Multivariate analysis and classification of 146 odor character descriptors

ABSTRACT

Introduction Smells can be described by assigning the words that come to mind when sniffing an odorous material. A great number of terms can be applied, but not all of them are independent and it is possible to establish groups of words often applied together when describing a smell. Such classification of olfactory descriptors is of scientific interest in order to better understand the dimensionality and structure of human olfactory perception space. For this purpose, compilations of olfactory profiles contain valuable information that may lead to certain consensus in odor classification.

Methods One of the most comprehensive odor databases is the Dravnieks' Atlas, which contains quantitative olfactory profiles for 160 samples. For each one, a large panel rated the applicability of 146 odor character descriptors on a numeric scale.

Results By applying principal component analysis to this Atlas, 105 descriptors were reorganized in 24 classes and 33 attributes were considered as odors intermediate of two or three categories. The similarities between classes were studied by means of a further multivariate analysis based on latent variables, which provides valuable information about the most salient dimensions of odor space.

Conclusions Consistent with other reported statistical analyses of olfactory databases, the perceptual space of odor character is multidimensional with about 20-30 dimensions, and it is better described as a continuum spectrum rather than as a segmented space.

Implications Attempts to classify all possible odor descriptors in a restricted number of classes appear to be inappropriate. Instead, 24 categories of related terms are proposed here, regarding the rest as intermediate smells, assuming that olfactory classes are not independent and follow certain hierarchy according to particular underlying dimensions.

Keywords: odor classification, sensory analysis, olfactory psychophysics

INTRODUCTION

Smells are difficult to describe, particularly for naive people. The most common approach consists of assigning the words that come to mind when smelling a substance, by describing how the perception resembles other common odors previously encountered and remembered (Stevenson and Boakes 2003). These terms are called odor character descriptors, attributes or notes, and the most useful ones in fragrance chemistry are those associated with the source of that smell. This procedure allows for a rapid generation of semantic olfactory profiles (Harper et al. 1968a), but training and some experience is required in order to assign the same words in the same way. The use of a panel is recommended because, otherwise, results can be influenced by personal subjectivity (Wise et al. 2000).

Psychophysical Studies about Classification of Odor Descriptors

A large amount of words can be applied to describe a smell, but not all of these terms are independent and it is possible to establish groups of descriptors often applied together. Many odor classification systems have been proposed, but none of them has yet gained wide acceptance. According to Sell (2004), the classification of smells is arbitrary because there are no agreed universal standards for odor character, all descriptors are associative, and apparent agreement can be misleading. A reported review of 28 perception-based classification studies of smells has discussed possible reasons to explain why these works have not yielded a generally accepted classification system yet (Kaeppler and Mueller 2013).

Few comprehensive semantic odor databases have been published. One study compiled 1396 pure substances from the handbook of Arctander (1969), the Fenaroli's manual (Burdock 2004) and other sources, resulting a database with 135 attributes which led to a descriptive model of olfactory perception space (Jaubert et al. 1986, 1987). Chastrette et al. (1988) selected 2467 compounds from the Arctander's handbook, and the use of different multivariate methods allowed a classification of 60 descriptors in 27 clusters, each containing 2–4 terms, while 14 remained as isolated notes. A similar study analyzed 1573 compounds from this database, and 126 attributes were classified in 19 categories by means of cluster analysis (Abe et al. 1990). Significant differences have been reported between Arctander's handbook and another commercial compilation of semantic odor profiles denoted as PMP-2001 (Pintore et al. 2006). Martínez-Mayorga et al. (2011) studied semantic descriptions of 4181 odorants taken from the Flavor-Base database. A further research study using this compilation has analyzed 251 odor notes for a set of 3508 molecules (Tromelin et al. 2017).

Firmenich S.A. used a panel of 7 perfumers to compile odor qualities of 628 pure compounds according to a restricted list of 32 descriptors, and a subsequent multivariate analysis classified these terms into 10 clusters (Chastrette et al. 1991). The commercial directory published by Sigma-Aldrich Fine Chemicals Company (2005), which will be referred to hereafter as SAFC catalog, contains semantic profiles for over 1700 materials. From this compilation, a reported study selected 278 descriptors for 851 compounds, and approximately 32 dimensions were identified using multidimensional scaling (Madany-Mamlouk et al. 2003; Madany-Mamlouk and Martinetz 2004). A subsequent analysis of the SAFC catalog classified 82 odor attributes into 17 classes (Zarzo and Stanton 2006).

Quantitative Odor Profiles

One method to characterize the differences between a set of odorants consists of assessing the perceptual similarity estimates of all possible pairs of the stimuli (Ravia et al. 2020). Another way consists of rating the odor similarity by direct comparison to a series of reference odorants (Schutz 1964; Wright and Michels 1964; Yoshida 1975). Amoore and Venstrom (1967) obtained the smell similarity to seven standards for a set of 107 aroma chemicals. Randebrock (1985) compiled olfactory profiles for 120 samples according to 29 descriptors. Using this method, Boelens and Haring (1981) asked a panel of 6 perfumers to smell 309 compounds and rate their similarities to 30 reference materials; this compilation will be referred to as B-H database. The subsequent application of principal component analysis (PCA) and factor analysis allowed for the classification of compounds into 14 clusters, though 27 groups were established in a further cluster analysis of the same database (Ennis et al. 1982). Relationships between 30 odor descriptors were discussed by Zarzo and Stanton (2009); this approach provides reliable information about the underlying dimensions in olfactory perception space, but it is restricted to a reduced number of references, since it would otherwise become time consuming and impractical.

An alternative method consists of smelling a substance and rating on a numeric scale the applicability or suitability of those descriptors that best apply from a fixed list. This approach avoids the subjectivity of open semantic descriptions observed in Arctander's and Fenaroli's handbooks. Based on this method, Harper et al. (1968b) obtained odor profiles for 53 stimuli according to the applicability of 44 qualities. In another work, observers rated each of 15 odorants according to 17 descriptors (Moskowitz and Gerbers 1974). Brud (1986) generated odor profiles based on 10 attributes, with applicability values from 0 to 4. In another sensory experiment, three panels rated the applicability of 11 descriptors, on a numeric scale, for a set of 40 odorants. A multivariate analysis of this dataset led to the

identification of an underlying construct that discriminated edible (i.e., odors recalling foodstuffs) vs. inedible smells (Zarzo 2008).

Dravnieks (1982) used a comprehensive list of 146 attributes, which requires more time than semantic profiles and may seem rather "noisy" because no reference materials were provided for each descriptor. Nonetheless, experimental results have shown that average odor profiles tended to stabilize using a large number of panelists. Reference-odorant methods seem *a priori* more accurate, but an experiment conducted with 49 subjects revealed that this approach had very good reproducibility (Dravnieks et al. 1978). In further research (Dravnieks 1985), a panel of about 120 individuals assessed 160 samples by scoring the applicability of each of the 146 descriptors, resulting in a database denoted as Dravnieks' Atlas. Different reported works have studied this database by applying PCA (Koulakov et al. 2011) or non-negative matrix factorization (NMF) (Castro et al. 2013). Keller and Vosshall (2004) have proposed a semantic classification of all Dravnieks' descriptors into 11 categories: *materials*, *chemicals*, *outdoors*, *fruits*, *foods*, *spices*, *foul*, *common*, *meats*, *vegetables*, and *body*. However, some of these classes comprise very different odor qualities (e.g., *fish*, *oily* and *celery* were included in the "meats" category).

Since the number of chemicals in Dravnieks' Atlas is somewhat limited, additional compounds were evaluated using the same descriptors with a panel of about 20 individuals (Jeltema and Southwick 1986), resulting in a compilation of 415 odorants. A test of reproducibility with 35 odorants showed a good correlation between Dravnieks' Atlas and the results from this reduced panel. A factorial analysis of this database allowed the classification of 119 descriptors into 17 clusters, though 10 attributes were considered in between two odor classes. This novel idea of regarding some descriptors as intermediate of two categories, which has received little attention, is consistent with the notion that our perceptual space of smells is not strictly segmented in well-defined olfactory categories. However, 10 intermediate descriptors out of 146 seems a low proportion. Most of the proposed clusters grouped semantically related terms, but some categories enclosed rather different odors (e.g., "animal - foul" comprised descriptors like *rancid*, *sour milk*, *animal*, *urine*, *blood*, *cheesy*, and *oily*). Another arguable issue is that *buttery* and *chocolate* were grouped together, which seems odd because their smell is clearly different. Likewise, *cooked meat* and *yeasty* were classified as "nutty", which is disputable.

Given these limitations of the reported classifications of Dravnieks' attributes (Jeltema and Southwick 1986; Keller and Vosshall 2004), a further effort was conducted here attempting to organize the 146 descriptors in the Atlas in meaningful categories in accordance with previous works about odor description and classification. For this purpose, a multivariate analysis of the Atlas was carried out in order to discuss the correlation structures between descriptors (i.e., odor descriptors normally assigned together in olfactory profiles), as a basis to establish clusters of related attributes and to categorize the rest as intermediate odors.

The classification of Dravnieks' list, which is constrained by the relatively reduced number of odorous stimuli contained in the Atlas, is of interest to better understand the number of perceptual dimensions of odor character. This issue is still a matter of debate because many studies assume about 20-30 dimensions (Chastrette et al. 1988; Abe et al. 1990; Madany-Mamlouk et al. 2003; Zarzo and Stanton 2006), but other authors postulate a lower dimensionality (Secundo et al. 2014). Another objective was to study the relationships between the resulting odor classes as a basis to better understand the most salient dimensions of odor space.

MATERIALS AND METHODS

Description of Dravnieks' Atlas

The present paper re-analyzes the olfactory Atlas compiled by Dravnieks (1985). Some information about this database is provided next because it is important to keep in mind how the sensory profiles were collected in order to properly discuss the results. This Atlas comprises 160 odorous stimuli (138 pure chemicals and some mixtures). Samples with a stronger smell were diluted and assessed at a lower concentration. Data were collected from 120–140 subjects. The panel was provided with a set of 146 descriptors, which will be referred to as the ASTM odor list because it was selected from a collection of 830 terms gathered by the Sensory Evaluation Committee of the American Society for Testing and Materials (ASTM) (Dravnieks et al. 1978; Dravnieks 1982). Each participant smelled the sample and assigned for each descriptor the score that best characterized the "degree of presence" of that odor note in the sample: 0 (absent), 1 (slightly), 2, 3 (moderate), 4 or 5 (extremely).

Numeric profiles contained in the Atlas can be arranged as two different matrices, so that the first step is to discuss which one seems more appropriate for the purpose of the present work. For each attribute and each odorant, Dravnieks (1985) calculated two values: the *average score* (AS) and the percentage of subjects who used the term, referred as *percentage of usage* (PU). Both parameters turned out to be strongly correlated (r = 0.988) because higher average scores correspond to those terms selected by most participants. In the case that two descriptors were assigned the same AS and very different PU percentages, it would make sense to use a higher corrected value for the descriptor most frequently chosen. Such correction was achieved by computing the geometric mean between the average score on a 0–100 scale (i.e., 20·AS) and PU percentages according to Equation 1, which was denoted as *percent applicability* (PA) (Dravnieks et al. 1978). Dravnieks (1985) assumed that such ratings were more reliable than AS, which justifies why PA values are displayed in the Atlas for each sample and each descriptor, instead of the average score. Nonetheless, in case of small panels, it might be more convenient to work with AS.

$$\mathrm{PA} = \sqrt{20 \cdot \mathrm{AS} \cdot \mathrm{PU}} \tag{1}$$

Data Pretreatment

Each page of the Atlas displays, for the list of 146 descriptors, the PU and PA values for one odorous sample. For clarity purposes, the highest PA values of each page are highlighted by means of a bar-chart built with asterisks, appearing next to the numeric value. Each asterisk accounts for 4% of the range, so that they appear next to descriptors with PA ≥ 4. In certain sense, these asterisks could be regarded as a simplified profile assuming implicitly a null score when PA < 4, which is intuitively appealing because the same threshold of significance was reported by Callegari et al. (1997) after comparing different olfactory databases. A similar process of setting the smallest values to zero was also applied by Boelens and Haring (1981) for odor aspect strengths compiled from a team of perfumers.

The following examples are proposed to discuss the random error of the panel. Pyridine, which yields the highest ratings of *heavy* (PA = 39.5) was also described as slightly *light* (PA = 2.8). This result is weird because both terms are semantically opposed, which suggests that the low value rated for *light* might be regarded as the random error of the panel. Thus, a null score could be considered for *light*. Conversely, limonene (PA_{light} = 32.1), which is the 5th highest value of this descriptor, was also rated as somewhat *heavy* (PA = 3.6). Another example is the case of spearmint oil, described as *cool* (PA = 46.0, 4th highest value)

and *warm* (PA = 3.1), which is contradictory because a smell clearly perceived as *cool* cannot be also described as slightly *warm*, given the antagonistic semantic character of both terms.

After scanning all pages of the Atlas, all PA values were digitalized by means of an OCR software and arranged as a "PA matrix" (raw unfiltered data) containing 160 observations (odorant samples, in rows) by 146 variables (odor qualities, in columns). Attempting to reduce this "noise" (random variation) in Dravnieks' Atlas, new variables were calculated by setting to zero those PA values below 4, based on the results of Callegari et al. (1997). To truncate something means to cut part of it off; hence, the new PA variables were referred to as "truncated" and denoted as $PA_{>4}$ because only values ≥4 are contained, being zero otherwise. This operation, which was carried out by means of an Excel spreadsheet by implementing the condition: if $x < 4$ then $x = 4$, leads to another dataset that was referred to as "$PA_{>4}$ matrix". In the example mentioned of spearmint oil, the "corrected" ratings for *cool* and *warm* would be: $PA_{>4}$ = 46.0 and 0, respectively.

In order to check if the threshold value PA=4 was the most appropriate, a new "$PA_{>5}$ matrix" was obtained by setting to zero all data of PA < 5 and, finally, another "$PA_{>3}$ matrix" was arranged likewise with the threshold PA=3. Thus, starting from the initial "PA matrix", three new ones were computed containing truncated variables, but only one of them had to be chosen for the detailed multivariate analysis. Attempting to decide which ones leads to a clearer identification of clusters, each matrix was analyzed by applying PCA. The contributions of variables in the formation of a given component are called loadings, being p[1] the loadings in the formation of the first principal component (PC1), p[2] are the loadings corresponding to PC2, etc.. The scatterplot of p[2] vs. p[1] will be referred to hereafter as PC1-2 loading plot, and so on for the rest of components. The loading plots corresponding to successive components (i.e., PC1-2, PC3-4, PC5-6...), for each of the four datasets (i.e., PA, PA>3, PA>4 and PA>5) were visually checked in order to identify clusters of correlated descriptors. The matrix leading to clusters more compact (i.e., with variables appearing closer to each other) was chosen and the others were disregarded.

