REVIEW ARTICLE

Odor Classification: A Review of Factors Influencing Perception-Based Odor Arrangements

Kathrin Kaeppler and Friedrich Mueller
Institute of Experimental Business Psychology, Leuphana University Lueneburg, Lueneburg, Germany

Correspondence to be sent to: Kathrin Kaeppler, Institute for Experimental Business Psychology, Leuphana University Lueneburg, Lueneburg, Germany. e-mail: kaeppler@leuphana.de
Accepted December 16, 2012

Abstract
Decoding the psychological dimensions of human odor perception has long been a central issue of olfactory research. Odor scientists as well as fragrance professionals have tried to establish comprehensive standards for the description, measurement, and prediction of odor quality characteristics. As odor percepts could not be linked to a few measurable physicochemical features of odorous compounds or physiological characteristics of the olfactory system, odor qualities have often been assessed by perception-based ratings. Although they have been applied for more than 5 decades, these psychological approaches have not yielded a comprehensive or generally accepted classification system yet. We assumed that design and methodology of these studies have largely prevented the development of unbiased odor arrangements. To address this issue, we reviewed 28 perception-based classification studies and found that their outcome has been largely determined by 4 influencing factors: 1) interindividual differences in perceptual and verbal abilities of subjects, 2) stimuli characteristics, 3) approaches of data collection, and 4) methods of data analysis. We discuss the effects of each factor in detail and illustrate how odor systems have reflected perceptual qualities as well as the conditions under which these have been assessed.

Key words: categorization, description, dimension, odor quality, odor space, olfaction

Introduction
For decades, scientists from various disciplines have been searching for an olfactory classification system to define a perceptual space and facilitate objective communication about odors. However, none of the yet proposed odor arrangements has gained wide acceptance or empirical confirmation. There is a lack of means to compare and describe odors accurately or estimate their degree of similarity with precision.

Fallacy of color analogy
To illustrate the aims of olfactory classification systems, researchers have often drawn an analogy to color perception, where classes, dimensions, and the perception space have been widely studied and well defined (Harper 1966; Harper et al. 1968; Chastrette 1998; Madany Mamlouk and Martinetz 2004; Haddad et al. 2008b). The dependency of perceived color quality on the wavelength of light and the color-specific sensibility of 3 receptor types in the human eye have facilitated the development of low-dimensional, neatly arranged color models. However, the assumption of a single comprehensive color system that is rooted in a natural arrangement of stimuli is a fallacy. A great number of color systems has been developed to accomplish distinct tasks at different levels of detail (for an overview, see Kuehni and Schwarz 2008). These systems have defined 1) color classes and appropriate labels, 2) color dimensions that characterize stimuli by their position on independent measures, or 3) color spaces that comprise an assumed entirety of perceivable colors along with meaningful dimensions to distinguish them. The structure of these systems has been anything but axiomatic: The number and character of dimensions or classes has varied with the purpose of each arrangement. Although a single, universal color scheme has not been established, nor sought after, the work of many odor researchers has been guided by this ideal conception. However, the potential purposes of odor systems are diverse: They range from the allocation of odors in classes with appropriate labels over the
identification of (hierarchical) relations between these classes and the features by which they may be distinguished to the establishing of an appropriate terminology, the depiction of blending rules, perceptual similarities, and finally the relations to physical, chemical, or functional criteria. Remarkably, in odor research, these aims have very often been specified only vaguely. Although numerous investigators have tried to establish classification systems to facilitate “differentiation, recognition, and identification” (Harper et al. 1968), most of them have, in fact, pursued different aims without stating them explicitly. Not surprisingly, neither an accepted system nor a reliable consensus on the basic principles of this arrangement has been reached. In order to attain meaningful outcomes, odor researchers are thus required to carefully address and define the functions of odor systems initially. The work of odor professionals and perfumers has shown how clearly defined tasks can yield valuable schemes of fragrance qualities (Köster 2002).

**General approaches to olfactory classification**

Given that an association between percepts and a single or a few physical parameters has not yet been found in olfaction (Turin and Yoshii 2003), odor scientists have relied on more subjective attempts in arranging odors. Early odor classification systems were largely based on individual expertise of botanists, chemists, or perfumers and have mainly ruled out perceptions of the sensory organ:

1. **Features of the sensory organ:** Several researchers have linked odor qualities to the function of olfactory receptors. Amoore (1967, 1977) assumed an increased detection threshold for specific odorants along with otherwise normal olfactory sensitivity as indication for the malfunction of a particular receptor type. He screened subjects for specific types of anosmia and defined a classification system based on 7 primary odors, each related to a distinct receptor type. Other studies used empirical cross-adaptation approaches to investigate the relation between odor classes and receptor types (Cain 1970; Todrank et al. 1991; Cain and Polak 1992; Pierce et al. 1993; Pierce et al. 1995). However, with the discovery of not less than 320 odor receptor types in humans (Glusman et al. 2001; Zozulya et al. 2001; Malnic 2004), the idea of a manageable system of primary odors has been largely discarded along with the attempt to establish olfactory classifications from physiological features of the olfactory system.

2. **Features of the sensory stimulus:** The chemical structure of an odorous compound strongly determines its perceived quality. There have been attempts to establish reliable structure–odor relationships (SOR) by linking perceptual properties to molecular vibration (Dyson 1938; Wright 1954; Wright and Sereniüs 1954; Wright and Michels 1964), molecular weight (Schiffman 1974b), functional group type and position (Uchida et al. 2000; Goëke 2002), molecule shape (Amoore 1963), electron donor (McGill and Kowalski 1977), acid–base character (Brower and Schafer 1975), chain length (Døving 1966), and other physicochemical parameters (for a review, see Rossiter 1996). However, all single measures have failed to reliably predict odor sensations or systematically explain odor perception so far. To address this issue, recent studies have revived early approaches (Amoore et al. 1967; Schiffman 1974a, 1974b; Schiffman et al. 1977) and attempted to include hundreds of physicochemical features in a single measure (Khan et al. 2007; Haddad et al. 2008a, 2008b). With this approach, Khan et al. (2007) successfully estimated odor pleasantness from a metric space of 1664 structural characteristics. Remarkably, Amoore (1971) was able to predict a single quality dimension from structural features already 30 years earlier. Hence, even modern computational approaches and the access to thousands of physicochemical odor attributes has not moved odor researchers closer to meaningful odor arrangements. This is little surprising: To provide valid outcomes, SOR approaches require what is actually under investigation—a reliable system of odor perceptions. SOR may actually benefit from the availability of perceptual odor spaces, while they are little promising in the development of basic classifications.

3. **Features of the sensory percept:** Henning (1916) was the first who directly classified olfactory percepts by arranging verbal odor descriptions. He presented 415 odorants to 6 participants and asked them to freely verbalize their perceptions. Based on a subjective summary of these verbal reports, Henning proposed 6 odor qualities and arranged them as corners of a prism. Although Henning’s model has been repeatedly tested and falsified (Dimmick 1922; MacDonald 1922; Findley 1924; Hazzard 1930), many studies have followed his approach and applied verbal reports of odor perception to established odor classifications. These studies have been largely based on the assumption that odors can be located in an n-dimensional space where their position illustrates their similarity to each other. To reveal the nature of these dimensions, odor scientists have collected data on the relation between odors and searched for an underlying structure by the means of multivariate statistical methods.

