



Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask

Sylvie Chollet^{a,*}, Maud Lelièvre^{a,b}, Hervé Abdi^c, Dominique Valentin^b

^a Institut Supérieur d'Agriculture, 59046 Lille Cedex, France

^b UMR CSG 5170 CNRS, INRA, Université de Bourgogne, 21000 Dijon, France

^c The University of Texas at Dallas, Richardson, TX 75083-0688, United States

ARTICLE INFO

Article history:

Received 23 November 2010

Received in revised form 7 February 2011

Accepted 16 February 2011

Available online 24 February 2011

Keywords:

Sorting task

Experts

Novices

Beer

ABSTRACT

In industries, the sensory characteristics of products are key points to control. The method commonly used to characterize and describe products is the conventional profile. This very efficient method requires a lot of time to train assessors and to teach them how to quantify the sensory characteristics of interest. Over the last few years, other faster and less restricting methods have been developed, such as free choice profile, flash profile, projective mapping or sorting tasks. Among these methods, the sorting task has recently become quite popular in sensory evaluation because of its simplicity: it only requires assessors to make groups of products perceived as similar. Previous studies have shown that this method produces sensory spaces similar to those obtained with conventional profiles but that the descriptions of the products are coarser than the descriptions yielded by sensory profiles. The aim of the present paper is to further evaluate the efficiency of the sorting task as a sensory tool. We present a series of studies highlighting the advantages and delineating the limits of the sorting task and illustrate advantages and limits using beer as the common type of stimuli. These studies underline the main issues encountered when designing sorting tasks. More precisely, we examine the potential of the sorting task to describe beer sensory characteristics, we determine the type of assessors able to perform a sorting task and we evaluate the stability of the results as well as some important methodological points (e.g. number of beers to be sorted, instructions given to the judges) that might impact the efficiency of the task.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

For industries, the sensory characteristics of products are essential criteria in various areas including R&D, quality control, and marketing. For example in product development, it is crucial to understand the sensory characteristics of a product in order to evaluate the relationship between raw material and/or process parameters and the quality of the product. Likewise, monitoring the sensory characteristics in routine control is essential to maintain and control product quality.

A classical way to describe these sensory characteristics is to select a small group of panelists and train them to identify and quantify the main sensory dimensions of the products. This type of method, called sensory profile, is quite efficient but also very expensive and time consuming and therefore most industries cannot routinely use this technique (Kemp, Hollowood, & Hort, 2009; Meilgaard, Civille, & Carr, 1999; Stone & Sidel, 1993). Thus, it is necessary to develop other sensory methods to obtain sensory

information about products. Among these new methods, the sorting task has been one of the most popular in the domain of product descriptions (see Abdi, Valentin, Chollet, & Chrea, 2007, for a review of sorting tasks applied on food and nonfood products). Several recent papers deal with this method, but they address only one aspect of this method, namely the comparison with conventional profile (Blancher et al., 2007; Cartier et al., 2006; Faye et al., 2004, 2006; Lelièvre, Chollet, Abdi, & Valentin, 2008, 2009; Saint-Eve, Paçi Kora, & Martin, 2004; Soufflet, Calonnier, & Dacremon, 2004; Tang & Heymann, 1999). The goal of this article is to synthesize the current state of knowledge about the sorting task and to delineate its main advantages and limits. After an overview of the different methodologies available to perform sensory descriptions (conventional profile, free choice profile, flash profile, projective mapping and sorting task) we will review the results from several sorting experiments carried out between 1998 and 2008 on different sets of beers. These experiments examined the potential of the sorting task: (1) to describe beer sensory characteristics (Experiments 1 and 2), (2) to describe the type of assessors able to perform sorting task (Experiment 3) and (3) to evaluate the stability of the results as well as the main factors (e.g. number of beers that are sorted, instructions given to the

* Corresponding author. Address: Institut Supérieur d'Agriculture, 48 Boulevard Vauban, 59046 Lille Cedex, France. Tel.: +33 3 28 38 48 01; fax: +33 3 28 38 48 47.
E-mail address: s.chollet@isa-lille.fr (S. Chollet).

judges) that might impact the efficiency of the task (Experiments 4–6).

1.1. Conventional profile

The most frequently used method to determine the sensory characteristics of a set of products is certainly the sensory profile. This method belongs to the quantitative descriptive methods and can be performed using different procedures. Among these procedures the most widely used for describing food products are the Quantitative Descriptive Analysis or QDA (Stone, Sidel, Oliver, Woolsey, & Singleton, 1974), the Quantitative Flavor Profiling (Stampanoni, 1994) and the Spectrum™ method (Munoz & Civille, 1992). All these procedures require a small number (6–18) of assessors who have been preselected for their good sensory abilities and trained to describe the products. Training includes several steps. First, assessors consensually generate a list of objective, unique, unambiguous, and independent terms that will be used to describe the products. Usually each term is associated to a physical or chemical reference and to a precise protocol of assessment. Then assessors are trained to rate, on a scale, the intensity of each attribute. Finally before using the panel for describing products, the assessors' performance is checked in terms of repeatability, discrimination and agreement. The quantitative data obtained for each attribute are generally analyzed using parametric statistics such as analysis of variance (ANOVA) and the relationship between attributes are often described with multivariate methods such as Principal Component Analysis (PCA).

