Food Quality and Preference 21 (2010) 1077–1087

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Comparison of PLS dummy variables and Fishbone method to determine optimal product characteristics from ideal profiles

Thierry Worch [a,*], Lauren Dooley [b], Jean-François Meullenet [b], Pieter H. Punter [a]

[a] OP&P Product Research, Burgemeester Reigerstraat 89, NL-3581 KP Utrecht, The Netherlands
[b] Department of Food Science, University of Arkansas, 2650 N. Young Avenue, Fayetteville, AR 72704, United States

ARTICLE INFO

Article history:
Received 31 August 2009
Received in revised form 31 July 2010
Accepted 10 August 2010
Available online 16 August 2010

Keywords:
PLS dummy variables
Fishbone plots
Ideal profile
Just About Right
Consumer liking

ABSTRACT

Sensory professionals mostly used trained or expert panels for diagnostic purposes and only use consumers for hedonic assessments. In market research, consumers are not only used for hedonic assessments, but also for product diagnostic purposes (as is the case with Just About Right procedure). In the Ideal Profile method, consumers are also used for both tasks. They rate the perceived and ideal intensities and the acceptance of a series of products. The two sets of information (product description and hedonic data) are then used for product improvement. In this paper, the results of two methods of analysis – PLS on dummy variables and Fishbone method – will be compared. The objectives of this study were (1) to compare PLS and Fishbone method to determine their similarity in predicting the impact of the attributes on overall liking, and (2) to determine if the methods would return similar or contrasting conclusions. These methodologies have been applied to a study concerning 12 commercially-available women's perfumes. Though the differences in model derivations caused some small dissimilarities, similar trends were found between products across methods for those perfumes far from the ideal. There was agreement regarding which attributes are too strong or too weak, and the order of importance of these attributes for liking. The two methods showed greater dissimilarity for perfumes that were already near the consumer's ideal.

1. Introduction

In the sensory world, it is common practice to use experts or trained panelists for the sensory description of products (i.e. quantitative descriptive analysis or QDA®), and consumers for the hedonic evaluation. It is believed that consumers are not capable of accurate sensory characterization of products without training. In the literature, many warnings are given concerning the use of consumers for the sensory description of products:

– "…as with any untrained panel, beyond the overall acceptance judgment there is no assurance that the responses are reliable or valid" (Stone & Sidel, 2004)
– "…consumers can only tell you what they like or dislike" (Lawless & Heymann, 1999)

However, QDA and associated methods, done with experts or trained panels, can be expensive and time consuming (both in terms of panelist training and testing time). More expeditious and still efficient methods with consumers (Abdi, Valentin, Chollet, & Chrea, 2007; Gazano, Ballay, Eladan, & Sieffermann, 2005; Healy & Miller, 1970; Nestrud & Lawless, 2008) have been presented in recent years, including Flash Profile (Sieffermann, 2002), Napping® (Pagès, 2005) and Free Sorting Tasks.

Husson, Le Dien, and Pagès (2001) showed that consumers meet the requirements of discrimination, consensus and reproducibility, Worch, Lê, and Punter (2009) found no significant differences between products profiled by trained or consumer panels, and Moskowitz (1996) showed that consumers can be used for the sensory description of sauces, and therefore "refutes the notion that consumers are incapable of validly rating the sensory aspects of products".

Hence, the use of consumers for profiling might be an alternative for experts or trained panelists.

Still, many authors doubt if consumers possess the capability to profile products and suggest using consumers only for the evaluation of attributes which are easily detectable and differentiated (Meilgaard, Civille, & Carr, 2007; Stone & Sidel, 2004).

In order to put the best possible product on the market, it is essential to understand consumer product perception and preferences, and relate hedonic responses to sensory product specifications. Several methods have been developed recently to define and characterize a "consumer ideal product". The general idea behind these methods is to extract liking information from consumers and link this information to the sensory characteristics of the

* Corresponding author.
E-mail address: thierry@opp.nl (T. Worch).

0950-3293/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2010.08.009

products obtained from trained or expert panels. The characteristics of the ideal product are estimated through statistical methods. Among these methods are external preference mapping (Greenhoff & MacFie, 1994) and unfolding (Coombs, 1976).

As the need for ideal product understanding becomes more important in the food industry, methods which directly measure ideals have been introduced (Issanchou, 2009; Meullenet, Xiong, & Findlay, 2007). The Just About Right (JAR) method, in which consumers are asked to rate a product's intensity relative to their ideal, uses an implicit ideal. In this case, consumers are asked to indicate – for a number of attributes – if the product they tasted is Just About Right, too little/much or far too little/much compared to their personal ideal product. When consumers can express the difference of the perceived intensity from an implicit ideal, it is assumed that they have a good representation and understanding of their personal ideal, and can rate it directly (Ideal Profile method). In this method, consumers are asked to rate both the perceived and the ideal intensities for each attribute and each product. At the end of a session, each consumer, who tested P products, will yield P perceived and P ideal intensities for each attribute.

Asking the ideal intensities for only one reference product or without a reference product is not recommended since ideals are influenced by the perceived intensities and we want the exact differences between perceived and ideal intensities for each consumer. Compared to the JAR scale, which only asks the deviations from ideal for each attribute and product combination, in the Ideal Profile method, both perceived and ideal intensities are asked directly (the JAR question 'is it just right, too much or too little' is replaced by the question 'how strong is it and what would be the ideal strength').

Disagreement exists in the sensory community as to whether or not consumers are capable of understanding and/or specifying their ideal. Some believe that consumers are unable to express their needs, are unaware of their needs or have a limited frame of reference (Stone & Sidel, 2004). Others are less strict and demonstrated limitations of the use of Just About Right, without completely rejecting it (Epler, Chambers, & Kemp, 1998). Others believe that consumers know their likes and dislikes and can explain the reasons for their opinions (Moskowitz, Munoz, & Gacula, 2003). Most market researchers belong to this latter school of thought.

Recently, Van Trijp, Punter, Mickartz, and Kruithof (2007) compared three methods to obtain ideal profiles: (1) JAR, (2) the conventional method (combining consumer liking with expert profiles) and (3) a variant (the Ideal Profile method presented in this paper) and found a high robustness of the ideals despite the methodological dissimilarities.

In this study, the authors agree with the assumption that consumers are able to describe products and specify their ideal. They compare two different methodologies that can be used to analyze JAR or Ideal Profile data, and which give guidance for product improvement, based on the deviations from the ideal levels and the relative importance of each attribute for liking. For the analysis of JAR data, PLS on dummy variables can be used (Xiong & Meullenet, 2006). For the analysis of Ideal Profile data, a method based on regression on principal components can be used (the Fishbone method).

2. Material and methods

2.1. Notation

In this document, the following notation is used (bold type being used for vectors):

y_{jpa}	intensity perceived by consumer j for the product p on attribute a
$\mathbf{y_{jp}}$: = $\{y_{jpa};a = 1{:}A\}$	vector of intensities perceived for all attributes by consumer j on product p
$\bar{y}_{.pa}$	averaged intensity perceived by all the consumers of the product p and the attribute a
z_{jpa}	ideal intensity rated by consumer j for the product p on attribute a
$\mathbf{z_{jp}}$:$\{z_{jpa};a = 1{:}A\}$	vector of ideal intensities rated for all attributes by consumer j on product p
$\bar{z}_{.pa}$	averaged ideal intensity rated by all the consumers of the product p and the attribute a
l_{jp}	overall liking given by the consumer j for product p
$\mathbf{l_p} = \{l_{jp};j = 1{:}J\}$	vector of liking ratings given by the J consumers on product p
$\mathbf{l} := \left\{ \begin{array}{c} l_{jp};j=1:J \\ p=1:P \end{array} \right\}$	vector of liking ratings given by the J consumers for the P products. In this case, $\mathbf{l}$ is appended vertically ($J \times P$ rows and 1 column).