Several authors have raised the question whether these perception-based efforts have been more successful than physiological or stimulus-centered approaches. Chastrette (1998, 2002) and Wise et al. (2000) reviewed a large body of research and reported that many of the proposed perception-based classification systems are vague or even
contradictory. From 5 decades of empirical classification research, Chastrette (1998, 2002) reported merely 4 basic conclusions: 1) the olfactory perception space is probably not hierarchically structured—its structure is generally weak, 2) it is rather high dimensional, 3) the labels of these odor dimensions remain arbitrary, and 4) odor classes may partly overlap. Wise et al. (2000) ascribed this lack of reliable insights to the subjective character of perception-based data collection methods, which “makes them anachronistic with modern methodology in experimental behavioral science” (p. 429). We assumed additional factors that have caused the conflicting results of classification studies. This paper reviews psychological classification studies published in the last 50 years to analyze the impact of 4 factors: 1) subjects, 2) stimuli characteristics, 3) approaches of data collection, and 4) methods of data analysis. Although several studies have addressed the impact of particular variables on odor arrangements (Yoshida 1975; Schiffman and Dackis 1976; Davis 1979; Jeltema and Southwick 1986; Chastrette et al. 1991; Higuchi et al. 2004), these effects have not been reported systematically so far. We examined papers listed in psychological databases that applied a perception-based approach to odor classification. Non-English publications, grey literature, and abstracts of symposia or conferences were excluded from a detailed review. Papers addressing methodological issues were not considered as classification studies. We identified 27 studies that complied with these criteria. Basic characteristics of these studies are summarized in Table 1. Their results clearly indicate that the proposed odor arrangements have varied considerably with respect to number and nature of olfactory dimensions. This paper discusses 4 possible factors that have affected the outcome of these studies and illustrates why consistency for olfactory systems has not been reached so far.

Factor 1: Subjects

The psychological approach to odor classification is mainly based on verbal descriptions of odor percepts. Hence, a valid olfactory classification requires the reliability of both perception and verbal expression. However, one should not simply assume 1) that olfactory perceptions are generally stable over time, 2) that different people perceive identical odorants in the same way, and 3) that different people verbalize their olfactory percepts consistently.

Intra- and interpersonal differences in odor perception

Some authors have addressed the “test–retest reliability” of perceptual ratings and found high correlations for experts as well as laymen, across different data collecting approaches and over short and medium time periods (Schutz 1964; Wright and Michels 1964; Dravnieks 1982; Jeltema and Southwick 1986; Lawless and Glatter 1990). However, reliability measures of odor ratings have usually been calculated from averaged group data and hence provide only little indication on the stability of an individual’s odor perception or ratings, respectively. Hence, the reproducibility of individual ratings may actually be much lower than the exceptionally high reliabilities reported by several authors (Cain et al. 1998).

Over very long time periods, odor ratings may be less reliable due to age-related changes in odor perception. Although there has been little longitudinal research, several cross-sectional studies have suggested a considerable influence of age on olfaction (Wysocki and Gilbert 1989; Russell et al. 1993; Corwin et al. 1995; Larsson et al. 2000): In 4 studies, participants received 6 microencapsulated odorants and were asked to describe each odor with only 1 attribute from an 11-item list. The results of about 2.4 million panelists aged between 10 and 90 years point toward a change in odor quality perception with increasing age. However, age-related differences were found to be odor specific: Wysocki and Gilbert (1989), for example, reported that adrostenone was identified correctly by only about 20–30% of the subjects in all age groups with a slight decrease in the sixth decade, whereas the identification rate of rose strongly declined from over 80% in third decade to less than 60% for panelists aged 80 and more. The influence of age on odor perception is not at all uniform across odors and hence difficult to control in olfactory studies. Elderly subjects might be excluded from classification studies to reduce the impact of physiological impairments. This can, however, not solve the problem of interindividual differences, as these also occur within age groups. Beyond age, gender (Cain 1982; Doty et al. 1985; Yousem et al. 1999; Keller et al. 2012), several other demographic variables (Corwin et al. 1995; Larsson et al. 2000; Larsson et al. 2004; Keller et al. 2012), certain diseases (Doty 1989), and psychiatric disorders (Atanasova et al. 2008) have been found to influence olfactory performance. A particular important influence is exerted by experience. It affects 1) basic perceptual ratings as well as 2) odor classifications. Several studies stated that odor quality perception is substantially shaped by experience and have illustrated this relation in cross-cultural comparisons (Pangborn et al. 1988; Wysocki et al. 1991; Ayabe-Kanamura et al. 1998; Song and Bell 1998; Chrea et al. 2004; Seo et al. 2011) as well as intracultural studies (Distel et al. 1999; Distel and Hudson 2001; Hudson and Distel 2002). Ayabe-Kanamura et al. (1998) compared the olfactory perception of 44 German and 40 Japanese subjects. Participants were asked to smell 18 everyday odorants (6 familiar to Japanese, 6 familiar to Germans, 6 familiar to both groups) and to judge them against several perceptual characteristics. For 10 odors, significant differences in familiarity ratings were found between both groups. Well-known odors were usually rated as more pleasant and more often as edible in each of the 2 populations. These results suggest that humans prefer the smells they have frequently experienced due to their culture-specific eating habits and hence demonstrate a substantial impact of cultural experience on
### Table 1  Overview of psychological classification studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Analysis of Subjects</th>
<th>Classification Method</th>
<th>Number of Attributes</th>
<th>Number of Reference Odors</th>
<th>Method for Data Analysis</th>
<th>Number of Dimensions</th>
<th>Pleasantness as Primary Dimension</th>
<th>Number of Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wright and Michels (1964)</td>
<td>—</td>
<td>n/a</td>
<td>45</td>
<td>—</td>
<td>EFA</td>
<td>8</td>
<td>No</td>
<td>—</td>
</tr>
<tr>
<td>Woskow (1968)</td>
<td>—</td>
<td>Laymen</td>
<td>25</td>
<td>—</td>
<td>MDS</td>
<td>3</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Wright and Michels (1964)</td>
<td>84</td>
<td>n/a</td>
<td>45</td>
<td>—</td>
<td>MDS</td>
<td>3</td>
<td>No interpretation</td>
<td>—</td>
</tr>
<tr>
<td>Cunningham and Crady (1971)</td>
<td>—</td>
<td>n/a</td>
<td>14</td>
<td>—</td>
<td>EFA</td>
<td>4</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Berglund et al. (1973)</td>
<td>—</td>
<td>Experienced laymen</td>
<td>21</td>
<td>—</td>
<td>PCA</td>
<td>3</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Schiffman (1974a, 1974b)</td>
<td>Wright and Michels (1964)</td>
<td>84</td>
<td>n/a</td>
<td>45</td>
<td>MDS</td>
<td>2</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Woskow (1964)</td>
<td>—</td>
<td>Laymen</td>
<td>25</td>
<td>—</td>
<td>MDS</td>
<td>2</td>
<td>No interpretation</td>
<td>—</td>
</tr>
<tr>
<td>Moskowitz and Gerbers (1974)</td>
<td>—</td>
<td>Experienced laymen</td>
<td>15</td>
<td>—</td>
<td>MDS</td>
<td>2</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Yoshida (1975)</td>
<td>—</td>
<td>Laymen</td>
<td>32</td>
<td>—</td>
<td>PCA</td>
<td>7</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Schiffman et al. (1977)</td>
<td>—</td>
<td>Laymen</td>
<td>19</td>
<td>—</td>
<td>MDS</td>
<td>2</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Coxon et al. (1978)</td>
<td>—</td>
<td>Laymen</td>
<td>23</td>
<td>9</td>
<td>MDS</td>
<td>5</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Boelens and Haring (1981)</td>
<td>—</td>
<td>Experts</td>
<td>309</td>
<td>—</td>
<td>PCA</td>
<td>15</td>
<td>No interpretation</td>
<td>—</td>
</tr>
<tr>
<td>Jeltema and Southwick (1986)</td>
<td>—</td>
<td>Laymen</td>
<td>35</td>
<td>A 146</td>
<td>EFA</td>
<td>17</td>
<td>No</td>
<td>27</td>
</tr>
<tr>
<td>Dravnieks (1985)</td>
<td>507</td>
<td>Experts</td>
<td>144</td>
<td>A 146</td>
<td>EFA</td>
<td>17</td>
<td>No</td>
<td>27</td>
</tr>
<tr>
<td>Study</td>
<td>Analysis of Subjects</td>
<td>Classification</td>
<td>Method</td>
<td>Number of attributes</td>
<td>Number of reference odors</td>
<td>Reference methods</td>
<td>Results</td>
<td>Number of clusters</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------</td>
<td>----------------</td>
<td>--------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Chastrette et al. (1988)</td>
<td>Expert</td>
<td>2467 OProf</td>
<td>CLA</td>
<td>—</td>
<td>—</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abe et al. (1990)</td>
<td>Expert</td>
<td>1573 OProf</td>
<td>CLA</td>
<td>—</td>
<td>—</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrasco and Ridout (1992)</td>
<td>Laymen</td>
<td>16 PSim</td>
<td>MDS</td>
<td>3</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stevens and O’Connell (1996)</td>
<td>Experts</td>
<td>15 S</td>
<td>MDS</td>
<td>3</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prost et al. (2001)</td>
<td>Laymen</td>
<td>40 A</td>
<td>CA</td>
<td>4</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugiyama et al. (2006)</td>
<td>Laymen</td>
<td>17 PSim</td>
<td>MDS</td>
<td>3</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khan et al. (2007)</td>
<td>Experts</td>
<td>144 A</td>
<td>PCA</td>
<td>4</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dalton et al. (2008)</td>
<td>Laymen</td>
<td>30 SemD</td>
<td>PCA</td>
<td>3</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zarzo (2008a)</td>
<td>Laymen</td>
<td>40 A</td>
<td>PCA</td>
<td>5</td>
<td>Yes</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zarzo (2008b)</td>
<td>Experts</td>
<td>309 RefO</td>
<td>PCA</td>
<td>4</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zarzo and Stanton (2009)</td>
<td>Experts</td>
<td>309 RefO</td>
<td>PCA</td>
<td>2</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thiboud (1991)</td>
<td>Expert</td>
<td>119 OProf</td>
<td>PCA</td>
<td>2</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n/a, not available; A, attributes; RefO, reference odors; PSim, pairwise similarity; OProf, odor profile; S, sorting; CLA, cluster analysis; DA, discriminant analysis; CA, correspondence analysis; SOM, self-organizing maps.
perceptual ratings of odors. Distel et al. (1999) applied the same approach and extended the German–Japanese sample by 39 Mexican subjects. Again, ratings in pleasantness and familiarity were found to correlate. Distel and Hudson (2001) and Hudson and Distel (2002) could replicate these cross-cultural findings in 2 German samples: To control for prior experience, subjects were either tested with odors they had rated as familiar or unfamiliar in an earlier session (Hudson and Distel 2002) or they were asked to identify the presented odorants (Distel and Hudson 2001). Similar to the cross-cultural studies, subjects’ knowledge of an odor yielded increased pleasantness and intensity ratings, confirming the experience dependency of odor quality judgments. Effects of experience have also been shown for the classification of odors (Ueno 1993; Chrea et al. 2004, 2005). Ueno (1993) asked 20 Japanese and 20 Nepalese (Sherpa) participants to sort 20 Japanese food flavors based on their perceived similarity. The data analysis revealed that different from the Japanese sample, Sherpa did not apply a distinct category for “fishy” odorants. Ueno ascribed these differences to culture-specific experiences, namely, the fact that Sherpa rarely come in contact with fish odors in their daily routine. A more comprehensive study with a similar approach was performed by Chrea et al. (2004, 2005): They investigated the perceptual categories of 3 cultural groups (United States, France, Vietnam). Participants were asked to sort 40 odorants based on their perceptual similarity in as many groups as they felt necessary. The results showed several culture-specific arrangements that were explained with differences in nutrition and domestic life. However, these differences were mainly found in the assignment of single odors to classes. The general structure of the 3 olfactory spaces was similar. Chrea et al. (2004, 2005) thus provided empirical evidence for the basic universality of odor perception that has been proposed by several authors (Carrasco and Ridout 1993; Currie et al. 1999; Dawes et al. 2004).