Sensory profile is the only method designed to analyze products with a high degree of reliability and precision. This method provides quantifiable and relevant information on the sensory characteristics of the products but requires trained assessors. Training may vary widely because it depends on the objectives of the study in terms of precision and sensitivity but it always requires a substantive amount of time and so sensory profile is always costly and time consuming. In particular, language development and calibration are likely to require a long time to develop. In fact, these two steps can last from a few weeks to several months. Other limitations come from the use of conventional profile in the industrial environment. Training in conventional profile is generally limited to a specific type of products and thus the vocabulary generated by the panel for a given type of product is specific to it and cannot be generalized to other products. As a consequence different panels need to be trained to describe different types of products (Bitnes, Rødbotten, Lea, Ueland, & Martens, 2007) but it is not always possible for companies' sensory analysts to set up as many panels as there are types of products to analyze. One way out of this problem is to form a panel that is able to analyze all types of products. But this kind of panel takes even more time to train and involves a long pre-training step to adapt the vocabulary to each type of products. Finally, this method is completely based on language and this creates potential comprehension and agreement problems. To alleviate these problems though, and to facilitate product description, some researchers have developed some consensual vocabulary for some families of products (see, e.g. for beers, Meilgaard, Dalglish, & Clapperton, 1979; for whiskies, Shortreed, Rickards, Swan, & Burtles, 1979, for wines, Guinard & Noble, 1986; for cheeses, Guerra, Méndez, Taboada, & Fernandez-Albalat, 1999). However even though these terminologies facilitate the communication, some comprehension and interpretation problems remain and in particular, panelists still require a long time to be "calibrated." So despite its qualities, but because it often requires too much time and resources, the sensory profile evaluation is often dropped out when results are urgently needed.

1.2. Alternative methods

In order to palliate some of the drawbacks of the conventional sensory profiling, some recent alternative sensory evaluation methods have been developed. These methods have the advantage of bypassing the training stage and thus might be an economical way of describing sensory properties.

1.2.1. Free choice profiling

One of the first alternatives that appeared in the literature was *free choice profiling*. Williams and Langron (1984) described a radically different approach to descriptive analysis in which no screening and training of assessors were required and in which assessors could use any words they wanted to describe and evaluate the products (Guy, Piggott, & Marie, 1989; Marshall & Kirby, 1988; Oreskovich, Klein, & Sutherland, 1991). The data obtained with free choice profiling were initially analyzed by generalized procrustes analysis (GPA, Gower, 1971) but multiple factor analysis (MFA, Abdi & Valentin, 2007a; Escoufier & Pagès, 1990) or any 3-way type multivariate analyses (e.g. STATIS, Abdi & Valentin, 2007b; Escoufier, 1980; Lavit, Escoufier, Sabatier, & Traissac, 1994) could also be used. These types of multivariate analyses give product maps similar to those obtained with PCA. The main difference is that on the map we find the specific terms of all the assessors rather than common terms. The main advantage of this technique is that it saves much time because it does not require training other than 1 h of explanation of the testing procedure to the assessors (i.e., generation of attributes, scoring of the attributes and use of the chosen scale). A second advantage of free profiling is that the assessors—which have not been trained—can still be regarded as representing naive consumers. However, free choice profiling is not problem-free. The large diversity of vocabulary used by the assessors makes the product map difficult to interpret. In order to provide reliable guidance for product researchers, the sensory analyst has to decide upon the meaning of each attribute. Therefore the resulting descriptions of the product sensory characteristics can come more from the sensory analyst than from the assessors.

1.2.2. Flash profile

Another alternative method is the *flash profile* (Dairou & Sieffermann, 2002) which is a combination of free choice profiling and a comparative evaluation of the whole product set. The flash profile was initially developed as a method providing a quick access to the relative sensory positioning of a set of products. The main advantage of this method is to provide a product map in a very short time because the phases of familiarization with the product space, attribute generation, and rating have been integrated into a single step. Assessors simply rank the products from the least intense to the most intense for each attribute that they have themselves chosen. This method forces assessors to focus on the perceived differences and to solely use discriminative attributes. Since the structure of the data is comparable to that of the free choice profile, the same multivariate analyses can be applied to both techniques. In some cases the simultaneous presentation of the whole set of products could be a drawback, when only one product is available at a time—as, for example, in control quality. Another weakness of the flash profile is to require expert assessors. According to Delarue and Sieffermann (2004), expert assessors have previously participated in several descriptive evaluation tasks and are able to understand panel leader's instructions and generate discriminative and non-hedonic attributes, even if these assessors do not need to be trained on a specific product set. Moreover, as in free choice profiling, it could be difficult to interpret the sensory characteristics of the products because of the diversity of the vocabulary (Dairou & Sieffermann, 2002).

1.2.3. Projective mapping or napping

In parallel in the 1980's and 1990's, Risvik and collaborators developed *projective mapping* (Kennedy & Heymann, 2009; Risvik, McEwan, Colwill, Rogers, & Lyon, 1994; Risvik, McEwan, & Rødbotten, 1997), also recently re-labeled—with an intriguing blend of French and English—“Napping” (“nappe” in French means “tablecloth” Pagès, 2003, 2005). In this method, assessors are asked to draw a map (which could be drawn on a tablecloth) in two dimensions and to position the products according to the similarity and dissimilarity between these products. The coordinates of each product on the map constitute the data. Originally, projective mapping data were analyzed with PCA but more recently Pagès (2003, 2005) proposed to use MFA because this technique takes into account the differences between assessors but as previously other equivalent methods could be used. As in flash profile the advantages and the drawbacks are linked to the comparative basis of this method: all the products have to be available at the same time. Moreover this method constrains the assessors to use two dimensions to discriminate between the products (Perrin et al., 2008).

1.2.4. Sorting task

The sorting task is a simple procedure for collecting similarity data in which each assessor groups together stimuli based on their perceived similarities. Sorting is based on categorization which is a natural cognitive process routinely used in everyday life and it does not require a quantitative response. The final objective of sorting task is to reveal—via statistical analyses—the structure of the product space and to interpret the underlying dimensions. Practically the assessors are in front of a set of products and are asked to compose different groups of products such that the products in a group are similar to each other. The groups should be homogenous and coherent. The sorting task can be stopped at this point or can be followed by a description step where assessors are asked to describe each group of products (for beers: Lelièvre et al., 2008, 2009; for other products: Blancher et al., 2007; Cartier et al., 2006; Faye et al., 2006; Lawless, Sheng, & Knoops, 1995; Lim & Lawless, 2005; Saint-Eve et al., 2004; Santosa, Abdi, & Guinard, 2010; Tang & Heymann, 1999). Concerning statistical interpretation, the collected data are distance matrices which can be analyzed with two main sets of methods. The first set gives (Euclidean) map representations and comprises techniques such as multidimensional scaling analysis (MDS: Schiffman, Reynolds, & Young, 1981), DISTATIS (Abdi, Valentin, O'Toole, & Edelman, 2005; Abdi et al., 2007), multiple correspondence analysis (MCA: Cadoret, Lê, & Pagès, 2009; Takane, 1981, 1982), common components and specific weights analysis (Qannari, Cariou, Teillet, & Schlich, 2009). The second set gives tree representation and comprises clustering techniques (Miller, 1969) and additive trees (Abdi, 1990).

