, then Gain is nearly always positive for a truncated solution. Although we calculate Gain from sum-of-squares calculations, Gain would be identical if it is calculated based on %VAF in only the relevant 2*C* rows in the respective matrices considered here.

3.2.4. Considerations related to standardizing variables

When variables in **X** are not directly comparable, then its columns are often standardized to mean zero and unit variance. In this case, columns in **X⊖X** also have a constant variance (Appendix A.1), so the sum of squares of columns in **X** and the sum of squares of columns in **X⊖X** are related by a scalar (Castura et al., 2023b). Since PCA of **X** and PCA of **X⊖X** both treat the variables as having equal weight, it is unnecessary to do a new standardization of columns of $\mathbf{X}\ominus\mathbf{X}$. The same is true if all paired comparisons are relevant because the sum of squares of columns in $\mathbf{\Delta}^*$ and the sum of squares of columns of $\mathbf{X}\ominus\mathbf{X}$ are identical. In this case, PCA of $\mathbf{\Delta}^*$ and PCA of $\mathbf{X}\ominus\mathbf{X}$ are equivalent.

However, if only a subset of paired comparisons is relevant, then the sum of squares in different columns of $\mathbf{\Delta}^*$ are not all identical; they depend on which paired comparisons are relevant, i.e. they depend on the $2C$ rows. In this case, the sum of squares is different for each variable. So, $\mathbf{X}$ with variance-standardized columns produces a $\mathbf{X}\ominus\mathbf{X}$ matrix with equal-variance columns, but a $\mathbf{\Delta}^*$ matrix with columns that have unequal variances. PCA of $\mathbf{\Delta}^*$ weights the variables unequally. Although it would be possible to variance-standardize the columns of $\mathbf{\Delta}^*$ before PCA, we do not do so. One reason is that the data in $\mathbf{\Delta}^*$ would no longer match the data in the relevant $2C$ rows in $\mathbf{X}\ominus\mathbf{X}$. Later, in Section 5.2, we will discuss the possibility of doing a new column standardization of $\mathbf{\Delta}^*$.

In this paper, we variance-standardize the columns of $\mathbf{X}$ to put the variables from the smoothie data set (Section 2.1.1) on an equal footing. Then we proceed with PCA of $\mathbf{\Delta}^*$ without first doing a new standardization of its columns.

3.3. Data sets with special structures

In the subsections that follow, we give two examples of data sets with special structures. For each example, we discuss how we obtain the matrix $\mathbf{\Delta}^*$ and how we obtain the optimal space $\mathbf{P}^*$ for investigating the relevant paired comparisons.

3.3.1. Data set with control and test products

In this example, the column-centered ($J\times M$) matrix $\mathbf{X}$ contains attribute intensity means from the sensory panel for one control formulation in the first row and J-1 test formulations in subsequent rows. For scale data, we begin by putting all attributes on an equal footing by standardizing the columns to unit variance (Næs et al., 2021). What is of primary interest are the comparisons between row 1 and each of the other rows, i.e., C=J-1 relevant paired comparisons. This control-test special structure occurs in the smoothie data set (Section 2.1).

Since each paired comparison is investigated via both of its twinned paired differences (Section 3.2), the ($2C$xM) matrix $\mathbf{\Delta}^*$ contains J-1 control-test comparisons ($\mathbf{x}_1$-$\mathbf{x}_2$, $\mathbf{x}_1$-$\mathbf{x}_3$, …, $\mathbf{x}_1$-$\mathbf{x}_J$) and J-1 control-test comparisons ($\mathbf{x}_2$-$\mathbf{x}_1$, $\mathbf{x}_3$-$\mathbf{x}_1$, …, $\mathbf{x}_J$-$\mathbf{x}_1$). The columns of these 2(J-1) rows sum naturally to zero. These relevant paired comparisons are explored in $\mathbf{P}^*$, which is obtained from PCA of $\mathbf{\Delta}^*$ in (7). Interpretation then focuses exclusively on control-test paired comparisons, not on test-test paired comparisons.

3.3.2. Data set with temporal sensory results

In this example, the column-centered ($J\times M$) matrix $\mathbf{X}$ has sensory panel results (e.g. citation rates) on M attributes for F formulations across S timepoints. There are where J=FS rows; each row is associated with both a formulation and a timepoint. This special structure is found in the yogurt data set (Section 2.2).

One approach, which we do not use, is to conduct one PCA per timepoint, i.e., conducting PCA of a column-centered submatrix for each timepoint of $\mathbf{X}$. If this were done, then the PCs would be defined by a different loading matrix at each timepoint. Consequently, citation rates that are identical but occur at different times could have different scores. Scores would need to be interpreted in ever-changing coordinates. For simplicity and interpretability, we prefer to explore paired comparisons of formulations within each timepoint in PCs that are consistent across timepoints.

Conventionally, this space is derived from a PCA of the column-centered citation rates for all formulations and times, as described in Section 1 and references therein. This space is optimal for investigating all paired comparisons, both within and across timepoints (Castura et al., 2023b). However, for temporal sensory data sets, paired comparisons across timepoints are often of lesser interest than the $F(F\text{-}1)/2$ unique paired comparisons of formulations within each of the S timepoints. If we focus only on paired comparisons within each timepoint then, by multiplication, there are $C=FS(F\text{-}1)/2$ relevant paired comparisons in total.

For example, if F=2 and S=2, then the matrix $\mathbf{\Delta}^*$ would have 4 rows: $\mathbf{x}_{f1,t1}$-$\mathbf{x}_{f2,t1}$, $\mathbf{x}_{f2,t1}$-$\mathbf{x}_{f1,t1}$, $\mathbf{x}_{f1,t2}$- $\mathbf{x}_{f2,t2}$, and $\mathbf{x}_{f2,t2}$- $\mathbf{x}_{f1,t2}$. Since each paired difference has a twin which has the minuend and subtrahend reversed, the entries for each pair sum to zero and columns in $\mathbf{\Delta}^*$ are centered naturally. The matrix $\mathbf{\Delta}^*$ contains relevant paired comparisons only. For the yogurt data set (Section 2.2), F=8 and S=56. There are 28 unique paired comparisons at each of the 56 timepoints and C=1568 paired comparisons in total. We investigate these C paired comparisons via the ($2C\times M$) matrix $\mathbf{\Delta}^*$, which contains $FS(F\text{-}1)$=3136 rows, the columns of which center naturally.

PCA of $\mathbf{\Delta}^*$ by (7) finds $\mathbf{P}^*$ which extracts the variance maximally from only these relevant paired comparisons in successive PCs. Interpretation of results can focus on comparing relevant pairs of formulations within each timepoint, from which variance is extracted maximally. Variance is not extracted maximally from the paired comparisons across timepoints, which are not the focus of interpretation.

3.4. Investigating uncertainty and paired differences

This section summarizes some existing methods for investigating paired comparisons after PCA; for further details, refer to Castura et al (2023a; 2023b). We construct confidence ellipsoids and obtain P values as described by Castura et al. (2023a), which we describe here for completeness.

3.4.1. The truncated total bootstrap (TTB) procedure

In the truncated total bootstrap (TTB; Castura et al., 2023b; Castura et al., 2022; Cadoret & Husson, 2013; Courcoux et al., 2012) method, a large number of virtual panels are composed using the real panel's results. Each virtual panel's raw data set is aggregated and analyzed in exactly the same manner as the real panel's data set. TTB-derived scores are obtained by using Procrustes rotation (Schönemann, 1966) to superimpose each virtual panel's truncated score matrix onto the real panel's truncated score matrix. Procrustes rotation is conducted without isotropic scaling to retain this source of variability (Castura et al., 2022). In this paper, we investigate the uncertainty of the paired comparisons in $\mathbf{X}\ominus\mathbf{X}$ and in $\mathbf{\Delta}^*$.

3.4.2. Constructing 95% confidence ellipsoids

If points are multinormally distributed in A dimensions, then its 100(1-α)% confidence ellipsoid is

$$\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} \leq \chi^2_{1-\alpha,A} \tag{10}$$

where $\mathbf{d}$ is an A-length vector of differences between cloud points and the cloud center, $\mathbf{S}$ is the covariance matrix for the cloud of points, and $\chi^2_{1-\alpha,A}$ is the (1-α)$^{\mathrm{th}}$ quantile of the χ^2 distribution with A degrees of freedom. The left-hand side of (10) is a squared Mahalanobis distance (Mardia et al., 1979; Mahalanobis, 1936) and the right-hand side is the critical value from the theoretical null distribution.

Since TTB-derived clouds can be asymmetric with many points near the mode but with long tails (Castura et al., 2023b), a 95% confidence ellipsoid obtained from (10) may enclose less than 95% of the cloud points. For each cloud, Castura et al. (2023a) calculate the squared Mahalanobis distance between all cloud points and the cloud center to obtain its empirical distribution Q. The 95% confidence ellipsoid

$$\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} \leq Q_{1-\alpha} \tag{11}$$

differs from (10) only in that the right-hand side is the (1-α)$^{\mathrm{th}}$ quantile of Q, i.e., $Q_{1-\alpha}$ is the critical value from the empirical, not theoretical, distribution. Since we use α=0.05, the ellipsoid formed by (11) contains precisely 95% of the cloud points.

We will use (11) to obtain confidence ellipsoids for paired comparisons based on the TTB results in $\mathbf{X}\ominus\mathbf{X}$ and in $\mathbf{\Delta}$*. These ellipsoids are used to visualize the uncertainty of paired comparisons in the PCs defined by $\mathbf{P}_A$ and $\mathbf{P}_A$* from (6) and (8), respectively.

3.4.3. Obtaining P values

Next, we make use of the TTB results to evaluate whether each of the relevant paired comparisons is significant. For each paired comparison, we will obtain the distribution Q (Section 3.4.2) based on the cloud of TTB-derived paired difference scores, then calculate the probability

$$\mathrm{P} = \Pr(\mathbf{d}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{d} \geq Q|\mathrm{H}_0) \tag{12}$$

where $\mathbf{d}$ is the squared Mahalanobis distance between the cloud center and the origin. A very small P value indicates that the squared Mahalanobis distance between the cloud center and the origin is as or more extreme than the cloud points to their centroid. In other words, a small P value indicates that a difference that is improbable to have occurred only by chance.

Although it is almost never the case that two products are truly identical, the screening is done pragmatically to draw attention to pairs having small P values.

We will use (12) to get P values for paired comparisons based on the real-panel results and the TTB results for both $\mathbf{X}\ominus\mathbf{X}$ and in $\mathbf{\Delta}$*. We will use these P values for screening purposes to draw extra attention to paired comparisons that seem to be well discriminated in $\mathbf{P}_A$ and $\mathbf{P}_A$*, respectively, so that systematic differences that are relatively large in comparison to the natural variation will not go unnoticed.

3.5. Statistical software

We conducted the analyses described above in Section 3 for the two data sets in two ways: first, based on the conventional analysis (PCA of $\mathbf{X}\ominus\mathbf{X}$, identical to all paired comparisons in PCA of $\mathbf{X}$), which investigates all paired comparisons (Section 3.2.1), then, second, taking into account the special structure of the data set (PCA of $\mathbf{\Delta}^*$; Section 3.2.2). In each case, the TTB procedure was conducted with B=15,000 virtual panels. All analyses were conducted in R version 4.2.2 (R Core Team, 2022).

4. Results

4.1. Smoothie results

4.1.1. PCA conducted conventionally based on all paired comparisons

From the raw smoothie data, we obtained a products-by-attributes matrix of real-panel means. Its columns were variance-standardized to obtain $\mathbf{X}$, which was crossdiff-unfolded to obtain $\mathbf{X}\ominus\mathbf{X}$. which submitted to PCA as in (5). The first four PCs extract 59.9%, 19.5%, 15.1%, and 2.6% of the variance from $\mathbf{X}$. We chose a three-component solution which extracts 97.0% of the total variance.