2.2. Materials

The materials used in this study are the same as presented in Worch et al. (2009). Twelve commercially-available luxury perfumes (see Table 1) were tested by 103 Dutch consumers (44 men and 59 women – 48 between 18 and 35 years old, 55 between 45 and 60 years old) at OP& P Product Research (Utrecht, the Netherlands). The women were daily luxury perfume users, and the men had a girlfriend or wife who used perfume regularly. Additionally, the men had to name at least two luxurious women perfumes. This criterion was added so they were direct or indirect users of this type of product.

Two products (Pure Poison and Shalimar) were replicated to test for panelist consistency (a total of 14 samples were tested). A balanced design and sequential monadic serving order was followed to account for order and carryover effects (MacFie, Bratchell, Greenhoff, & Vallis, 1989). The study was performed on two separate days, with seven products being tested each day. Perfumes were sprayed onto cotton pads, placed in lidded Styrofoam® cups and replaced hourly.

The consumers rated both perceived and ideal intensities on 21 attributes: Odor intensity, Freshness, Jasmin, Rose, Camomille, Fresh lemon, Vanilla, Mandarin/Orange, Anis, Sweet fruit/Melon, Honey, Caramel, Spicy, Woody, Leather, Nutty/Almond, Musk, Animal, Earthy, Incense and Green. A 100 mm unstructured line scale,

Table 1
List of the products.

Product	Type
Angel	Eau de Parfum
Cinema	Eau de Parfum
Pleasures	Eau de Parfum
Aromatics Elixir	Eau de Parfum
Lolita Lempicka	Eau de Parfum
Chanel N5	Eau de Parfum
L'Instant	Eau de Parfum
J'Adore (EP)	Eau de Parfum
J'Adore (ET)	Eau de Toilette
Pure Poison	Eau de Parfum
Shalimar	Eau de Toilette
Coco mademoiselle	Eau de Parfum

with marks at 10% and 90% was used. After testing each product, the consumer rated the overall liking on a structured 9-point scale.

2.3. Methods

For the PLS on dummy variables and for the Fishbone method, the aim is to estimate, for each attribute, the possible gain in liking if that attribute was at its ideal level. Hence, we can define for each product:

$$Liking_{loss} = Liking_{ideal} - Liking_{measured} \tag{1}$$

Where

$Liking_{loss}$ is the potential loss in liking due to the deviations of the product tested from its ideal;

$Liking_{ideal}$ is the liking of the ideal product (where liking is maximized);

$Liking_{measured}$ is the appreciation of the product tested by the different consumers.

In the Eq. (1), $Liking_{loss}$ is the important parameter to minimize: it measures the difference between the actual liking and the liking of the ideal product.

2.3.1. PLS on dummy variables

In Xiong and Meullenet, 2006 proposed the use of PLS on dummy variables to analyze JAR data. In the Ideal Profile method, data can be transformed into JAR values by taking (for each consumer) the difference between perceived and ideal intensities for each product and each attribute ($d_{jpa} = y_{jpa} - z_{jpa}$). PLS is applied on each product separately, which creates a product-specific model, meaning that each product is given its own unique guidance on improvement. Since each product is optimized separately, the analysis takes only subject variability into consideration.

Algorithmically, PLS on dummy variables is done in three steps. First, for a given product p, the original dataset is organized into dummy variables. To do so, the difference d_{jpa} is calculated. Depending on its sign, the difference is assigned to the category "too little" (attribute −), or the category "too much" (attribute +). Hence, for each attribute, two columns (the so-called dummy variables) with J rows (J being the number of consumers) are created, one taking the positive values of the difference, and one taking the negative values of the difference. The values from the other column are set to 0 (attribute – when the difference is positive, or attribute + when the difference is negative).

Dummy variables have better prediction properties for the overall liking (Fig. 1) than the original variables. The construction of the dummy variables and the decision rules are presented in Table 2. Compared to binary dummy variables, which only take values of 0 and 1, the dummy variables used here take either the value of the difference d_{jpa}, or 0.

Second, a PLS regression explaining the overall liking $\boldsymbol{l_p}$ in function of the dummy variables is performed. A descending step by step selection of the model is done through jackknifing (Martens & Martens, 2000) until only significant dummy variables remain in the model. The weights β^{+}_{pa} and β^{-}_{pa} associated to the significant dummy variables, which have an impact on liking, are then extracted. For a given attribute a, the comparison of the two regression weights β^{+}_{pa} and β^{-}_{pa} helps determining on which region the attribute is more detrimental to overall liking. For instance, if the coefficient β^{+}_{pa} is significantly different from 0 while β^{-}_{pa} is not, the product considered is described as having too much of the attribute a. This explains why the liking is not optimal for this product.

Third, the potential gain in liking (called "mean drop" in Xiong & Meullenet (2006)) associated with each dummy variable is calculated by multiplying the dummy variables by their associated weights β^{+}_{pa} and β^{-}_{pa} (Eq. (2)).

$$\begin{cases} gain\ in\ liking^{+}_{pa} = \beta^{+}_{pa} * dummy\ \ variable^{+}_{pa} \\ gain\ in\ liking^{-}_{pa} = \beta^{-}_{pa} * dummy\ \ variable^{-}_{pa} \end{cases} \tag{2}$$

Graphically, only the gain in liking associated to the significant dummy variables is shown.

2.3.2. Fishbone method

Compared to the PLS on dummy variables, the Fishbone method (Punter, 2008; Punter & Worch, 2009) is not "product-specific". The impact of each attribute on overall liking is estimated by taking all the products simultaneously in the regression model. Hence, the variability involved in the analysis is both product and consumer related.

Algorithmically, the Fishbone method is done in four steps. First, a principal component analysis (PCA) – run on the correlation matrix – is performed on the *product × subject × attribute* matrix. In this case, one row is one product described by one subject

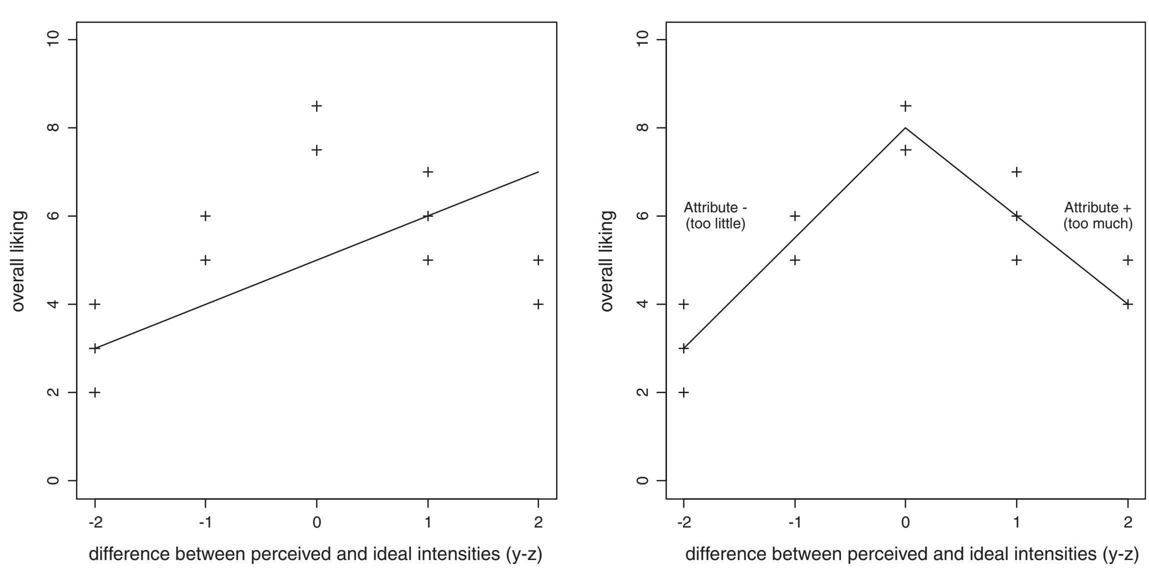

Fig. 1. Prediction of the liking when using the original variables (left) or the dummy variables (right).