In summary, olfactory ratings appear to be stable over short periods of time. Interestingly, Keller et al. (2012) reported that within-individual variability does not increase with longer time intervals. The variance between 2 measures is largely attributable to sniff-to-sniff changes, that is, processes in the range of seconds or minutes.

However, across the life span, olfactory perception may alter with physiological changes and cause intra- and interpersonal differences. In addition to physiological effects, experience accounts for interpersonal differences in odor perception or evaluation, respectively. Nevertheless, one can assume a basic universality in odor perception for people at comparable ages, with similar cultural backgrounds, and without olfactory deficiencies. Intercultural research has shown that culturally acquired experience mainly affects the evaluation of familiar versus unfamiliar odors rather than perceptual processes in general.

Not surprisingly, intra- and interindividual variances have also been observed for basic perceptual ratings in other sensory modalities, as in color vision (Viénot 1980; North and Fairchild 1993; Alfvin and Fairchild 1997). Nevertheless, color systems have often been based on perceptual data. An example is the widely applied color metric established by the International Commission on Illumination (CIE) in 1931 (CIE 1932) and 1964 (CIE 1964), respectively. Both CIE color spaces were established from color-matching experiments conducted by Wright (1929), Guild (1931), Stiles and Burch (1959), and Speranskaya (1959). Remarkably, Wright (1929) and Stiles and Burch (1959) reported considerable differences in the color-matching functions of the observers they had tested. Wright (1929) discarded the results of 10 participants due to “inaccuracy and unreliability” (p. 152). This general variance in color ratings has, however, neither prevented the development of the CIE color system nor induced a general debate on the applicability and validity of perception-based color systems.

Harper et al. (1968) summarized that an olfactory classification should be based on “some (specifiable) degree of agreement between different people” (p. 114) to be effective. Hence, odor researchers are advised to control for basic sources of variability, namely age, gender, and culture, both in the recruitment of participants and the analysis of (group) data. Apart from this, they may accept variance as a fundamental characteristic of perception that may not be factored out in classification systems.