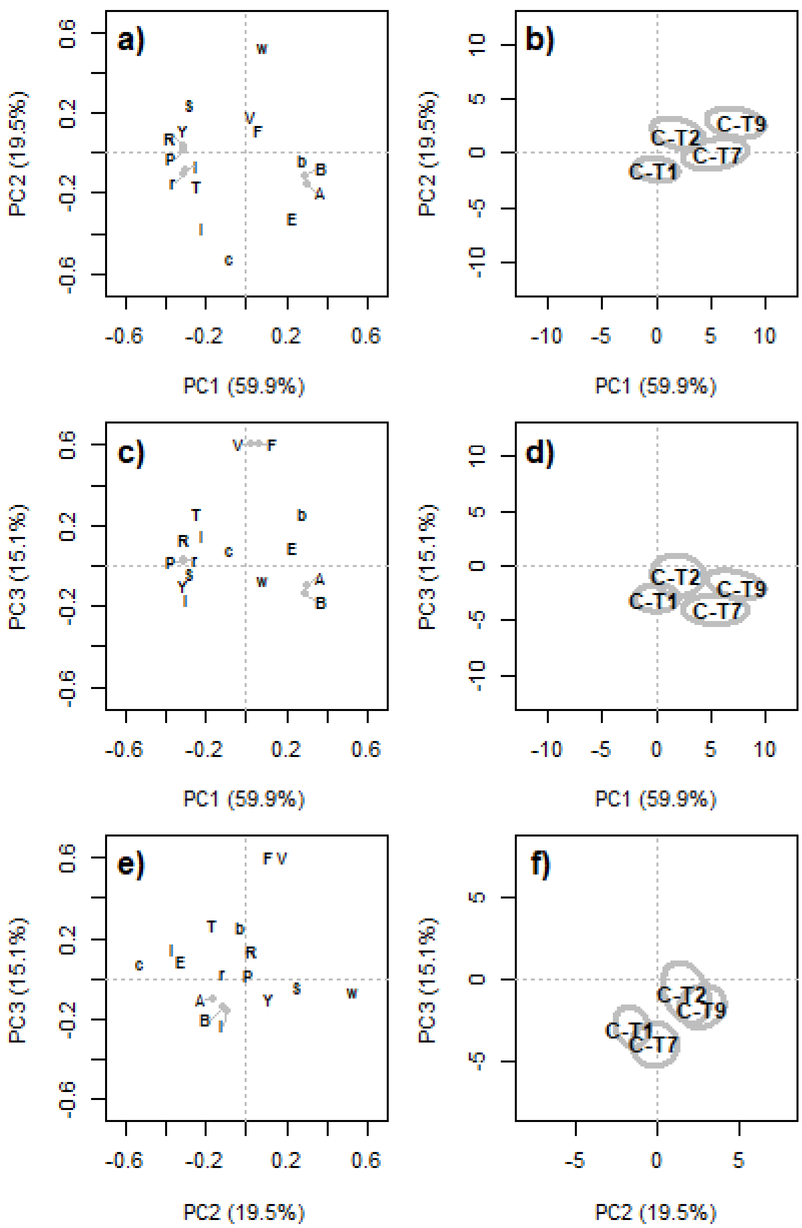
Loading plots (Fig. 2, panels a, c, e) show that PC1 contrasts artificial, bitter, pungent, astringent, and artificial sensations with fruity, sweet, and acidic taste sensations. PC2 contrasts a white colour vs colour strength. PC3 contrasts low vs high viscosity and fullness. The control smoothie was associated with fruit and sweetness, a whiter colour, and lower viscosity and fullness sensations, which were expected based on its formulation (Table 1). Test smoothies tended to have off-flavours related to their formulations. Beetroot powder was added to Test smoothies 2, 3, and 6 through 9, which were perceived to have a more intense colour, on average, than the smoothies without this ingredient.

The real-panel PCA results were typical of the virtual-panel PCA results in terms of %VAF (Suppl. Table S1, eComponent), so results from these virtual panels were used to obtain TTB-derived scores from which we obtained 95% confidence ellipsoids for the smoothies and all paired comparisons. Confidence ellipsoids for the paired comparisons are shown in the space obtained from PCA based on all paired comparisons (top row of Suppl. Fig. S2). The control-test smoothie pairs tended to be well discriminated in the plane of PC1 vs PC3. But the confidence ellipsoids for some pairs overlap zero in PC2. Differences between test-test pairs were more pronounced in PC2 than control-test pairs.

<<FIG 2 APPROXIMATELY HERE>>

Fig. 2. Plots from PCA of all paired comparisons of smoothie formulations. Results are visualized via loading plots (left column) and paired difference score plots showing only four of the nine relevant paired comparisons (right column) in the planes of PC1 vs PC2 (top row; a and b), PC1 vs PC3 (middle row; c and d), and PC2 vs PC3 (bottom row; e and f; note that axes have a different scale in f). Attributes: odour intensity [i], *fruit/berry odour* [b], *artificial odour* [r], *colour strength* [c], *whiteness* [w], *taste intensity* [I],

417 *acidity* [A], *sweetness* [E], *sourness* [S], *bitterness* [T], *fruit/berry flavour* [B], *artificial flavour* [R], *fullness*
418 [F], *viscosity* [V], *astringency* [Y], and *pungency* [P]. *(See Table 1 for smoothie formulation details.)*



419

420

421 PCA of **X**⊖**X** as in (5) extracts 70.1%, 8.6%, and 17.6% of the relevant sum of squares in one, two, and
422 three PCs. The plane of PC1 vs PC3 extracts 87.7% of the relevant sum of squares, which is more than
423 the 78.7% extracted in the PC1 vs PC2 plane. Although PC2 extracts more sum of squares than PC3, a
424 larger proportion of the sum of squares extracted in PC2 are related to test-test paired comparisons,

whereas a larger proportion of the sum of squares extracted by PC3 are related to control-test paired comparisons. This explains why the control-test smoothie pairs are better discriminated in PCA based on all paired comparisons in the plane of PC1 vs PC3 (top row of Suppl. Fig. S2).

P values, which were used to evaluate the separation of test smoothies from the control smoothie, indicate that each of the test formulations were discriminated from the control formulation ($P<0.01$). Table 3 shows P values based on this analysis; P values based on PCA accounting for the special data structure, which will be discussed later in Section 4.1.2, are also presented to facilitate comparison.

<<TABLE 3 APPROXIMATELY HERE>>

Table 3. P values for evaluating test smoothie formulations are discriminated from the control smoothie formulation (vs control) derived from the TTB procedure after PCA based on all paired comparisons (*PCA of* $\mathbf{X}\ominus\mathbf{X}$) *and PCA accounting for the special data structure* (PCA of $\mathbf{\Delta}^*$). *(See Table 1 for smoothie formulation details.)*

	PCA of $\mathbf{X}\ominus\mathbf{X}$	PCA of $\mathbf{\Delta}^*$
Test 1	0.0007	<0.0001
Test 2	0.0025	<0.0001
Test 3	0.0014	<0.0001
Test 4	0.0011	0.0001
Test 5	0.0003	<0.0001
Test 6	0.0001	<0.0001
Test 7	0.0005	<0.0001
Test 8	0.0026	0.0003
Test 9	0.0001	<0.0001

4.1.2. PCA accounting for the special data structure based on relevant paired comparisons

Starting from the column-standardized matrix $\mathbf{X}$, we obtained the matrix $\mathbf{\Delta}^*$ with $2C=18$ rows corresponding to the $C=9$ unique control-test paired comparisons, as described in Section 3.3.1. PCA of $\mathbf{\Delta}^*$ as in (7) extracts 80.8%, 9.3%, 7.5%, and 1.1% of the variance in the first four PCs. We would probably find a two-component solution sufficient here, but to allow for a direct comparison with the results just presented, we truncated the solution to $A=3$ PCs, which extract 97.5% of the variance in $\mathbf{\Delta}^*$. Its truncated loading matrix is denoted $\mathbf{P}_A^*$.

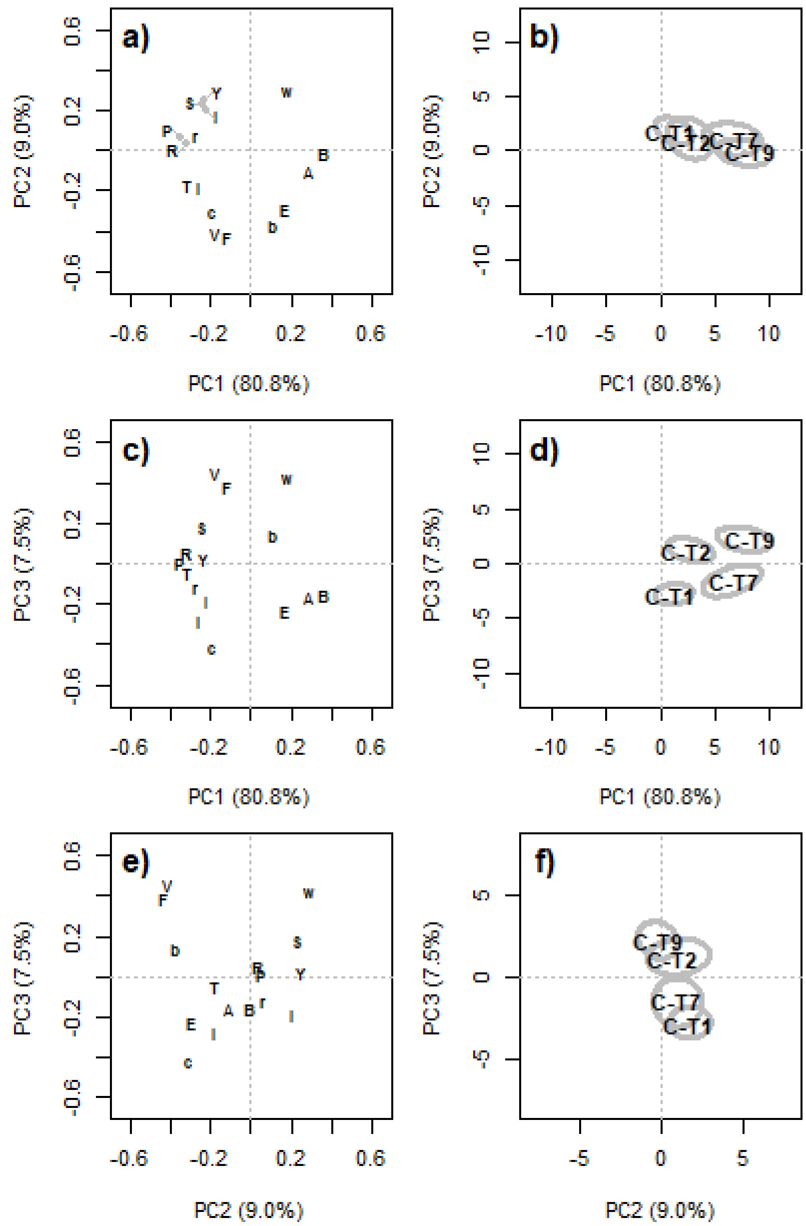
Since $\mathbf{\Delta}^*$ only contains only control-test paired comparisons, PCA of $\mathbf{\Delta}^*$ extracts the largest possible proportion of the relevant sum of squares. The percentage of the relevant sum of squares that is extracted in the PCA of $\mathbf{\Delta}^*$ matches the %VAF, so the percentage of the relevant sum of squares extracted is 90.1% in the first two PCs and 97.5% in the first three PCs. Using PCA of $\mathbf{\Delta}^*$ rather than PCA of all paired comparisons delivers a Gain of 15% in one PC, 14% in the first two PCs, but only 1% in the first three PCs. This result quantifies the benefit of using PCA of $\mathbf{\Delta}^*$ over PCA of $\mathbf{X}\ominus\mathbf{X}$.