Table 2
Decision rule for the construction of the dummy variables, with y_{jpa}(resp.z_{jpa}) the perceived (resp.ideal) intensity rated by consumer j for product p and attribute a.

Panelist	Product	Perc. Att	Ideal Att	Difference	Attribute +	Attribute −
1	p	77.0	49.4	29.6	29.6	0.0
2	p	43.3	36.4	6.9	6.9	0.0
...	...	...	...	...	...	...
j	p	y_{jpa}	z_{jpa}	$d_{jpa} = y_{jpa} - z_{jpa}$	$\begin{cases} d_{jpa} & \text{if } d_{jpa} > 0 \\ 0 & \text{if } d_{jpa} < 0 \end{cases}$	$\begin{cases} 0 & \text{if } d_{jpa} < 0 \\ d_{jpa} & \text{if } d_{jpa} > 0 \end{cases}$
...	...	...	...	...	...	...
J	p	35.4	57.0	−21.6	0.0	−21.6

(hence, we have $P \times J$ rows, P and J being the number of products and subjects respectively).

Second, the overall liking $\boldsymbol{l}$ is regressed on the dimensions of the PCA (Eq. (3)). Backward deletion is performed until only significant dimensions are kept in the model.

$$l_{jp} = \alpha_1 Dim_1 + \alpha_2 Dim_2 + \ldots + \alpha_i Dim_i + \ldots + \alpha_1 Dim_1 \tag{3}$$

Where

l_{jp} is the liking of product p by consumer j

α_i is the weight of the dimension i on liking (α_i = 0 for not significant dimensions)

The weight β_a of each attribute on overall liking is then estimated using Eq. (4).

$$\beta_a = \sum_{i=1}^{I} \alpha_i * loading_{ai} \tag{4}$$

Where

β_a is the weight of the attribute a on liking ;

α_i is the weight of the dimension i on liking;

$loading_{ai}$ is the loading of the attribute a on dimension i.

Third, for each product p, the difference (noted rd_{pa}) between the averaged ideal $\bar{z}_{.pa}$ and the averaged perceived $\bar{y}_{.pa}$ intensities is calculated for each attribute a. This difference is made relative to the actual perceived intensity of the product. It is expressed in percentage of change that must be obtained to make the attribute ideal (Eq. (5)).

$$rd_{pa} = 100 * (\bar{z}_{.pa} - \bar{y}_{.pa})/\bar{y}_{.pa} \tag{5}$$

For each attribute, the relative difference rd_{pa} is weighted in function of its potential impact on liking. To do so, the relative difference is multiplied by the attributes' weight β_a. We obtain the relative difference corrected $rd_{cor_{pa}}$ (Eq. (6)).

$$rd_{cor_{pa}} = \beta_a * rd_{pa} \tag{6}$$

For each product, the possible overall gain in liking is computed by subtracting the averaged overall liking of the actual product to the liking of the ideal product. By default, the liking of the ideal product is set as the maximum value of the liking scale used (nine in the perfume example). The difference obtained is made relative to the liking of the product tested and is expressed as percentage (Eq. (7)).

$$ogl_p = 100 * (9 - \bar{l}_{.p})/\bar{l}_{.p} \tag{7}$$

Finally, the potential gain in liking pgl_{pa} for each product and each attribute is calculated by matching the relative difference corrected with the overall gain in liking. For a given product p, the sum of the relative difference corrected over the Aattributes is forced to be equal to the overall gain in liking. The scaling factor γ_p used is presented Eq. (8) and its calculation is highlighted in the example presented below:

$$pgl_{pa} = \gamma_p * rd_{cor_{pa}}$$

With

$$\gamma_p = ogl_p \Big/ \sum_{a=1}^{A} rd_{cor_{pa}} \tag{8}$$

Example: Let's consider a product described on three attributes Att1, Att2 and Att3.

On a 9-point hedonic scale, this product had a liking rating of 6. To increase from 6 to 9, liking can be increased by 50%(=100*((9 − 6)/6)).

The relative difference corrected for the three attributes is presented in Table 3 (left).

In order to match the overall gain in liking (50%) with the sum of the relative difference corrected (25%), we force the two percentages to be equal. Hence, a scaling factor equal to 2 (=50/25) is applied to the relative difference.

For interpretation purposes, the effects of the attributes are considered independent. If the intensity of an attribute a for a product p is changed from its perceived level $\bar{y}_{.pa}$ to its ideal level $\bar{z}_{.pa}$, without changing the intensity of the other attributes, then liking would increase with pgl_{pa}%. But as the attributes are often highly correlated, the effects of the attributes are somehow linked. Therefore, the conclusions drawn here should only be taken as guidance to improvement, and not as recipe.

For each attribute and each product, the difference between ideal and perceived intensities ($\bar{z}_{.pa} - \bar{y}_{.pa}$) and the potential gains in liking pgl_{pa} are shown in a graph (as this graph looks like a fishbone, it has been named "Fishbone plot"). In the Fishbone plots, attributes are arranged in increasing order in function of the potential gain in liking (represented as histogram on the graphic). In order to simplify the graphic, only the attributes with a possible impact on liking of 2% or higher are shown. As a complement, the difference between ideal and perceived intensities is also shown on the graph (diamonds).

Table 3
Calculation of the potential gain in liking from the Relative difference corrected.

	Relative difference corrected (%)		Potential gain in Liking pgl_{pa} (%)
Att1	5	Att1	10
Att2	8	Att2	16
Att3	12	Att3	24
Sum	25	Sum	50

2.3.3. Comparison of the results of the PLS and the Fishbone methods

In both methods, the main information extracted concerns the potential impact of each attribute on overall liking. This is shown as a potential gain in liking if that attribute was at its ideal level and is either expressed as potential gain (PLS) or as percentage (pgl_{pa} in the Fishbone method). In PLS, these potential gains can be summed together over all attributes in order to estimate the potential liking of the ideal product. In the Fishbone method, the ideal liking rating is assumed to be 9 on a 9-point scale, which is why it is expressed as a percentage.

The correlation coefficient is then calculated between the vectors of potential gain in liking related to the different attributes for PLS and for the Fishbone method.

Note that in PLS, each attribute is related to two coefficients β^{+}_{pa} and β^{-}_{pa}. As only one coefficient (noted β_{pa}) per attribute is necessary, three different cases can be defined:

– the two dummy variables are not significant: the variable is considered as not important and the coefficient β_{pa} is set to 0;
– only one of the two dummy variables is significant: the variable is considered important and the coefficient β_{pa} takes the weight β^{+}_{pa} or $\beta_{pa^{-}}$ associated to the significant dummy variable;
– both dummy variables are significant: the variable is considered important (see Xiong & Meullenet (2006) for an example on how to treat this particular case). In the perfume dataset, we are not concerned with such a situation since none of the paired regression coefficients were significant at the same time.