The sorting task is simple and easy to perform but raises several practical and methodological issues that we explore in a series of experiments described below.

2. Experimentations

2.1. General descriptions of materiel and methods

2.1.1. Beers

Six sets of beers were used in the different experimentations (Table 1). All the beers were presented in brown plastic tumblers and served between 8 and 10 °C under red light to mask the possible color differences between them and to avoid that assessors use this type of information in the sorting task. We used “blind tasting” because previous work has showed that the color of the beers has

such an importance in these tasks that chemo-sensory characteristics were neglected by most assessors (Lelièvre et al., 2009).

2.1.2. Assessors

The first two experiments were carried out with trained assessors, the other four experiments with two types of assessors: trained and untrained assessors. Trained assessors were enrolled in a training program designed to produce beer experts. They were trained 1 h a week to detect and identify flavors¹ in beer and to evaluate the intensity of general compounds² on a non-structured linear scale. The untrained assessors were beer consumers who did not have a formal training or experience in the description of beer flavors. Table 2 presents a description of the assessors' groups.

2.1.3. Sorting task procedure

The sorting task procedure was the same for all the experiments. Assessors were presented with the entire set of beers. The order of presentation of the samples was randomized prior to presentation and so was different for each assessor. Assessors started to taste all the beers one at a time. Then the assessors were asked to put together the beers that seemed similar to them. No criterion was provided to perform the sorting task. Assessors were free to make as many groups as they wanted and to put as many beers as they wanted in each group. They were allowed to take as much time as they wanted.

Assessors took part individually in the experiment in a single session. The experiments were conducted in separate booths. Mineral water and bread were available for assessors to rinse between samples. Assessors could spit out beers if they wanted.

2.1.4. Data analysis

For each assessor the results of the sorting task were encoded in individual distance matrices where the rows and the columns are beers. A value of 0 between a row and a column indicates that the assessor put these two beers together whereas a value of 1 indicates that the beers were not put together. Several methods to analyze these sorting data were used in the different experimentations: additive trees, MDS and DISTATIS (see Fig. 1).

2.1.4.1. Additive trees. Additive trees are applied on the global distance matrix and gave tree representations (Abdi, 1990). Additive trees are used to represent products as “leaves” on the tree so that the distance on the tree between two leaves reflects the similarity between the products. This method is particularly efficient to reveal the structure of the data and to evaluate its robustness.

2.1.4.2. Multidimensional scaling analysis. MDS is a multivariate statistical technique that analyzes the similarity relationships among stimuli by representing these stimuli as points on a map (Abdi, 2007b; Schiffman et al., 1981). So in a product map the products are represented by points which are positioned such that the distances between pairs of points reflect as well as possible the distances between the pairs of products: two products which have been often sorted together by the assessors are close on this representation and two products which have rarely been sorted together are far apart. MDS is, in general, performed on the global similarity matrix which is the sum of all the individual matrices.

2.1.4.3. Distatis. DISTATIS is a recent method which, contrary to the first two methods, takes into account individual sorting data as this method is performed directly on individual distance matrices

¹ The flavors used in training were almond, banana, butter, caramel, cabbage, cheese, lilac, metallic, honey, musty, bread, cardboard, phenol, apple and sulfite.

² The general compounds are bitterness, astringency, sweetness, alcohol, hop, malt, fruity, floral, spicy, sparklingness and lingering.

Table 1
Beers for each experiment.

Set 1	Set 2	Set 3	Set 4	Set 5
20 beers	12 lager beers	24 beers	12 beers	9 beers
Hoppy beers	Bitter beers	a. Ale beers	Malted beers	Blond beers
John Martin's	F	Lefte blonde	Adelscott	Pelfoth blonde
Meteor Pils,		Loburg	Welscotch	Chti blonde
Hommel bier		Tuborg [*]		Lefte blonde
1664 Fleur de Houblon ¹				
High-alcohol beers	Light, sweet, less bitter	a. Amber beers	Alcoholic beers	Amber
Bush	D	Tradition anglaise	Ekü 28	Pelforth ambrée
Bière du Démon		Jenlain	Bush	Chti ambrée
Ekü 28		George Killian's [*]	Belzebuth	Lefte ambrée
Belzebuth ¹				
Stout beers	High alcoholic degree	a. Dark beers	Tasty beers	Dark
Extra Stout	B	Lefte radieuse	Chimay Rouge	Pelforth brune
Pony Stout	C	Lefte brune	Maredsous	Chti brune
Louwaege's		Pelforth [*]	Kastell beer	Lefte brune
Leroy Stout ¹				
Gueuze beers	Banana flavor	b. Bitter beers	Bitter beers	
Saint-Louis	E	Bit Burger	Duvel	
Linderman's		Kaiser Pils	Elephant beer	
Chapeau		Ur Pils [*]	Tradition allemande	
Mort Subite ¹			Pils Urquell	
Alcohol-free beers	Without distinctive characteristics	b. Alcoholic beers		
Kronenbourg alcohol free	A	Ekü 28		
Celta		Bush		
Tourtel		Bière du Démon [*]		
Buckler ¹				
	A with an aroma	b. Malted beers		
	Honey	Adelscott		
	Caramel	Welscotch		
	Banana	Bière de l'Evenement [*]		
	Metallic			
	Bread			
	Lilac			

For set 1, either the entire set was used (20 beers), or a part of this set composed of only 15 beers: one beer of each group was removed (noted¹).

The set 3 was divided in two sets of 12 beers, noted 3-a and 3-b.