Descriptive Statistics

Once the most suitable matrix was decided, a descriptive statistical analysis was conducted using two parameters: the summation of scores for each row of the matrix (SS_i: Sum of Scores for a given odorant), and the column averages ($\bar{x}_j$: odor descriptor average). Next, for each descriptor, the set of 10 variables yielding the strongest correlation were identified (i.e., highest values of Pearson's correlation coefficient) and they were sorted in a decreasing order. This information was used to interpret the similarities of particular attributes as a complement to the multivariate analyses. In this case with 160 observations, the correlation becomes statistically significant ($\alpha = 0.1\%$) if $r > 0.28$.

Principal Component Analysis

PCA was carried out by means of the software SIMCA-P 10.0 (www.umetrics.com). Variables were autoscaled (i.e., mean-centered and scaled to unit variance) prior to analysis. Loading plots were visually examined in order to detect clusters of similar descriptors. Once a group was clearly identified, it was regarded as an odor class if such category was consistent with similar olfactory studies. It might be argued that the visual inspection of loading plots aimed at deciding odor clusters is somewhat arbitrary, taking into account that such examination was just carried out by the author. Nonetheless, this was not the only criterion, since semantic similarities were also taken into consideration (e.g., *fried chicken* and *oily-fatty* are associated because, obviously, oil is used for frying). When it was not possible to identify

any other cluster due to the weak correlation of the remaining variables not classified yet, those descriptors from the clusters already established were set aside and a new PCA was fitted in order to discuss further correlation structures. The procedure was repeated until all attributes were cataloged. The same methodology was successfully applied to classify 82 odor attributes contained in the SAFC catalog (Zarzo and Stanton 2006).

In summary, the approach was to refit PCA models again and again after removing those variables classified in the previous models. A set of descriptors was regarded as an olfactory category according to three criteria: they should be positively correlated, semantically associated, and the classification should be consistent with previous studies. In case of descriptors with a doubtful classification due to a weak correlation with the rest of variables and because the semantic relationships are ambiguous, their classification was discussed according to the most similar notes (i.e., by identifying those attributes yielding the highest correlation with that descriptor), and based on relationships reported in similar studies. If a term could not be assigned undoubtedly to any of the proposed classes, it was regarded as an intermediate odor.

Given the long list of descriptors to be classified, they cannot be presented in a single table and it was necessary to split the list in two parts. One criterion could be according to the hedonic character (i.e., one list for the terms referring to pleasant smells and another for the less enjoyable ones). Another option is to split the list in two groups: "edible" (i.e., odors recalling foodstuffs) or "inedible". This second option was chosen because such contrasting polarity has also been identified in other statistical analyses of olfactory profiles (Zarzo 2008), which is reasonable because one of the basic functions of olfaction that probably played a key role throughout evolution was to assess the edible character of goods.

Once an olfactory category was established for a group of descriptors, the representative latent variable (i.e., the most salient underlying construct) was computed by fitting a PCA with the terms contained in that class and by obtaining the projection of observations over PC1. These latent variables, one per category, were used in certain cases to discuss the proper classification of doubtful descriptors. Finally, once all attributes were catalogued, a further PCA was carried out with the matrix of latent variables, and the loading plots were checked in order to discuss the relationships between categories and to interpret underlying dimensions in olfactory perception space.

RESULTS AND DISCUSSION

Data Pretreatment

By comparing the PC1-2 loading plot from the "PA matrix" vs. the "$PA_{>4}$ matrix", it turns out that the former leads to points rather scattered, while the latter shows a concentration of about 50 descriptors close to the center (Suppl. Fig. 1), which facilitates the interpretation of results. Loading plots for PC5-6 and further PCs also revealed significant divergences between both matrixes, and clusters appeared better defined using the "$PA_{>4}$ matrix". This result suggests that the lower ratings in odor profiles might be regarded as random error of the sensory panel, and the procedure of setting to zero the lowest PA values seems to be effective for dealing with this noise.

Loading plots derived from the "$PA_{>4}$ matrix" were visually compared with those from the $PA_{>5}$ and $PA_{>3}$ matrixes, and similar results were obtained. Actually, p[1] loadings from $PA_{>4}$ are strongly correlated with those from $PA_{>3}$ (r = 0.994) and $PA_{>5}$ (r = 0.997). The same occurs if p[2] loadings from $PA_{>4}$ are compared with those from $PA_{>3}$ and $PA_{>5}$ (r = 0.993).

Nonetheless, clusters seemed slightly more compact using the "$PA_{>4}$ matrix". Interestingly, Callegari et al. (1997) obtained the same threshold of significance after comparing different olfactory databases, and the bar-charts depicted in the Atlas to highlight the relevant PA values also assume implicitly the same 4% threshold. The following six $PA_{>4}$ descriptors yielded null values for all odorants and they were disregarded: *apple*, *laurel leaves*, *beery*, *rope*, *fresh eggs*, and *soupy*.

When all variables are measured in the same scale as it is the case here, they are usually mean-centered prior to applying PCA by subtracting the mean odor profile, so that all centered variables have a null average. By using such data pretreatment, PCs are influenced by variables with highest variance, which are those with greater mean in this case given the correlation between variance and average of $PA_{>4}$ variables (r = 0.82). It was observed that the 10 variables with highest average were the ones appearing more distant from the center of the PC1-2 plot, allowing for the identification of those descriptors most frequently chosen by panelists. However, for the purpose of properly discussing the similarities and dissimilarities between descriptors, all variables were mean-centered and scaled to unitary variance prior to PCA in order to prevent components from being influenced by those variables with highest variance.

Descriptive Statistics

After discarding the six $PA_{>4}$ variables mentioned with all null values, $\bar{x}_j$ ranges from 0.03 to 23.7 and follows a positively skewed distribution that can be modeled as log-normal. Descriptors yielding the highest average (indicated in parentheses) are the following: *fragrant* (23.7), *aromatic* (20.7), *sweet* (18.7), *heavy* (15.1), *light* (14.1), *woody-resinous* (12.9), *sickening* (10.8), *sharp-pungent-acid* (10), *perfumery* (9.6), *chemical* (8.9), and *warm* (8.6). Except for *woody-resinous*, these terms refer to rather general and unspecific smells. The lowest mean (0.03) corresponds to *burnt milk* and *chalky*, with all null data except for one odorant. The classification of the former descriptor is possible because there are other "burnt" attributes, but it becomes uncertain for the latter and it was disregarded.

The parameter SS_i (i.e., sum of scores for a given odorant) ranges from 141 to 882 with an average of 365, and it follows approximately a normal distribution. This wide range suggests that the number of descriptors applied by the panel for a given sample differed considerably, perhaps because some odors were very specific and matched with few descriptors, while in other cases the panel found more attributes in the list applicable for the smell, resulting in a higher SS_i.

Identification of the Highest Correlation Coefficients

An unexpected correlation coefficient r = 1.0 was found between *caraway* and *cheesy*, and it turned out that the original PA and PU values of both descriptors were identical for all samples, which evidences an error in the Atlas. Such variables yield the highest correlation with: *minty-peppermint*, *anise-licorice*, *cool*, *tea leaves*, *dill*, and *black pepper*, which are odors related with *herbaceous*. Thus, it was assumed that the values for *caraway* were correct, and they were probably also assigned by mistake to *cheesy*. Consequently, the latter was discarded from further consideration. Coincident profiles were also found between *soupy* and *apple*, which were also given exactly the same PA and PU values, but neither of them was included in the present study because $PA_{>4}$ scores are null for both descriptors.

Model 1: PCA with 138 Variables

After discarding *chalky*, *cheesy* and the six null $PA_{>4}$ variables, PCA was applied to the remaining 138 descriptors. The software SIMCA-P computes the amount of variance explained by cross-validation (Q^2). Components with a positive Q^2 are often assumed to account for the relevant information, which basically applies for components up to PC9 in this case (Suppl. Table 1). Similarly, Castro et al. (2013) focused on 10 basic factors. Nonetheless, the software is more restrictive and recommends focusing the attention on components with $Q^2 \geq 0.013$, which is fulfilled up to PC5. This result was somewhat unexpected because the dimensionality of odor space is assumed to be about 20-30 (Chastrette et al. 1988; Abe et al. 1990; Madany-Mamlouk et al. 2003; Zarzo and Stanton 2006), though some authors postulate a lower dimensionality (Secundo et al. 2014). The reason might be the limited number of odorous stimuli, which does not allow for a detailed characterization of the broad olfactory spectrum. Actually, some odor qualities are just represented by a few samples. Another criterion commonly used is to focus on those PCs with an eigenvalue >1, which is satisfied by 34 PCs. By carefully inspecting the loading plots for PC9-10, PC11-12, and so on, it was found that components beyond PC14 reflected relationships between descriptors quite difficult to interpret (i.e., not associated semantically and not consistent with the literature). The summary overview of components up to PC15 is shown in Suppl. Table 1. The loading plots corresponding to PCs up to PC15 were visually inspected and, based on those descriptors appearing at the extremes (i.e., by looking at their labels) and according to the discussion of the results, an interpretation is indicated in the last column of Suppl. Table 1. Such interpretation is preliminary or tentative in the sense that some of the underlying factors cannot be properly regarded as a polarity of two opposed odor categories.

Different studies of the Atlas have interpreted PC1 as the hedonic dimension (Zarzo 2008, 2011; Koulakov et al. 2011; Castro et al. 2013; Secundo et al. 2014) because it discriminates pleasant vs. unpleasant descriptors. Thus, p[1] loadings could be regarded as hedonic scores of the descriptors. Such values, which are indicated in Tables 2–3 (H_{sc}), are correlated (r = 0.74) with the hedonic tones obtained experimentally from a large panel (Dravnieks 1984). PC1 explains 13.2% of the data variability (Suppl. Table 1), but a higher value (≈30%) turns out working with centered variables, as reported by Secundo et al. (2014). However, the pretreatment applied here is more convenient for the purpose of discussing the correlation structures among variables.

Other studies have also reported that pleasantness is the most salient dimension when a wide range of odors is assessed (Woskow 1968; Schiffman 1974; Coxon et al. 1978; Davis 1979; Licon et al. 2018). Khan et al. (2007) selected nine odorants from the Atlas, and it was found that the pairwise distance between two odorants along PC1 was correlated with the pairwise distance in odor pleasantness perceived by a sensory panel.

Terms referring to the most unpleasant odors appear in the PC1-2 loading plot close to each other, and one cluster could be established with them. However, they account for rather different smells like *sour-vinegar, putrid, sweaty,* or *urine*. As the target of the present study was to group descriptors based on odor character, the classification of these unpleasant terms will be discussed in further PCA models once other clusters are identified.

PC2 could be regarded as a "chemical" dimension because the highest p[2] loadings correspond to odors sharing certain chemical note. By visually checking different loading plots for PC2 and further components, it was found that the plot for PC2-13 revealed a compact cluster formed by chemical-solvent descriptors (Fig. 1), while PC13 discriminates those with a burnt-smoky note (p[2] > 0.07, p[13] < -0.1) with respect to woody or

camphoraceous terms (p[2] > 0.07, p[13] > 0.04). A reported classification of Dravnieks' attributes (Keller and Vosshall 2004) has also proposed a group called "chemicals" which basically accounts for the same terms revealed by PC2.

In the cataloging of olfactory terms reported by Urdapilleta et al. (2006), odors recalling objects were divided into three categories: "nature", "food" and "civilization". The first two ones do not seem to be salient dimension in the Atlas, but the latter can be matched with the cluster reflected by Fig. 1. Regarding "civilization" smells, the authors subdivided them as: body/bathroom products (e.g., soap, shampoo, shower gel), skin products, anesthesia at the dentist's, pharmaceutical/hospital (e.g., alcohol, antiseptic, cough syrup), and house/indoor (e.g., cleaning products for dishwashing, softener, or animal repellent). This classification fits reasonably with the cluster highlighted in Fig. 1, which suggests that odors described as *chemical* evoke some "artificial" note. Our sense of smell has evolved to recognize odors encountered in natural environments, so that non-natural odorants typical of human civilization are perceived as chemical. Accordingly, Keller and Vosshall (2016) found that *chemical* is a generic descriptor applied frequently to unfamiliar odorants, which tend to be rated as neither pleasant nor unpleasant. Hence, PC2 could be labeled as a chemical vs. natural dimension of odor character.

[FIGURE 1 ABOUT HERE]

A "chemical-solvent" class (Table 1) was set up with the 10 descriptors grouped in Fig. 1 except *medicinal*, which was considered between two classes as described below. Another study using the ASTM odor list proposed a "solvent" cluster like in the present work, and *medicinal* was categorized between this class and *minty* (Jeltema and Southwick 1986). This "solvent" cluster included *turpentine, tar,* and *metallic*, but such terms were regarded here as intermediate.

The descriptor *etherish-anesthetic* appears within the "chemical-solvent" cluster in Fig. 1 and it is the one with highest loading in PC2. However, different studies have found a similarity between *ethereal* and *fruity* (Chastrette et al. 1988, 1991; Zarzo and Stanton 2006), but such descriptors are dissimilar in the Atlas (r = -0.19). Zwaardemaker (1925) proposed "ethereal" as an odor class in place of "fruity".

[FIGURE 2 ABOUT HERE

PC5 discriminates fruity vs. perfumery descriptors (Fig. 2), which appear as distinct categories. The former ones are perceived as "edible", while perfumery scents could be regarded as "inedible". According to Harper et al. (1968a), the fact that pleasant aromas in perfumes are not acceptable in foods was first proposed in classical antiquity by Theophrastus (born 370 BC). Based on the loading plot for PC5-9 (Fig. 2), a cluster was established for all non-citrus fruity descriptors (Table 2). *Banana* is located closer to the center of this plot because the sample rated with highest scores for this attribute (isoamyl acetate) smells like solvent. The strongest similarity of *cantaloupe melon* corresponds to *cucumber*, which indicates that the panel regarded the former as less fruity with certain vegetable note, which is intuitively appealing because cantaloupe (*Cucumis melo* var. *reticulatus*) and cucumber (*Cucumis sativus* L.) are plants of the same botanical family (Cucurbitaceae). The lower fruity character of *melon* was also found in the SAFC database (Zarzo and Stanton 2006).

Fruity terms yield the highest positive p[5] loadings, but PC9 discriminates between citrus and non-citrus fruits (Fig. 2). Accordingly, a "citrus" cluster was set up with *citrus, lemon, orange,* and *grapefruit* (Table 2). *Citrus* and *fruity* are somewhat correlated in the

Atlas (r = 0.40) but not in the B-H database (r = 0.06, p = 0.26), which justifies that several authors have established each of them as independent classes (Rimmel, 1895; Jennings-White 1984; Jeltema and Southwick 1986; Zarzo and Stanton 2006). Chastrette et al. (1991) found a similarity between *citrus* and *herbaceous*, while *fruity* resulted related to *ethereal*. Abe et al. (1990) partitioned the "citrus" cluster in two groups: *lemony* was regarded as more herbaceous and fresh, while *orange*, *mandarin*, and *orange-blossom* were described more similar to *fruity* and *floral*. This difference between *lemon* and the rest of citrus terms was not identified here given the compact citrus cluster observed in the PC10-11 plot (Fig. 2).