Loadings from the PCA accounting for the special structure ($\mathbf{P}_A$*; Fig. 3 a, c, e) have some resemblance to the $\mathbf{P}_A$ loading coefficients in PC1, PC3, and PC2, respectively ($\mathbf{P}_A$; Fig. 2 a, c, e).

Virtual panel results, which were analyzed in the same manner as the real-panel results, resemble the real-panel results in terms of %VAF (Suppl. Table S2, eComponent). Using the TTB-derived scores, we obtain 95% confidence ellipsoids. The bottom row of Suppl. Fig. S2 (eComponent) shows each paired comparison and projections of its 95% confidence ellipsoid on the three planes of PCs in the space of $\mathbf{P}_A$*. Paired comparisons are discriminated if they are separated from the origin in at least one plane. All nine test smoothie formulations are discriminated from the control smoothie. Discrimination is especially good in the PC1 vs PC2 plane. Every confidence region excludes the origin in at least one plane of PCs. In Fig. 3, we show the same four control-test pairs that are visualized in Fig. 2. Plots in Fig. 3 and the bottom row of Suppl. Fig. S2 show that the control-test smoothie formulations are discriminated in all three planes of PCs.

<<FIG 3 APPROXIMATELY HERE>>

Fig. 3. Plots from PCA of relevant paired comparisons of smoothie formulations. Results are visualized via loading plots (left column) and paired difference score plots showing only four of the nine relevant paired comparisons (right column) in the planes of PC1 vs PC2 (top row; a and b), PC1 vs PC3 (middle row; c and d), and PC2 vs PC3 (bottom row; e and f; note that axes have a different scale in f) onto which the 95% confidence ellipsoids for the paired difference scores are projected. Attributes: odour intensity [i], *fruit/berry odour* [b], *artificial odour* [r], *colour strength* [c], *whiteness* [w], *taste intensity* [I], *acidity* [A], *sweetness* [E], *sourness* [S], *bitterness* [T], *fruit/berry flavour* [B], *artificial flavour* [R], *fullness* [F], *viscosity* [V], *astringency* [Y], and *pungency* [P]. *(See Table 1 for smoothie formulation details.)*



All test smoothie formulations were discriminated from the control smoothie (Table 3). The P values based on both PCA of **Δ*** are small for all of the relevant paired comparisons (P<0.01). The control smoothie had higher fruity, sweet, and white colour intensities than the test smoothies, for which the intensities of artificial, bitterness, colour, viscosity, and fullness were higher (Suppl. Fig. S2).

The relevant paired comparisons are all well discriminated in both PCA of **Δ*** and PCA of **X**⊖**X**, but PCA of **Δ*** extracts a larger proportion of the sum of squares from the relevant paired comparisons, so there

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).

is some Gain from investigating the relevant paired comparisons in the directions of $\mathbf{P}_A$* instead of $\mathbf{P}_A$. These results show that $\mathbf{P}_A$* is a better space than $\mathbf{P}_A$ for investigating the control-test paired comparisons.

4.2. Yogurt results

4.2.1. PCA conducted conventionally based on all paired comparisons

Yogurt data were processed as described in Section 2.1.2. The real-panel citation rates were obtained with all combinations of yogurts and timepoints in rows. Attributes, in columns, were centered. PCA of **X** and PCA of $\mathbf{X} \ominus \mathbf{X}$ as in (2) and (5) yield identical loading matrices (**P**) and in both cases their first four PCs extract 49.0%, 25.8%, 12.6%, and 8.3% of the variance. We chose to investigate a three-component solution, which in both PCA solutions extracts 87.4% of the variance.

The truncated loading matrix ($\mathbf{P}_A$) is visualized in three loading plots (Fig. 4, panels a, c, e). The attribute loading coefficients in PC1 all share the same sign. PC1 can be considered to be a mean citation rate dimension, in which rates are zero or nearly zero at the beginning and end of the evaluation, and peak in early- to mid-evaluation, which is the same pattern described by Castura et al. (2016b), as was discussed in Section 1. PC2 contrasts *gritty* vs *sandy* textural perceptions. PC3 contrasts perceptions of *thin* vs *thick*.

The virtual-panel PCA results resemble the real-panel results in terms of %VAF (Suppl. Table S3, eComponent), so were used to obtain the TTB-derived results, from which we obtained the 95% confidence ellipsoids for each of the *C*=28 unique paired comparisons of the eight yogurts at each of the 56 timepoints.

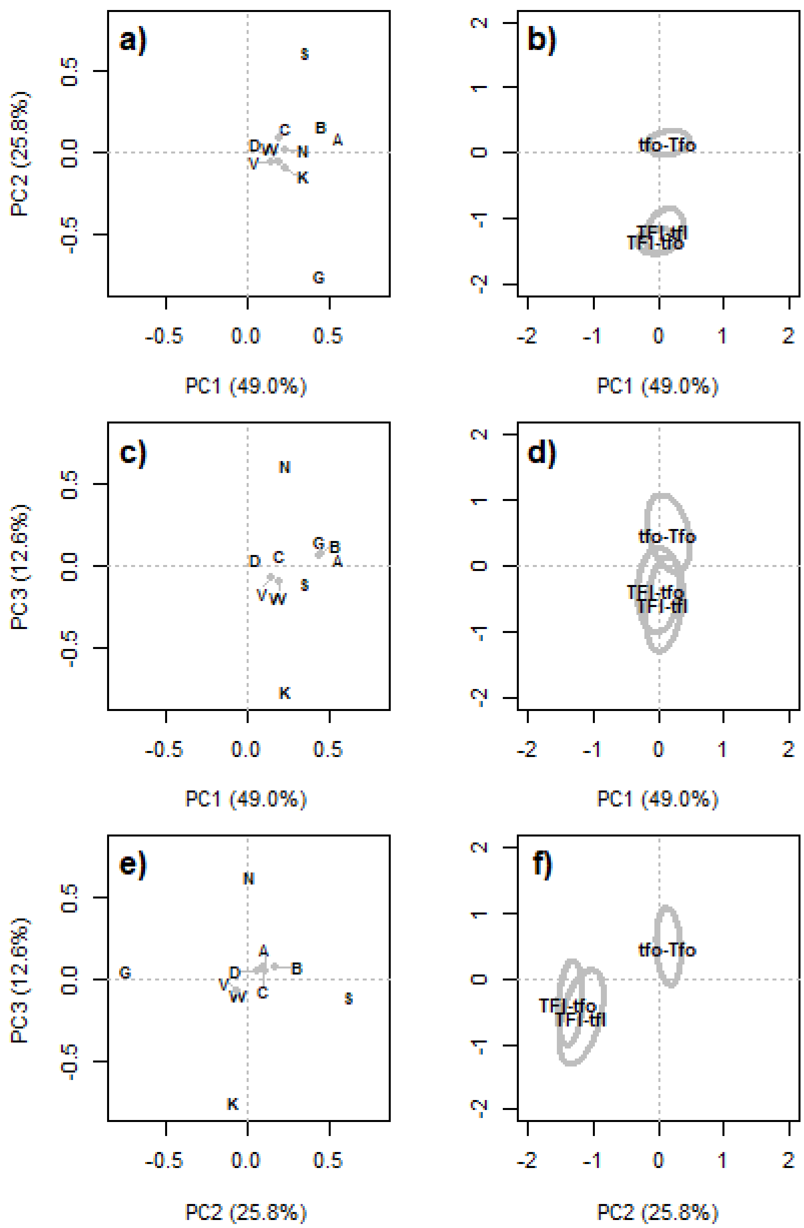
Three of the 28 yogurt pairs are visualized at 10 s after PCA (Fig. 4, b, d, f). We chose to visualize these pairs because they show a range of formulation differences. All three of these pairs differ in viscosity. The TFl vs tfl pair also differs in particles size, whereas the TFl vs tfo pair differs in all three design factors (Table 2). In spite of these differences, these yogurt pairs are not discriminated in PC1, as indicated by the overlap of the origin in PC1 by their ellipsoids. The formulations tfo and Tfo differ only in viscosity; the 95% confidence ellipsoid for this paired comparison is only just visually separated from the origin in the PC2 vs PC3 plane (Fig. 4f). The other two paired comparisons are well discriminated in this plane.

Results for all 28 pairs are shown in biplots in the top row of Suppl. Video S1 (eComponent). The biplots show the cloud of TTB-derived paired difference scores and projections of the 95% confidence ellipsoid onto each plane. Loading vectors shorter than 0.1 in a plane are hidden for improved legibility; loadings are magnified to double size. These plots show that the yogurt pairs are much better discriminated in the PC2 vs PC3 plane early in the evaluation than in planes that include PC1. The panel described yogurts as either *thin* or *thick* according to their viscosity level (Table 2). Yogurts were described relatively often as *gritty* when formulated using flakes and as *sandy* when formulated using flour.

<<FIG 4 APPROXIMATELY HERE>>

518 *Fig. 4. Plots from PCA of all paired comparisons of yogurt formulations at 10 s. Results are visualized via*
519 *loading plots (left column) and paired difference score plots for three of the 28 relevant paired*
520 *comparisons (right column) in the planes of PC1 vs PC2 (top row; a and b), PC1 vs PC3 (middle row; c and*
521 *d), and PC2 vs PC3 (bottom row; e and f) onto which the 95% confidence ellipsoids for the paired*
522 *difference scores are projected. Attributes: acidic* [A], *bitter* [B], *cloying* [C], *dry* [D], *gritty* [G], *sandy* [S],
523 *sweet* [W], *thick* [K], *thin* [N], and *vanilla* [V]. *(Yogurt formulations shown:* thick with flakes and low
524 flavour intensity [TFl]; thin with flour and low flavour intensity [tfl]; thin with flour and optimal flavour
525 intensity [tfo]; thick with flakes and optimal flavour intensity [Tfo].*)*

526



Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).

P values for the 28 paired comparisons at each timepoint were used to screen the results that might go unnoticed if interpretation relied solely on visually inspecting the results in Suppl. Video S1. Table 4 presents only the smallest P value per paired comparison. The yogurt pairs tended to be best discriminated within the first 10 s (Table 4); any yogurt pair that was not discriminated in the first 10 s was not discriminated at all. All yogurts having formulation differences in either viscosity or particle size were successfully discriminated by the panel. Yogurts with formulation differences in both viscosity and particle size were especially well discriminated. But yogurts formulated with different flavour levels were not as well discriminated.

<<TABLE 4 APPROXIMATELY HERE>>

Table 4. Yogurt results were investigated based on PCA of all paired comparisons (*PCA of* **X**⊖**X**) and *PCA accounting for the special data structure (PCA of* **Δ****). P values from each solution were obtained to investigate whether the yogurt formulations were discriminated at each time point, where the time 0 s coincided with the initial response in each evaluation. P values for the relevant paired comparisons (within each timepoint) are shown at times when each pair of formulations was best discriminated. (See Table 2 for details on the yogurt formulations.)*

	PCA of **X**⊖**X**		PCA of **Δ***	
	P value	Time (s)	P value	Time (s)
tFl-TFl	0.0007	2	<0.0001	1
tFl-tfl	<0.0001	2	<0.0001	2
tFl-Tfl	<0.0001	2	<0.0001	0
tFl-tFo	0.1549	7	0.1314	7
tFl-TFo	<0.0001	6	<0.0001	1
tFl-tfo	<0.0001	2	<0.0001	2
tFl-Tfo	<0.0001	1	<0.0001	0
TFl-tfl	<0.0001	1	<0.0001	0
TFl-Tfl	<0.0001	6	<0.0001	5
TFl-tFo	0.0077	0	<0.0001	0
TFl-TFo	0.1185	0	0.0637	1
TFl-tfo	<0.0001	2	<0.0001	0
TFl-Tfo	<0.0001	4	<0.0001	3
tfl-Tfl	0.0033	3	<0.0001	0
tfl-tFo	<0.0001	2	<0.0001	1
tfl-TFo	<0.0001	1	<0.0001	0
tfl-tfo	0.095	10	0.054	13
tfl-Tfo	0.0001	5	<0.0001	0
Tfl-tFo	<0.0001	5	<0.0001	0
Tfl-TFo	<0.0001	5	<0.0001	3

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).