Interpersonal differences in odor terminology

A basic requirement for language-based classifications is that people express their percepts similarly and apply verbal descriptions of odors in a similar way. Several authors who addressed the interrater reliability of verbal odor ratings reported high consistencies for both panelists with the same experience level (Dravnieks et al. 1978; Dravnieks 1982) as well as between trained and naive subjects (Jeltema and Southwick 1986). However, other studies reported differences in verbal ratings of experts and laymen and indicated that linguistic expressions of odor perceptions are inconsistent: 1) When compared with laymen, experts use further and more specific descriptors to verbalize their perceptions (Lawless 1984; Solomon 1990, 1997). 2) Different from nonprofessionals, experts can give verbal descriptions for odors that are matched with an appropriate stimuli by other experts (Lawless 1984; Solomon 1990). Some authors noted that these differences should be ascribed to enhanced perceptual skills (Parr et al. 2004). Others suggested that experience primarily affects the verbal and cognitive processing of odors (Hughson and Boakes 2001, 2002; Valentin et al. 2007). Training might enhance both perceptual and verbal skills. Its impact on language is especially strong not least because of the sparse olfactory terminology of untrained subjects. Harper and colleagues (1968) characterized the language people use to capture odors as “a borrowed one” (p. 84), “a language of substances and things”
Nonprofessionals even lack proper odor names and hence usually specify odors by their source. This may be the chemical substance (“amyl acetate”) or—more likely—the object that emanates a specific smell (“banana”) (Dubois and Rouby 2002). However, odors are poor retrieval cues for verbal labels. Laymen usually have major difficulties in naming even familiar odors correctly and identification rates rarely exceed 50% (Desor and Beauchamp 1974; Cain 1979; de Wijk and Cain 1994; Cain and Potts 1996; Cain et al. 1998). At the same time, odors are powerful cues for episodic memories (Chu and Downes 2000). Odor-related autobiographical memories can be recalled even without odor identification (Herz and Cupchik 1992). Thus, when people are unable to identify an odor, they normally express their olfactory perceptions by experiences they have gained with it: places or situations (“Christmas”), activities (“cleaning,” “baking”), effects (“relaxing”), or—on the most basic level—hedonic ratings (“pleasant”) (Schleidt et al. 1988; Dubois 2000; Rouby and Bensafi 2002). To facilitate a satisfactory communication despite this inaccuracy of everyday language, perfumers and fragrance companies have established a professional terminology. Along with this terminology, odor professionals have acquired cognitive categories that allow them to perceive a continuous space of odors in discrete conceptual categories. That is, experts are skilled in a categorical perception (CP) of odors. CP improves the discrimination among perceptual objects when these objects are assigned to different rather than the same categories. CP is a fundamental process in perception: It was first observed for color vision and has since been found in various perceptual domains (Harnad 1987). There has been a constant matter of scientific debate whether mental categories are innate and thus universal or learned and therefore experience dependent. Unquestionably, the perceptual classes applied by odor professionals are acquired. However, the degree to which they reflect “natural” odor categories or are completely arbitrary remains questionable. Various findings indicate that superior experience yields in odor arrangements that are distinct from the perceptive systems of nonprofessionals: 1) When data has been gathered from either experts or laymen, odor arrangements have varied with respect to the prevalence of a pleasantness factor. In the classification studies we reviewed, ratings by laymen have often yielded a hedonic dimension (Woskow 1968; Berglund et al. 1973; Moskowitz and Gerbers 1974; Yoshida 1975; Schifman et al. 1977; Coxon et al. 1978; Carrasco and Ridout 1993; Chrea et al. 2004, 2005), whereas those by odor professionals have usually not applied pleasantness as comparison criterion (Ennis et al. 1982; Jeltema and Southwick 1986; Zarzo and Stanton 2006; Zarzo 2008b; Zarzo and Stanton 2009). 2) When attribute lists have been provided by experts and applied by laymen, terms have been understood and used differently by the untrained subjects (Lawless 1984; Solomon 1990, 1997). 3) In nonverbal classification procedures, professional terminology has affected the interpretation of results when researchers imposed their acquired system on the data. Nonverbal data sets lack a verbal reference frame for the interpretation. Language and expectations may hence exert a particularly strong influence. Nevertheless, we also found arbitrary interpretations in the study of Jeltema and Southwick (1986) that was based on a verbal classification approach. One of the 17 dimensions they found was related to attributes like fresh green vegetable, crushed grass, green pepper, herbal, green, musty, earthy, moldy, celery. Jeltema and Southwick (1986) labeled it green—a term that has been commonly used in professional odor language (Edwards 2012; Sigma-Aldrich Company 2011). The dimension could, however, be labeled with terms that refer to the semantic arrangement of the attributes like garden, vegetable, fresh, organic, or ecological. To uncover the criteria applied by subjects, researchers might ask them to provide verbal labels for their-nonverbal arrangements (Stevens and O’Connell 1996; Chrea et al. 2004). This approach can both facilitate the interpretation process and help to uncover nonperceptual strategies of odor classification.

The inappropriateness of everyday language for the description of odor perceptions has initiated the development of an expert vocabulary. This terminology has facilitated a more objective communication on odors. At the same time, it may have affected odor arrangements when the linguistic or perceptual categories of laymen have been captured and possibly blurred by professional terms.

**Factor 2: Odorants**

**Quality and quantity**

In any study, researchers determine the scope of their results by defining the sample they assess. This has also been true for olfactory classifications: The selection of odors has determined the structure and meaning of odor arrangements. Hence, odors should have been selected to represent the full extent of olfactory space. However, as the organization of this space is under investigation, the matter of representativeness is vague and classification studies have dealt differently with this issue: Several studies of the works we reviewed did not report selection criteria at all (Wright and Michels 1964; Woskow 1968; Jeltema and Southwick 1986; Stevens and O’Connell 1996; Dalton et al. 2008), others chose odors according to a specific physicochemical criterion (Coxon et al. 1978) or presented compounds as diverse as possible from a perceptual or a chemical perspective (Berglund et al. 1973; Moskowitz and Gerbers 1974; Schifman et al. 1977). Some studies followed the approach of earlier works (Cunningham and Crady 1971; Yoshida 1975; Carrasco and Ridout 1993) or even selected odors in accordance with existing classification systems (Prost et al. 2001; Sugiyama et al. 2006). Without an objective selection criterion, the presentation of qualitatively and chemically various compounds seems most reasonable. Nevertheless, most odors applied in
classification studies belong to very specific quality categories, namely food (Gilbert and Greenberg 1992; Calkin and Jellinek 1994), flowers, and cosmetics. Given that the presentation of familiar odors to nonprofessionals is meant to facilitate the already demanding task of odor evaluation, this is comprehensible. Odor researchers should nevertheless consider that the exclusion or underrepresentation of specific odor classes, especially of unpleasant odors, will yield biased classification systems. This constraint to odor arrangements has only rarely been discussed in classification studies.

In order to represent the different odor qualities appropriately, a minimum number of odors is required. The studies we reviewed usually applied around 30 odorants (Table 1). With this sample size, researchers have usually found a trade-off between methodological requirements and practicability aspects. But it may be questioned whether this number is sufficient to represent a presumably high-dimensional olfactory space. Several studies have, therefore, analyzed existing data sets that have been established by odor professionals and comprise between more than 100 (Dravnieks 1985; Thiboud 1991) and several thousand (Arctander 1994; Sigma-Aldrich Company 2011) odorants. However, a large number of compounds is not necessarily more representative than a well-selected smaller stimuli set.

**Intensity**

In many classification studies, participants have been explicitly instructed to ignore potential differences in intensity when evaluating odors. This approach has been based on the assumption that intensity represents a distinct perceptual dimension—comparable, for instance, with color perception. However, in olfaction, the quality and intensity of a compound interact considerably and a shift in one dimension is often accompanied by a shift in another dimension: Whereas a color keeps its basic quality (blue) with increasing or decreasing intensity (light blue, dark blue), odors often change their quality with higher or lower concentrations. Thus, subjects might have difficulties in ignoring intensity effects—simply because they directly affect the sensation of quality. Gross-Isserhoff and Lancet (1988) found quality changes for 8 odors in a study with nonprofessional panelists. Subjects were asked to decide whether pairwise presented odors were identical. Although subjects were able to correctly identify pairs of the same odor in identical dilutions in over 90% of the trials, errors increased considerably when the very same odor was presented in different concentrations. In a more recent study, Laing et al. (2003) assessed 5 odorants at 7 different concentrations. They asked subjects to rate each sample against 145 descriptors and found a quality change with increasing intensity for 4 of the 5 tested odors. Hence, intensity is very likely neither a separate dimension outside a quality space nor congruent with a single quality dimension inside this space (Henion 1971). In a number of the classification studies we reviewed, intensity effects were controlled. In these studies, odors were presented in concentrations that had been rated as equally intense in a pretest (Berglund et al. 1973; Moskowitz and Gerbers 1974; Stevens and O’Connell 1996; Dalton et al. 2008). However, in various other studies, intensity effects were considered only marginally or not at all (Wright and Michels 1964; Wokow 1968; Cunningham and Crady 1971; Yoshida 1975; Schiffman et al. 1977; Coxon et al. 1978; Jeltma and Southwick 1986; Carrasco and Ridout 1993; Chrea et al. 2004; Sugiyama et al. 2006). This lack of control might have produced variance in the data that has been falsely ascribed to odor quality (Berglund et al. 1973). For future research, scientists should not only control for intensity effects by presenting compounds at equally intense dilutions. They should also keep in mind that the quality of some odors cannot be fully represented at a single intensity level and that these odors will not have a single distinct position in an olfactory space.