[FIGURE 3 ABOUT HERE]; [TABLES 1, 2 ABOUT HERE]

A direction of data variability is observed in Fig. 3 (p[5] < -0.1, p[7] < -0.05) which could be denoted as "perfumery" dimension because it accounts for descriptors commonly applied to describe fragrances. *Perfumery* and *cologne* were classified as general/unspecific terms (Suppl. Table 2) because they relate to more than one odor class. The same criterion was adopted for *soapy* because soaps can be scented with different smells and this attribute yields the highest correlation with *cologne* (r = 0.44) and *perfumery* (r = 0.43).

The attributes *floral*, *rose* and *violets* were grouped as a "floral" category (Table 1) given their semantic relationship and proximity in Fig. 3. Although *rose* and *violet* have been considered as independent odor classes (Rimmel 1895; Willis 1944; Cerbelaud 1951; Jennings-White 1984), they are strongly correlated in the Atlas (r = 0.90) probably because panelists were not experts in the characterization of flowery scents. *Rose* appeared as the best representative of floral terms in the SAFC catalog (Zarzo and Stanton 2006), and rose oil was the preferred reference for *floral* among perfumers (Brud 1986). By contrast, *violet* yielded certain similarity with *leafy* in the Arctander database (Chastrette et al. 1988), and it was classified in between of *floral* and *green* (Zarzo and Stanton 2006).

Flowery scents are commonly encountered in women's fragrances, while lavender notes are typical of the fougère category, which is the most successful family of men's perfumes (Zarzo and Stanton 2009). In the Atlas, *lavender* yields the highest similarity with *floral* (r = 0.84) and *cologne* (r = 0.82), which justifies why Jeltema and Southwick (1986) incorporated *lavender* within the "floral" cluster. However, this attribute was classified by Abe et al. (1990) within the "herbaceous" category with other notes like *clary-sage*, *hay* or *tea*. Consistent with this criterion, *lavender* yields a moderate correlation with *vegetable* (r = 0.23, p < 0.0001) in the B-H database, but not with *floral* (r = -0.12, p = 0.035). Therefore, *lavender* was considered as intermediate of "floral" and "herbaceous" (Table 3).

[TABLE 3 ABOUT HERE]

Musk and *dry-powdery* were grouped as another class because they appear next to each other in Figs. 2–3, and such similarity has also been reported by Chastrette et al. (1991), who grouped together *musk*, *powdery*, *animal*, and *phenolic*. The strongest correlation of *musk* corresponds to *cologne* (r = 0.54), which explains its proximity to the "floral" cluster (Figs. 2–3), though they are distinct smells. A similarity between *musk* and *animal* has been reported in the Arctander database (Chastrette et al. 1988), but such correlation is not significant in the Atlas (p = 0.13). Incense also appears close to these descriptors due to its correlation with *perfumery* (r = 0.60). Setting aside the unspecific terms, *incense* yields the highest similarity with *cedarwood*, *floral* and *musk* (r = 0.50, 0.46 and 0.45, respectively); hence, it was regarded as midway of such classes.

According to Fig. 3, PC7 basically discriminates spicy vs. perfumery scents. Spicy descriptors also appear chose to each other in the loading plot for PC13-14 (Fig. 4). Thus, a "spicy" category was established with *spicy, cinnamon, clove,* and *tea leaves* (Table 2). Such cluster comprised by *spicy, cinnamon* and *clove* was also proposed by Jeltema and Southwick (1986). Different authors have considered cinnamon as the best representative material for *spicy* (Jennings-White 1984; Abe et al. 1990). However, a reported survey revealed that perfumers preferred clove bud oil and eugenol rather than cinnamon bark oil as a reference for *spicy* (Brud 1986). Boelens and Haring (1981) chose eugenol as standard for this attribute.

Black pepper was also included in this "spicy" category (Table 1) because this material is one of the most common spices added to cuisines around the world. This descriptor yields the highest correlation with *seasoning (for meat)* (r = 0.31) and *garlic-onion* (r = 0.27). However, unexpectedly, the correlation with *spicy* is very weak (r = 0.19, p = 0.015) probably because only three samples in the Atlas presented this smell, which seems too low. The strongest similarity of *raisins* corresponds to *tea leaves, spicy, clove, cinnamon,* and *fruity*; hence, it was classified between "spicy" and "fruity" (Table 3).

PC4 discriminates balsamic vs. green odors, while PC8 could be labeled as a balsamic vs. sulfidic dimension (Suppl. Table 1). The SAFC catalog contains a "balsamic" category comprised by: *vanilla, chocolate, caramel, honey, balsam, sweet, cinnamon,* and *anise*. As most of them appear grouped together in the loading plot for PC4-8 (Fig. 4), a "balsamic" cluster could be defined with *vanilla, chocolate, malty, caramel, maple syrup, molasses, honey, coffee,* and *burnt milk*. However, the PC13-14 plot (Fig. 4) reveals clear differences within this group. Accordingly, a "balsamic-vanilla" class was established with *vanilla, chocolate* and *malty*, while another "balsamic-caramel" subcluster was set up with *caramel, maple syrup,* and *molasses* (Table 2). Although *cinnamon* is classified as balsamic in the SAFC catalog given its sweet character, it clearly appears as spicy in the Atlas (Fig. 4).

Honey is usually classified as balsamic, which agrees with Fig. 4, but different studies suggest certain floral character (Chastrette et al. 1988, 1991). Actually, *honey* and *rose* presented the highest correlation among all possible pairs of descriptors in the SAFC database (Zarzo and Stanton 2006). Given that *honey* yields the strongest similarity with *vanilla* (r = 0.44) but the correlation with *floral* is also statistically significant (r = 0.27, p = 0.0006), it was classified as midway of "balsamic-vanilla" and "floral" (Table 3). *Coffee* and *burnt milk* also appear within the balsamic cluster in Fig. 4, but both were regarded as intermediate odors as commented below because their smell is not so sweet.

By definition, *balsam* refers to various fragrant exudations from certain trees, like Peru balsam, which was the third preferred reference for *balsamic* in a survey conducted among perfumers (Brud 1986). This material smells balsamic, sweet, and reminiscent of vanilla (Rimmel, 1895). Accordingly, balsamic odors are often defined as *sweet, soft,* and *warm,* which agrees with the positive correlation between *balsamic* and *sweet* in the B-H database (r = 0.34) as well as in the SAFC catalog (Zarzo and Stanton 2006). Consistent with this criterion, *sweet* and *warm* appear close to the "balsamic" cluster in Fig. 4 due to their correlation with *vanilla* (r = 0.32 and 0.21, respectively). In the Arctander database, *balsamic* yields the highest similarity with *vanillin* (Chastrette et al. 1988), and this odorant has been considered as a standard for *balsamic* from long ago (Rimmel 1895; Zwaardemaker 1925). However, Brud (1986) found that only 4% of perfumers chose *vanillin* as a reference for balsamic scents (Brud 1986).

[FIGURE 4 ABOUT HERE]

The following 7 terms form a compact cluster in Fig. 4: *fresh green vegetables, crushed grass, crushed weeds, green pepper, raw potato, beany,* and *hay.* They account for the smell of raw vegetables and freshly cut grass, which is called as "green" in perfumery, and the same name was assigned here (Table 2). *Raw cucumber* was also regarded as "green", though it appears closer to the center given its similarity with the fruity attribute *cantaloupe*. The strongest correlation of *herbal-green-cut grass* corresponds to *crushed grass, crushed weeds* and *fresh green vegetables*. Hence, it was classified as "green", though this attribute might be misleading because *herbal* and *green* are generally considered as independent odor categories (Brud 1986; Chastrette et al. 1988; Abe et al. 1990; Zarzo and Stanton 2006).

Jeltema and Southwick (1986) also regarded *celery* as "green", but it yields the highest similarity with *molasses* and *maple syrup* because one odorant contained in the Atlas (celeriax®) smells like celery, sweet and maple-like. If this observation is removed, *celery* is mainly correlated with *green pepper* (r = 0.71) and *raw potato* (r = 0.67), which supports the classification of *celery* as "green". The descriptor *geranium leaves* was categorized between "green" and "floral" (Table 3) because it yields the strongest correlation with *crushed grass* (r = 0.52), *crushed weeds* (r = 0.51) and *rose* (r = 0.50). Such relationship with *floral* is also apparent in the SAFC catalog (Zarzo and Stanton 2006). *Cooked vegetables* and *musty-earthy-moldy* were also classified by Jeltema and Southwick (1986) as "green", but such attributes were regarded here as intermediate odors as discussed below.

PC6 could be regarded as a nutty vs. sulfidic polarity. The loading plot for PC6-8 (Suppl. Fig. 2) suggests that another cluster may be established with *garlic-onion, household gas, meaty (cooked), seasoning (for meat), sulfidic,* and *burnt rubber*. This group could be named "alliaceous" as proposed by different authors (Linnaeus 1756; Lovell 1923; Zwaardemaker 1925; Jennings-White 1984; Zarzo and Stanton 2006) because such odors share a common note of garlic (*Allium sativum* L.) and onion (*Allium cepa* L.). In order to further analyze the similarities between these descriptors, the set of 5 attributes yielding the highest correlation with each of them are indicated in Suppl. Table 3. *Sulfidic* appears in the first or second position, which implies that it seems to be the best representative of the group. This term refers to the molecules containing sulfur. Such compounds can be detected at very low concentrations and they are frequently added in household gas, which is odorless, in order to be detected in case of gas leaks, which explains the strong correlation of *sulfidic* vs. *household gas* (r = 0.85). Hence, both terms were regarded as an independent category (Table 1). *Burnt rubber* was considered in between of "burnt" and "sulfidic" (Table 3) because this descriptor yields the strongest similarity with *sulfidic* (r = 0.75), *household gas* (r = 0.72), and *burnt-smoky* (r = 0.53). The classification of other related attributes is commented below.

Model 2: PCA with 78 Variables

Except PC1, PC3, and PC12, the rest of components up to PC14 are depicted in Figs. 1–4 and Suppl. Fig. 2 (model 1, 138 variables). No clear additional clusters were identified by checking the loading plots with other components. Moreover, the large amount of descriptors in the plots makes it somewhat difficult to identify groups of correlated variables which are semantically related. Hence, the set of 60 descriptors already classified was set aside and a new PCA was fitted with the remaining 78 variables, attempting to have a clearer picture of the residual correlation structure. Based on the descriptors appearing on the extremes of each axis, PC2 could be labeled as burnt vs. light, while PC6 discriminates burnt vs. earthy scents. The loading plot for PC2-6 (Fig. 5) reveals a group of terms sharing a burnt or smoky note.

Such cluster could be named as "empyreumatic" or "pyrogenous", which is an odor category introduced long ago, giving toasted bread and roasted coffee as examples (von Haller, 1763). This class comprises a wide spectrum ranging from inedible to edible smells: *burnt, tar, fresh tobacco smoke, phenolic, leather, toasted, roasted coffee,* and *nutty* (Harper et al. 1968a). A reported analysis of Dravnieks' Atlas found that the 8th factor was related with these odors (Castro et al. 2013).

Creosote, tar and *leather* appear within this cluster in the PC2-6 plot, but PC3 discriminates them with respect to the other empyreumatic attributes probably because such smells present certain chemical note, as reflected by Fig. 1. Accordingly, a "burnt" class was established with: *burnt-smoky*, *burnt paper*, *burnt candle, stale tobacco smoke,* and *sooty* (Table 1). Other authors have also considered "burnt" as an independent category (Henning 1916; Crocker and Henderson 1927; Cerbelaud 1951; Jennings-White 1984; Jeltema and Southwick 1986). *Creosote* is the generic name of a variety of chemical mixtures obtained by distilling wood tar or coal tar, which explains the correlation with *tar* (r = 0.84). Both descriptors were considered in between the "chemical" and "burnt" categories (Table 3), given the similarities with such clusters revealed by Figs. 1 and 5, respectively. *Leather* was regarded midway of "burnt", "chemical-solvent", and *medicinal* because it yields the strongest correlation with *creosote, tar, disinfectant,* and *medicinal*. A similar criterion was adopted by Abe et al. (1990). Given that *new rubber* is mainly correlated with *tar* (r = 0.52) and *leather* (r = 0.46), this descriptor was classified as midway of such smells (Table 3).

[FIGURE 5 ABOUT HERE]

Cork, coffee and *burnt milk* also appear in Fig. 5 as "burnt" attributes. The former yields the highest correlation with *sooty* (r = 0.50), *burnt candle* (r = 0.40), *leather* (r = 0.39), and *oakwood* (r = 0.36). Given that cork refers to the layer of dead tissue below the bark in woody plants, such burnt character is arguable and it was ignored. Hence, *cork* might be regarded as midway of *leather* and "woody" (Table 3), which partly agrees with the classification of *cork* as "woody" proposed by Jeltema and Southwick (1986). Abe et al. (1990) found that *coffee* was similar to *burnt* given its toasted smell and it was classified with other terms like *caramelic, garlic*, *onion,* and *sulfureous*. This attribute was regarded as intermediate of "balsamic-caramel" and "burnt*"* (Table 3) because it yields the strongest correlation with *burnt milk, maple syrup,* and *caramel*. The same categorization was decided for *burnt milk* because both descriptors appear close to each other in Figs. 4–5.

Regarding *fresh tobacco smoke*, it appears in Fig. 5 very close to the center because it only presents one datum different to zero. Given that Jeltema and Southwick (1986) categorized this descriptor within the "nutty" cluster, it was classified as midway of "burnt" and "nutty" (Table 3), though this criterion might be arguable.

PC3 discriminates nutty vs. medicinal odors, while PC4 could be labeled as *sulfidic* vs. *woody*. The loading plot for PC3-4 (Suppl. Fig. 3) reveals a group of three descriptors with an edible character sharing a sulfidic note (Suppl. Table 3): *garlic-onion, seasoning for meat,* and *meaty-cooked*. Their similarities are appealing because garlic and onion, which contain sulfurous odorants, are commonly used as seasoning for meat. It was decided to consider *meaty (cooked)* and *seasoning for meat* as an independent category denoted as "cooked meat" (Table 2), while *garlic-onion* was regarded as midway of this class and "sulfidic". The latter is stronger correlated with *garlic-onion* (r = 0.78) than in the case of *meaty (cooked)* (r = 0.53), which supports this criterion. *Cooked vegetables* was classified as intermediate of the

categories "green" and "cooked meat" because it yields the highest correlation with their corresponding latent variables (r = 0.50 and 0.51, respectively).

Model 3: PCA with 61 Variables

Starting from the previous model with 78 variables, the set of 17 descriptors already classified was put aside, and a new PCA was fitted. PC1 discriminates pleasant vs. unpleasant descriptors. Particularly, it turned out that *fragrant, aromatic, light, and sweet* appeared with the highest p[1] loadings, i.e., as the most pleasant descriptors contained in the model (Fig. 6). Such variables are among the ones with highest average (Suppl. Table 2), which indicates that the panel rated them with a high frequency. They were included in the category of general/unspecific pleasant descriptors (Suppl. Table 2) because the differences within this group are not clearly deduced from the multivariate analysis.