When an attribute is important, it implies that the intensities of at least one of the corresponding dummy variables deviate from zero and are not Just About Right. When an attribute is not important, it implies that either the corresponding pairs of dummy variables are zero or Just About Right, or that the attribute does not affect the overall liking.

2.3.4. Validations of the results obtained with the two methods

In order to check the validity of the results, the link between the current and the ideal products must be studied. This can be done by examining the position of the ideal products in the current product space. To do so, the current product space is created by applying PCA to the *product* $\times$ *attribute* matrix $\bar{y} = \{\bar{y}_{.pa}, p = a : P\&a = 1 : A\}$. In this product space, the averaged ideal products $\bar{z} = \{\bar{z}_{.pa}, p = a : P\&a = 1 : A\}$ are projected as supplementary entities (Escofier & Pagès, 2008). The attribute representation related to this product space provides an overview of the attributes to be increased/decreased in intensity to position the current product closer to its ideal. The conclusions drawn here are then compared to the raw values of the product and its ideal.

When the relationship between both methods is weak, a decision must be made as to which method may be most appropriate.

2.3.5. Measurement of the repeatability of the methods (duplicated products)

Another way to validate the two methods is to assess the validity of the conclusions they draw. With replicated products, one can compare the conclusions from one product with its replicate. The two conclusions should converge and return similar recommendations.

As previously stated, the correlation coefficient is measured between the vector of potential gain related to the *A* attributes for a product and its replicate, for each method.

3. Results and discussion

Before applying the methodology described previously to the perfume dataset, the quality of the data must be checked. The methodology used here to validate the data is the one proposed by Van Trijp et al. (2007). It consists of (1) checking that the products are different in terms of their attribute levels and overall liking and (2) ensuring the reliability of ideal levels across the products. A one-way analysis of variance measuring the product effect on each variable (perceived intensity, ideal intensity and overall liking) can be performed. The first point is validated if the product effect is significant for all the variables involved (perceived intensity and overall liking). The second point is validated if the product effect is not significant for the variables involved (ideal intensity). The results (Table 4) indicate that the product effect is significant for all the intensity attributes and the overall liking at 5%, except for jasmin (*p*-value = 0.138), camomille (*p*-value = 0.838) and anis (*p*-value = 0.264). Concerning the ideal attributes, the product effect is never significant at 5%, except for the ideal freshness (*p*-value = 0.010). The dataset is considered valid.

3.1. Results

Table 5 shows the correlation coefficients measured for each product between the gain in liking obtained from the Fishbone and PLS on dummy variables method. It ranges from 0.60 (Cinema, Lolita Lempicka and Shalimar) showing a strong linear link between the results from both methods, to −0.33 (J'Adore (EP)), indicating no linear link between the results from both methods. The average correlation coefficient is 0.30.

3.1.1. Case of a product showing agreement between both methods: Angel

Fig. 2a and b shows the results for Angel from PLS on dummy variables and from the Fishbone method, respectively.

In Fig. 2a, for all attributes except camomille, one dummy variable is significant, meaning that it is not at its Just About Right level. The most important attributes to modify (i.e. those that would increase liking the most) are freshness, earthy, odor intensity, green, woody and fresh lemon (the attributes are ordered in function of their increasing impact on liking). Moreover, the sum of the

Table 4
p-Values related to the product effect in the one-way anova used for the validation of the Ideal Profile data.

	Perceived intensity	Ideal intensity
Intensity	0.000	0.616
Freshness	0.000	0.010
Jasmin	0.138	0.098
Rose	0.001	0.250
Camomille	0.838	0.994
Fresh_lemon	0.000	0.448
Vanilla	0.000	0.061
Citrus	0.012	0.999
Anis	0.264	0.997
Sweet_fruit	0.000	0.639
Honey	0.000	0.988
Caramel	0.000	0.558
Spicy	0.000	0.974
Woody	0.000	0.074
Leather	0.000	0.159
Nutty	0.000	0.557
Musk	0.000	0.835
Animal	0.000	0.983
Earthy	0.000	0.645
Incense	0.000	0.819
Green	0.000	0.932
Liking	0.000	

Table 5
Correlation coefficient measured between the potential gain in liking for each attribute obtained from the two methods. For L'Instant, no model can be found in the PLS on dummy variables.

Product	*R*
Angel	0.52
Aromatics Elixir	0.35
Chanel N5	0.53
Cinema	0.60
Coco Mademoiselle	0.30
J'Adore (EP)	−0.33
J'Adore (ET)	0.42
L'Instant	Not available
Lolita Lempicka	0.60
Pleasures	0.17
Pure Poison	−0.21
Pure Poison2	0.39
Shalimar	0.60
Shalimar2	0.47

potential gain in liking shows that if all attributes were at their ideal levels, overall liking for Angel would increase from 4.5 to 6.0.

Fig. 2b is a bit stricter, as only 17 out of 21 attributes were selected. The most important attributes from the Fishbone method are fresh lemon, green, freshness, earthy and jasmin. Aside from providing the potential increase in liking, it also states how much each attribute should be modified to attain the ideal level (represented in the graphical display as "diamonds"). For some attributes, such as odor intensity, a large change in terms of intensity is necessary (around 12 points in score), but this does not necessarily induce a big gain in overall liking (only around 4% gain). For other attributes such as jasmin, a smaller change (around five points in score) will generate more gain in liking (around 8%).

In this case, similar attributes have the highest impact on overall liking. The only exception concerns odor intensity which seems to be one of the most important attributes in PLS, and one of the less important (but still relevant) attributes in the Fishbone method.

3.1.2. Case of a product showing disagreement between both methods: J'Adore (EP)

Fig. 3a and b shows the results for J'Adore (EP) from the PLS on dummy variables and from the Fishbone methods, respectively.

In PLS on dummy variables (Fig. 3a), only 9 out of 21 attributes show an impact on the overall liking. Among these nine attributes

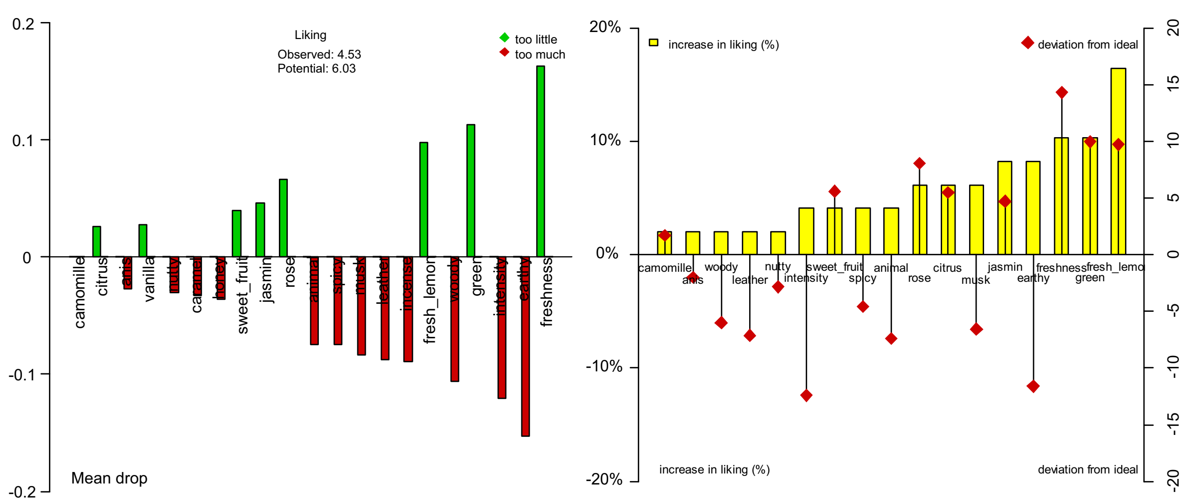

Fig. 2. (a and b): PLS on dummy variables (left, a) and Fishbone Plots (right, b) results for Angel.