**Verbal cues**

Experience has been shown to influence quality perception by providing facts on the identity, function, or effect of an odor. Several studies have shown that this information can also originate from contextual cues like the color (Zellner et al. 1991; Gilbert et al. 1996; Lavin and Lawless 1998; Morrot et al. 2001; Sakai 2005) or verbal label attached to an odor. Herz and von Clef (2001) investigated the influence of verbal labels on odor descriptions by presenting identical odors with different labels in 2 several test sessions (violet leaf as “fresh cucumber” or “mildew”). Among other rating tasks, subjects were asked to report a memory evoked by each smell, to describe its function, and to generate a name for what they supposed the odors could be. Raters without knowledge of the experimental design evaluated whether these reports differed between the trials: Depending on the compound, 50–83% of the subjects actually changed their description as a function of verbal context. However, although some odors are susceptible to verbal influences, others have distinct perceptual features that are less affected by context information. From the 5 odors presented, menthol and pine oil were considerably less affected by the different labels than violet leaf, patchouli, and a 1:1 mixture of isovaleric and butyric acid. In addition to the direct impact of labels’ odor quality, numerous studies have confirmed an effect of verbal information on hedonic ratings (Moskowitz 1979; Lorig and Roberts 1990; Distel and Hudson 2001; Herz and von Clef 2001; Herz 2003; Lundström et al. 2006; Bensafi et al. 2007; Djordjevic et al. 2008). This, in turn, might influence intensity as well as quality features (Figure 1).

The impact of visual or verbal information has not been a methodological issue in classification studies as odors have been presented in neutral carriers without any labeling. However, along with the judgment of perceptual properties, subjects have very likely made assumptions what the source of an odor is and thus attached a label to them. If these
labels have been wrong or varied between subjects or trials, they have caused variance in a classification system falsely ascribed to odor quality.

Contextual effects

In addition to visual and verbal cues, the set of presented odors might act as reference frame itself and affect the evaluation of every single compound. Among various contextual biases in sensory judgments, contrast effects are probably the most common (Lawless and Heymann 2010). They occur when the perception of a stimulus characteristic is affected by the strength of this property in surrounding stimuli and shifted in the direction away from this context. Hulshoff Pol et al. (1998) demonstrated the effect for odor intensity ratings where an odor was judged weaker in the context of stronger stimuli and stronger in the context of weaker stimuli. Lawless (1991) and Lawless et al. (1991) confirmed the impact of perceptual context on quality judgments. They presented a citrus–woody odor (dihydromyrcenol) in a session with either prototypically citrus or prototypically woody odor and instructed participants to rate the stimuli against several quality descriptors. The ambiguous odor was rated as more woody in the citrus context and as being more citrus in the presence of woody compounds. Context effects also appear when only a single compound in a set is replaced: Kurtz et al. (2000) collected dissimilarity ratings for a set of 4 fixed odors (licorice, mint, mothballs, rose) and either vinegar or

Figure 1  Research on mutual effects of odor characteristics. Arrows indicate the direction of relations assessed.
rubbing alcohol. They found significant differences for ratings of the fixed compounds between both conditions with odors being rated as more similar in the presence of a very disparate smell (vinegar). Contrast effects are an ordinary mechanism of perception and appear to be unavoidable in sensory studies. However, because they decrease the reliability of judgments, they are highly undesirable for the development of comprehensive odor arrangements. To stabilize ratings, several of the reviewed classification studies have counterbalanced their stimuli sets (Wright and Michels 1964; Woskow 1968; Berglund et al. 1973; Coxon et al. 1978; Prost et al. 2001; Chrea et al. 2004) or even fully randomized the presentation order for each subject (Moskowitz and Gerbers 1974; Carrasco and Ridout 1993). These strategies have been good practice in sensory research for decades but whether they actually counteract contrast effects remains questionable. Lawless and Heymann (2010) stressed that balancing and randomization can counter order effects by changing the direct context of each odor. However, the overall context defined by the set of odors remains unchanged.

Naturally, a classification system is determined by the quality and number of odors presented. Hence, odors have usually been considered as representative samples of the odor space. However, in many of the reviewed studies, we have found an overrepresentation of specific quality classes that has undoubtedly yielded in fragmentary and biased odor arrangements. Remarkably, odor quality is not only a matter of the stimuli offered but also of how these are presented and to whom. The perceived quality of an odor is anything but a fixed characteristic that can be fully controlled by a careful selection. It rather changes with other odor characteristics, contextual information, and personal experiences. Even most basic aspects like an odor carrier and dilution (air, liquid) or the duration of its presentation may affect the perceptual evaluation. The entire list of interference factors is impossible to consider, but odor researchers are asked to both choose the very limited number of test odors thoroughly and to control for biases—not primarily to fully eliminate them, but to improve the understanding and valid interpretation of outcomes (Lawless and Heymann 2010).

**Factor 3: Method of data collection**

To collect perception-based data, researchers have usually applied approaches based on verbal or nonverbal judgments of odor characteristics. That is, odors have either been rated against a list of attributes or against other odors. However, due to the immense effort of these direct attempts, several classification systems have been established from odor profiles or based on the secondary analysis of previous data sets.

**Verbal profiling**

In verbal profiling approaches, odors are rated against a predefined list of verbal descriptors or semantic differentials (Cunningham and Crady 1971; Coxon et al. 1978; Jeltema and Southwick 1986; Prost et al. 2001; Dalton et al. 2008). Subjects are asked to express their olfactory sensations verbally, but instead of actively generating verbal descriptions, they evaluate odors against fixed references. A verbal approach restricts subjects to the qualitative aspects of an odor (that have been presented by the researcher). However, it also prevents panelists from deciding on the relevance of given attributes or applying individual comparison criteria. Several authors stressed that attribute lists should contain terms that are representative of the olfactory space as well as not associated (Gregson and Mitchell 1974; Civille and Lawless 1986). We examined the attribute lists applied in the verbal classification approaches of Pilgrim and Schutz (1957), Dravnieks (1985), Prost et al. (2001), and Zarzo (2008a). Out of 175 different descriptors, the vast majority pointed to odor sources (84.3%), 8.4% represented sense-specific qualities (fragrant, aromatic, rancid), and 6.7% described nonolfactory percepts (dry, heavy, sweet); 1 descriptor referred to pleasantness, 1 to an odor effect (full list is available from first author). From the different categories, only source labels refer to real, distinct percepts and hence seem to provide most applicable rating standards—especially for untrained panelists. However, these labels require subjects to compare an actual sensation (test odor) with an imagined odor (verbal label) and especially nonprofessionals have often reported difficulties in imaging odors (Stevenson and Case 2005). Hence, when subjects have been instructed to compare an odor with an attribute as “birch bark” (Dravnieks 1985), they might have pictured odor sources or appropriate situations (walk in the woods, collect mushrooms) rather than a distinct smell. Odor scientists should thus keep in mind how a list of verbal descriptors always provides a definition of what odor quality is (Moskowitz and Gerbers 1974). These definitions might over- or underrepresent certain perceptual dimensions and could be processed differently by different subjects.

**Similarity ratings**

**Pairwise similarity**

The pairwise similarity is evaluated in each possible dyadic combination of a set of odors on numerical or visual rating scales (Woskow 1968; Berglund et al. 1973; Gregson and Mitchell 1974; Moskowitz and Gerbers 1974; Schiffman and Dackis 1976; Schiffman et al. 1977; Jeltema and Southwick 1986; Carrasco and Ridout 1993). As the classification of \( n \) stimuli requires \( n(n-1)/2 \) comparisons, the method is highly time consuming and considerably restricts the selection of test compounds. Similarity ratings are independent of a verbal reference system provided by the researcher. They rather allow panelists to apply an individual definition of relevant quality features. However, these criteria usually remain unknown. Instead of evaluating odor similarities
with respect to fixed and distinct criteria, ratings are rather based on a single (outstanding) characteristic or a general perceptual impression. Davis (1979) and Gregson (1972) showed that panelists actually differ in the comparison strategies they apply. Not surprisingly, several studies reported a poor agreement among subjects in pairwise ratings of odors (Yoshida 1964; Gregson 1972; Berglund et al. 1973). Hence, nonverbally established classification systems may reflect perceptual dimensions as well as a subject’s lack of clear comparison criteria in the test situation.