"Aromatic" and "fragrant" were long ago proposed as two independent classes of pleasant odors, being the latter comprised by floral and balsamic scents (Linnaeus 1756; Zwaardemaker 1925). However, the strongest similarity of *aromatic* in the Atlas corresponds to *fragrant* (r = 0.85), and the latter is not correlated with the latent variable from the balsamic cluster (p = 0.45). This result evidences that panelists were not trained experts in odor description and classification. Nonetheless, floral scents are commonly encountered in fragrances, which would explain why *floral* yields a correlation with *fragrant* (r = 0.76) higher than with *aromatic* (r = 0.46).

PC3 could be labeled as a *camphoraceous* vs. *sweet* axis. The loading plot for PC1-3 reveals a cluster formed by camphoraceous and woody descriptors, but PC2 discriminates between them (Fig. 6). Likewise, Abe et al. (1990) grouped together *woody, pine, camphor,* and *minty*, but Chastrette et al. (1991) found that *woody* was slightly disimilar to *piney, camphoraceous* or *minty*. Jennings-White (1984) have also postulated "woody" as an odor class. Accordingly, a "woody" category was established with *woody-resinous, cedarwood, bark,* and *oakwood*. An independent "woody" class comprised by the same descriptors and *cork* was proposed by Jeltema and Southwick (1986). In the B-H database, *woody* yields the highest correlation with *dusty* (r = 0.62), *earthy* (r = 0.39), *balsamic* (r = 0.37), and *powdery* (r = 0.31). Likewise, *woody-resinous* is correlated in the Atlas with *dry-powdery* (r = 0.45) and *musty-earthy-moldy* (r = 0.43), though not with the latent variable of balsamic descriptors (p = 0.72).

According to Fig. 6, a "camphoraceous" cluster was established with *camphor, mothballs,* and *eucalyptus*. Although *turpentine* and *medicinal* appear very close to these terms, the former was classified as intermediate of "chemical-solvent", "camphoraceous" and "woody" given the correlation with the latent variables from those clusters (r = 0.67, 0.65 and 0.44, respectively). Analogously, *medicinal* was regarded as midway of "chemical-solvent" and "camphoraceous" because it yields the highest correlation with *etherish* (r = 0.81), *disinfectant* (r = 0.80), and *camphor* (r = 0.76). A similar criterion was adopted in the statistical analysis of the SAFC catalog (Zarzo and Stanton 2006), while Jeltema and Southwick (1986) regarded *medicinal* in between of the "solvent" cluster and *minty*.

Camphoraceous and *minty* are often grouped together (Jeltema and Southwick 1986; Chastrette et al. 1988; Mamlouk et al. 2003; Zarzo and Stanton 2006), but some authors have proposed each of them as independent classes (Rimmel 1895; Jennings-White 1984). *Minty-peppermint* and *cool-cooling* appear next to each other in Fig. 6 given their correlation (r = 0.82), which is appealing because *minty* is often used as a trigeminal cooling flavor. Both

descriptors were classified as intermediate of "camphoraceous" and "herbaceous" (Table 3) because the latent variable extracted from them yields the strongest correlation with the latent variable of such clusters (r = 0.52 and 0.36, respectively). Consistent with this criterion, Chastrette et al. (1991) found a similarity between *minty* and *hay* (i.e., dried grass).

Musty-earthy-moldy yields the highest correlation with the latent variables from the categories "green" (r = 0.52) and "woody" (r = 0.35), which agrees with the proximity to such categories in Figs. 4 and 6. Accordingly, it was categorized in between of these classes. A similar result is derived from the B-H database given the correlation of *earthy* with respect to *vegetable* (r = 0.57) and *woody* (r = 0.39). *Musty*, *earthy*, and *moldy* appear as a single attribute in the Atlas given their similarity (Zarzo and Stanton 2006), which is intuitively appealing because mushroom, mold or moss are developed in wet environments, and some authors have proposed *earthy-fungoid* as an independent odor class (Klein 1947).

[FIGURE 6 ABOUT HERE]

Based on Suppl. Fig. 2, a "nutty" cluster was defined with *nutty*, *peanut butter*, *grainy*, and *popcorn*. *Almond* was classified as "nutty" by Zarzo and Stanton (2006), but such similarity was not reported by Jeltema and Southwick (1986), and Abe et al. (1990) grouped it with *bread*. Given that *almond* yields the highest similarity with *nutty* (r = 0.45), *cherry* (r = 0.45), *warm* (r = 0.41), and *vanilla* (r = 0.31), it was regarded in between of "nutty" and *vanilla* (Table 3). Such similarities are consistent with the position of *almond* in Figs. 5−7.

Different studies have considered *coconut* as a nutty odor (Chastrette et al. 1988; Jeltema and Southwick 1986; Mamlouk et al. 2003; Zarzo and Stanton 2006), but this term is not correlated with any descriptor in the "nutty" class. This result is conditioned by the presence of aldehyde C-18, which smells distinctly sweet. If this odorant is removed, *coconut* yields the highest correlation with *vanilla* (r = 0.39) and *almond* (r = 0.25), which is intuitively appealing because Jeltema and Southwick (1986) regarded *coconut* and *almond* as two independent related descriptors (Table 3). Thus, both terms were classified alike.

PC2 discriminates nutty vs. cool scents, while PC7 could be named as a buttery vs. warm polarity (Fig. 7). Accordingly, another category denoted as "buttery" was established with *buttery* and *bakery (fresh bread)*. Jeltema and Southwick (1986) grouped both of them with other balsamic odors; however, *buttery* is weakly correlated with *caramel* (r = 0.33) or *vanilla* (r = 0.25), and it appeared as a rather independent descriptor in the B-H database (Zarzo and Stanton 2009). Abe et al. (1990) found a similarity between *buttery* and *oily*, but such correlation is not significant in the Atlas (r = 0.18, p = 0.02). *Warm* appears very close to this "buttery" cluster (Fig. 7), but it was regarded as a general/unspecific term.

The descriptor *oily-fatty* is semantically related with *fried chicken*, which has only one value different from zero and, consequently, appears closer to the center of the plot. Some authors have suggested *oily* as an independent odor class (Jennings-White 1984), and Chastrette et al. (1988) proposed a cluster with *fatty*, *oily*, and *waxy*. Accordingly, *oily-fatty* and *fried chicken* were grouped as a "fatty" category (Table 2). The former is positively correlated with the latent variables from the classes "nutty" (r = 0.38) and "cooked meat" (r = 0.31), which agrees with the edible character shared by these smells.

Stale yields the highest correlation with *rancid* (r = 0.68) and *musty-earthy-moldy* (r = 0.64), which is interesting because *stale* is semantically related with both attributes and, hence, it was classified as intermediate of them (Table 3). The location of *stale* in Fig. 7

agrees with such similarities given the proximity to *musty-earthy-moldy*, and taking into account that rancid smells are produced by chemical oxidation of fats.

[FIGURE 7 ABOUT HERE]

Model 4: PCA with 33 Variables

After pulling apart the set of 28 descriptors just classified, a new PCA model was carried out with the 33 remaining variables. It was found that PC4 is basically determined by *cardboard* and *wet paper*, which are correlated descriptors (r = 0.60) because they share certain earthy note. The former was categorized in between of *musty-earthy* and "woody*"* because it yields the strongest similarity with *wet paper, musty-earthy-moldy, bark, cork,* and *woody-resinous*. The same criterion was adopted for *wet paper*.

PC1 reveals hedonic aspects, while PC3 could be labeled as bitter vs. nasty odors (Suppl. Fig. 4), and a cluster was highlighted formed by: *heavy, bitter, sickening, metallic,* and *sharp-pungent-acid.* The first three terms were grouped as "sickening" (i.e., nauseating) because their hedonic scores and tones are negative (Table 1), revealing a bad smell. The term *heavy* is semantically related with *intense*, but the former was applied to the most unpleasant smells, probably regardless of their intensity. It might be speculated that panelists perhaps tended to assign higher scores to the stronger smells, and different studies have reported that one intensive dimension (weak vs. strong odor sensation) is often salient in the description of odors (Woskow 1968; Davis 1979). However, there is not enough evidence about this issue in the Atlas probably because intense odorants were evaluated at a lower concentration. Moreover, panelists were requested to assess the applicability or "degree of presence" of the different descriptors, and not the "odor aspect strengths", a concept used in other studies (Boelens and Haring 1981) that might be associated with intensity.

Given that *sharp-pungent-acid* yields the strongest correlation with *sickening, heavy* and *sour-vinegar* (r = 0.73, 0.67, and 0.63, respectively), it was considered as midway of "sickening" and "sour". By contrast, Jeltema and Southwick (1986) incorporated this descriptor in the "animal-foul" cluster. These authors classified *metallic* within the "solvent" class, but it was regarded here midway of "sickening" and "chemical" (Table 3) because it presents the same correlation with *sharp-pungent* and *chemical* (r = 0.42).

Based on Suppl. Fig. 4, another cluster was established with *dill, caraway* and *anise (licorice),* which makes sense since these plants belong to the same botanical family (Umbelliferae). This category was denoted as "*herbaceous*" because these plants are commonly referred as "herbs" and used for flavoring or aromatic purposes. This is usually the interpretation of *herbaceous* in perfumery as an odor class independent to *green / vegetable*, though vegetables are also herbaceous plants (i.e., non-woody annual) from a botanical standpoint. *Anise* is listed in the SAFC catalog under the balsamic category due to its sweet character, but this descriptor is weakly correlated with *sweet* in the Atlas (r = 0.19, p = 0.019).

Mushroom yields the highest correlation with *dill* (r = 0.41), *crushed weeds* (r = 0.34), and *musty-earthy-moldy* (r = 0.32). Therefore, it was classified it as intermediate of *herbal-green* and *musty-earthy-moldy*, which makes sense because mushrooms are edible fungi. Chastrette et al. (1988) found a similarity between *mushroom* and *earthy*, while Abe et al. (1990) classified the former as herbaceous.

Model 5: PCA with 22 Variables

From the previous model with 33 variables, another PCA was fitted after setting aside those 11 descriptors just classified. It was found that *mouse* and *seminal (sperm-like)* yield the highest p[3] loadings, but an accurate classification is not possible because all their values are zero except for one odorant, which was primarily described as *sickening*, *heavy*, and *sweaty* (i.e., an unpleasant animalic smell). Thus, both descriptors were classified within the "animal" cluster, as described below, and the model was repeated after removing such variables.

PC2 discriminates fishy vs. sour scents. *Fishy* and *kippery (smoked fish)* appear next to each other in Suppl. Fig. 5 due to their strong correlation (r = 0.94). Accordingly, both were grouped as a "fishy" category. *Ammonia* was classified in between of "chemical-solvent" and "fishy" because it yields the highest correlation with the latent variables from such clusters (r = 0.37 and 0.36, respectively). This similarity with *fishy* has a chemical basis: the presence of amine groups in a molecule usually confers it a fishy odor, being trimethylamine the compound mainly responsible for this characteristic smell. *Urine* and *cat urine* were considered as intermediate of *animal* and *ammonia* because they appear in Suppl. Fig. 5 midway of both descriptors. This similarity is intuitively appealing due to the presence of urea (carbamide) in urine, which is decomposed in carbonic acid and ammonia when degraded.

Sauerkraut, sour milk and *sour-vinegar* are semantically related and present negative p[2] loadings. The former appears closer to the center because it contains non-zero values for only two samples. It was found that *yeasty* yields the highest correlation with *sour milk* (r = 0.74), and the same occurs with *fermented (rotten) fruit* (r = 0.82). Hence, a "sour" category was established with these five descriptors.

PC1 still reflects hedonic aspects. The most negative p[1] loadings correspond to *putrid-foul-decayed, rancid, fecal (like manure),* and *cadaverous*. This group comprises the most unpleasant smells, and it could be denoted as "fetid", which is an odor class long ago proposed (Linnaeus, 1756). Considering that these odors are basically related with degradation (e.g., rancid smells appear by chemical oxidation of oil and fats), this class was called "fetid-decay". *Sewer* was also included here because it refers to decomposition.

Another "animal" cluster was established with *animal, wet wool*, and *sweaty*. *Mouse* and *seminal* were also included here, as commented above, as well as *dirty linen* because odors reminiscent of human body are considered as animalic. Finally, *blood (raw meat)* was classified midway of "fishy", "fetid", and "animal" because it yields the highest correlation with the latent variables from these clusters (r = 0.63, 0.59, and 0.47, respectively).

In summary, the ASTM list of 146 descriptors was classified as following: 6 of them were removed for presenting all null values of $PA_{>4}$ variables, as well as *cheesy* and *chalky*; 8 were regarded as general/unspecific odors (*fragrant, aromatic, sweet, light, perfumery, cologne, warm*, and *soapy*), 97 were classified in 23 categories and 33 were considered as intermediate of two or three odors. Consistent with this result, Castro et al. (2013) found that a 25 dimensional representation of the perceptual data in Dravnieks' Atlas was the most accurate achievable with NMF. Callegari et al. (1997) reported that 25 well-chosen descriptors seem sufficient to represent faithfully the perceptual olfactory space. Another study has compiled semantic descriptions from publicly available odor databases comprising in total 3016 molecules and 526 perceptual descriptors, which were grouped in 17 clusters (Kumar et al. 2015). The fact that mixtures containing many (>20) diverse odorants tend to

smell alike, a phenomenon that has been termed 'olfactory white', suggests that the dimensionality of odor percepts may be around 20 or less (Meister 2015).

Study of Similarities Between the Proposed Odor Classes

In order to characterize the relationships between the 23 odor categories, a matrix was arranged containing 23 latent variables, one per class, as well as the unspecific terms *sweet* and *warm*. Each latent variable was computed by fitting a PCA with the descriptors included in each class, except those containing just one non-zero value, and obtaining the projection of 160 odorants over PC1. The percentage of data variance (R^2_X) explained by PC1 (Suppl. Table 4) indicates if the resulting latent variable was a good representative of the odor class (i.e., high values correspond to groups of descriptors strongly correlated). The R^2_X for "fatty" is missing because *fried chicken* was disregarded and, hence, no latent variable can be extracted.

The most independent categories will be those weakly correlated with the others, which correspond to "herbaceous", "green", and "fishy" (Suppl. Table 4). The former yields the highest correlation with *warm* (r = 0.18) due to the sweet-spicy character of *anise*, though most herbaceous materials present a cool-camphoraceous note. The independent character of "green" is arguable because it yields certain similarity with vegetable, citrus, watery, and refreshing scents (Zarzo and Stanton 2009). The latent variable from "green" has a weak correlation with "woody" (r = 0.20, p = 0.012) but not with *sweet* (p = 0.17). Conversely, in the B-H database, *green* yields a negative correlation with both descriptors (r = -0.14 and -0.40, respectively). These discrepancies might be partly explained by the lack of reference materials for each descriptor during the sensory profiling, or maybe due to the reduced number of "green" odorants in the Atlas (just four compounds yielded the highest ratings of *weeds*, *crushed grass*, and *herbal-green*).