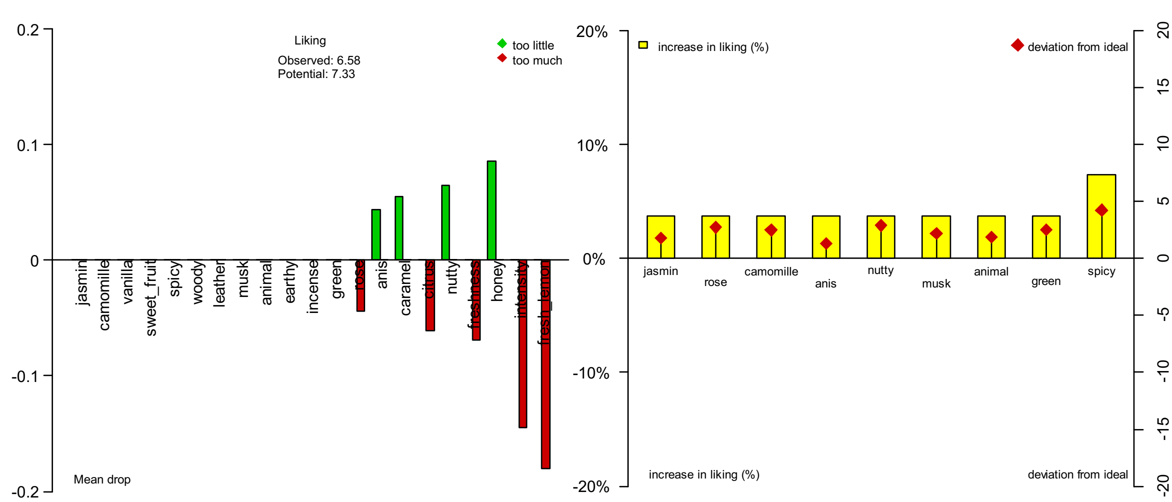

Fig. 3. (a and b): PLS on dummy variables (left, a) and Fishbone Plots (right, b) results for the J'Adore (EP).

are fresh lemon, odor intensity, honey and freshness. The sum of the potential gain in liking shows that changing the attributes to their ideal level would only increase the overall liking from 6.6 to 7.3.

The Fishbone method (Fig. 3b) also shows that 9 out of 21 attributes affect the overall liking, including spicy, green, animal, musk, nutty, anis, camomille, rose and jasmin. These attributes all deviate from the ideal level by having "too little" in the product (in the graphic, the diamonds all point to the positive side, indicating that the differences $\bar{z}_{.pa} - \bar{y}_{.pa}$ are all positive).

In this case, only four attributes are similar between the two methods (honey, nutty, rose and anis). The conclusions drawn from these data would be different depending of which method was used.

3.1.3. Comparison of the PLS and Fishbone results

For Angel, the correlation coefficient between the estimated potential gain in liking for each attribute from both methods is relatively high (0.52, see Fig. 4).

As most of the attributes are close to the first bisectrix, we can conclude that they agree on the estimation of the potential gain in liking calculated for each attribute. The only minor disagreement concerns the attributes "incense" (more important in PLS) and "fresh lemon" (more important in the Fishbone plots).

For J'Adore (EP), there is no similarity between the two methods, as the correlation coefficient between the potential gain in liking for each attribute is −0.33 (see Fig. 5). Indeed, the two methods point out different attributes. In PLS, honey, freshness, nutty, citrus and caramel are the most important attributes to change, while spicy, green, animal, vanilla and woody are the most important for the Fishbone method. The only common attributes are honey, nutty, citrus and rose.

More generally, for Angel, Chanel N5, Cinema, Lolita Lempicka and Shalimar, there is a link between the results from both methods ($r > 0.5$), while for J'Adore (EP) and Pure Poison, the methods seem to present different conclusions ($r < 0$). (Table 5).

3.1.4. Validation of the results obtained with the two methods

In order to validate the conclusions drawn by the two different methods, the product space (Fig. 6 and Fig. 7) is evaluated. PCA is

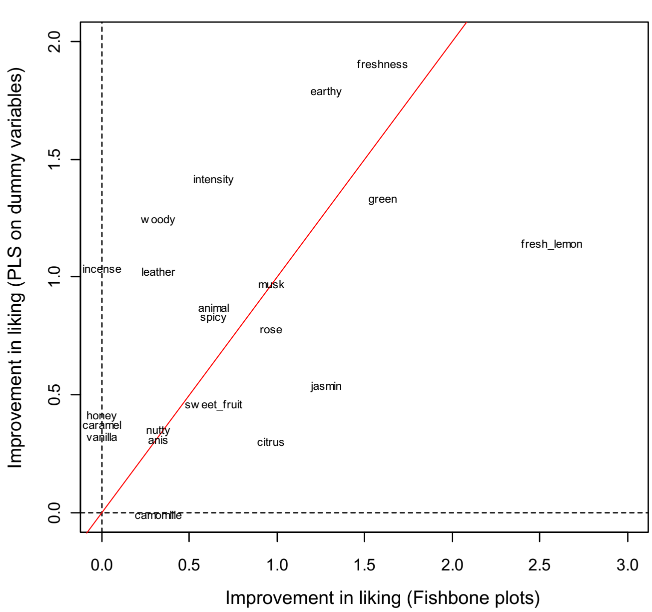

Fig. 4. Relationship between the potential improvement in liking by attribute suggested by the two methods for Angel.

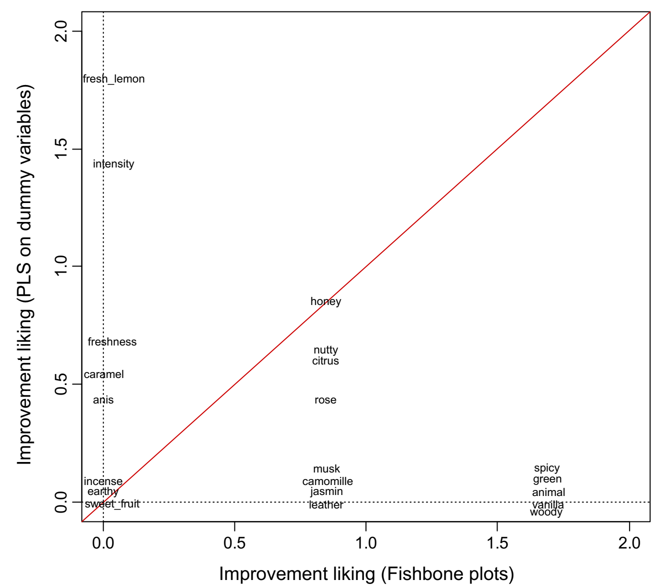

Fig. 5. Relationship between the potential improvement in liking by attribute suggested by the two methods for J'Adore (EP).

performed on the *products* × *attributes* table: the first dimension of the PCA opposes the perfumes with fresh and flower notes (J'Adore (EP) and (ET) and Pleasures) to the perfumes with oriental notes (the two replicates of Shalimar and Angel); the second dimension opposes the perfumes with vanilla, honey and caramel notes (Angel, Lolita Lempicka) to the perfume with spicy and intense notes (Aromatics Elixir).