Sorting

Lawless (1989) applied a time-efficient alternative to pairwise similarity ratings: He adopted a method used in personality research (Rosenberg and Park Kim 1975) and asked subjects to sort odorants based on their similarity in as many groups as they considered necessary. An index of similarity is derived across all panelists from counting the joint occurrence of any possible pair of odors in the same group. Hence, a gradation among similarity ratings emerges from the agreement or disagreement between subjects. However, this index is not only affected by the allocation of odors to groups but also by the number of groups created. Subjects might differ in the number of groups they form due to different mental strategies of sorting. Nevertheless, odor sorting provides important advantages: It is less time consuming than direct similarity ratings and minimizes perceptual fatigue. For these reasons, recent studies have usually collected nonverbal similarity data based on this approach (Lawless and Glatter 1990; MacRae et al. 1990, 1992; Stevens and O’Connell 1996; Dubois 2000; Chrea et al. 2004; Higuchi et al. 2004).

Some authors have asked participants to provide labels for samples that have been characterized with about 270 different attributes. Zarzo and Stanton (2009) noted that panelists actually differ in the comparison strategies they apply. Thiboud (1991) distinguished objective (olfactory quality) from subjective attributes (individual associations on origin, function, or effect). However, as all odor descriptors and contain verbal descriptions for hundreds or thousands of odors. They provide information on odor similarities that are calculated from the co-occurrence of attributes on different odors. Hence, several classification systems have been based on these data sets (Chastrette et al. 1988; Abe et al. 1990; Madany Mamlouk et al. 2003; Madany Mamlouk and Martinetz 2004; Zarzo and Stanton 2006; Zarzo and Stanton 2009).

One of the most comprehensive odor profiles was developed by Arctander (1969). It contains as much as 3102 odor samples that have been characterized with about 270 different attributes. However, subjectivity is a basic constraint of the data set as each odor was described by Arctander himself. Empirical evidence for the distortion of the data was provided by Pintore et al. (2006) who found considerable disagreements between Arctander’s work and a commercial database. A second catalog that has been originated from the expertise of a single expert was established by Thiboud (1991): 119 compounds are described by 3–4 main and several secondary notes; a total of 85 different descriptors were applied. Thiboud (1991) distinguished objective (olfactory quality) from subjective attributes (individual associations on origin, function, or effect). However, as all odor descriptions have been based on a single expert’s opinion and lack an external validation criterion, this distinction has been rather theoretical. A third database that has been applied in several classification studies is published and regularly updated by Sigma-Aldrich. The latest version of the catalog (Sigma-Aldrich Company 2011) comprises profiles of more than 1600 aroma raw materials that have been characterized with 82 attributes. Zarzo and Stanton (2006) noted that these profiles have been generally acquired from literature, odor expert reports, or other odor profiles, respectively.
Hence, it remains largely unknown how exactly each odor has been characterized and by whom. With the *Atlas of Odor Character Profiles*, Dravnieks (1985) published an extensive database from expert ratings: He carefully developed a 146-attribute list (Dravnieks 1975; Dravnieks et al. 1978) and asked a total of 507 perfumers and odor scientists (120–140 experts per odor) to evaluate 160 odors against it. Given that the data set has been based on a verbal profiling approach applied by trained raters, it is not surprisingly that odor descriptions have been found to be highly reliable (Dravnieks 1982).

Several researchers have addressed the impact of data collection approaches on classification systems. Verbal approaches have sometimes been criticized for being biased by linguistic references that may be understood differently and hence decrease interrater agreement. Schiffman and Dackis (1976) and Schiffman et al. (1977) compared direct similarity ratings to judgments on semantic differentials. In both studies, they found that the similarity data “was virtually identical for each subject” (Schiffman et al. 1977), whereas most “semantic differential ratings tended to differ—more so for nonverbal techniques” (Yoshida 1964; Gregson 1972; Berglund et al. 1973). Still others noted that verbal and nonverbal approaches generally yield comparable results (Stevens and O’Connell 1996; Moskowitz and Gerbers 1970) that does not average data across subjects. However, other studies demonstrated high interrater agreement for verbal ratings (Dravnieks 1982; Jeltma and Southwick 1986) and considerable interindividual variety for nonverbal techniques (Yoshida 1964; Gregson 1972; Berglund et al. 1973). Still others noted that verbal and nonverbal approaches generally yield comparable results (Stevens and O’Connell 1996; Moskowitz and Gerbers 1974; Higuchi et al. 2004). Higuchi and collaborators (2004) explained this agreement with the quality of their attribute list; high reliabilities found for Dravnieks’ extensive list support this assumption. These findings suggest that verbal approaches are highly affected by the soundness of descriptors rather than generally inferior to nonverbal methods. It is debatable whether verbal attributes can appropriately reflect the quality features applied by nonprofessionals. They, however, provide a reference frame for odor evaluations that might be especially important to guide untrained subjects. Nonverbal approaches require raters to decide on comparison criteria themselves. These will reflect natural perceptual dimensions more appropriately than predefined verbal descriptors but they may change for every new pair of odors and they may remain largely unknown—to the participants as well as to the researcher. Beyond that, one might question whether similarity ratings and sorting procedures are truly free of verbal influences. Even though the verbal mediation of olfactory processes has not been definitely resolved, the comparison of 2 compounds may require the mental search through verbally represented criteria. A nonverbal odor comparison thus involves perceptual as well as verbal processes. Hence, when people differ in the perception and verbalization of odors, variance in the data set cannot be reliably attributed to an (assumed) inaccuracy of language-based methods.

### Factor 4: Method of data analysis

Olfactory classifications have usually been based on the assumption that odors are projections in an n-dimensional space where their position comprehensively describes the perception they elicit. Thus, classification studies have usually applied analysis approaches that either search for a parsimonious but meaningful dimensionality of the data set, like exploratory factor analysis (EFA), principal component analysis (PCA), and multidimensional scaling (MDS), or summarize odors to more homogenous groups, like cluster analysis.

#### Multidimensional scaling

MDS refers to a group of exploratory data analysis approaches that attempt to visualize the underlying structure of high-dimensional data sets. It represents objects as points of a (preferably) low-dimensional space in a way that interpoint distances best match the measured (dis)similarities of associated objects. The more similar objects are, the closer their points are located in this space and vice versa. In classification studies, these objects have usually been odors and the spatial map a representation of the olfactory space. In the studies we reviewed, similarity measures were either obtained directly from pairwise similarity ratings (Woskow 1968; Døving 1970; Schiffman 1974a, 1974b; Yoshida 1975; Schiffman et al. 1977; Carrasco and Ridout 1993) and sorting tasks (Stevens and O’Connell 1996; Chrea et al. 2004; Sugiyama et al. 2006) or calculated from odor profiles (Madany Mamlouk et al. 2003; Madany Mamlouk and Martinetz 2004) and attribute ratings (Coxon et al. 1978) by correlations. When Lawless (1989) first applied a sorting task to generate MDS data, he raised the question whether different data collection approaches yield comparable olfactory spaces. Several studies addressed this issue and found differences in MDS solutions caused by different scale levels of the data (Rao and Kaltz 1971; Humphreys 1982; Bijnol and Wedel 1999). Classification studies have usually calculated MDS spaces from averaged group data (Woskow 1968; Døving 1970; Schiffman 1974a, 1974b; Coxon et al. 1978) because interrater variance has been treated as noise. This assumption, however, has disregarded that there are differences between subjects that should manifest in some way in an odor classification. Ashby et al. (1994) assessed how averaging considerably changes the underlying structure of a data set. In a simulation study, they demonstrated a good fit of averaged data to standard MDS models, whereas these models failed to represent the data of any individual subject appropriately. They emphasized that similarity measures should be analyzed either on subject level or treated with MDS procedures like INDSCAL (Carroll and Chang 1970) that does not average data across subjects. However, only few of the classification studies we reviewed, applied an individual scaling approach to verify their group results or at
least considered the impact of averaging explicitly (Yoshida 1975; Schiffman et al. 1977; Carrasco and Ridout 1993). Hence, odor arrangements have possibly factored out differences between subjects with the application of classical MDS models. In addition, the major aim of MDS to visually represent complex data sets has likely biased the results of several studies.