Given the edible character of "cooked meat", its similarity with "sickening" (r = 0.38) is peculiar, but most odorants described as *meaty-cooked* or *seasoning for meat* were rated as *sickening* and *heavy*. The correlation between *fatty* and "sickening" (r = 0.43) was also unexpected, which might be explained by the fact that the odorant rated with highest fatty character (2-trans-4-trans-decadienal) was also described as heavy, rancid and sickening.

By applying PCA to the matrix of latent variables, the PC1-2 plot (Fig. 8) accounts for 29.5% of the data variance and reflects the main relationships between odor classes. It turns out that PC1 still accounts for hedonic aspects, given the correlation (r = 0.76, p < 0.0001) between p[1] loadings and the hedonic tones reported by Dravnieks (1984) (values in Suppl. Table 4). "Chemical" and "camphoraceous" present the highest loadings in PC3, which agrees with the close proximity of both latent variables to the plot center. It is noteworthy that the chemical factor is still salient despite of having just 25 variables in this multivariate model.

PC2 can be interpreted as a *warm* vs. *fresh* dimension because descriptors with p[2] > 0.2 have also been regarded as warm in other reported studies while, conversely, those with p[2] < -0.2 (i.e., *citrus*, *sour* and *fruity*) are commonly perceived as fresh (Zarzo and Stanton 2009; Zarzo 2013). Spicy and balsamic scents yield certain similarity (Abe et al. 1990; Chastrette et al. 1991) probably due to their warm character (Suppl. Table 4). The two balsamic subclasses are obviously related (r = 0.43) and appear next to each other in Fig. 8. The proximity of *fatty, cooked meat* and *fishy* is intuitively appealing given their edible quality.

The contrasting polarity of *warm* vs. *citrus* reflected by PC2 has a direct correspondence with PC1 of the B-H database (Zarzo and Stanton 2009). Most of 30 descriptors contained in this directory, taking into account their corresponding reference materials, can be matched tentatively with Dravnieks' attributes. *Aldehyde* was linked with *rancid* because the standard considered for this attribute smells somewhat rancid (Zarzo and Stanton 2009). Likewise, *balsamic, spicy, erogenic, coniferous* and *vegetable*, from the B-H compilation, were matched with *incense, clove, musk, pine oil* (i.e., *turpentine*) and *hay*, respectively. *Sweet* (heliotropin) was linked with *sweet/almond* (i.e., midway of both in the Atlas) based on the smell of this material. Five additional attributes (*nutty, cinnamon, bitter, dill,* and *leather*) were also considered taking into account their warm/fresh character based on Fig. 5 of Zarzo and Stanton (2009). Strikingly, for the resulting set of 30 descriptors from the Atlas, their correlation coefficient with *warm* (r_{warm}) is strongly correlated with the p[1] loadings of equivalent attributes in the B-H database (Fig. 8: r = -0.85, which becomes -0.91 if the outlying attributes *lavender* and *animal* are disregarded). Similar results were reported by Zarzo (2013).

[FIGURE 8 ABOUT HERE]

Some unpleasant descriptors and edible terms like *fishy* or *cooked meat* are obviously absent in perfumery studies. Several works in this context have reported that one factor discriminates attributes mainly encountered in feminine fragrances vs. those targeted to men. Attempting to investigate if this underlying dimension is also contained in the Atlas, 13 descriptors often used in perfumery were selected and, next, a PCA model was fitted. One dashed line was drawn in the PC1-2 loading plot (Fig. 9) between *citrus* and *warm*, given their opposite character. Terms on the right side (i.e., *sweet, floral, cinnamon, vanilla,* and *fruity*) are commonly applied to women's fragrances, while those on the left (i.e., *woody, leather, musty-earthy,* and *green*) are more typical of men's (Zarzo and Stanton 2009).

Floral refers to feminine scents and it yields the most negative correlation with *stale* (r = -0.40), which was classified midway of *musty-earthy* and *rancid* (Table 3). Interestingly, this polarity of floral vs. mossy-earthy scents is apparent in the Odor Effects Diagram and the B-H database (see Fig. 5 of Zarzo and Stanton 2009). A final PCA was fitted with *floral*, *musty-earthy* and related terms (*rose, violets, stale*, and *putrid-foul*). The projection of odorants over PC1 (R^2_X = 52%) is a latent variable that was denoted as Y_{FEM} because it could be interpreted as an indirect estimation of the feminine odor character. The correlation coefficient between Y_{FEM} and a given descriptor (r_{fem}) is some sort of assessment about the feminine/masculine character of that term. As the same interpretation has been given to PC2 of the B-H database, the approach just mentioned was also applied by matching descriptors in this database with those in the Atlas. It was found that p[2] loadings from the B-H database are correlated (r = 0.53, p = 0.002) with r_{fem} values (Fig. 9), though a stronger relationship (r = 0.67) becomes after disregarding *aldehyde* and *lavender*. The latter is an apparent outlier given the predilection of lavender notes in men's fragrances of the fougère category (Zarzo and Stanton 2009). Both dimensions (i.e., fresh/warm and masculine/feminine) account for the so-called Odor Effects Diagram, which is well known in perfumery, but this is probably the first study reporting that such factors are also salient in the Atlas.

[FIGURE 9 ABOUT HERE]

CONCLUSIONS

Based on the multivariate analysis of Dravnieks' Atlas, 105 descriptors were classified in 24 categories, 10 of non-food smells (43 attributes, Table 1), 13 classes of edible odors (54 descriptors, Table 2), and another group of generic olfactory terms (Suppl. Table 2). The total

number of categories proposed is consistent with other reported statistical analyses of odor profile databases suggesting about 20-30 dimensions of odor character (Chastrette et al. 1988; Abe et al. 1990; Callegari et al. 1997; Madany-Mamlouk et al. 2003; Zarzo and Stanton 2006, Castro et al. 2013). This modern scientific perspective contrasts with a more empirical approach proposed several decades ago based on a relatively small number of odor classes or dimensions, ranging from four to nine (Henning 1916; Lovell 1923; Zwaardemaker 1925; Crocker and Henderson 1927; Klein 1947; Amoore 1962). Nonetheless, this issue is still a matter of debate, and some authors are supporting that the perceptual space of odors may have rather few dimensions (Secundo et al. 2014).

Jeltema and Southwick (1986) obtained olfactory profiles for 415 odorants according to the 146 descriptors contained in the Atlas, 119 of which were classified in 17 odor classes, which is consistent with the belief that our perceptual space of smells is highly dimensional. It should be remarked that 10 attributes were considered in between two odor classes, which is a novel idea consistent with the notion that human olfactory space is better explained as a spectrum, rather than strictly segmented in well-defined odor categories. However, these authors just proposed 10 out of the 119 attributes as intermediate, which seems a low proportion. By contrast, 33 descriptors were classified here as intermediary of two or three classes.

Starting from the first PCA model with 138 variables, by means of a careful inspection of the loading plots corresponding to the relevant components and taking into account previous reported studies, 10 odor categories were proposed initially: “chemical-solvent”, “fruity”, “citrus”, “floral”, “musk”, “spicy”, “green”, “sulfidic”, “balsamic-vanilla”, and “balsamic-caramel”. The last two ones are obviously linked, and some relationship also exists, for example, between “citrus” and “fruity”, which refer to fruits. Likewise, “floral” and “musk” share certain cosmetic / perfumery character. Understanding the similarities between olfactory classes is of interest for providing certain standards of communication between perfumers. As no further clusters could be clearly determined from this first model, it was necessary to remove the set of descriptors already classified and perform a new PCA. The four additional models carried out reveal that the remaining correlation structures are progressively less and less informative. Thus, an important conclusion derived from the present research is that the set of 160 odorants contained in the Atlas is a rather limited sample of the huge spectrum of odorous materials, and does not allow an accurate characterization of the dimensionality of olfactory space.

It might be argued that many of the decisions about the clusters established seem to be essentially “educated guesses” based upon previous analyses and semantic associations, rather than well-defined clusters derived from the multivariate analysis. This is truly a restriction of the present work, and further research is encouraged using a larger number of odorous stimuli that should be representative of the wide spectrum of odor space. Nonetheless, based on the limitations derived from the relatively reduced number of odorants contained in the Atlas, this is probably “the best that we can do” aimed at classifying the ASTM odor list.

The taxonomy proposed by Jeltema and Southwick (1986) and the present one are indicated in Tables 1-3. It can be observed that both are rather coincident, but some differences are remarkable. For example, 19 unpleasant descriptors were regarded by Jeltema and Southwick (1986) as “animal-foul”, but they were grouped here basically into 4 categories: “animal”, “sickening”, “fetid-decay”, and “sour”. Moreover, *buttery* was included with the balsamic attributes, which is arguable. The classification proposed here (Tables 1-3)

has higher face validity than the one suggested by Keller and Vosshall (2004) based on just 11 categories, some of which grouped clearly different smells (e.g., *floral*, *musky*, *bitter*, *sweaty*, and *cool* were regarded as "common").

The work of Castro et al. (2013) has identified 10 underlying factors in the Atlas, but their interpretation is unclear in some cases. For example, *sour*, *earthy* and *putrid* were classified here in different categories, but they appeared as important in the fourth factor. Likewise, the ninth factor comprised rather different odors like *garlic*, *burnt* and *putrid*. Moreover, many descriptors are shared in common by different factors, which complicate the interpretation of results. By contrast, the organization of odor classes proposed here seems more consistent with previous studies of odor description and classification.

Attempting to better understand the relationships among descriptors, the similarities between classes were further studied, and three underlying perceptual dimensions were discussed: (i) odor attributes can be classified according to their hedonic quality, (ii) based on their masculine vs. feminine character, and (iii) another dimension discriminates *warm* vs. fresh scents like *citrus*. The same factors have been reported by multivariate analysis of other olfactory databases.

Reference materials were not considered by Dravnieks in the Atlas compilation given the long list of descriptors, which might partly explain some discrepancies as discussed above. For example, *lavender* and *floral* yield a strong positive correlation in the Atlas (r = 0.84, p < 0.0001), but not in the B-H database (r = -0.12, p = 0.03). This type of inconsistencies has not been discussed in detail in previous reported studies of the Atlas. Dravnieks (1985) compiled the sensory information from a large panel, attempting to reduce the noise. However, instead of focusing the efforts in collecting the assessments from many individuals, if the panel had been previously trained about the fundamentals of odor description and classification, probably similar profiles would have resulted with a reduced panel. Thus, a trained panel and a greater number of odorants are necessary for an accurate characterization of our olfactory perception space. For this purpose, it seems that quantitative odor profiles provide more accurate information compared with semantic approaches. Hence, the use of such methods is encouraged, though they require more time.

Data availability The data can be accessed via this link: https://doi.org/10.1520/DS61-EB

Supplementary Information (four tables and five figures). Suppl. Table 1: PCA summary overview. Suppl. Table 2: Descriptors in the Atlas classified as general / unspecific pleasant odors. Suppl. Table 3: Similarities between odor descriptors with a sulfidic (garlic-like) note. Suppl. Table 4: Similarities between the 23 odor classes indicated in Tables 1 and 2. Suppl. Fig. 1: Loading plot corresponding to PC1-2 from the "PA matrix" and from the "$PA_{>4}$ matrix". Suppl. Fig. 2: Loading plot for PC6-8 (model 1). Suppl. Fig. 3: Loading plot for PC3-4 (model 2). Suppl. Fig. 4: Loading plot for PC1-3 and PC3-5 (model 4). Suppl. Fig. 5: Loading plot for PC1-2 (model 5).

REFERENCES

Abe H, Kanaya S, Komukai T, et al (1990) Systemization of semantic descriptions of odors. Anal Chim Acta 239:73-85. https://doi.org/10.1016/S0003-2670(00)83837-4

Amoore J E (1962) The stereochemical theory of olfaction. I. Identification of seven primary odours. Proc Sci Sect Toilet Goods Assoc 37(suppl):1-13

Amoore JE, Venstrom D (1967) Correlations between stereochemical assessments and organoleptic analysis of odorous compounds. In: Hayashi T (ed) Olfaction and taste II. Pergamon Press, Oxford, pp 3-17

Arctander S (1969) Perfume and flavor chemicals (aroma chemicals), Vol. I and II. Arctander S publisher, Montclair, NJ

Boelens H, Haring HG (1981) Molecular structure and olfactive quality [Internal report]. Naarden International, Bussum, The Netherlands

Brud WS (1986) Words versus odours: how perfumers communicate. Perf Flavor 11:27-44

Burdock GA (Ed.) (2004) Fenaroli's handbook of flavor ingredients, 5th edn. CRC Press, Boca Raton, FL

Callegari P, Rouault J, Laffort P (1997) Olfactory quality: from descriptor profiles to similarities. Chem Senses 22:1-8. https://doi.org/10.1093/chemse/22.1.1

Castro JB, Ramanathan A, Chennubhotla CS (2013) Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. PLoS One 8:e73289. https://doi.org/10.1371/journal.pone.0073289

Cerbelaud R (1951) Formulaire de parfumerie. Opera, Paris

Chastrette M, Elmouaffek A, Sauvegrain P (1988) A multidimensional statistical study of similarities between 74 notes used in perfumery. Chem Senses 13:295-305. https://doi.org/10.1093/chemse/13.2.295

Chastrette M, de Saint Laumer JY, Sauvegrain P (1991) Analysis of a system of description of odors by means of four different multivariate statistical methods. Chem Senses 16: 81-93. https://doi.org/10.1093/chemse/16.1.81

Crocker EC, Henderson LF (1927) Analysis and classification of odours: an effort to develop a workable method. Amer Perf Essent Oil Rev 22:325-356

Coxon JM, Gregson RAM, Paddick RG (1978) Multidimensional scaling of perceived odour of bicyclo[2.2.1]heptane, 1,7,7-trimethylbicyclo[2.2.1]heptane and cyclohexane derivatives. Chem Senses 3:431-441. https://doi.org/10.1093/chemse/3.4.431

Davis RG (1979) Olfactory perceptual space models compared by quantitative methods. Chem Senses 4:21-33. https://doi.org/10.1093/chemse/4.1.21

Dravnieks A (1982) Odor quality: semantically generated multidimensional profiles are stable. Science 218:799-801. https://doi.org/10.1126/science.7134974

Dravnieks A (1985) Atlas of odor character profiles, data Series DS 61. American Society for Testing and Materials, Philadelphia. https://doi.org/10.1520/DS61-EB

Dravnieks A, Bock FC, Powers JJ et al (1978) Comparison of odors directly and through profiling. Chem Senses 3:191-225. https://doi.org/10.1093/chemse/3.2.191

Dravnieks A, Masurat T, Lamm RA (1984) Hedonics of odors and odor descriptors. J Air Pollut Control Assoc 34:752-755. https://doi.org/10.1080/00022470.1984.10465810

Ennis DM, Boelens H, Haring H et al (1982) Multivariate analysis in sensory evaluation. Food Technol 36:83-90

Harper R, Bate-Smith EC, Land DG (1968a) Odour description and odour classification: a multidisciplinary examination. Elsevier, New York.