The projection of the ideal perfumes in this product space illustrates that the ideals are located near J'Adore (EP), Cinema and L'Instant. It appears that the ideal products have fresh, sweet fruit and citrus notes, with some vanilla and honey odor. With respect to the correlation between attributes, this also corresponds to low ratings in the odor intensity and spicy notes.

The direct comparison of the profiles for a product and its ideal is possible through the use of spider plots. To bring Angel closer to its ideal (Fig. 8), fruit and flower notes (sweet fruit, fresh lemon, rose, green, freshness) should be increased and oriental notes (incense, earthy, animal, spicy, leather, musk, etc.) and the odor intensity should be decreased. These conclusions correspond to those stated in Section 3.1.1.

The comparison of the perceived and ideal profiles for J'Adore (EP) (Fig. 9) shows that the current product is already close to its ideal, the only difference being the vanilla, nutty and spicy notes (needs more) and the odor intensity (needs less). In Section 3.1.2, we have shown that the conclusions drawn from the two methods disagree, the only agreement being on honey, nutty, citrus and rose. This disagreement is not important, since both conclusions are realistic. Indeed, as the attributes are highly correlated (Fig. 7), decreasing the perception of freshness (defined as important in the PLS) is equivalent to increasing the perception of spicy (defined as important in the Fishbone).

One hypothesis to explain the disagreement between the two methods is due to the similarity between the current product and its ideal. When a product is close to its ideal, it is more difficult to create a clear hierarchy in terms of importance of the different attributes on liking, as they all might have the same impact. In this case, it is more difficult to define a model explaining the liking. For instance, for J'Adore (EP), the difference between the perceived and the ideal intensities is almost constant among the attributes (around two points in intensity), which makes the hierarchy

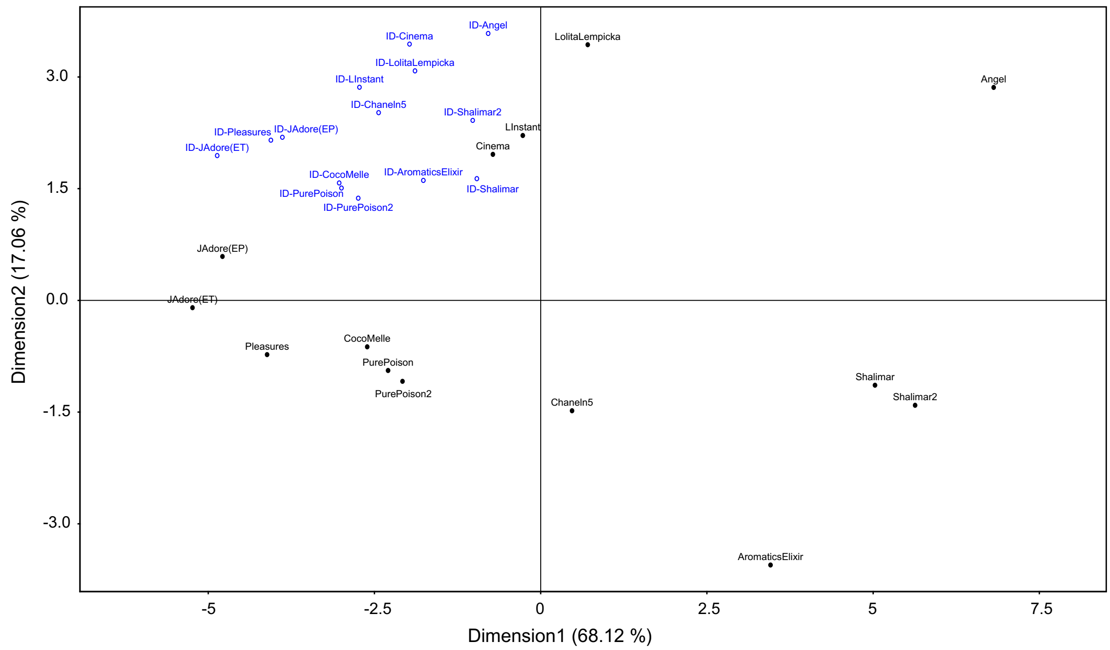

Fig. 6. Product space obtained by PCA with projection as illustrative of the ideal products (points with a label starting with "ID-").

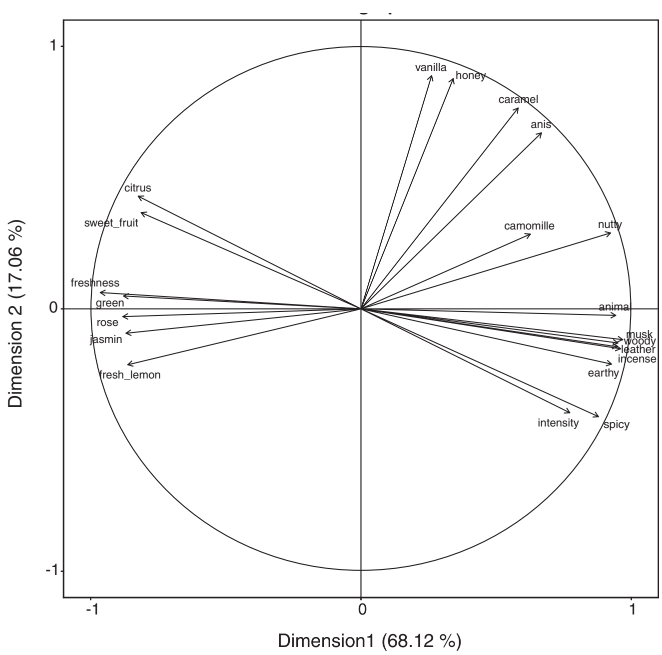

Fig. 7. Variable representation associated to the product space obtained by PCA.

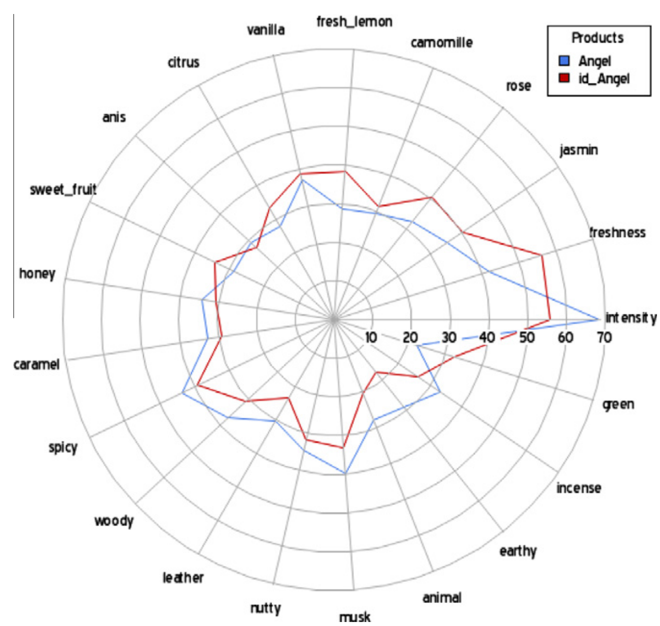

Fig. 8. Comparison of the perceived and ideal profiles for Angel.

difficult to create. As for Angel, the difference between the perceived and the ideal intensities is larger (and more heterogeneous), the hierarchy is easier to create and the model is easier to define (which leads to a better agreement between methods). Moreover, the calculation of the correlation coefficient does not take into account the link existing between attributes.