Although the quality of an MDS solution is usually assessed by its stress value (Kruskal 1964) or the squared correlation index (\(R^2\)), in practice both criteria have frequently been traded off against low dimensionality. Most classification studies that applied MDS actually reported 2- or 3-dimensional olfactory spaces (Table 1). However, theoretical assumptions (Harper et al. 1968; Chastrette 2002) and other methods of analysis like EFA and PCA have suggested considerably more perceptual dimensions (Table 1). Jeltema and Southwick (1986) noted that MDS might provide less useful results than PCA or EFA because “MDS dimensions cover multiple sensory dimensions which were pulled apart by factor analysis” (p. 133). Independent from their accuracy, low-dimensional solutions have often facilitated visually driven interpretations that have been based on the (visual) allocation of odors along the displayed dimensions (Moskowitz and Gerbers 1974; Schiffman 1974a; Yoshida 1975; Coxon et al. 1978). This approach, however, lacks objective criteria and is likely biased by expectations on (over)simplified data structures. Several studies have thus attempted to minimize subjectivity by performing further analyses on the results (Stevens and O’Connell 1996; Chrea et al. 2004) or regressing MDS solutions against attribute ratings acquired in separate experimental trials (Woskow 1968; Schiffman et al. 1977; Carrasco and Ridout 1993; Higuchi et al. 2004). However, this approach requires both the soundness of applied descriptors as well as the comparability of verbal and nonverbal procedures with respect to their results. It thus remains questionable whether it has truly yielded more objective results.

Despite these constraints, MDS has provided several advantages for classification studies as it processes ordinal raw data to multidimensional maps and thus yields quantitative information on odor similarities without assuming linear relations or multivariate normality.

**Principal component analysis and exploratory factor analysis**

The general purpose of PCA is to reduce the dimensionality of a complex data set that comprises values for \(n\) objects on \(p\) interrelated variables. In order to reveal a simplified structure, new uncorrelated variables—the principal components (PCs)—are calculated. PCs are linear combinations of the original variables and explain the entire variance of the data in successively decreasing proportions. In other words, an original data represents \(n\) points in a \(p\)-dimensional space. PCA searches for \(m\) (\(m < p\)) dimensions that provide an alternative description of the data points minus the redundancy expressed by intercorrelations between the original variables. From the studies we reviewed, 10 analyzed their data with PCA (Berglund et al. 1973; Yoshida 1975; Boelens and Haring 1981; Ennis et al. 1982; Zarzo and Stanton 2006; Khan et al. 2007; Dalton et al. 2008; Zarzo 2008a, 2008b; Zarzo and Stanton 2009) and 3 reported the application of EFA (Wright and Michels 1964; Cunningham and Crady 1971; Jeltema and Southwick 1986). This prevalence of PCA over EFA does not just apply to classification studies but has also been found in other areas of psychological research (Ford et al. 1986; Fabrigar et al. 1999; Conway and Huffcutt 2003; Costello and Osborne 2005). This preference has not necessarily been justified. Very often, there has even been a lack of differentiation between both methods: In the literature as well as in statistical software packages, PCA has often been considered as (default) extraction method of EFA (Ford et al. 1986). However, both approaches are based on distinct mathematical assumptions even though often yield fairly similar results. EFA assumes a common factor model and searches for those latent factors that cause the correlation of original variables. Each manifest variable is expressed as linear combination of common factors that explain the shared variance plus factors of unique variance and measurement error. PCA is not based on a statistical model and uses the entire variance of the original variables to calculate PCs. These PCs account for shared variance as well as unique variance and measurement error. Hence, although EFA primarily searches for interpretable dimensions, PCA is aiming at a computational data reduction. However, in classification studies, PCs have usually been interpreted as perceptual dimensions and their labels have sometimes even exceeded the meaning of the original variables. Zarzo (2008b) and Zarzo and Stanton (2009) labeled a dimension *feminine versus masculine* that discriminated *floral, fruity* from *earthy, dusty* odors. Although similarity ratings had been collected from odor experts, the appropriateness of these labels remains questionable. Other investigators have successfully reduced the impact of subjectivity on interpretations by underpinning their assumptions with independent criteria (Khan et al. 2007). Despite these measures, expectations remain a central issue of most classification studies: Especially the search for a pleasantness factor has considerably affected numerous odor arrangements.

To our knowledge, no study has yet compared the applicability of PCA and EFA to classification data. However, classification studies have usually not explained why they applied either method. One reason for preferring PCA over EFA may be to determine the position of odors in an olfactory space in addition to identifying its relevant dimensions. This is, however, not easily attained by EFA: For a given odor, several factor scores (positions in the odor space) can be calculated that perfectly fit the factor model but vary considerably. This issue has been discussed as factor indeterminacy problem in the common factor model (for a review, see...
have decided on a final number of clusters (Døving 1970; Chastrette et al. 1988; Abe et al. 1990) but without reporting their decision criteria appropriately or at all.

In summary, cluster analysis has played a minor part in olfactory research. Classification studies have strongly focused on the number and character of perceptual dimensions and thus mainly applied PCA, EFA, or MDS on their data. Many of the reviewed studies—especially early works—missed to report and explain their choices throughout the analysis process sufficiently. This has not only complicated the replication of these studies but also a comprehensive understanding and debate on their attempts. How substantially these decisions affect the meaningfulness of outcomes has been shown by studies that applied the same analysis approach to identical data sets and yet yielded considerably different results: Døving (1970) and Schiffman (1974a, 1974b) applied MDS to a data set established by Woskow (1964) and found a 4- and 2-dimensional olfactory space, respectively. The data set of Boelens and Haring was analyzed with PCA in several studies (Boelens and Haring 1981; Ennis et al. 1982; Zarzo 2008b; Zarzo and Stanton 2009) and yielded 2–15 perceptual dimensions. To allow an informed choice on the appropriateness of classification results, odor researchers are required to both provide detailed information on their analysis approaches and consider potential limitations more thoroughly.

Naturally, the use of a specific method is determined by the research question and data obtained. Interestingly, recent studies have predominantly performed PCA on their data. We have questioned the adequacy of a purely computational data reduction as an approach to meaningful perceptual dimensions. Based on the research reviewed, we give preference to MDS and EFA for the analysis of nonverbal data sets and profile data, respectively. Generally, researchers should consider that methods successfully applied in “higher senses” may not be appropriate for odor research questions. Especially, the predominant averaging of data in classification studies has ignored the specific characteristics of odor perception (Köster 2002) and yielded biased and incomplete arrangements.

Conclusion

Odor classifications—especially early studies—have often been guided by the efforts and notions of color systems, where a manageable number of receptor types and relevant physical dimensions have constituted neat, low-dimensional arrangements of stimuli and comparatively few primary colors (Harper et al. 1968). Color classifications have condensed a complex perceptual reality to a few dimensions along with a minimum loss of relevant information. Odor professionals have searched for systems with the same lucidity. But to this day none of them has sustained scientific scrutiny. Perceptual–verbal approaches of odor classification have been especially prone to errors and biases. We identified
a vast number of influencing factors that can be grouped to 4 main categories with respect to interindividual differences, stimuli characteristics, methods of classification, and methods of analysis. We could neither confirm nor disprove the general findings on structure and dimensionality stated by Chastrette (2002). Especially the number and character of perceptual dimensions remains a matter of debate. Although Chastrette (2002) assumed a high-dimensional olfactory space, most of the reviewed studies reported between 2 and 4 perceptual dimensions. Investigators have largely agreed that among these, one dimension reflects the pleasantness of odor perceptions: In 14 of the 28 reviewed studies, pleasantness was found to be a dominant factor of odor perception.