Harper R, Bate-Smith EC, Land DG et al (1968b) A glossary of odour stimuli and their qualities. Perf Essent Oil Rec 59:22-37

Henning H (1916) Der geruch. Barth, Leipzig, Germany

Jaubert JN, Gordon G, Doré JC (1986) Classification of odours and their sensorial perception. Quintessenza 5:27-42
Jaubert JN, Gordon G, Doré JC (1987) Une organisation du champ des odeurs. II. Modèle descriptif de l'organisation de l'espace odorant. Parfum Cosmét Arômes 78:71-82
Jeltema MA, Southwick EW (1986) Evaluation and applications of odor profiling. J Sensory Stud 1:123-136. https://doi.org/10.1111/j.1745-459X.1986.tb00165.x
Jennings-White C (1984) Human primary odors. Perfum Flavor 9:46-58
Kaeppler K, Mueller F (2013) Odor classification: A review of factors influencing perception-based odor arrangements. Chem Senses 38:189-209. https://doi.org/10.1093/chemse/bjs141
Keller A, Vosshall LB (2004) Human olfactory psychophysics. Current Biol 14:875-878. https://www.cell.com/current-biology/pdf/S0960-9822(04)00780-8.pdf
Keller A, Vosshall LB (2016) Olfactory perception of chemically diverse molecules. BMC Neurosci 17:55. https://doi.org/10.1186/s12868-016-0287-2
Khan RM, Luk CH, Flinker A et al (2007) Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. J Neurosci 27:10015-10023. https://doi.org/10.1523/JNEUROSCI.1158-07.2007
Klein S (1947) Primary odour element classification. Amer Perfum Essent Oil Rev 50:453-454
Koulakov AA, Kolterman BE, Enikolopov AG et al (2011) In search of the structure of human olfactory space. Front Syst Neurosci 5:1-8. https://doi.org/10.3389/fnsys.2011.00065
Kumar R, Kaur R, Auffarth B et al (2015) Understanding the odour spaces: A step towards solving olfactory stimulus-percept problem. PLoS ONE 10:e0141263. https://doi.org/10.1371/journal.pone.0141263
Licon CC, Manesse C, Dantec M et al (2018) Pleasantness and trigeminal sensations as salient dimensions in organizing the semantic and physiological spaces of odors. Sci Rep 8:8444. https://doi.org/10.1038/s41598-018-26510-5
Linnaeus C (1756) Odores medicamentorum. In: Amoenitates academicae, Vol. 3. Lars Salvius, Stockholm, pp 183-201
Lovell JH (1923) Classification of flower odours. Amer Bee J 63:392-394
Madany-Mamlouk A, Chee-Ruiter C, Hofmann UG et al (2003) Quantifying olfactory perception: mapping olfactory perception space by using multidimensional scaling and self-organizing maps. Neurocomput 52-54:591-597. https://doi.org/10.1016/S0925-2312(02)00805-6
Madany-Mamlouk A, Martinetz T (2004) On the dimensions of the olfactory perception space. Neurocomput 58-60:1019-1025. https://doi.org/10.1016/j.neucom.2004.01.161
Martínez-Mayorga K, Peppard TL, Yongye AB et al (2011) Characterization of a comprehensive flavor database. J Chemom 25:550-560. https://doi.org/10.1002/cem.1399
Meister M (2015) On the dimensionality of odor space. Elife 4:e07865. https://doi.org/10.7554/eLife.07865.001
Moskowitz HR, Gerbers CL (1974) Dimensional salience of odors. Ann NY Acad Sci 237:1-16
Pintore M, Wechman C, Sicard G et al (2006) Comparing the information content of two large olfactory databases. J Chem Inf Model 46:32-38. https://doi.org/10.1021/ci0502505
Randebrock RE (1985) Geruch und konstitution. Eine beweisführung für die molekulartheorie des geruches. I: Eine objektivierbare sprache für den geruch. Parfüm Kosmet 66:435-443

Ravia A, Snitz K, Honigstein D, Finkel M, Zirler R, Perl O, Secundo L, Laudamiel C, Harel D, Sobel N (2020) A measure of smell enables the creation of olfactory metamers. Nature 588:118-123. https://doi.org/10.1038/s41586-020-2891-7

Rimmel E (1895) The base of perfumes. London. (French translation 1990, Le livre des parfums. Les éditions 1900, Paris)

Schiffman SS (1974) Physicochemical correlates of olfactory quality. Science 185:112-117. https://www.jstor.org/stable/1738579

Schutz HG (1964) A matching-standards method for characterizing odor qualities. Ann NY Acad Sci 116:517-526. https://doi.org/10.1111/j.1749-6632.1964.tb45081.x

Secundo L, Snitz K, Sobel N (2014) The perceptual logic of smell. Curr Opin Neurobiol 25:107-115. https://doi.org/10.1016/j.conb.2013.12.010

Sell CS (2004). Scent through the looking glass. Chem Biodivers 1:1899-1920. https://doi.org/10.1002/cbdv.200490146

Sigma–Aldrich Fine Chemicals Company (2005) Flavors and fragrances 2005-2006 catalog. Author, Milwaukee, WI

Stevenson RJ, Boakes RA (2003) A mnemonic theory of odor perception. Psychol Rev 110:340–364. https://doi.org/10.1037/0033-295X.110.2.340

Tromelin A, Chabanet C, Audouze K et al (2017) Multivariate statistical analysis of a large odorants database aimed at revealing similarities and links between odorants and odors. Flav Fragr J 33:106-126. https://doi.org/10.1002/ffj.3430

Urdapilleta I, Giboreau A, Manetta C et al (2006) The mental context for the description of odors: a semantic space. Eur Rev Appl Psychol 56:261-271. https://doi.org/10.1016/j.erap.2005.09.013

von Haller A (1763) Classes odorum. In: Elementa physiologiae, Vol. 5. Grasset, Lausanne, pp 162–168

Willis JH (1944) Flower perfumes and their classification. Victorian Naturalist, 61, 131-136

Wise PM, Olsson MJ, Cain WS (2000) Quantification of odor quality. Chem Senses 25: 429-443. https://doi.org/10.1093/chemse/25.4.429

Woskow MH (1968) Multidimensional scaling of odors. In Tanyolac, N. (ed.), Theories of odors and odor measurement. Robert College Research Center, Istanbul, pp 147-188

Wright RH, Michels KM (1964) Evaluation of far infrared relation to odors by a standard similarity method. Ann NY Acad Sci 116:535-551. https://doi.org/10.1111/j.1749-6632.1964.tb45083.x

Yoshida M (1975) Psychometric classification of odors. Chem Senses 1:443-464. https://doi.org/10.1093/chemse/1.4.443

Zarzo M (2008) Psychologic dimensions in the perception of everyday odors: pleasantness and edibility. J Sensory Stud 23:354-376. https://doi.org/10.1111/j.1745-459X.2008.00160.x

Zarzo M (2011) Hedonic judgments of chemical compounds are correlated with molecular size. Sensors 11:3667-3686. https://doi.org/10.3390/s110403667

Zarzo M (2013) What is a fresh scent in perfumery? Perceptual freshness is correlated with substantivity. Sensors 13:463-483. https://doi.org/10.3390/s130100463

Zarzo M, Stanton DT (2006) Identification of latent variables in a semantic odor profile database using principal component analysis. Chem Senses 31:713-724. https://doi.org/10.1093/chemse/bjl013

Zarzo M, Stanton DT (2009) Understanding the underlying dimensions in perfumers' odor perception space as a basis for developing meaningful odor maps. Atten Percept Psychophys 71:225-247. https://doi.org/10.3758/APP.71.2.225

Zwaardemaker H (1925) Classification des odeurs. In: L'odorat. Doin, Paris, pp 178-224

Table 1 Proposed categories of 43 odor character descriptors contained in the Atlas corresponding to inedible (non-food) smells, and taxonomy proposed by Jeltema and Southwick (1986) for these descriptors

Odor descriptor in Dravnieks' Atlas[a]	Category	Taxonomy J-S[b]	$\bar{x}_j$ [c]	H_{sc} [d]	H_{tone} [e]	r_{max} [f]	Most similar descriptor
Floral	floral	floral	7.97	14	2.79	0.92	perfumery
Rose	floral	floral	2.40	8	3.08	0.90	violets
Violets	floral	floral	0.80	8	2.68	0.90	rose
Musk	musk	-	1.52	3	0.21	0.54	cologne
Dry, powdery	musk	-	3.39	3	-0.07	0.45	woody, resinous
Woody, resinous	woody	woody	12.94	3	0.94	0.72	cedarwood
Cedarwood	woody	woody	4.31	5	2.11	0.72	woody, resinous
Bark, birch bark	woody	woody	1.43	1	1.18	0.68	woody, resinous
Oakwood, cognac	woody	woody	0.22	1	1.23	0.39	etherish, anesthetic
Camphor	camphoraceous	cool-minty	3.88	4	-0.55	0.88	eucalyptus
Mothballs	camphoraceous	cool-minty / solvent	1.92	-1	-1.25	0.43	camphor
Eucalyptus	camphoraceous	cool-minty	1.73	4	0.99	0.88	camphor
Chemical	chemical–solvent	solvent	8.87	-6	-1.64	0.79	cleaning fluid
Cleaning fluid	chemical–solvent	solvent	2.10	-1	-1.69	0.79	etherish, anesthetic
Disinfectant, carbolic	chemical–solvent	solvent	4.18	0	-1.60	0.80	medicinal
Etherish, anesthetic	chemical–solvent	solvent	5.33	0	-1.54	0.81	medicinal
Gasoline, solvent	chemical–solvent	solvent	2.75	-4	-1.16	0.79	cleaning fluid
Kerosene	chemical–solvent	solvent	0.45	-1	-1.67	0.67	varnish
Varnish	chemical–solvent	solvent	1.43	1	-0.85	0.83	paint
Paint	chemical–solvent	solvent	3.37	0	-0.75	0.83	varnish
Alcoholic	chemical–solvent	solvent	1.42	0	-0.47	0.74	etherish, anesthetic
Nail polish remover	chemical–solvent	solvent	1.15	1	-0.81	0.75	varnish
Burnt, smoky	burnt	burnt / nutty	5.58	-10	-1.53	0.76	burnt paper
Burnt paper	burnt	burnt	0.60	-4	-1.47	0.76	burnt, smoky
Burnt candle	burnt	burnt	0.17	0	-0.08	0.40	cork
Stale tobacco smoke	burnt	-	0.26	-3	-2.83	0.49	peanut butter
Sooty	burnt	burnt	0.05	-1	-1.69	0.50	cork
Household gas	sulfidic	sulfidic	0.87	-10	-2.30	0.85	sulfidic
Sulfidic	sulfidic	sulfidic	1.70	-15	-2.45	0.85	household gas
Animal	animal	animal-foul	1.28	-16	-1.13	0.79	fecal (manure)
Sweaty	animal	animal-foul	5.89	-19	-2.53	0.87	rancid
Wet wool, wet dog	animal	animal-foul	0.61	-14	-2.28	0.68	animal
Dirty linen	animal	animal-foul	0.67	-14	-2.55	0.80	fecal (manure)
Mouse	animal	animal / nutty	0.03	-2	-2.20	0.33	mothballs
Seminal (sperm-like)	animal	-	0.14	-0.2	-1.04	0.24	dirty linen
Sickening	sickening	animal-foul	10.83	-21	-3.34	0.94	putrid, foul, decay
Heavy	sickening	animal-foul	15.05	-16	-0.79	0.68	sickening
Bitter	sickening	-	0.66	-12	-1.38	0.55	sharp, pungent, acid
Putrid, foul, decayed	fetid–decay	animal-foul	4.59	-20	-3.74	0.94	sickening
Sewer	fetid–decay	animal-foul / sulfidic	1.20	-17	-3.68	0.85	cadaverous
Rancid	fetid–decay	animal-foul	4.91	-20	-3.15	0.89	putrid, foul, decay
Cadaverous	fetid–decay	animal-foul	1.10	-17	-3.75	0.88	putrid, foul, decay
Fecal (like manure)	fetid–decay	animal-foul	1.87	-16	-3.36	0.82	putrid, foul, decay

[a]Categories are sorted approximately in decreasing hedonic character.
[b]The symbol "-" indicates that the descriptor was not classified by Jeltema and Southwick (1986).
[c]Average of the variable ($PA_{>4}$ values) for the 160 odorant samples (equation 3).
[d]Hedonic score computed as: $100 \cdot p[1]$ (loadings in PC1 from model 1).
[e]Hedonic tones obtained experimentally by Dravnieks (1984); for both H_{sc} and H_{tone}, negative values correspond to unpleasant odors and positive values to pleasant scents.
[f]Pearson's correlation coefficient of the most similar odor character attribute (indicated in the next column).