Tfl-tfo	0.0067	8	<0.0001	2
Tfl-Tfo	0.1032	3	0.0313	3
tFo-TFo	0.0119	5	<0.0001	2
tFo-tfo	<0.0001	2	<0.0001	2
tFo-Tfo	<0.0001	2	<0.0001	0
TFo-tfo	<0.0001	2	<0.0001	2
TFo-Tfo	<0.0001	3	<0.0001	4
tfo-Tfo	0.0005	8	<0.0001	1

These results, together with visualization in Fig. 4 and Suppl. Video S1 show that yogurts are best discriminated in the PC2 vs PC3 plane (Fig. 4 and in Suppl. Fig. S3), which accounts for only 38.4% of the total sum of squares. PC1 extracts a larger proportion of the total sum of squares (49.0%), but nearly all of the sum of squares that it extracts is related to differences across, instead of within, timepoints. In fact, PC1 extracts only 1.5% of the *relevant* sum of squares, compared to 53.7% in PC2 and 33.8% in PC3. This finding is consistent with our observation that within-timepoint discrimination of the yogurt pairs is best in PC2 and PC3.

4.2.2. A PCA of yogurt results focusing on relevant paired comparisons

The matrix $\mathbf{\Delta}^*$ with 3136 rows corresponding to 28 unique paired comparisons within 56 time points (Section 2.3.2) was submitted to PCA as in (7). The first four PCs extract 56.0%, 28.1%, 5.8%, and 4.1% of the variance from this column-centered matrix. Although a two-component PCA solution might be sufficient, to facilitate direct comparisons with the results presented above we selected a three-component solution which extracts 89.9% of the variance from $\mathbf{\Delta}^*$.

To visualize the truncated loading matrix ($\mathbf{P}_A^*$), we present three loading plots (Fig. 5, panels a, c, e). PC1 contrasts *gritty* vs *sandy* textural perceptions. PC2 contrasts perceptions of *thin* vs *thick*. PC3 contrasts *sweet* vs *acidic* and *bitter* perceptions. The loading plot for $\mathbf{P}_A^*$ in the PC1 vs PC2 plane (Fig. 5a) resembles the loading plot for $\mathbf{P}_A$ in the PC2 vs PC3 plane (Fig. 4e). $\mathbf{P}_A^*$ does not have a component similar to PC1 from $\mathbf{P}_A$, which captures mainly across-timepoint variability (which is not of primary interest) but negligible within-timepoint variability (which is of primary interest).

The virtual-panel PCA results resemble the real-panel results in terms of %VAF (Suppl. Table S4, eComponent). The TTB-derived scores were used to obtain 95% confidence ellipsoids for the C=28 unique paired comparisons of the eight yogurts as described in Section 2.3.2 and Section 3.3.2.

In the previous section, we presented results for three pairs of yogurts at 10 s in $\mathbf{P}_A$ (Fig. 4 in Section 4.2.1). Now, we investigate these same pairs at the same timepoint, but for results in $\mathbf{P}_A^*$ (Fig. 5, b, d, f). As would be expected based on the loadings, the results and interpretations that we get in the PC1 vs PC2 plane (Fig. 5b) are similar to the results and interpretations that we get from the PC2 vs PC3 plane based on $\mathbf{P}_A$ (Fig. 4f). Discrimination of the tfo vs Tfo pair, which differs only in viscosity, is borderline, whereas the other two pairs, which differ in more than one design factor, are well discriminated.

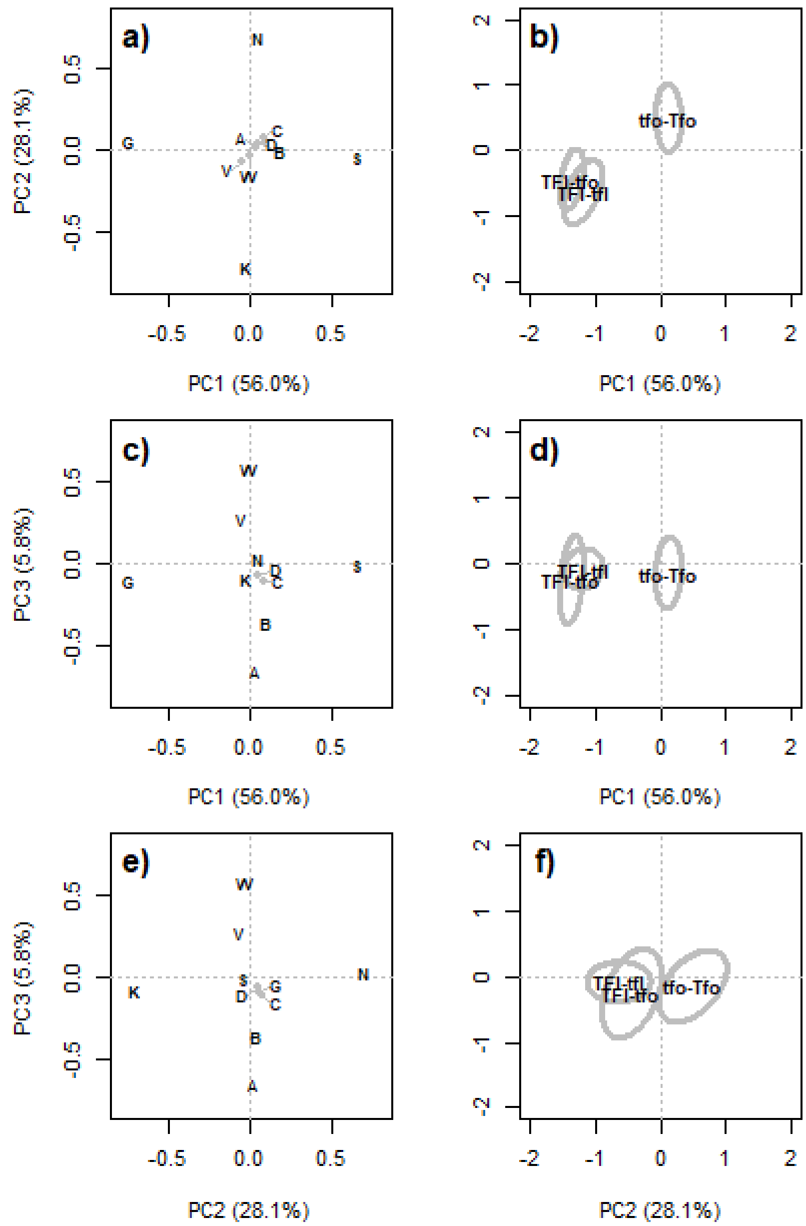
Suppl. Video S1 (bottom row; eComponent) shows biplot for the three PCs planes in $\mathbf{P}_A$*. These plots show each of the 28 paired comparisons over time. Paired differences in the PC1 vs PC2 plane of $\mathbf{P}_A$* often resembled paired differences in the PC2 vs PC3 plane of $\mathbf{P}_A$. *Thin* and *thick* perceptions were associated with thin and thick viscosity levels, respectively (Table 2). *Gritty* and *sandy* perceptions were associated with particles sizes of flake and flour, respectively. But the PC1 vs PC3 and PC2 vs PC3 planes of $\mathbf{P}_A$* also show that low-flavour yogurts tend to be described as more often as *bitter* and *acidic* than the optimal-flavour yogurts, which tend to be described more often as *sweet* (Suppl. Video S1, bottom row).

To screen these results, we calculated P values for the 28 paired comparisons at each of the timepoints. The timepoint at which each paired comparison was best discriminated is shown in Table 4. The time at which differentiation was best occurred earlier in the analysis of relevant paired comparisons than in the conventional analysis based on all paired comparisons. Again, yogurt pairs were best discriminated if they differed in all three design factors (viscosity, particle size, flavour intensity). Most yogurt pairs were discriminated; the four yogurt pairs that differed only in their flavour formulations were not discriminated. Generally, the P values from the analysis that accounts for the special structure of the data (Section 3.1.2) were smaller, and thus more discriminating, than the P values from the PCA of all paired comparisons (Section 3.1.1).

<<FIG 5 APPROXIMATELY HERE>>

Fig. 5. Plots from PCA of relevant paired comparisons of yogurt formulations at 10 s. Results are visualized via loading plots (left column) and paired difference score plots for three of the 28 relevant paired comparisons (right column) in the planes of PC1 vs PC2 (top row; a and b), PC1 vs PC3 (middle row; c and d), and PC2 vs PC3 (bottom row; e and f) onto which the 95% confidence ellipsoids for the paired difference scores are projected. Attributes: acidic [A], *bitter* [B], *cloying* [C], *dry* [D], *gritty* [G], *sandy* [S], *sweet* [W], *thick* [K], *thin* [N], and *vanilla* [V]. *(Yogurt formulations shown: thick with flakes and low flavour intensity* [TFl]; *thin with flour and low flavour intensity* [tfl]; *thin with flour and optimal flavour intensity* [tfo]; *thick with flakes and optimal flavour intensity* [Tfo].*)*

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).



PCA of $\boldsymbol{\Delta}^*$ extracts sum of squares maximally only from relevant paired comparisons because $\boldsymbol{\Delta}^*$ only contains relevant paired comparisons. The percentage of the relevant (within-timepoint) sum of squares in a PC coincides with the percentage of sum of squares extracted in that PC. The percentage of relevant sum of squares extracted by the PCA of $\boldsymbol{\Delta}^*$ is 84.1% in the PC1 vs PC2 plane and 89.7% in the first three PCs. Since PCA of $\mathbf{X}\ominus\mathbf{X}$ extracts relevant sum of squares mainly in PC2 and PC3, but only negligible relevant sum of squares in PC1, the Gain from using PCA of $\boldsymbol{\Delta}^*$ is more than 3500% in one PC, 52% in the first two PCs, but only 0.7% in the first three PCs.

These Gain results indicates that $\mathbf{P}_A$* is a better space than $\mathbf{P}_A$ for investigating the relevant paired comparisons, particularly in PC1 and PC2. Here, as in Section 4.1.2, $\mathbf{P}_A$* extracts a larger proportion of the sum of squares from the relevant paired comparisons and provides a space in which the relevant pairs are better discriminated.

5. Discussion

5.1. Different PCA solutions with different statements of significance for paired comparisons

One reason that PCA is widely used is that it optimally compresses variance in the original variables into only a few PCs, which can then be investigated visually. In the present paper, we show how PCA can extract variance optimally from only a subset of relevant paired comparisons. The PCA of relevant paired comparisons yields a different space that better discriminates these pairs than PCA based on all paired comparisons.

Many researchers judging the importance of each PC by its %VAF, which assumes that results are best investigated in the PC1 vs PC2 plane. But our results show that this practice can be misleading if only a subset of the paired comparisons are relevant. For example, in the conventional analysis of the smoothie data set, relevant pairs are best investigated in the PC1 vs PC3 plane (Section 4.1.1). In the conventional analysis of the yogurt data set, the relevant pairs are best investigated in the PC2 vs PC3 plane because almost all of the variance extracted in PC1 is related to between-timepoint paired comparisons, which were not considered to be relevant (Section 4.2.1). The results also showed that conducting PCA accounting for the special data structure has benefits, but the size of the benefit differs depending on the data.