In PLS, the closeness between perceived and ideal profiles can lead to a particular case where no-model is found, as no dummy variables are significant (as is the case for L'Instant). As all products are used simultaneously in the model, this is not observed with the Fishbone method.

3.1.5. Reproducibility of the results

In this study, two products (Pure Poison and Shalimar) were replicated. This enables us to measure the consistency of each method by comparing the results from the two replicates together.

3.1.5.1. Case of Shalimar. Fig. 10a and b shows that for Shalimar and its replicate, both methods are reproducible. On each Figure, all attributes are close to the first bisectrix, meaning that from one replicate to another, the two methods give the same importance to the attributes (on the Fishbone method, only freshness and fresh lemon are slightly shifted showing that these attributes have a slightly bigger impact on liking for the first replicate than for the

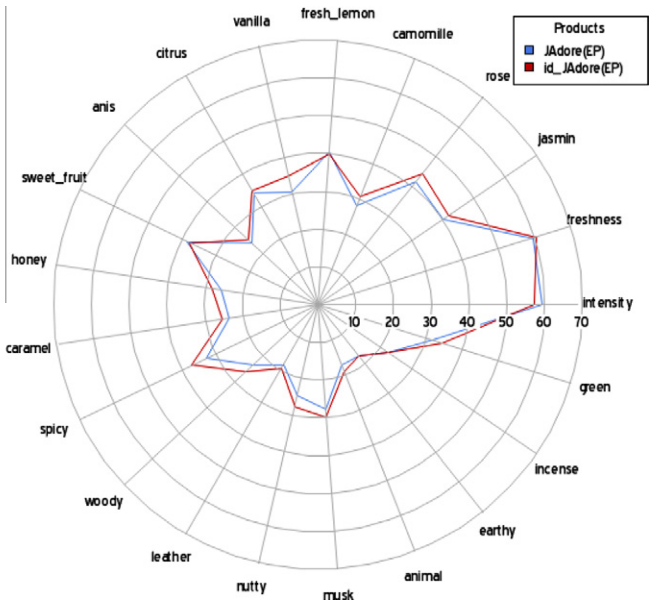

Fig. 9. Comparison of the perceived and ideal profiles for J'Adore (EP).

coefficient of −0.12 is measured between both replicates. Still, the variable representation (Fig. 7) shows similar patterns as the attributes highlighted for the two replicates are highly correlated. Hence, the conclusions are not in complete disagreement.

For the Fishbone method (Fig. 11b), the results are more consistent. The correlation coefficient is 0.78. Green, jasmin, sweet fruit, earthy and freshness seem to be the most important attributes in both cases.

With respect to the product space (Fig. 6), it appears that Pure Poison is close to its ideal. As for J'Adore (EP) (Section 3.1.2), this closeness makes it difficult to define a model with the PLS on dummy variables method. Hence, the results are different from one replicate to the other (as only small changes are necessary to make Pure Poison ideal). Concerning the Fishbone method, the models defined for the two replicates are more similar than with the PLS on dummy variables, and the conclusions are similar. This is probably due to the way the model is created. In the Fishbone method, all products are considered simultaneously while for PLS, it is product-specific. In this manner, the Fishbone method is more "stable", but the model associated to this method over fits the data. Indeed, in the Fishbone method, the model defined might over-estimate the impact of some attribute on overall liking (in this case, several attributes seem to improve a product that is already close to ideal).

second). The correlation coefficient between the results obtained with the two replicates is 0.93 for the PLS on dummy variables (Fig. 10a) and 0.86 for the Fishbone method (Fig. 10b). Overall, the results are consistent for the replicates.

Moreover, the same attributes (freshness, earthy, fresh lemon and green) are important in both methods, the only difference being that for the Fishbone method, sweet fruit and citrus appear to be important while for the PLS it is not the case. With respect to the product space (Fig. 6), Shalimar is far from its ideal. These results appear to confirm the statement made in Section 3.1.4.

3.1.5.2. Case of Pure Poison. In Fig. 11a and b, for Pure Poison, the results are less reproducible than for Shalimar. PLS on dummy variables (Fig. 11a) indicates that each replicate is only defined by a few drivers of liking and from one replicate to another, the drivers of liking differ. For the first replicate, odor intensity is the main driver of liking while for the second replicate, it's freshness, green and musk (odor intensity no longer influences liking). A correlation

3.2. Discussion and synthesis

Globally, the results are consistent from one analysis to another, especially when the products tested are far from the ideal. When the products are already close to the ideal, the definition of a model is not always possible with the PLS on dummy variables (i.e. no significant dummy variables can be found) while the Fishbone method tends to over-estimate the importance of some attributes on overall liking. In that case, both methods will show disagreement. Indeed, the correlation coefficient measured between the results obtained from the two methods is usually low. But it has to be said that the correlation coefficient does not take into account the link existing between the attributes.

The advantages and disadvantages of both methods are summarized Table 6.

For further analysis, it may be interesting to compare PLS and Fishbone methods by making both methods product-specific. This would avoid the over-estimation of the impact of some attributes

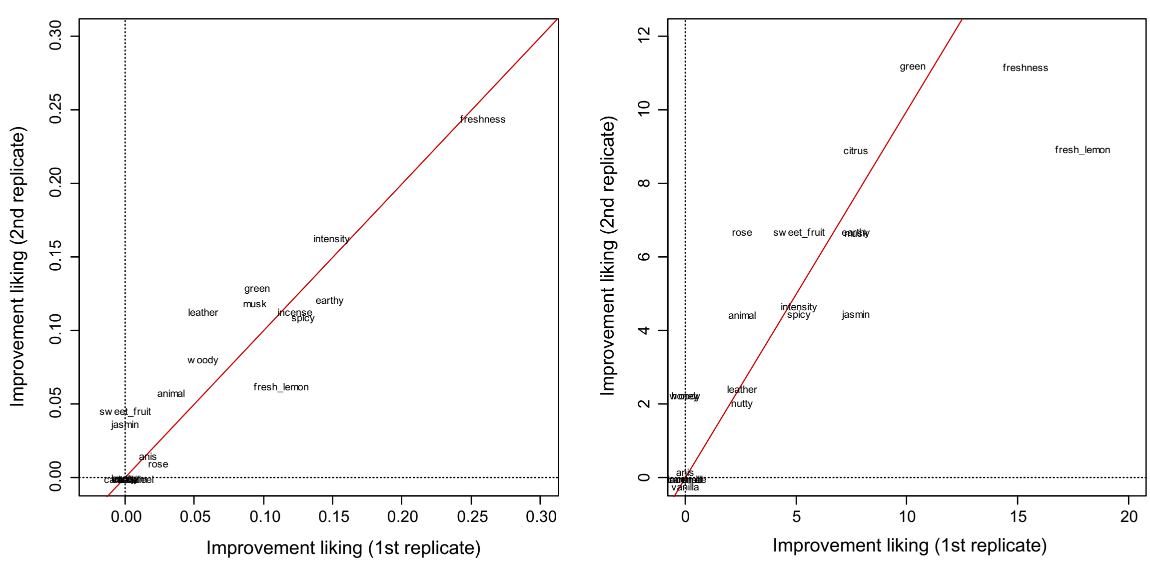

Fig. 10. (a and b): Reproducibility of the results obtained with PLS (left, a) and Fishbone (right, b) for Shalimar.