**Prevalence of pleasantness in classification studies**

The prevalence of a hedonic dimension in olfactory spaces may have different reasons. One might argue that pleasantness primarily reflects an inappropriateness of everyday language that causes a lack of description standards for odor perceptions (in an experimental setting). In other words: In absence of clear descriptors and ratings standards, panellists may confine themselves to the most basic attribute of an odor—its pleasantness. Evidence for this assumption has been provided by the results of nonverbal classification studies: All but one of the reviewed studies that applied pairwise similarity ratings revealed a primary pleasantness dimension, whereas verbal methods found a hedonic factor considerably less frequent. Beyond that, experts have only rarely stated pleasantness as substantial perceptual quality of odors, whereas studies with nonprofessionals have almost always yielded a hedonic dimension. However, whether ratings by odor professionals reflect the “true” character of an odor space more precisely than those of laymen, is questionable: On the one hand, odor experts are highly experienced in odor evaluations and barely misled by the absence of a (verbal) reference frame. On the other hand, their terminology explicitly excludes hedonic ratings; perfumers are simply trained to disregard the hedonic tone of odors. Pleasantness may have also been introduced to classification systems by researcher expectations. Many investigators have focused on confirming the existence of a hedonic factor that has been based on theoretical considerations or proposed by previous studies.

There has, however, been evidence for the assumption that pleasantness is a significant aspect of odor perception. Firstly, the neuronal processing of odors and emotions are partly overlapping in limbic structures (Gottfried 2006). Secondly, the close connection is rooted in a point of human evolution when odors primarily provided information on food, mates, natural predators, or kinship—in other words, when they informed on what to approach and what to avoid. Herz (2005) proposed that this evolutionary meaning has caused the weak connection of odors and language. First and foremost our ancestors had to learn how to respond to odors. Names or other perceptual features did, however, not provide essential information. More recent, several studies assessed emotional response to odors more closely (Chrea et al. 2009; Ferdenzi et al. 2011; Delplanque et al. 2012) and actually provided evidence for this assumption. Affective responses to odors do not mirror basic human emotions but rather reflect the “role of olfaction in well-being, social interaction, danger prevention, arousal or relaxation sensations” (Chrea et al. 2009). These findings suggest that pleasantness is a generic factor of olfactory perception, which is subdivided in more specific facets related to the functions and effects of odors on humans.

The prevalence of pleasantness in perception-based olfactory systems may be caused by the “helplessness” of (untrained) subjects in odor rating tasks, it may have been introduced to experimental settings by assumptions of odor researchers or reflect the true importance of pleasantness in odor perception. Currently, a combination of all factors is likely, but more research is needed to properly judge the role of hedonic in olfaction.

**Influencing variables**

We illustrated how each perception-based odor arrangement has been determined by basic characteristics of study design, sampling, and data analysis. Hence, odor systems reflect the relation of perceptual qualities as well as the conditions under which these qualities have been assessed. Even if a sufficiently representative yet relatively small set of odors can be found, the sensations of these odors are anything but accurate reflections of actual stimuli. Subjects classify a mental representation of an odor that is shaped by various interactions between odor characteristics as well as the impact of interindividual differences in age, knowledge, culture, and so on. When panelists agree on their perceptions, the applied methods of classification may still overrate or underemphasize certain quality aspects. Finally, different approaches of data analysis yield in results that might not appropriately reflect the mental odor categories of laymen when they are interpreted by the means of professional terminology.

Although odor researchers have been faced with these issues for more than 5 decades, neither a debate on the general appropriateness of perception-based methods nor adjustments of the applied approaches have been initiated. Instead, the scientific interest in a perceptual arrangement of odors has decreased in the last years. Perception-based attempts have not delivered the anticipated results and the topic of odor classification appears to be off the table. More recent studies have focused on specific domains of this space (Chrea et al. 2009; Ferdenzi et al. 2011; Delplanque et al. 2012) or applied neuroimaging techniques to eventually uncover the rules of odor coding and arrangement (Gottfried et al. 2006; Howard et al. 2009). However, the “more objective” methods are incapable of discriminating the perceptual aspects of olfaction. Psychological classifications are still
needed to understand the outcomes of these studies. Odor researchers should thus not lose sight of perception-based classification and focus on new approaches for establishing them. Impulses may be provided by the evolutionary functions of odors or their effects on human behavior (Holland et al. 2005; Liljenquist et al. 2010). Beyond this, more general principles of object categorization may be applied to odor perception. The concept of CP, for example, has been useful in understanding the processing of visual and auditory stimuli. CP explains how the mental arrangement of objects in cognitive classes affects the perception of their similarity and their mental processing, respectively. With respect to odor classification, it may facilitate the identification of classes and their relations to one another. As mental classes provide information on the distinctiveness of stimuli that overrule their actual physical similarity, observers should make faster judgments and fewer mistakes in discrimination tasks for odors from different mental categories. The acuity of their answers should peak at the boundary between 2 adjacent groups. Discriminability may also vary within a category given that some odors are perceived as more typical than others (Chrea et al. 2005). Psychophysical or behavioral approaches will, however, not directly address the origin of these categories. Whether odor categories are innate or acquired depends on the influence of language on odor processing. Hence, the unique and complex interaction between language and olfaction should be assessed more thoroughly. Several authors noted that odor classifications have been particularly affected by the linguistic or semantic arrangements of (supposed) odor sources rather than the sensory characteristics of odors (Chastrette et al. 1988; Lawless 1989; Prost et al. 2001; Chrea et al. 2005). Some have argued that odors are generally processed perceptually and only arranged semantically when (verbal or visual) identifiers are available (Chrea et al. 2005; Herz 2005). However, when subjects search for criteria to compare, sort or evaluate odors, they will primarily look for information on odor sources. If this information becomes available from contextual cues or memory, it will dictate perceptual ratings (Lorig and Roberts 1990; Distel and Hudson 2001; Herz and von Clef 2001; Herz 2003; Rolls et al. 2003; Lundström et al. 2006; Bensafi et al. 2007; Djordjevic et al. 2008). We thus raise the question whether any olfactory system will be unbiased by the linguistic classification of odor sources, or even more so, if olfactory systems may be in fact linguistic arrangements. Hence, the linguistic taxonomy of odor sources might be a close (if not the closest) approximation to the mental–perceptual arrangement of odors. To address this assumption, further research on the interdependency of semantic categorizations on perceptual ratings is certainly needed.

References


Psychol Res. 6:115–124.

1:443–464.

Yousem DM, Maldjian JA, Siddiqi F, Hummel T, Alsop DC, Geckle RJ, Bilker
WB, Doty RL. 1999. Gender effects on odor-stimulated functional mag-

Zarzo M. 2008a. Psychological dimensions in the perception of everyday

Zarzo M. 2008b. Relevant psychological dimensions in the perceptual space

Zarzo M, Stanton DT. 2006. Identification of latent variables in a semantic
odor profile database using principal component analysis. Chem
Senses. 31:713–724.

Zarzo M, Stanton DT. 2009. Understanding the underlying dimensions in
perfumers’ odor perception space as a basis for developing meaningful

Zellner DA, Bartoli AM, Eckard R. 1991. Influence of color on odor identifi-

Zozulya S, Echeverri F, Nguyen T. 2001. The human olfactory receptor rep-
ertoire. Genome Biol. 2:research0018.1–research0018.12.