It might seem peculiar to have a data set (**X**) for which we conduct two PCAs, one based on all paired comparisons ($\mathbf{X}\ominus\mathbf{X}$), one based on selected paired comparisons ($\mathbf{\Delta}$*), which give different results leading to different statements of significance for the paired comparisons. A reason this occurs is that one or more variables that contribute to forming one space may be mostly left out of the other space. For example, we found that visual attributes were more important for separating test-test smoothie paired comparisons than the control-test smoothie paired comparisons (Section 4.1). We also find that PCA of all yogurt paired comparisons produces a PC1 that extracts mostly variability across timepoints (Section 4.2.1), whereas the PCA of only relevant (i.e. within-timepoint) paired comparisons exacts mostly variability of the yogurt paired comparisons within timepoints (Section 4.2.2). If within-timepoint differences are of primary interest, then the latter analysis is more appropriate.

5.2. The issue of variance-standardization of columns of $\mathbf{\Delta}$* *before PCA*

Variance-standardizing the columns of the smoothie data set (Section 4.1) puts the variables in **X** on equal footing. Column variances of $\mathbf{\Delta}$* depend on which paired comparisons are included (Section 3.2.4). If its columns are variance-standardized to obtain $\mathbf{\Delta}$†, then PCA of $\mathbf{\Delta}$† yields a loading matrix **P**† that equals neither **P** nor **P***. Since $\mathbf{\Delta}$* and $\mathbf{\Delta}$† have the same dimensions, %VAF can be used to compared their respective PCA solutions. In Appendix A.2, we show that the %VAF in the directions of $\mathbf{P}_A$† can be larger or smaller than the %VAF in the directions of $\mathbf{P}_A^*$. We cannot make a general statement about which approach is superior using %VAF as a criterion because which one is superior *depends on the data*.

It could be argued that $\mathbf{P}_A^\dagger$ is appropriate because it performs PCA with all variables on an equal footing. One reason that we find $\mathbf{P}_A^*$ appropriate is that variables had already been put on an equal footing when $\mathbf{X}$ was column-standardized. When all paired comparisons are relevant, we would avoid variance-standardizing columns of $\mathbf{X}\ominus\mathbf{X}$ or $\mathbf{\Delta}^*$ since their PCA results would no longer be connected to the row differences in PCA of $\mathbf{X}$, which are shown in (4). This argument might be extended also to the case where only a subset of paired comparisons is relevant.

A reason that we did not do a new column standardization is given in Section 3.2.4: it ensures that the relevant ($2C \times M$) in $\mathbf{X}\ominus\mathbf{X}$ and the ($2C \times M$) matrix $\mathbf{\Delta}^*$ have identical data, so the variances in their respective columns are equal. Their respective sums of squares are also equal. The relevant sum of squares in the directions of $\mathbf{P}_A^*$ is never less that in the directions of $\mathbf{P}_A$. But the relevant sum of squares obtained from $\mathbf{P}_A^\dagger$ might be smaller or larger than the relevant sum of squares obtained from $\mathbf{P}_A$. This matter probably deserves further study. Ultimately, the decision to use $\mathbf{\Delta}^*$ or $\mathbf{\Delta}^\dagger$ rests with the researcher and is typical of judgments that must be made during data analysis. We advocate that such decisions be made to answer the relevant research questions in the most appropriate manner.

5.3. Variable filtering before PCA

For the smoothie data set, we conducted a two-way analysis of variance per variable, then dropped variables that did not have a significant formulation (i.e. treatment or product) effect (Section 2.1.1). This variable filtering step is especially important if variables will be variance-standardized: it avoids weighting all variables equally regardless of whether the differences arise systematically or by chance alone. But, since only on the control-test paired comparisons were considered relevant, Dunnett's many-to-one multiple comparison test procedure (Dunnett, 1964) might have been conducted per variable instead of analysis of variance. In this case, we would retain the variables having at least one test formulation that differed significantly from the control formulation, and drop variables with no significant control-test differences.

5.4. Gain from PCA of relevant paired comparisons in temporal sensory data sets

PCA of all paired comparisons in the yogurt data (Section 2.2) via (5) yielded a PC1 that extracted about half of the total variance but only a trivial proportion of the relevant variance (Section 4.2.1). Castura et al. (2016b) report exactly this phenomenon: their PC1 extracted an even larger percentage of variance (85%). The reason is that they used a longer evaluation period (170 s) in which perceptions were tracked to extinction. Since their citation rates started at zero and ended at (nearly) zero, the variability across timepoints was very large. By contrast, the yogurt TCATA data in Section 2.2 were left-trimmed, so every evaluation started with the first attribute checked; the yogurt evaluations were both shorter and had a shorter duration of (nearly) zero citation rates, which reduced the relative proportion of variability across timepoints. Perhaps the most appropriate use of PCA of all paired comparisons in temporal sensory studies include the goal of understanding the common temporal signature (Meyners & Castura, 2019). In the PCA of all paired comparisons, within-timepoint variability can still be investigated in subsequent components, accepting the %VAF in these component might be slight. If the focus is to understand within-timepoint variability, then PCA of relevant paired comparisons is advised. Or, if the goal is to learn as much as possible, then it could be very useful to conduct PCA of $\mathbf{\Delta}^*$ and PCA of $\mathbf{X} \ominus \mathbf{X}$. Nothing precludes obtaining and interpreting both of these solutions.

5.5. Other data sets with special structures

The examples of special structures (control-test paired comparisons, temporal sensory data) that we have presented in this manuscript are not intended to be exhaustive, only to illustrate the approach. In other data sets, the approach could be used to investigate other special structures, such as within-group comparisons or reference-test comparisons with multiple references. In fact, whenever a product (row) of $\mathbf{X}$ is dropped before conducting PCA of $\mathbf{X}$, its solution will be equivalent to PCA of $\mathbf{\Delta}^*$ in which the relevant pairs are all paired comparisons except the pairs involving the dropped product.

The approaches that we have described in this manuscript could also be extended to other multivariate analyses, such as multiple factor analysis and generalized Procrustes analysis. There, as here, the size of the benefit from taking into account the special structure of the data over a conventional analysis will be different for different data sets, with Gain being larger in cases where the relevant paired comparisons differ markedly on attributes that are unimportant for paired comparisons that are not of interest.

6. Conclusions

This paper focuses on how paired comparisons can be investigated within PCA. Often, all paired comparisons are of interest. When this is the case, the variance in these paired comparisons can be investigated optimally in the same space as the original data. But in some cases, data sets have a special structure where not all paired comparisons are of interest. Two such examples are a data set with control-test comparisons, and a data set with temporal sensory data in which within-timepoint paired comparisons are more important than across-timepoint paired comparisons. When only a subset of paired comparisons is of interest, then the variance in this subset is not investigated optimally in the PCA space obtained from all paired comparisons.

After showing how the variance in the relevant paired comparisons can be investigated optimally in PCA, we proposed how to construct confidence regions, how to visualize results, and how to screen the results to ensure that significant differences are not missed. In two example data sets with a special structures, we show that PCA of only relevant paired comparisons extracts a larger proportion of the relevant sum of squares than PCA based on all paired comparisons. Gain, which quantifies the benefit from using PCA accounting for the special structure, depends on the data set. In the temporal sensory data set, Gain was humungous: PCA accounting for the special structure extracts 56.0% of the relevant sum of squares in PC1 vs 1.5% in PCA based on all paired comparisons. In the data set from the trained sensory panel that evaluated the smoothie formulations, Gain was comparatively modest, but still large: PCA accounting for the special structure extracts 80.8% of the relevant sum of squares in PC1 vs 70.1% in PCA based on all paired comparisons. In both data sets, the relevant paired comparisons were better separated in the analyses that accounted for the special structure. The methods proposed in this manuscript can be adapted to investigate other data sets with other special structures and to other multivariate analyses.

Researchers are also free to obtain complementary PCA solutions, one from PCA conducted conventionally, identical to PCA based on all paired comparisons, and one from PCA accounting for the special structure, where insights from each solution are combined to maximize what can be learned from the study. More broadly, the findings presented here can also serve as a reminder that decisions in data analysis should not always be run by established conventions, but should instead be guided by which research questions need to be answered.

Appendix A

A.1. Properties of a matrix of relevant paired comparisons

The goal of this appendix is to demonstrate that the sample covariance matrix for $\mathbf{\Delta}^*$ (denoted $\mathbf{\Sigma}_{\mathbf{\Delta}^*}$) is *not* related by a scalar to the sample covariance matrix for $\mathbf{X}\ominus\mathbf{X}$ (denoted $\mathbf{\Sigma}_{\mathbf{X}\ominus\mathbf{X}}$). We make this demonstration by counterexample by showing two matrices $\mathbf{X}\ominus\mathbf{X}$ and $\mathbf{\Delta}^*$ that are not related only by a scalar:

$$\mathbf{X} = \begin{bmatrix} 2 & -1 & -3 \\ -4 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}, \qquad \mathbf{X} \ominus \mathbf{X} = \begin{bmatrix} 0 & 0 & 0 \\ 6 & 0 & -5 \\ 0 & -3 & -2 \\ -6 & 0 & 5 \\ 0 & 0 & 0 \\ -6 & -3 & 3 \\ 0 & 3 & 2 \\ 6 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{\Delta}^* = \begin{bmatrix} 6 & 0 & -5 \\ 0 & -3 & -2 \\ -6 & 0 & 5 \\ 0 & 3 & 2 \end{bmatrix},$$

i.e., only the paired comparisons of rows 1 vs 2 and rows 1 vs 3 are considered to be relevant when constructing $\mathbf{\Delta}^*$. Each matrix is column centered. The sample covariance matrices of these matrices are

$$\mathbf{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 12 & 3 & -8 \\ 3 & 3 & -1/2 \\ -8 & -1/2 & 19/3 \end{bmatrix}, \quad \mathbf{\Sigma}_{\mathbf{X}\ominus\mathbf{X}} = \begin{bmatrix} 18 & 9/2 & -12 \\ 9/2 & 9/2 & -3/4 \\ -12 & -3/4 & 19/2 \end{bmatrix}, \qquad \mathbf{\Sigma}_{\mathbf{\Delta}^*} = \begin{bmatrix} 24 & 0 & -20 \\ 0 & 6 & 4 \\ -20 & 4 & 58/3 \end{bmatrix}.$$

The relationship between the first and middle sample covariance matrices depends on the number of rows and not the data (Castura et al., 2023b), where in this case,

$$\mathbf{\Sigma}_{\mathbf{X}}/\mathbf{\Sigma}_{\mathbf{X}\ominus\mathbf{X}} = \begin{bmatrix} 2/3 & 2/3 & 2/3 \\ 2/3 & 2/3 & 2/3 \\ 2/3 & 2/3 & 2/3 \end{bmatrix}.$$

The relationship between the last and middle sample covariance matrices depends on the data. In this case,

$$\mathbf{\Sigma}_{\mathbf{\Delta}^*}/\mathbf{\Sigma}_{\mathbf{X}\ominus\mathbf{X}} = \begin{bmatrix} 1.33 & 0 & -1.67 \\ 0 & 1.33 & -0.53 \\ 1.67 & -0.53 & 2.04 \end{bmatrix}.$$

Since $\mathbf{\Sigma}_{\mathbf{X}}$ and $\mathbf{\Sigma}_{\mathbf{X}\ominus\mathbf{X}}$ differ only by a scalar, SVD (1) of **X** and SVD of **X**⊖**X** yield the same right singular vectors **P** (Castura et al., 2023b). But the sample covariance matrices $\mathbf{\Sigma}_{\mathbf{\Delta}^*}$ and $\mathbf{\Sigma}_{\mathbf{X}\ominus\mathbf{X}}$ do not differ only by a scalar; their differences depend on the data. Thus, if SVD of **X**⊖**X** yields the right singular vectors **P** and SVD of **Δ*** yields the right singular vectors **P***, then **P**≠**P***.