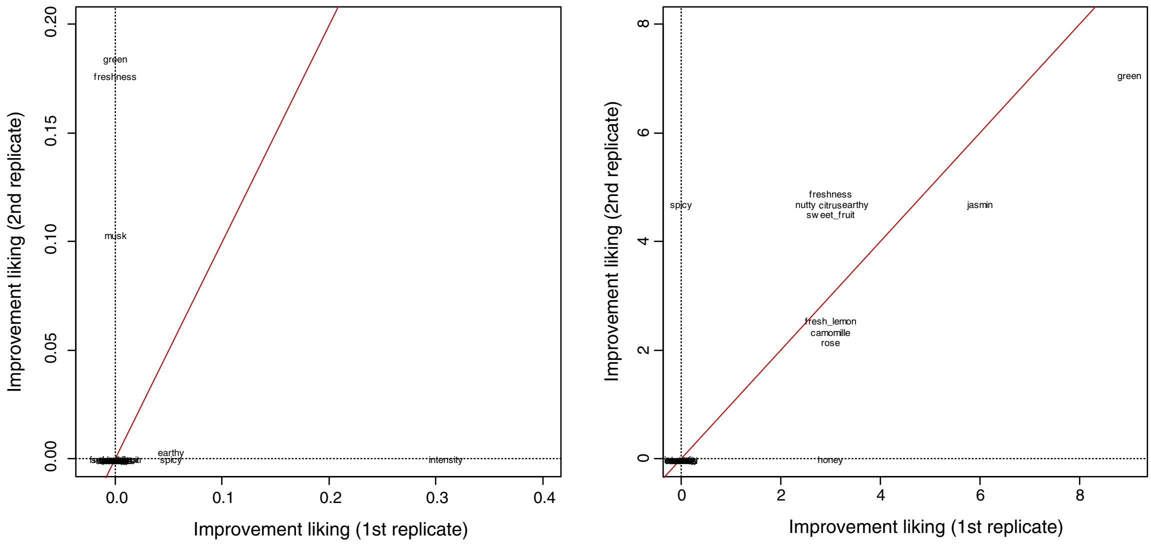

Fig. 11. (a and b): Reproducibility of the results obtained with PLS (left, a) and Fishbone (right, b) for Pure Poison.

Table 6
Advantages and disadvantages of the two analyses.

	PLS dummy variables	Fishbone method
Advantages	The model is product-related, and product-specific improvements are made	The model is general for a set of products
	The model estimates attribute weights separately for each product	The general idea for improvement across all products in the product set; product adjustments are made based on a global ideal
Disadvantages	Less idea about what is generally important for a set of products	The attributes' weights are fixed for all products, although they might change from on product to another
	Difficulties to find model when products are close to ideal	It can over-estimate the impact of some attributes on overall liking when the products are close to ideal
When should we use it?	To define the drivers of liking for one particular product	To define the drivers of liking for a set of products

on overall liking for the Fishbone, and would provide a more fair comparison when the products are close to ideal.

Finally, in this example, we did not check for clusters of consumers with respect to their liking. As the ideal profiles and the overall liking are strongly linked, defining clusters might enrich the analysis, as we could give different guidance on improvement depending on the final target consumers.

Acknowledgments

The authors would like to thank the reviewers for their interesting and valuable comments and suggestions.

References

Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference, 18*, 1–16.

Coombs, C. H. (1976). *A theory of data*. New York: Wiley and Sons Inc..

Escofier, B. & Pagès, J. (2008). *Analyses factorielles simples et multiples: Objectifs, méthodes et interprétations*. 4ème éd. Paris: Dunod.

Epler, S., Chambers, E., IV, & Kemp, K. E. (1998). Hedonic scales are a better predictor than Just-About-Right scales of optimal sweetness in lemonade. *Journal of Sensory Studies, 13*, 191–197.

Gazano, G., Ballay, S., Eladan, N., & Sieffermann, J.M. (2005). Flash profile and flagrance research: Using the words of the naïve consumers to better grasp the perfume's universe. In *ESOMAR fragrance research conference*. New York, NY, 15–17 May 2005.

Greenhoff, K., & MacFie, H. J. H. (1994). Preference mapping in practice. In H. J. H. MacFie & D. M. H. Thomson (Eds.), *Measurement of food preferences* (pp. 137–166). Glasgow: Blackie Academic and Professional.

Healy, A., & Miller, G. A. (1970). The verb as the main determinant of the sentence meaning. *Psychometric Science, 20*, 372.

Husson, F., Le Dien, S., & Pagès, J. (2001). Which value can be granted to sensory profiles given by consumers? Methodology and results. *Food Quality and Preference, 12*, 291–296.

Issanchou, S. (2009). *Détermination directe d'un idéal sensoriel in evaluation sensorielle* (3ème éd., pp. 225–234). Lavoisier: Manuel méthodologique.

Lawless, H. T., & Heymann, H. (1999). *Sensory evaluation of food: Principles and practices*. New York: Kluwer.

MacFie, H., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies, 4*, 129–148.

Martens, H., & Martens, M. (2000). Modified Jack-knife estimation of parameter uncertainty in bilinear modeling by partial least squares regression (PLSR). *Food Quality and Preference, 11*, 5–16.

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (2007). *Sensory evaluation techniques* (4th ed.). Boca Raton, FL: CRC Press.

Meullenet, J. F., Xiong, R., & Findlay, C. J. (2007). Analysis of just about right data. In *Multivariate and probabilistic analyses of sensory science problem* (pp. 207–235). Blackwell Publishing, IFT Press.

Moskowitz, H. R. (1996). Experts versus consumers: A comparison. *Journal of Sensory Studies, 11*, 19–37.

Moskowitz, H. R., Munoz, A. M., & Gacula, M. C. (2003). Descriptive panels/experts versus consumers. In *Viewpoints and controversies in sensory science*. Trumbull, CT: Food & Nutrition Press, Inc.

Nestrud, M. A., & Lawless, H. T. (2008). Perceptual mapping of citrus juices using projective mapping and profiling data from culinary professionals and consumers. *Food Quality and Preference, 19*, 431–438.

Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference, 12*, 297–309.

Punter, P. H. (2008). Bridging the gap between R& D and marketing. The ideal profile method. In Presented at the *Society of sensory professionals meeting*. Cincinnati, Ohio, USA, November 2008.

Punter, P. H. & Worch, T. (2009). The ideal profile method: Combining classical profiling with JAR methodology. In *SPISE 2009 proceeding: Food consumer insights in Asia*, Ho-Chi-Minh-City, Vietnam.

Sieffermann, J. M. (2002). Flash profiling. A new method of sensory descriptive analysis. In *AIFST 35th convention*, Sidney, Australia, July 21–24.

Stone, H., & Sidel, J. L. (2004). *Sensory evaluation practices* (3rd ed.). London, UK: Elsevier.

Van Trijp, H. C., Punter, P. H., Mickartz, F., & Kruithof, L. (2007). The quest for the ideal product: Comparing different methods and approaches. *Food Quality and Preference, 18*, 729–740.

Worch, T. W., Lê, S., & Punter, P. (2009). How reliable are consumers? Comparison of sensory profiles from consumers and experts. *Food Quality and Preference, 21*, 309–318.

Xiong, R., & Meullenet, J. F. (2006). A PLS dummy variable approach to assess the impact of JAR attributes on liking. *Food Quality and Preference, 17*, 188–198.