^{*} Indicates a beer presented twice.

(Abdi, Valentin, O'Toole, & Edelman, 2005; Abdi et al., 2007). DISTATIS is a three-way generalization of classical multidimensional scaling and like MDS it provides a map of the products. This map—called the “compromise map”—integrates the assessors distance matrices in the most efficient way. Like in MDS, in this map, the proximity between two points reflects their similarity. Moreover DISTATIS provides information about assessors' agreement because it also shows how each assessor positioned the products relative to the compromise map. DISTATIS then uses this information to compute statistical confidence ellipses around products. In addition DISTATIS provides an MDS-like map of the assessors which can be used to analyze the structure of the group of assessors. DISTATIS is very useful when we need to analyze the results of an experiment in which several groups of assessors sort a set of beers.

2.2. Experiment 1: Can we use a sorting task with verbalization to describe beer sensory properties?

In the sorting task assessors are asked to concentrate on the global similarities between the products, whereas in profile, assessors should analyze their perceptions with specific attributes whose choice strongly influence their discriminating ability (Blancher et al., 2007). Some authors (Bárcenas, Pérez Elortondo, & Albisu, 2003; Chauhan & Harper, 1986; Saint-Eve et al., 2004) even question one of the basic principles of conventional profile based on the independent analysis of the sensory properties. According to these authors, the profile cannot reveal interactions. Moreover, the sorting task is not, *a priori*, a descriptive method because the verbalization step is performed after the group formation. Thus, the

description concerns a group of products and not an individual product. To sum up, different strategies are involved in conventional profiles and sorting tasks and we can expect that descriptions of beers obtained with a sorting task followed by a verbalization should be less precise than those obtained with a conventional profile. To evaluate this hypothesis we compared the descriptions of a set of beers obtained with the conventional profile and the sorting task for trained assessors.

2.2.1. Procedure

2.2.1.1. Sorting task with verbalization. This task consisted in two steps. In the first step nine trained assessors were provided with 15 beers of set 1 and were asked to sort the beers as described in Section 2.1.3. In the second step, after performing the sorting task, the trained assessors were asked to describe each group of beers with a word, preceded or not with a term indicating the intensity such as “little,” “medium” or “very.”

2.2.1.2. Profile. The nine assessors were asked to quantify on a 10 cm linear scale eight sensory characteristics of the beers: bitter, sweet, acid, astringent, alcohol, hop, malt, and after-taste. To minimize carry-over effects due to memory between the two tasks, the profile was carried out one month after the sorting task. Between these two tasks assessors were trained with other types of beers.

2.2.2. Data analysis

Data analysis consisted in two steps. In the first step, the frequency of citation of each word used in the sorting task was computed for each beer and analyzed with correspondence analysis (CA). The profile data were analyzed with a PCA. In the second step,

Table 2
Composition of the assessors' groups for each experiment. (abv.: y.o. = years old).

Beer set	Assessors	Age and sex	Training
Set 1	9 trained assessors	9 men Mean age: 18.5 y.o. Range age: 18–19 y.o.	17 h
	16 untrained assessors	5 women and 11 men Mean age: 22 y.o. Range age: 19–30 y.o.	/
Set 2	22 trained assessors	6 women and 16 men Mean age: 38.9 y.o. Range age: 19–64 y.o.	11 h
	18 untrained assessors	5 women and 13 men Mean age: 35.2 y.o. Range age: 22–62 y.o.	/
Set 3	9 trained assessors	2 women and 7 men Mean age: 33.4 y.o. Range age: 23–54 y.o.	75 h
	9 untrained assessors	2 women and 7 men Mean age: 23.4 y.o. Range age: 22–26 y.o.	/
Set 4	15 untrained assessors at the beginning becoming 15 trained assessors at the end of the experiment	3 women and 12 men Mean age: 37.6 y.o. Range age: 25–60 y.o.	0 h at the beginning 15 h at the end
Set 5	13 trained assessors	5 women and 17 men Mean age: 34.9 y.o. Range age: 25–53 y.o.	3.4 years (1 h/week)
	18 untrained assessors (A)	5 women and 13 men Mean age: 29.3 y.o. Range age: 20–70 y.o.	/
	19 untrained assessors (B)	6 women and 13 men Mean age: 26.6 y.o. Range age: 22–56 y.o.	/
	18 untrained assessors (C)	19 women and 9 men Mean age: 24.6 y.o. Range age: 21–31 y.o.	/