A.2. Percentage of variance extracted in the first PCs of **Δ*** *and* **Δ**† *depends on the data*

The goal of this appendix is to show that the %VAF in the first *A* PCs of a matrix of relevant paired comparisons (**Δ***) can be larger or smaller than the %VAF in the first *A* PCs of a matrix of relevant paired comparisons with variance-standardized columns (**Δ**†). We set *A*=1 for simplicity, then show that %VAF in PC1 of **Δ*** can larger or smaller than the %VAF in PC1 of **Δ**†, depending on the data.

Case 1. We show that PC1 of **Δ*** can extract *more* variance than PC1 of **Δ**†. For simplicity, we base this demonstration on a column-centered (*3*×*3*) matrix **X**. We treat 1 vs 2 and 1 vs 3 as the relevant paired comparisons.

From a particular column-centered results matrix (**X**), we obtain the matrices **Δ*** and **Δ**†, where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 4/3 \\ -1 & -2 & 10/3 \\ 0 & 1 & -14/3 \end{bmatrix}, \qquad \mathbf{\Delta}^* = \begin{bmatrix} 2 & 3 & -2 \\ 1 & 0 & 6 \\ -2 & -3 & 2 \\ -1 & 0 & -6 \end{bmatrix}, \text{ and } \mathbf{\Delta}^\dagger = \begin{bmatrix} 1.10 & 1.22 & -039 \\ 0.55 & 0 & 1.16 \\ -1.10 & -1.22 & 0.39 \\ -0.55 & 0 & -1.16 \end{bmatrix}.$$

PC1 extracts 76.2% of the variance in **Δ***. PC1 of extracts 63.7% of the variance in **Δ**†. This demonstrates a case where PC1 extracts more variance from **Δ*** than from **Δ**†.

Case 2. We show that PC1 of **Δ*** can extract *less* variance than PC1 of **Δ**†. Again, we start with a (*3*×*3*) column-centered matrix **X**, treat 1 vs 2 and 1 vs 3 as the relevant paired comparisons.

From a particular column-centered results matrix (**X**), we obtain the matrices **Δ*** and **Δ**†, where

$$\mathbf{X} = \begin{bmatrix} 2/3 & -1/3 & -2 \\ 2/3 & -7/3 & 2 \\ -4/3 & 8/3 & 0 \end{bmatrix}, \qquad \mathbf{\Delta}^* = \begin{bmatrix} 0 & 2 & -4 \\ 2 & -3 & -2 \\ 0 & -2 & 4 \\ -2 & 3 & 2 \end{bmatrix}, \qquad \mathbf{\Delta}^\dagger = \begin{bmatrix} 0 & 0.68 & -1.10 \\ 1.22 & -1.02 & -0.55 \\ 0 & -0.68 & 1.10 \\ -1.22 & 1.02 & 0.55 \end{bmatrix},$$

PC1 ($\mathbf{P}_A$*) extracts 56.8% of the variance in **Δ***. PC1 ($\mathbf{P}_A$†) extracts 63.2% of the variance in **Δ**†. This demonstrates a case where PC1 extracts more variance from **Δ**† than from **Δ***.

Case 1 gives an example where *A* PCs extract *more* variance from **Δ*** than from **Δ**†. Case 2 gives an example where *A* PCs extract *less* variance from **Δ*** than from **Δ**†. Taken together, these cases show that the %VAF in the first *A* PCs of **Δ*** can be larger or smaller than the %VAF in the first *A* PCs of **Δ**†, depending on the data.

Acknowledgements

Authors TN and PV acknowledge financial support from the Research Council of Norway and the Norwegian Fund for Research Fees for Agricultural Products (FFL) through the project "FoodForFuture" (Project number 314318; 2021-2024).

References

Abdi, H., & Williams, L.J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 433-459. https://doi.org/10.1002/wics.101

Asioli, D., Nguyen, Q.C., Varela, P., & Næs, T. (2022). Comparison of different ways of handling L-shaped data for integrating sensory and consumer information. *Food Quality and Preference*, *96*, 104426. https://doi.org/10.1016/j.foodqual.2021.104426

Babamoradi, H., van den Berg, F., & Rinnan, Å. (2013). Bootstrap based confidence limits in principal component analysis—A case study. *Chemometrics and Intelligent Laboratory Systems*, *120*, 97-105. https://doi.org/10.1016/j.chemolab.2012.10.007

Berget, I., Castura, J.C., Ares, G., Næs, T., & Varela, P. (2020). Exploring the common and unique variability in TDS and TCATA data–A comparison using canonical correlation and orthogonalization. *Food Quality and Preference*, *79*, 103790. https://doi.org/10.1016/j.foodqual.2019.103790

Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality and Preference*, *28*, 106–115. https://doi.org/10.1016/j.foodqual.2012.09.005

Castura, J.C. (2018). Dynamics of consumer perception. In G. Ares & P. Varela (eds.): *Methods in Consumer Research*, Volume 1 (pp. 211-240). Duxford, UK: Woodhead Publishing.

Castura, J.C. (2020). Investigating temporal sensory data via a graph theoretic approach. *Food Quality and Preference*, *79*, 103787. https://doi.org/10.1016/j.foodqual.2019.103787

Castura, J.C., Antúnez, L., Giménez, A., & Ares, G. (2016a). Temporal Check-All-That-Apply (TCATA): A novel dynamic method for characterizing products. *Food Quality and Preference*, *47*, 79-90. https://doi.org/10.1016/j.foodqual.2015.06.017

Castura, J.C., Baker, A.K., & Ross, C.F. (2016b). Using contrails and animated sequences to visualize uncertainty in dynamic sensory profiles obtained from temporal check-all-that-apply (TCATA) data. *Food Quality and Preference*, *54*, 90-100. https://doi.org/10.1016/j.foodqual.2016.06.011

Castura, J.C., Rutledge, D.N., Ross, C.F., & Næs, T. (2022). Discriminability and uncertainty in principal component analysis (PCA) of temporal check-all-that-apply (TCATA) data. *Food Quality and Preference*, *96*, 104370. https://doi.org/10.1016/j.foodqual.2021.104370

Castura, J.C., Varela, P., & Næs, T. (2023a). Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis. *Food Quality and Preference*, in press. https://doi.org/10.1016/j.foodqual.2023.104843

Castura, J.C., Varela, P., & Næs, T. (2023b). Investigating paired comparisons after principal component analysis. *Food Quality and Preference*, 106, 104814. https://doi.org/10.1016/j.foodqual.2023.104814

Courcoux, P., Qannari, E.M., Taylor, Y., Buck, D., & Greenhoff, K. (2012). Taxonomic free sorting. *Food Quality and Preference*, 23, 30–35. https://doi.org/10.1016/j.foodqual.2011.04.001

Dunnett, C.W. (1964). New tables for multiple comparisons with a control. *Biometrics*, 20, 482-491. https://doi.org/10.2307/2528490

Efron, B., & Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*. New York: CRC Press.

Galler, M., Næs, T., Almli, V.L., & Varela, P. (2020). How children approach a CATA test influences the outcome. Insights on ticking styles from two case studies with 6–9-year old children. *Food Quality and* Preference, 86, 104009. https://doi.org/10.1016/j.foodqual.2020.104009

Gonzalez-Estanol, K., Cliceri, D., Biasioli, F., & Stieger, M. (2022). Differences in dynamic sensory perception between reformulated hazelnut chocolate spreads decrease when spreads are consumed with breads and wafers. *Food Quality and Preference*, *98*, 104532. https://doi.org/10.1016/j.foodqual.2022.104532

Hort, J., Kemp, S.E., & Hollowood, T. (Eds.). (2017). *Time-Dependent Measures of Perception in Sensory Evaluation*. Chichester, UK: John Wiley & Sons.

Husson, F., Lê, S., & Pagès, J. (2005). Confidence ellipse for the sensory profiles obtained by principal component analysis. *Food Quality and Preference*, *16*, 245-250. https://doi.org/10.1016/j.foodqual.2004.04.019

Kiers, H.A.L., & Groenen, P.J.F. (2006). Visualizing Dependence of Bootstrap Confidence Intervals for Methods Yielding Spatial Configurations. In: S. Zani, A. Cerioli, M. Riani, & M. Vichi (eds.) *Data Analysis, Classification and the Forward Search. Studies in Classification, Data Analysis, and Knowledge Organization*. Heidelberg: Springer. https://doi.org/10.1007/3-540-35978-8_14

Lebart, L. (2007). Which Bootstrap for Principal Axes Methods? In P. Brito, G. Cucumel, P. Bertrand, & F. de Carvalho (Eds.): *Selected Contributions in Data Analysis and Classification* (pp. 581–588). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73560-1_55

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).

Mahalanobis, P.C. (1936). On the Generalised Distance in Statistics. *Sankhya A*, 80, 1–7 (2018). https://doi.org/10.1007/s13171-019-00164-5

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). *Multivariate Analysis*. Toronto: Academic Press.

McMahon, K.M., Culver, C., Castura, J.C., & Ross, C.F. (2017). Perception of carbonation in sparkling wines using descriptive analysis (DA) and temporal check-all-that-apply (TCATA). *Food Quality and Preference*, *59*, 14-26. https://doi.org/10.1016/j.foodqual.2017.01.017

Meyners, M. (2020). Temporal methods: Are we comparing apples and oranges? *Food Quality and Preference*, *79*, 103615. https://doi.org/10.1016/j.foodqual.2018.11.022

Meyners, M., & Castura, J.C. (2019). Did assessors select attributes by chance alone in your TDS study, and how relevant is it to know? *Food Research International*, *119*, 571-583. https://doi.org/10.1016/j.foodres.2018.10.035

Næs, T., Tomic, O., Endrizzi, I., & Varela, P. (2021). Principal components analysis of descriptive sensory data: Reflections, challenges, and suggestions. *Journal of Sensory Studies*, *36*, e12692. https://doi.org/10.1111/joss.12692

Nguyen, Q.C., Liland, K. H., Tomic, O., Tarrega, A., Varela, P., & Næs, T. (2020a). SO-PLS as an alternative approach for handling multi-dimensionality in modelling different aspects of consumer expectations. *Food Research International*, *133*, 109189. https://doi.org/10.1016/j.foodres.2020.109189

Nguyen, Q.C., Næs, T., & Varela, P. (2018). When the choice of the temporal method does make a difference: TCATA, TDS and TDS by modality for characterizing semi-solid foods. *Food Quality and Preference*, *66*, 95-106. https://doi.org/10.1016/j.foodqual.2018.01.002

Nguyen, Q.C., Næs, T., Almøy, T., & Varela, P. (2020b). Portion size selection as related to product and consumer characteristics studied by PLS path modelling. *Food Quality and Preference*, *79*, 103613. https://doi.org/10.1016/j.foodqual.2018.11.020

Nguyen, Q.C., & Varela, P. (2021). Identifying temporal drivers of liking and satiation based on temporal sensory descriptions and consumer ratings. *Food Quality and Preference*, *89*, 104143. https://doi.org/10.1016/j.foodqual.2020.104143

Nguyen, H., & Wismer, W.V. (2022). Temporal sensory profiles of regular and sodium-reduced foods elicited by temporal dominance of sensations (TDS) and temporal check-all-that-apply (TCATA). *Foods*, *11*, 457. https://doi.org/10.3390/foods11030457

Poveromo, A.R., & Hopfer, H. (2019). Temporal check-all-that-apply (TCATA) reveals matrix interaction effects on flavor perception in a model wine matrix. *Foods*, 8(12), 641. https://doi.org/10.3390/foods8120641

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reyes, M.M., Castura, J.C., & Hayes, J.E. (2017). Characterizing dynamic sensory properties of nutritive and nonnutritive sweeteners with temporal check-all-that-apply. *Journal of Sensory Studies*, *32*, e12270. https://doi.org/10.1111/joss.12270

Schönemann, P.H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*, 1–10. https://doi.org/10.1007/BF02289451

Schumaker, M.R., Diako, C., Castura, J.C., Edwards, C.G., & Ross, C.F. (2019). Influence of wine composition on consumer perception and acceptance of Brettanomyces metabolites using temporal check-all-that-apply methodology. *Food Research International*, *116*, 963-972. https://doi.org/10.1016/j.foodres.2018.09.034

Sharma, M., & Duizer, L. (2019). Characterizing the dynamic textural properties of hydrocolloids in pureed foods—A comparison between TDS and TCATA. *Foods*, *8*, 184. https://doi.org/10.3390/foods8060184

eComponent

Suppl. Table S1. PCA of the matrix of all paired comparisons (Section 3.1.1) of smoothies based on results from the real panel and from the virtual panels composed by the TTB procedure.