Table 2 Proposed categories of 54 odor character descriptors in the Atlas corresponding to edible smells, and taxonomy proposed by Jeltema and Southwick (1986)

Odor descriptor in Dravnieks' Atlas	Category	Taxonomy J-S	$\bar{x}_j$ [a]	H_{sc}	H_{tone}	r_{max}	Most similar descriptor
Fruity (non-citrus)	fruity	fruit	6.98	11	2.23	0.79	cherry
Cherry (berry)	fruity	fruit	1.98	8	2.55	0.79	fruity
Pear	fruity	fruit	0.40	5	2.26	0.79	pineapple
Peach (fruit)	fruity	fruit	0.41	6	2.67	0.77	strawberry
Strawberry	fruity	fruit	1.08	6	2.93	0.77	fruity
Grape juice	fruity	fruit	0.80	6	2.07	0.76	strawberry
Pineapple	fruity	fruit	0.88	5	2.59	0.79	pear
Cantaloupe melon	fruity	fruit	0.80	5	2.41	0.54	cucumber
Banana	fruity	fruit	0.45	2	2.00	0.46	nail polish remover
Citrus	citrus	citrus	3.34	8	2.72	0.94	lemon
Lemon	citrus	citrus	2.72	7	2.50	0.94	citrus
Orange	citrus	citrus	1.42	7	2.86	0.90	citrus
Grapefruit	citrus	citrus	0.36	4	1.95	0.77	lemon
Spicy	spicy	spicy	6.59	7	1.99	0.79	clove
Cinnamon	spicy	spicy	2.23	4	2.54	0.79	spicy
Clove	spicy	spicy	1.62	4	1.67	0.79	spicy
Tea leaves	spicy	-	0.13	3	1.40	0.65	clove
Black pepper	spicy	-	0.22	-1	0.19	0.31	seasoning for meat
Vanilla	balsamic - vanilla	brown / almond	0.93	3	2.57	0.90	chocolate
Chocolate	balsamic - vanilla	brown	0.25	1	2.78	0.90	vanilla
Malty	balsamic - vanilla	brown	0.36	0	1.05	0.83	chocolate
Caramel	balsamic - caramel	brown	0.59	0	2.32	0.91	maple syrup
Maple syrup	balsamic - caramel	brown	0.62	0	2.26	0.96	molasses
Molasses	balsamic - caramel	brown	0.23	-1	1.00	0.96	maple syrup
Dill	herbaceous	-	0.20	-1	0.87	0.42	mushroom
Caraway	herbaceous	caraway-anise	0.12	1	1.06	0.37	minty
Anise (licorice)	herbaceous	caraway-anise	1.98	3	1.21	0.34	aromatic
Fresh green vegetables	green	green	0.62	-2	2.19	0.86	crushed weeds
Crushed grass	green	green	0.83	-1	1.34	0.92	crushed weeds
Crushed weeds	green	green	0.82	-2	-0.21	0.92	crushed grass
Herbal, green, cut grass	green	green	7.78	9	2.14	0.63	crushed grass
Green pepper	green	green	0.21	-2	1.39	0.88	raw potato
Raw potato	green	green	0.17	-2	0.26	0.93	beany
Beany	green	green	0.03	-2	0.54	0.93	raw potato
Hay	green	-	0.18	-1	1.31	0.68	green pepper
Raw cucumber	green	green	0.19	0	1.30	0.54	cantaloupe
Celery	green	green	0.29	0	1.36	0.67	molasses
Buttery, fresh butter	buttery	brown	0.36	-2	2.04	0.83	popcorn
Bakery (fresh bread)	buttery	brown	0.07	-2	3.53	0.45	popcorn
Nutty (walnut, etc.)	nutty	nutty	1.64	-3	1.92	0.83	peanut butter
Peanut butter	nutty	nutty	1.05	-3	1.99	0.83	nutty
Grainy (as grain)	nutty	nutty	0.29	-3	0.63	0.67	popcorn
Popcorn	nutty	nutty	0.95	-3	2.47	0.83	buttery
Meaty (cooked)	cooked meat	nutty	0.55	-9	2.34	0.53	sulfidic
Seasoning (for meat)	cooked meat	-	0.47	-6	1.27	0.74	garlic, onion
Oily, fatty	fatty	animal-foul	6.57	-13	-1.41	0.63	stale
Fried chicken	fatty	-	0.06	-2	2.53	0.52	oily, fatty
Fishy	fishy	fishy / animal	1.49	-7	-1.98	0.94	kippery
Kippery (smoked)	fishy	fishy / animal	0.43	-6	-0.69	0.94	fishy
Sour, vinegar	sour	animal-foul	5.12	-16	-1.26	0.79	rancid
Sour milk	sour	animal-foul	0.67	-11	-2.91	0.82	rotten fruit
Sauerkraut	sour	animal-foul	0.09	-6	-0.60	0.47	cooked vegetables
Fermented (rotten) fruit	sour	fruit	0.34	-11	-2.76	0.82	sour milk
Yeasty	sour	nutty	0.50	-13	-0.07	0.74	sour milk

[a]See footnote of Table 1 for definition of $\bar{x}_j$, hedonic scores (H_{sc}), hedonic tones (H_{tone}), and r_{max}.

Table 3 Classification of 33 descriptors in the Atlas as intermediate odors, and taxonomy proposed by Jeltema and Southwick (1986) for these descriptors

Odor descriptor in Dravnieks' Atlas	Proposed classification [a]	Taxonomy J-S	$\bar{x}_j$ [b]	H_{sc}	H_{tone}	r_{max}	Most similar descriptor
Almond	"nutty" - *vanilla*	coconut-almond	3.60	3	2.01	0.45	nutty
Ammonia	"fishy" - "chemical solvent"	-	0.64	-7	-2.47	0.47	alcoholic
Blood, raw meat	"animal" - "fetid" - "fishy"	animal-foul	0.59	-13	-1.64	0.65	cadaverous
Burnt milk	"balsamic caramel" - "burnt"	burnt	0.03	-1	-2.19	0.87	coffee
Burnt rubber	"burnt" - "sulfidic"	burnt	0.61	-10	-3.01	0.75	sulfidic
Cardboard	*musty, earthy* - "woody"	-	0.57	-3	-0.54	0.60	wet paper
Cat urine	"animal" - *ammonia*	animal-foul	0.45	-12	-3.64	0.74	urine
Coconut	"nutty" - *vanilla*	coconut-almond	0.73	2	1.93	0.30	vanilla
Coffee	"balsamic caramel" - "burnt"	brown	0.19	-1	2.33	0.87	burnt milk
Cooked vegetables	"green" - "cooked meat"	green	0.31	-7	1.58	0.59	green pepper
Cool, cooling	"camphoraceous" - "herbaceous"	cool-minty	8.41	13	1.53	0.82	minty, peppermint
Cork	*leather* - "woody"	woody	0.19	-1	0.19	0.50	sooty
Creosote	"burnt" - "chemical solvent"	burnt / solvent	1.22	-5	-1.35	0.84	tar
Fresh tobacco smoke	"burnt" - "nutty"	nutty	0.04	0	-0.66	0.30	stale tobacco smoke
Garlic, onion	"sulfidic" - "cooked meat"	sulfidic	2.64	-9	-0.17	0.78	sulfidic
Geranium leaves	"floral" - "green"	-	0.38	2	0.57	0.52	crushed grass
Honey	"balsamic vanilla" - "floral"	brown	0.41	4	2.08	0.44	vanilla
Incense	"woody" - "musk" - "floral"	-	2.35	12	1.01	0.64	fragrant
Lavender	"floral" - "herbaceous"	floral	1.13	10	2.25	0.84	floral
Leather	"burnt" - "chem. solvent" - *medicinal*	leather-rubber	0.46	-2	1.30	0.72	creosote
Medicinal	"chem. solvent" - "camphorac."	solvent / minty	6.55	2	-0.89	0.81	etherish
Metallic	"sickening" - "chemical solvent"	solvent	1.07	-6	-0.94	0.42	sharp, pungent, acid
Minty, peppermint	"camphoraceous" - "herbaceous"	cool-minty	4.33	6	2.50	0.82	cool, cooling
Mushroom	*herbal, green - musty, earthy*	-	0.41	-1	0.52	0.42	dill
Musty, earthy, moldy	"green" - "woody"	green	8.04	-11	-1.94	0.64	stale
New rubber	"chem. solvent" - "burnt" - *leather*	solvent / leather-rubber	0.57	-4	-0.96	0.52	tar
Raisins	"spicy" - "fruity"	fruit	0.14	2	1.56	0.40	tea leaves
Sharp, pungent, acid	"sickening" - "sour"	animal-foul	10.00	-16	-2.34	0.73	sickening
Stale	*rancid - musty, earthy*	animal-foul	2.16	-17	-2.04	0.68	rancid
Tar	"burnt" - "chemical solvent"	solvent	1.30	-6	-1.63	0.84	creosote
Turpentine (pine oil)	"chem. solv." - "camphor." - "woody"	solvent	3.75	4	-0.73	0.73	varnish
Urine	"animal" - *ammonia*	animal-foul	0.92	-12	-3.34	0.74	cat urine
Wet paper	*musty, earthy* - "woody"	-	0.21	-1	-0.94	0.60	cardboard

[a]The names within inverted commas correspond to odor categories, and those in italics correspond to odor descriptors. The same criterion has been used throughout the text.

[b]See footnote of Table 1 for definition of $\bar{x}_j$, hedonic scores (H_{sc}), hedonic tones (H_{tone}), and r_{max}.

Fig 1 Loading plot corresponding to PC13 (chocolate↔cedarwood) vs. PC2 (fruity↔etherish), from the PCA of the "$PA_{>4}$ matrix" (138 variables included in the model). The "chemical-solvent" cluster is highlighted

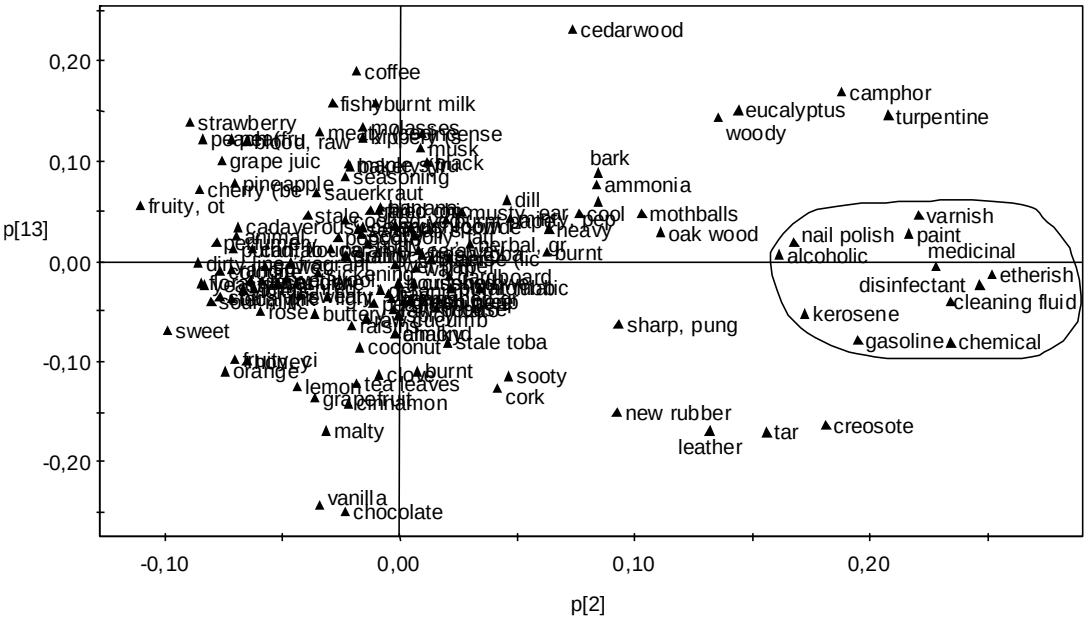

Fig 2 Loading plot from the "$PA_{>4}$ matrix" (138 variables included in the model): PC9 (cherry↔grapefruit) vs. PC5 (lavender↔pear) (*left*) and PC11 (lemon↔kippery) vs. PC10 (almond↔grapefruit) (*right*). The "fruity" and "*citrus"* clusters are highlighted. The dashed cluster encloses odors related with perfumery

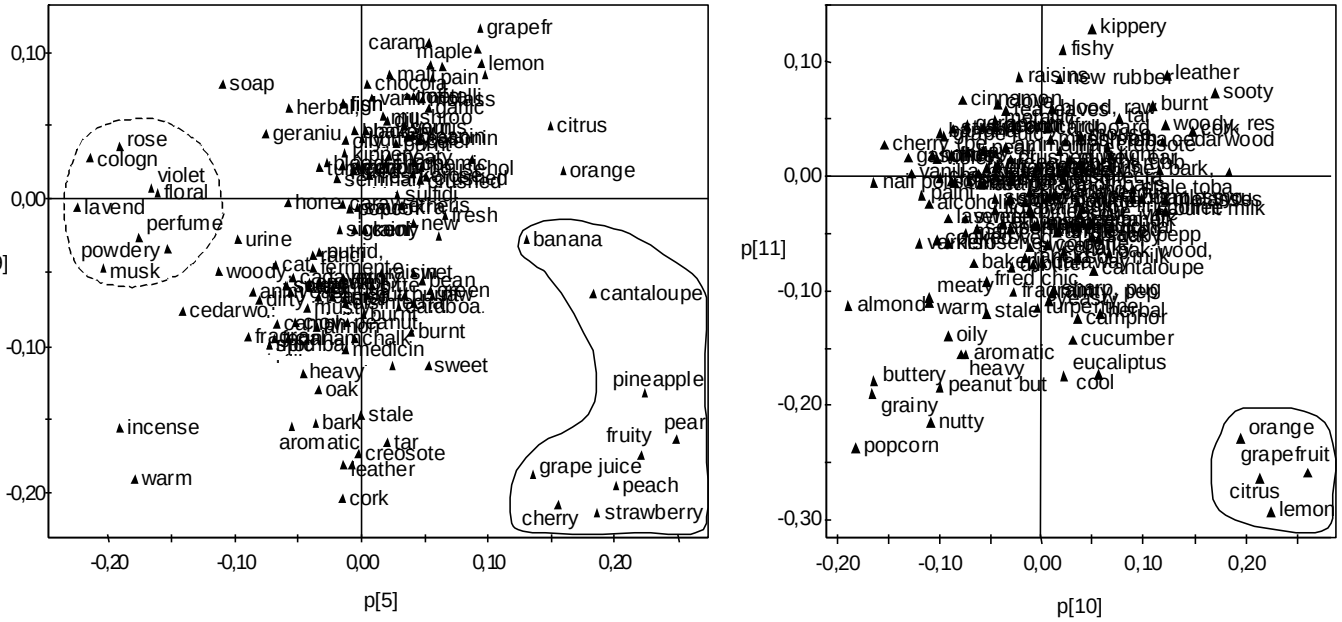

Fig 3 Loading plot corresponding to PC7 (lavender↔spicy) vs. PC5 (lavender↔pear), from the "$PA_{>4}$ matrix" (138 variables included in the model). The clusters "floral", "spicy", and "fruity" are highlighted. The dashed line corresponds to the perfumery dimension

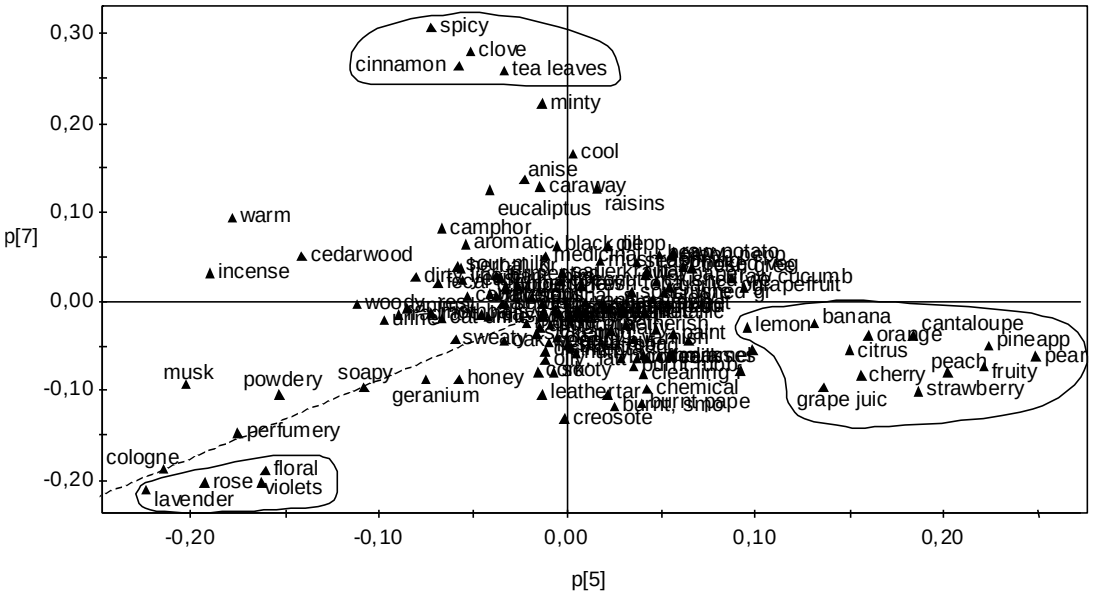

Fig 4 Loading plot from the "$PA_{>4}$ matrix" (138 variables included in the model): PC8 (caramel↔garlic) vs. PC4 (herbal↔caramel) (*left*) and PC14 (chocolate↔clove) vs. PC13 (chocolate↔eucalyptus) (*right*). The clusters "green", "balsamic", and "spicy" are highlighted. The "balsamic" cluster in PC4-8 appears subdivided in two sets in PC13-14