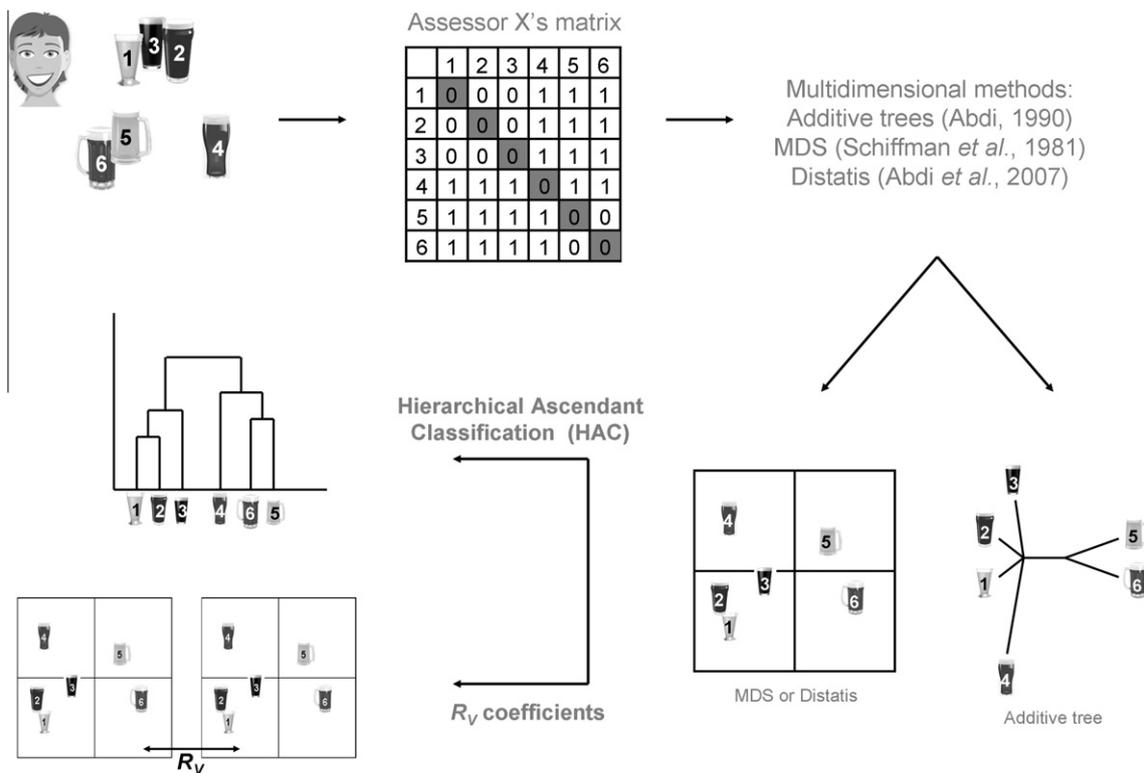


Fig. 1. Schema of statistical tools used to analyze sorting task data.

a hierarchical cluster analysis (HCA) using the Ward criteria was applied to the factorial coordinates of the beers in the spaces defined by the PCA and CA. The clusters identified by truncating the tree diagrams were consolidated by aggregation around mobile centers. The attributes best defining the resulting clusters were identified by computing their probability of characterizing a cluster (Lebart, Morineau, & Piron, 1995).

2.2.3. Results

Figs. 2 and 3 present a comparison of clusters and descriptions coming from the sorting task and profile. The same number of beer groups (4) is observed for the two methods. Two beer groups are identical: the stout beers and the alcohol-free beers. In the sorting task the Gueuze beers constitute a group by themselves whereas beers with high degree of alcohol are grouped with hoppy beers. By contrast, the profile data reveal a group of beers with a high degree of alcohol but do not separate Gueuze from hoppy beers. Moreover the maps obtained in sorting task and in profile are significantly similar (as evaluated by the R_V coefficient: $R_V = .336$, $p < .05$).

Concerning the vocabulary, the sorting task provided a somewhat larger number of terms (19 in the sorting task and 16 in the profile). However among the terms from the sorting task, 26% were imprecise and uninformative terms such as “insipid” or

“light.” Only the term “sweet” was used in the profile and sorting tasks to describe the same groups of beers.

2.2.4. Discussion

Similar product maps are obtained with the sorting task and the profile with trained assessors. This result is consistent with the works of Blancher et al., 2007; Cartier et al., 2006; Faye et al., 2006, 2004; Saint-Eve et al., 2004; Tang & Heymann, 1999. Nevertheless it seems that the profile provides more precise and especially more easily interpretable descriptions than the sorting task with verbalization. However the sorting task provides global information about basic, salient and common characteristics. Taken into account these differences between sorting and profile, in order to describe precisely and reliably complex products such as beers an obvious question is: Do we need conventional profiling (with its training phase) or can we modify the sorting task to improve its description quality?

2.3. Experiment 2: Can we improve the efficiency of a sorting task with verbalization with a list of terms and a rating?

As it was exposed previously, compared with standard profiling the sorting task gives similar results in terms of beer clustering but the vocabulary interpretation is more difficult. This difficulty arises

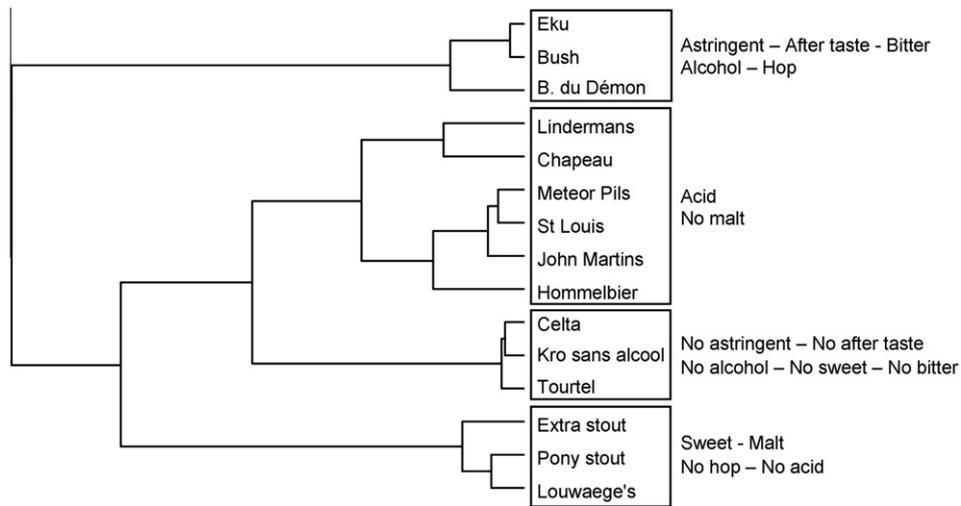


Fig. 2. HCA computed on factorial coordinates of the beers in the space defined by PCA of profile. The attributes best defining each cluster are also presented.