Panel(s)	**%VAF**	**First 3 PCs**	**PC1**	**PC2**	**PC3**	**PC4**
Real	Result	94.4	59.9	19.5	15.1	2.6
	95% LCL	80.2	40.1	16.0	10.9	2.4
Virtual	Mean	89.0	54.1	20.6	14.3	5.1
	95% UCL	94.2	63.7	26.2	18.2	10.5

Suppl. Table S2. PCA of the matrix of selected (control-test) paired comparisons (Section 3.3.1) of smoothies based on results from the real panel and from the virtual panels composed by the TTB procedure.

Panel(s)	**%VAF**	**First 3 PCs**	**PC1**	**PC2**	**PC3**	**PC4**
Real	Result	97.5	80.8	9.3	7.5	1.1
	95% LCL	89.7	62.7	5.9	4.4	1.0
Virtual	Mean	94.8	76.6	10.6	7.7	2.4
	95% UCL	97.6	86.5	16.6	12.1	5.4

Suppl. Table S3. PCA of the matrix of all paired comparisons (Section 3.1.1) of yogurt-time combinations based on results from the real panel and from the virtual panels composed by the TTB procedure.

Panel(s)	Result	Retained PCs	PC1	PC2	PC3	PC4
Real		87.4	49.0	25.8	12.6	8.3
	95% LCL	80.4	41.5	21.1	8.5	5.8
Virtual	Mean	84.6	47.3	24.6	12.7	8.4
	95% UCL	87.4	53.2	27.9	17.2	11.4

Suppl. Table S4. PCA of the matrix of selected paired comparisons (Section 3.3.2) of yogurts within timepoints based on results from the real panel and from the virtual panels composed by the TTB procedure.

Panel(s)	Result	Retained PCs	PC1	PC2	PC3	PC4
Real		89.9	56.0	28.1	5.8	4.1
	95% LCL	79.5	41.9	18.3	6.0	3.8
Virtual	Mean	84.5	50.0	26.1	8.4	5.5
	95% UCL	88.1	56.8	33.7	11.9	7.8

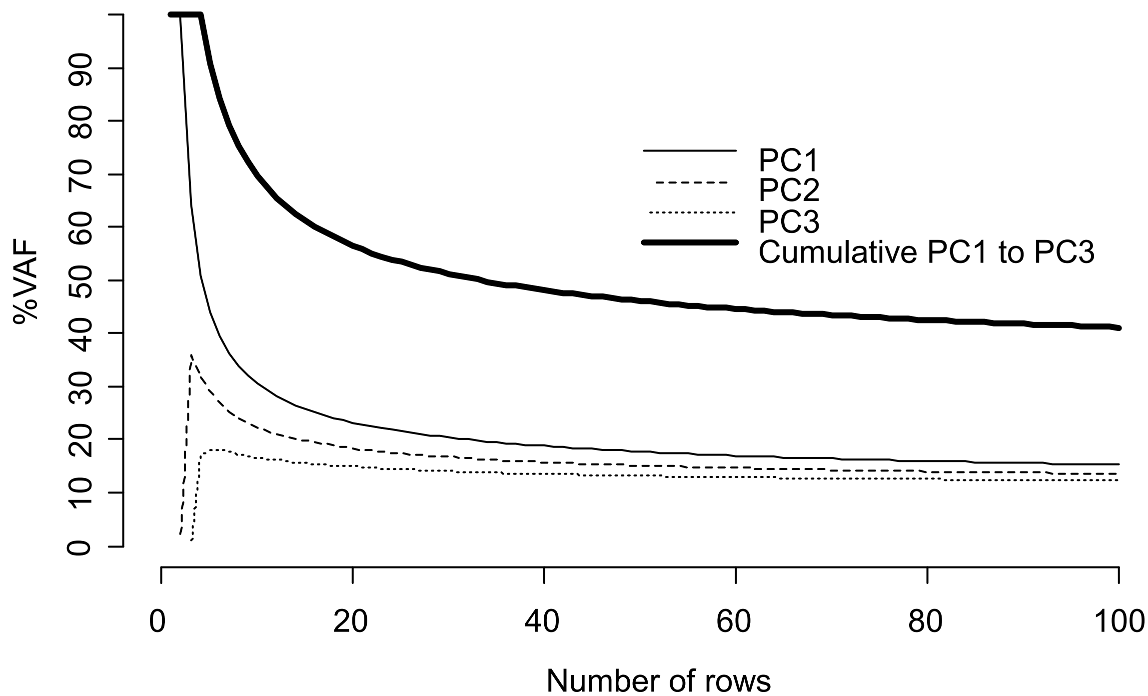
Suppl. Video S1. PCA plots of paired comparisons of yogurt formulations over time in PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right) based on PCA based on all paired comparisons (top row), which is consistent with PCA conducted conventionally, and based on PCA conducted on paired comparisons within each timepoint (bottom row). Projections of the TTB-derived paired difference scores and projections of the 95% confidence ellipsoids are shown on each plane for all yogurt paired comparisons (see Table 2 for details on yogurt formulations):

Pair	Start time	Pair	Start time	Pair	Start time	Pair	Start time
0:04	tFl-TFl	5:47	TFl-tfl	11:30	tfl-tFo	17:13	Tfl-Tfo
0:53	tFl-tfl	6:36	TFl-Tfl	12:19	tfl-TFo	18:02	tFo-TFo
1:42	tFl-Tfl	7:25	TFl-tFo	13:08	tfl-tfo	18:51	tFo-tfo
2:31	tFl-tFo	8:14	TFl-TFo	13:57	tfl-Tfo	19:40	tFo-Tfo
3:20	tFl-TFo	9:03	TFl-tfo	14:46	Tfl-tFo	20:29	TFo-tfo
4:09	tFl-tfo	9:52	TFl-Tfo	15:35	Tfl-TFo	21:18	TFo-Tfo
4:58	tFl-Tfo	10:41	tfl-Tfl	16:24	Tfl-tfo	22:07	tfo-Tfo

In each biplot, only attribute loading vectors longer than 0.1 are shown with abbreviations: acidic [A], bitter [B], cloying [C], dry [D], gritty [G], sandy [S], sweet [W], thick [K], thin [N], and vanilla [V].

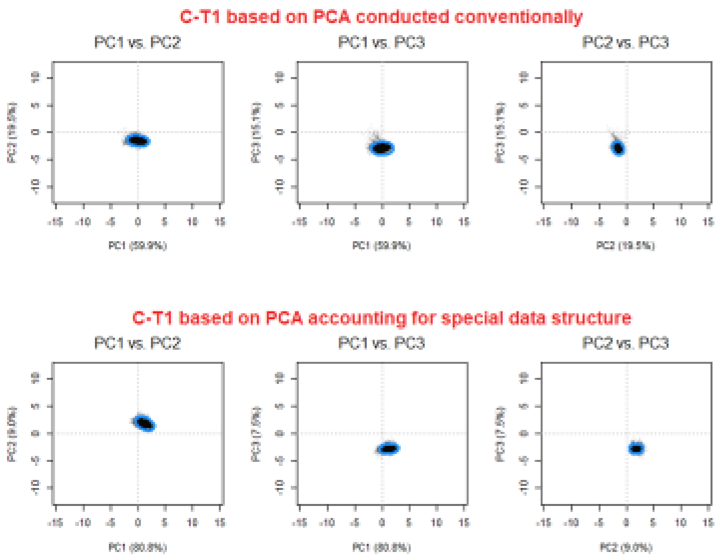
PREVIEW LINK FOR SUPPL. VIDEO 1: [SupplVideo1_Yogurt.mp4]

Suppl. Fig. S1. Data matrices, each with 10 columns and a specific number of rows (R in 1, 2, … 100), were obtained with matrix elements sampled from the standard normal distribution. For each number of rows (x axis), 5000 matrices were obtained. Columns in each matrix were centered and variance-standardized, then submitted to PCA. The percentage of variance accounted for (%VAF; y axis) in the first PC (solid line), in the second PC (dashed line), in the third PC (dotted line), and cumulatively in first three PCs (heavy solid line) are shown. When there is only 1 row, the matrix rank is 1, and 100% of the variance is extracted in one PC. As R increases, the cumulative %VAF in three PCs decreases. These results show that %VAF as calculated conventionally is inappropriate for making direct comparisons of PCA solutions that are derived from matrices having a different number of rows.



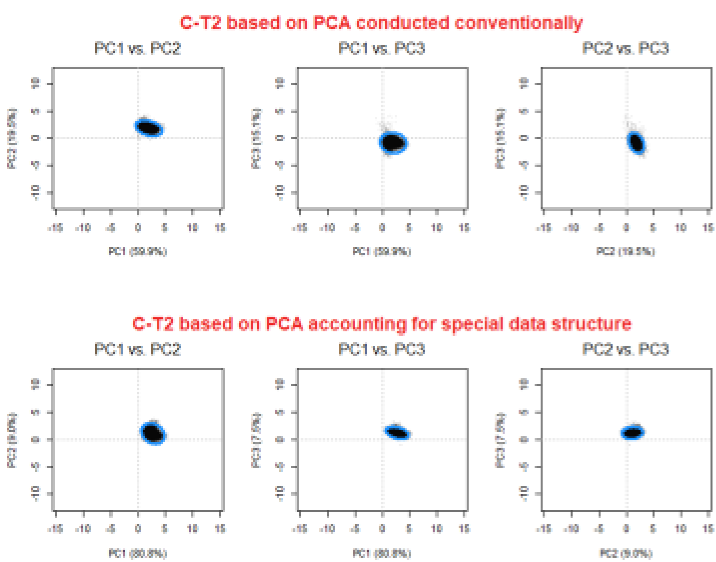
Suppl. Fig. S2. PCA results for the smoothie data set. Top row: plots from PCA based on all paired comparisons. Bottom row: plots from PCA accounting for the special structure. Left column: PC1 vs PC2; center column: PC1 vs PC3; right column: PC2 vs PC3. Score plots (panels a to i) show 95% confidence ellipsoids for Control vs Test smoothie paired comparisons projected onto the plane. (See Table 1 for details on the smoothie formulations.) Loading plots (panel j) show contributions of the sensory attributes. For improved legibility in the loading plots, only attribute vectors longer than 0.1 are shown. (Attribute abbreviations: odour intensity [i], fruit/berry odour [b], artificial odour [r], colour strength [c], whiteness [w], taste intensity [I], acidity [A], sweetness [E], sourness [S], bitterness [T], fruit/berry flavour [B], artificial flavour [R], fullness [F], viscosity [V], astringency [Y], and pungency [P].)