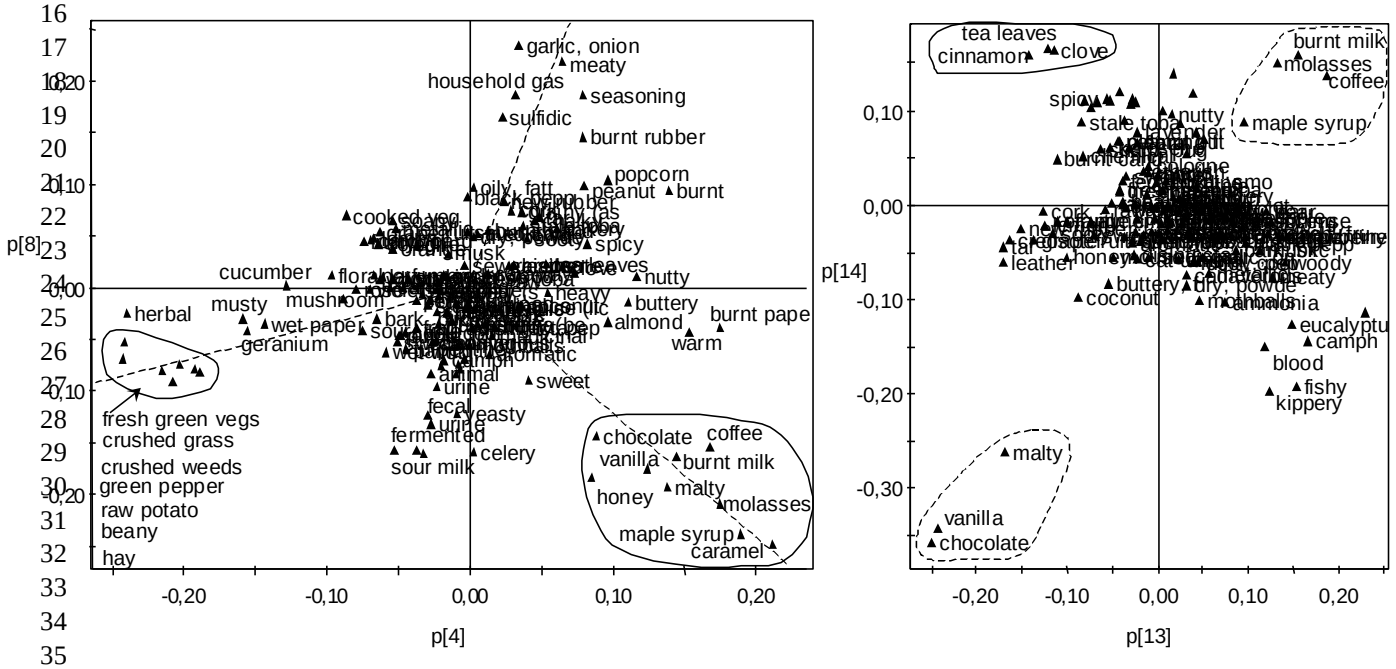

Fig 5 Loading plot from the "$PA_{>4}$ matrix" (78 variables included in the model): PC6 (burnt paper↔wet paper) vs. PC2 (sweet↔burnt) (*left*) and PC6 vs. PC3 (nutty↔medicinal) (*right*). Descriptors appearing close to each other and classified equally were joined together. Odor categories "burnt", "nutty", and "camphoraceous" are highlighted

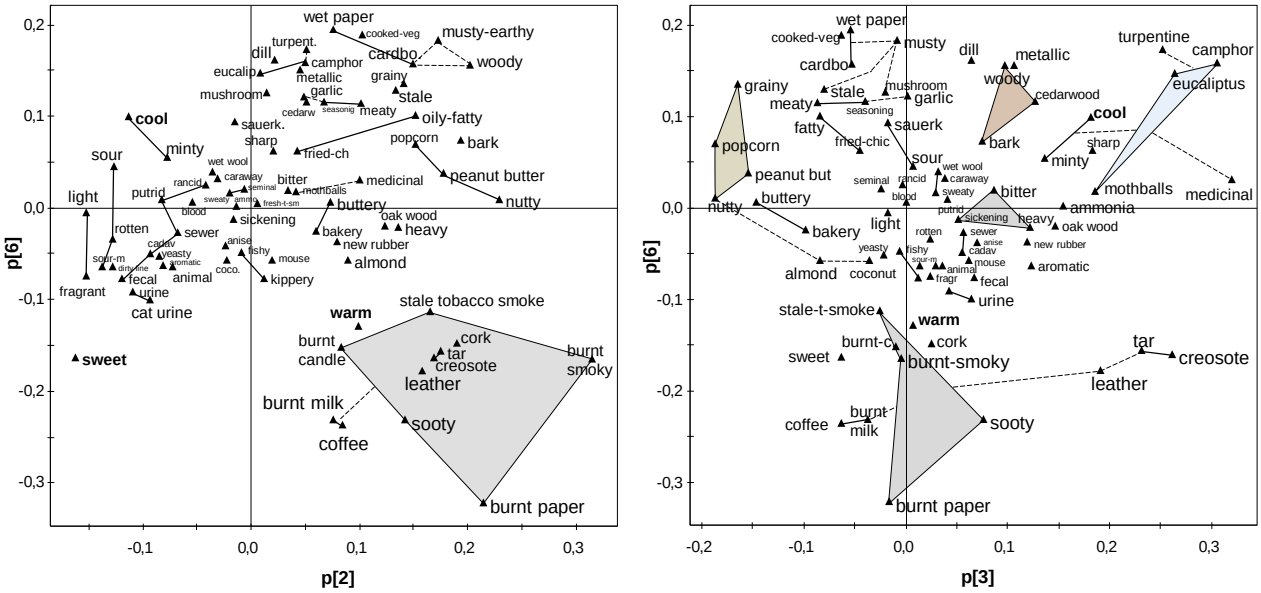

Fig 6 Loading plot from the "$PA_{>4}$ matrix" (61 variables included in the model): PC3 (sweet↔camphor) vs. PC1 (putrid↔light) (*left*) and PC3 vs. PC2 (nutty↔cool) (*right*). Descriptors appearing close to each other and classified equally were joined together. Odor categories "camphoraceous" and "woody" are highlighted

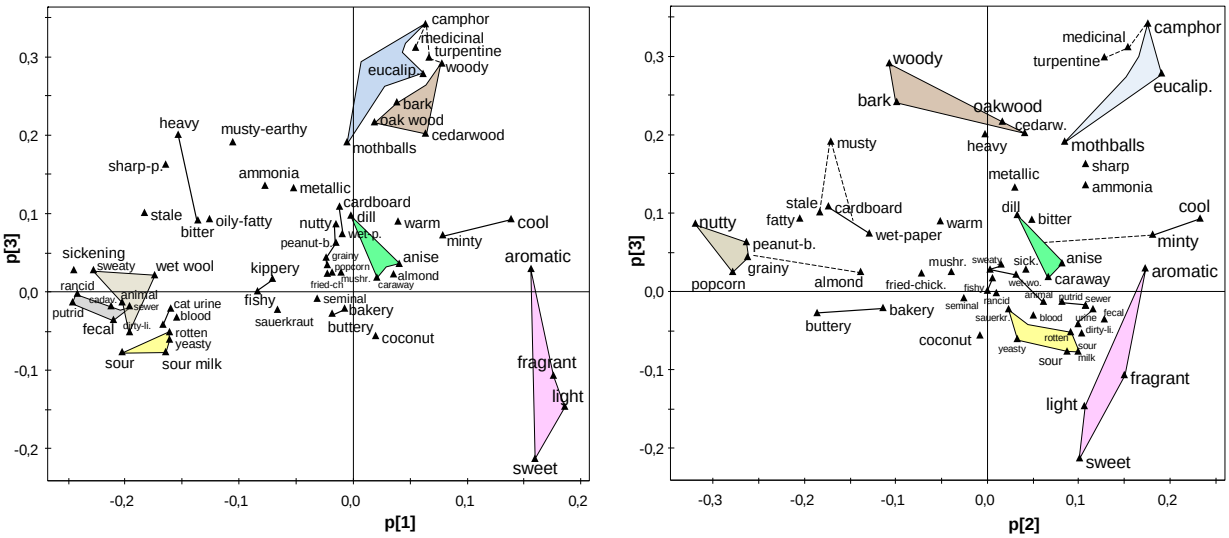

Fig 7 Loading plots from the "PA$_{>4}$ matrix" (61 variables included in the model): PC2 (nutty↔cool) vs. PC4 (warm↔woody) (*left*) and PC2 vs. PC7 (buttery↔warm) (*right*). Descriptors appearing close to each other and classified equally were joined together. Odor categories "nutty" and "buttery" are highlighted

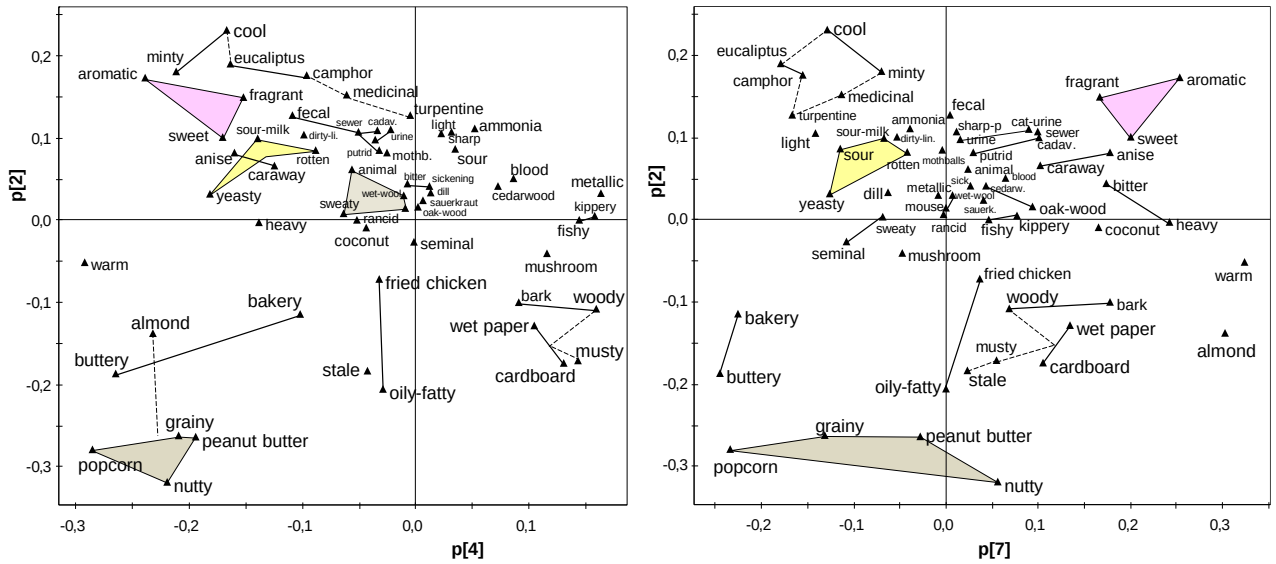

Fig 8 Loading plot for PC2 (fruity↔nutty) vs. PC1 (fetid↔sweet), from the matrix with 23 latent variables (one per odor category) as well as the descriptors *sweet* and *warm* (*left*); scatterplot of r_{warm} (correlation coefficient with *warm*) vs. p[1] loading from the B-H database (*right*). Variables (*left*): odor descriptors (labels in italics), or latent variables (otherwise). Some related classes are highlighted; the dashed cluster encloses categories sharing certain warm character

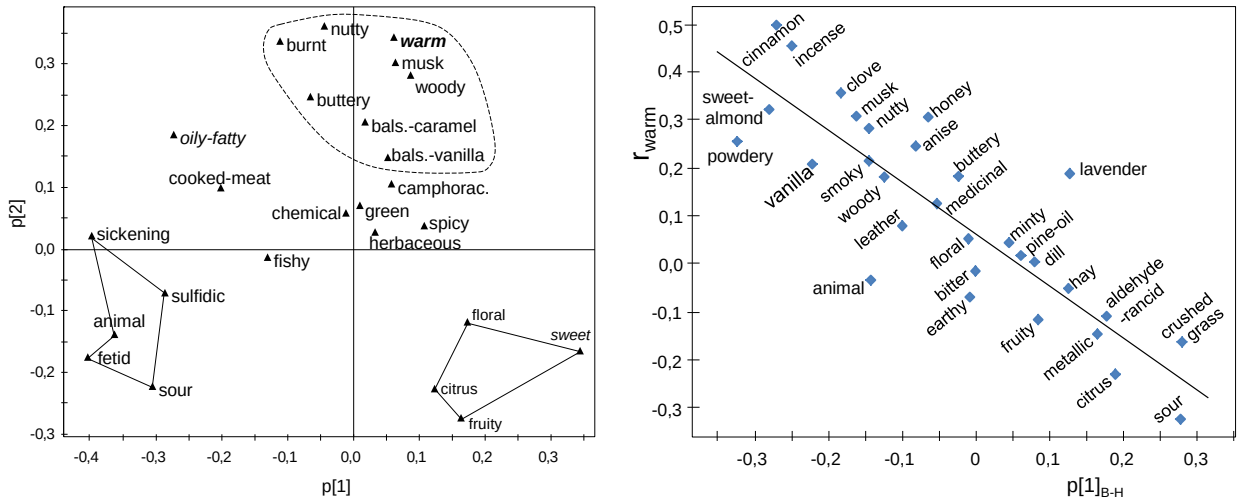

Fig 9 Loading plot for PC2 (citrus↔musk) vs. PC1 (musty↔sweet), considering 13 descriptors commonly used in perfumery (*left*); scatterplot of r_{fem} (correlation coefficient with the latent variable that discriminates *floral* vs. *earthy*) vs. p[2] loading from the B-H database (*right*). The dashed line discriminates approximately masculine vs. feminine descriptors