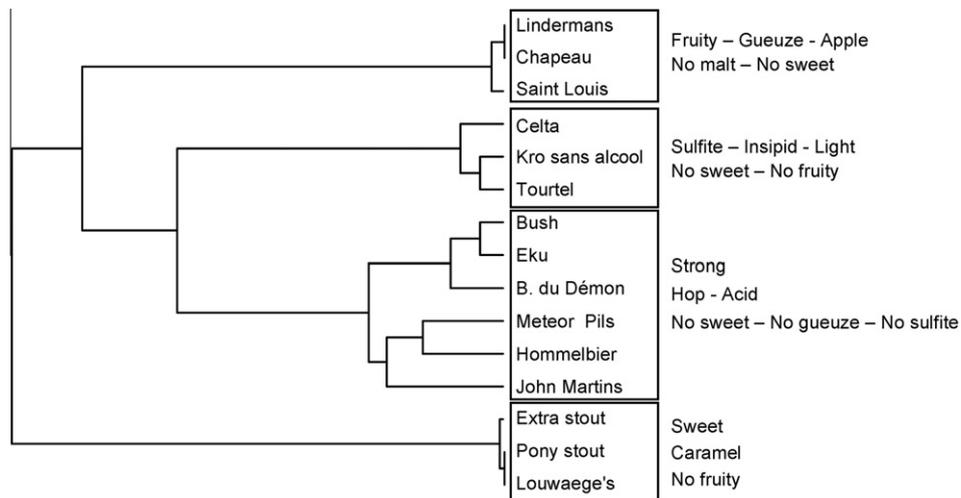


Fig. 3. HCA computed on factorial coordinates of the beers in the space defined by CA of sorting task. The attributes best defining each cluster are also presented.

from the imprecision of the assessors' terms and from the lack of common terms among the assessors. In addition, the statistical analysis of the vocabulary is more delicate because the assessors provide descriptive words but do not rate their intensity. Therefore the analysis uses frequencies instead of the quantitative data (means, standard deviation, etc. per attribute) obtained from the profile. In order to overcome this problem we devised a modification of the sorting task which includes a special verbalization procedure. This modification consists in providing a list of attributes to the trained assessors and asking them to quantitatively rate these attributes. We ran the following experiment to evaluate the efficiency of this new procedure by comparing the results to those of the profile and the sorting task without list.

2.3.1. Procedure

After performing the sorting task with the nine beers of set 5, trained assessors were asked to describe each group with some words, according to two conditions. In the first condition assessors were free to use their own words. In the second condition assessors had to choose their words from a list of terms. This list was extracted from the Flavor Wheel of the International Terminology System for Beer (Meilgaard et al., 1979) which comprises 44 terms. In both conditions assessors were told to use no more than five words per group of beers and to indicate the intensity of the descriptors using a four-point scale labeled: "not," "a little," "medium" and "very."

2.3.2. Data analysis

Each intensity term was converted into a score to obtain an intensity score for each term quoted to describe the groups of beers using the following coding scheme: "not" = 0, "a little" = 1, "medium" = 2 and "very" = 3. Then in order to analyze the vocabulary, we computed the geometric mean for each quoted term and each beer (see Lelièvre et al., 2008 for more details). Since these data

are quantitative the profile and sorting task can now be analyzed exactly in the same way with a PCA followed by a HAC on the beer's coordinates derived from the PCA.

2.3.3. Results

Figs. 4–6 present the clusters and descriptions coming from the profile, sorting task without list, and sorting task with a list. We observe three clusters for the profile, four for the sorting task without list and five for sorting task with a list. Just one group is strictly identical in the three methods (amber Pelforth). In the sorting task without list, a group of beers from Leffe brewery clearly emerges but not in the profile where the three Leffe beers are in the same cluster as dark Pelforth, amber Chti and dark Chti. In the profile two blond beers are together (blond Pelforth and blond Chti) and in the sorting task without list these two beers are grouped with the amber Chti. In the sorting task with a list, the three Leffe are not clustered together, the dark Leffe being with the dark Pelforth. The R_V coefficient computed between the product coordinates on PCA of profile and sorting tasks shows a greater similitude between profile and sorting task without list ($R_V = 0.574, p < .05$) than between profile and sorting task with a list ($R_V = .260, p > .05$).

Concerning the vocabulary, as in the previous experiment, fewer terms were used to describe the beer groups in profile (9) than in the sorting task (15 for sorting task without list and 13 for sorting task with a list). The provided list does not change the number of terms used to describe beer groups. Only the term "yellow fruit" is used in both the profile and the sorting task without list to describe the same beer group. Only two terms are common to the profile and the sorting task with list ("sulfitate" and "caramel"), but they are not used to describe the same beers.

2.3.4. Discussion

It seems that the sorting task without list and the profile give similar results in term of categories of beers. These results concur

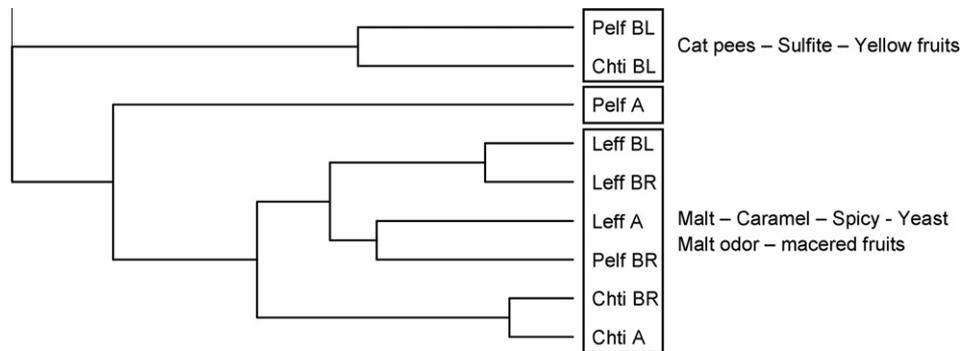


Fig. 4. HCA computed on factorial coordinates of the beers in the space defined by PCA of profile. The attributes best defining each cluster are also presented.