938



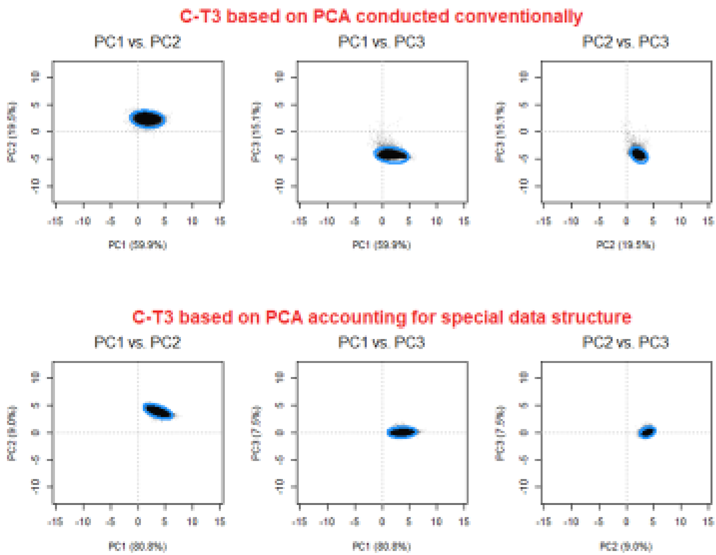
Suppl. Fig. S2a. PCA results are shown for Control vs Test 1 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).

939

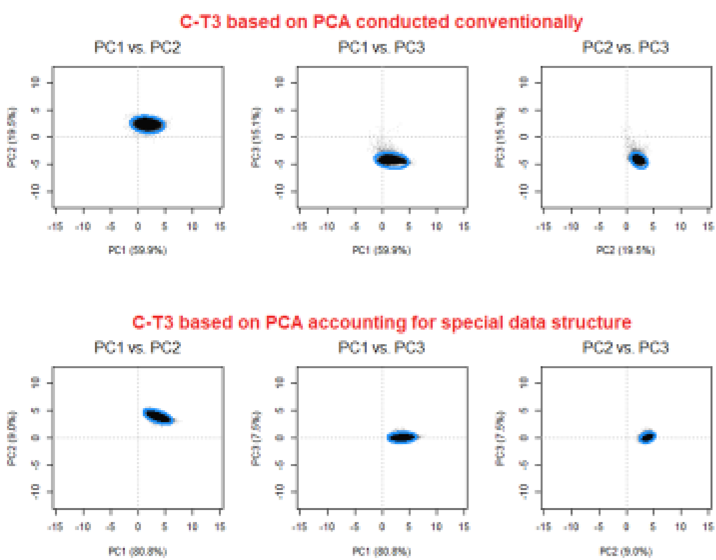


Suppl. Fig. S2b. PCA results are shown for Control vs Test 2 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).

940

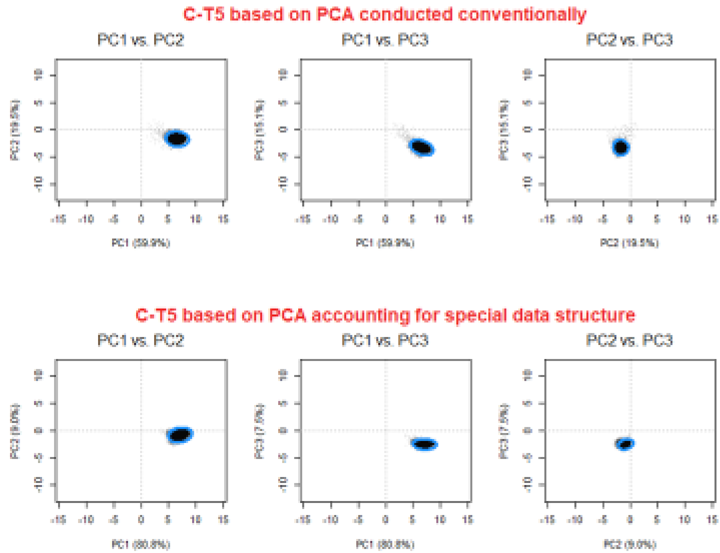


Suppl. Fig. S2c. PCA results are shown for Control vs Test 3 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).

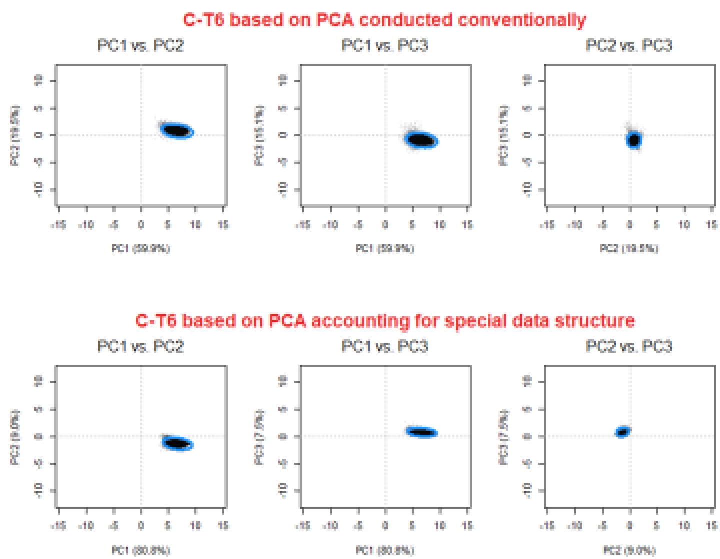


Suppl. Fig. S2c. PCA results are shown for Control vs Test 3 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).

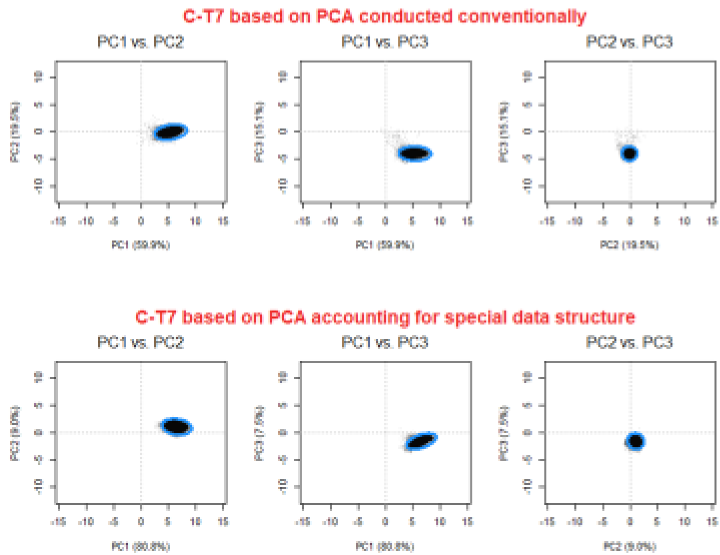
Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).



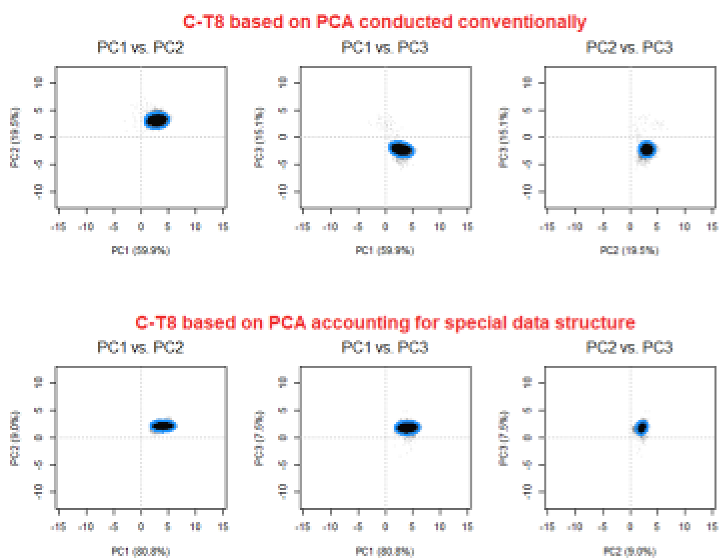
Suppl. Fig. S2e. PCA results are shown for Control vs Test 5 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).



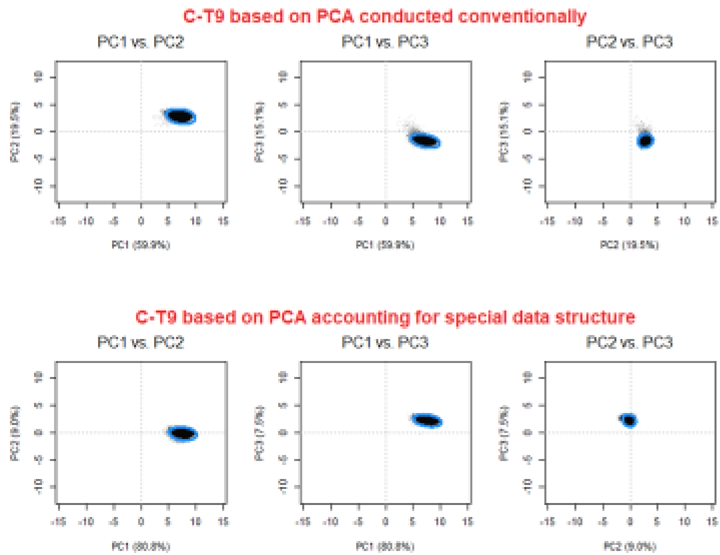
Suppl. Fig. S2f. PCA results are shown for Control vs Test 6 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).



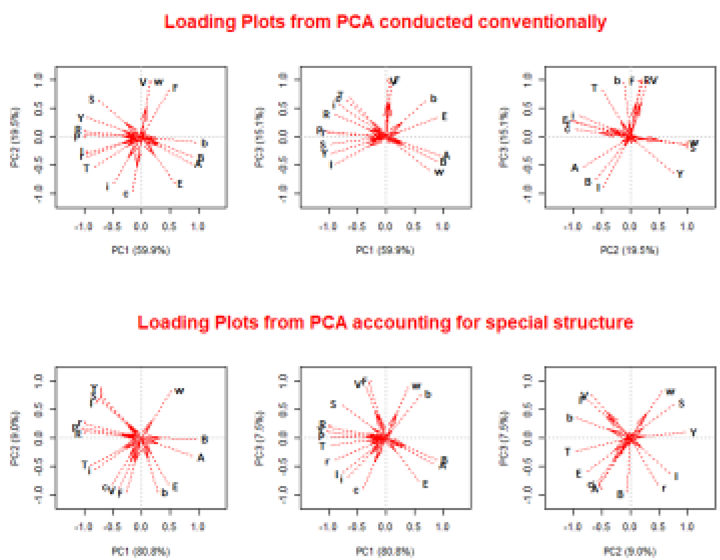
Suppl. Fig. S2g. PCA results are shown for Control vs Test 7 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).



Suppl. Fig. S2h. PCA results are shown for Control vs Test 8 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).



Suppl. Fig. S2i. PCA results are shown for Control vs Test 9 smoothie paired comparison. Score plots show projections of the 95% confidence ellipsoids for all control-test paired comparisons based on PCA based on all paired comparisons, which is consistent with PCA conducted conventionally (top row), and PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right).



Suppl. Fig. S2j. Loading plots from PCA conducted conventionally (top row) and from PCA accounting for the special structure (bottom row) in the planes of PC1 vs PC2 (left panel), PC1 vs PC3 (center), and PC2 vs PC3 (right). For improved legibility, only attribute vectors longer than 0.1 are shown. (Attribute abbreviations: odour intensity [i], fruit/berry odour [b], artificial odour [r], colour strength [c], whiteness [w], taste intensity [I], acidity [A], sweetness [E], sourness [S], bitterness [T], fruit/berry flavour [B], artificial flavour [R], fullness [F], viscosity [V], astringency [Y], and pungency [P].)

Pre-print. Manuscript submitted to *Food Quality and Preference* on March 17, 2023 (under review).