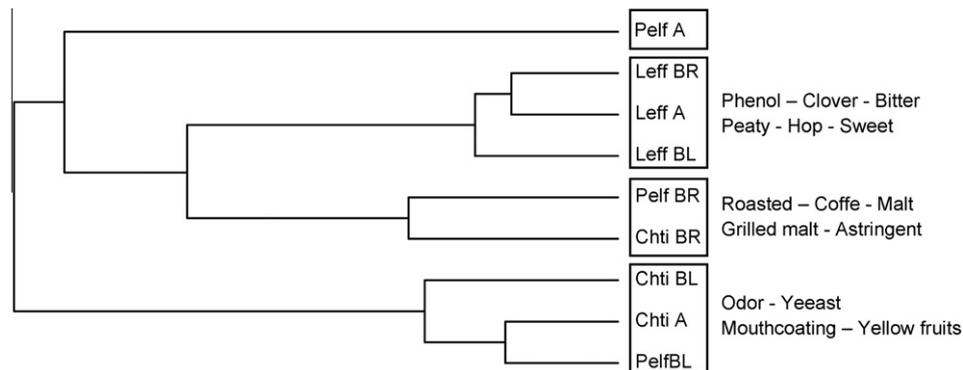


Fig. 5. HCA computed on factorial coordinates of the beers in the space defined by PCA of sorting without list. The attributes best defining each cluster are also presented.

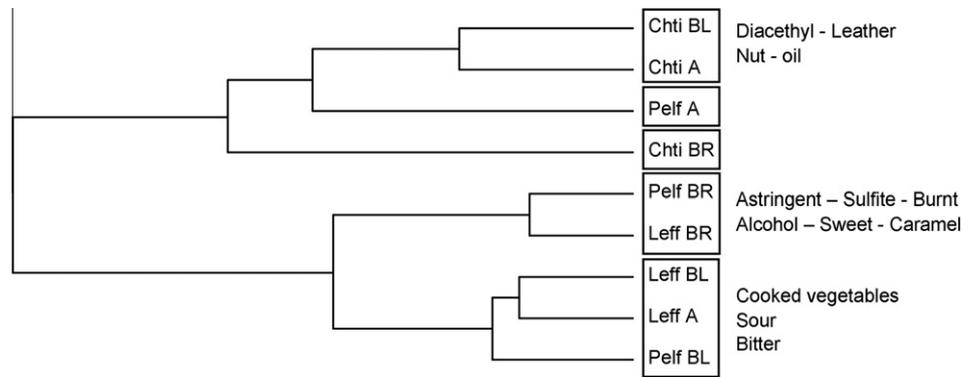


Fig. 6. HCA computed on factorial coordinates of the beers in the space defined by PCA of sorting with a list. The attributes best defining each cluster are also presented.

with conclusions reached from sorting tasks performed on other kinds of products (Blancher et al., 2007; Cartier et al., 2006; Faye et al., 2004, 2006; Saint-Eve et al., 2004; Soufflet et al., 2004; Tang & Heymann, 1999). However the sorting task with a list leads to results different from those of the profile. So it seems that using a list did not help the trained assessors. Two reasons could explain this result. First it is possible that our list was too long (44 terms), and, indeed Hughson and Boakes (2002) showed that a short list (14 terms) helped assessors in a matching task. Second it could be that the terms provided were different from the terms used in training. For example, trained assessors described blond Chti as butter in the sorting task without list but did not use the diacetyl term (which is associated with the butter flavor) in the sorting task with a list. Some authors such as Rainey (1986), Civille and Lawless (1986), Stampanoni (1994) or Chollet and Valentin (2001) have already underlined the importance of using a common terminology based on references. Finally it is interesting to add that asking assessors to rate the descriptive terms facilitates the statistical interpretation (contrary to free choice profiling or flash profile) because this allows using parametric statistical methods as with the profile data.

2.4. Experiment 3: Should we use trained or untrained assessors to perform sorting task?

Because it involves the comparison of a set of products the sorting task rely both on assessors' discrimination and short term memory abilities. Previous works showed that these two abilities are linked to participants' expertise level. For example, Avancini de Almeida, Cubero, and O'Mahony (1999) report that trained panelists are less impaired by a delay between samples in a discrimination task than novices, suggesting that they might have developed better perceptual memory abilities than novices. Likewise, Chollet, Valentin, and Abdi (2005) found that trained assessors outperformed untrained assessors on discriminative tasks, but only with beers learned during training. Taken together these results lead us to conjecture that the sorting task might also be contingent to expertise level. In particular the superior discrimination and memory abilities of experts might lead them to different categorization schemes than novices. To test this hypothesis we compared the results of different sorts obtained with trained and untrained assessors.

2.4.1. Procedure

We compared the results of six sorting tasks (sets 1, 2, 3-a, 3-b, 4 and 5) obtained by different groups of untrained and trained assessors. The procedure was the same as described in part 2.3.

2.4.2. Data analysis

For each set of beers and for each group of assessors we first counted the number of beer groups made by trained and untrained

assessors in the different sorting tasks. Then a MDS was performed on the distance matrices. In order to compare the configurations obtained for trained and untrained assessors we computed R_V coefficients between the product coordinates obtained from each configuration. The R_V coefficient (Escoufier, 1973) measures the similarity between two configurations and can be interpreted in a manner analogous to a squared correlation coefficient but its statistical test requires a specific procedure (see Abdi, 2007a, 2010, for details).

2.4.3. Results

A series of Student t -tests showed that, except for one set of beers (set 2), the numbers of groups performed by trained and untrained assessors were not significantly different. Nevertheless, on the whole, trained assessors tend to use more groups than untrained assessors. To confirm this trend we pooled the data of the six tests and computed a Student t -test. This test confirmed that trained assessors make more beer groups than untrained assessors [$t(159) = 2.85, p < .01$].

The R_V coefficients computed between the product coordinates obtained from the MDS configurations of trained and untrained assessors for the six different sorting tasks were significant for four sorting tasks indicating that trained and untrained MDS configurations are not significantly different, except for Sets 2 and 3-a. The difference between trained and untrained assessors is not surprising for Set 2 because this set is composed of very similar lager beers (the same lager beer with an added aroma at low concentration). For Set 3-a, this difference is more surprising. An additive tree analysis of the sorting data shows that trained assessors' results are very structured (78.3% of variance) whereas for untrained assessors trees are less easily interpretable and less structured (47.8% of variance).

To better estimate the quality of untrained assessors' results, we compared the results of three groups of untrained assessors with those of a group of trained assessors on Set 5. The position of each beer for each group of assessors has been plotted on the same maps using DISTATIS (Fig. 7).

Fig. 7 reveals two key points. *First* the three different groups of 18 untrained assessors behave similarly: for each beer, the points denoted A, B and C are very close to each other. This means that these three different groups of 18 untrained assessors gave similar results in the sorting task. *Second* untrained assessors' results are similar to those of trained assessors.

2.4.4. Discussion

To sum up, all these results show that for most sets of beers, untrained assessors were able to perform the sorting task as well as trained assessors but that trained assessors tended to use more groups than novices. It is also interesting to note the fact that all three groups of novices provided similar results, a pattern that

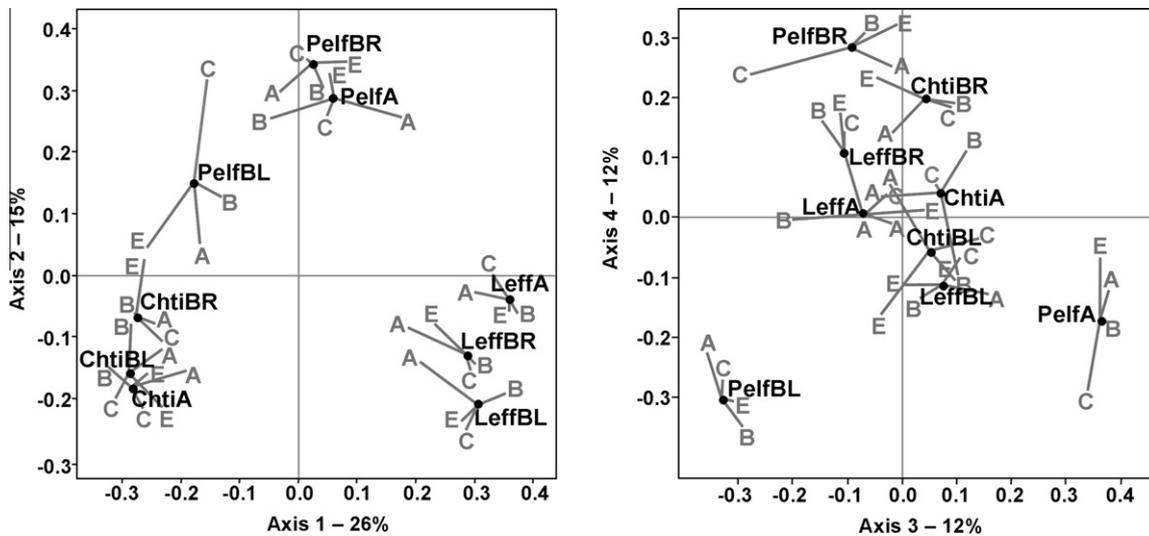


Fig. 7. Four dimensional compromise maps for 13 trained (E) and 3 groups of 18 untrained assessors (A, B, C) for the set 5.

shows that the sorting task with novices provides stable results. Even if trained assessors outperform untrained assessors in discriminative and memory tasks (Chollet et al., 2005; Chollet & Valentin, 2006; Valentin, Chollet, Béal, & Patris, 2007), these two factors do not seem to prevent untrained assessors from performing efficiently sorting tasks.

2.5. Experiment 4: How robust is the sorting task?

A good measure needs to have good validity and reliability. For the profile this is ensured by monitoring the individuals as well as the panel as a whole (Meilgaard et al., 1999). In addition, for the profile, different aspects of performance are checked such as the reproducibility, discrimination and agreement between assessors. To estimate the quality of sorting task as a sensory tool, we carried out a series of experiments: *First* we compared the results of two sorting tasks realized by the same group of assessors; *second* we looked if two samples from the same beer were classified in the same group, and *finally* we estimated the assessors' variability in the sorting task.

2.5.1. Procedure

2.5.1.1. Duplicated sorting task. Two sorting tasks using the same group of assessors (trained or untrained assessors) were carried out and the group results were compared. For each set of beers (Sets 2, 3-a, 3-b and 5), both repetitions were carried out in the same session; there was about 15 min between the repetitions.

2.5.1.2. Duplicated samples. Trained and untrained assessors carried out two sorting tasks with two sets of 12 beers (3-a and 3-b) where three beers were duplicated.

2.5.1.3. Variability. The participants' agreement in the sorting task was first evaluated by measuring the between-individual stability for the first repetition of Set 5 for trained and untrained assessors. Second, using a bootstrap procedure (Efron & Tibshirani, 1993) on the data, we examined the stability of the categorizations over 68 sorts (Lelièvre et al., 2009).

2.5.2. Data analysis

2.5.2.1. Duplicated sorting task. We first performed an MDS on the sorting data for each repetition of the sets of beers of each group

of assessors (trained and untrained assessors). Then we computed R_V coefficients between the product coordinates of each repetition.

2.5.2.2. Duplicated samples. We computed the percentage of assessors grouping both beers in the same group.

2.5.2.3. Variability. R_V coefficients were computed between the individual matrices of each assessor and the rest of her/his group for the first repetition.

2.5.3. Results

2.5.3.1. Duplicated sorting task. The R_V coefficients between the MDS configurations for the four sets of beers show that, except for Set 2, the configurations from the two different repetitions are similar. The results for Set 2 are not surprising because it is composed of very similar lager beers (the same lager beer with an added aroma at low concentration) which makes the task particularly difficult. Moreover, for this same set of beers we found a difference between the configuration of trained and untrained assessors (see previous section). We can conjecture that novices got tired during the second sort. For trained assessors, another explanation could be advanced: it is possible that trained assessors changed deliberately their criteria between the first and the second sort because they recognized the beers. But globally, as shown in other studies on different types of products (Cartier et al., 2006; Lawless & Glatter, 1990; MacRae, Rawcliffe, Howgate, & Geelhoed, 1992), we can conclude that group results (trained and untrained assessors) are repeatable with the sorting task.

2.5.3.2. Duplicated samples. For the first set of beers (Set 3-a), 40.7% of trained and untrained assessors categorized the repeated beers in the same group and 51.8% of trained assessors and 40.7% of untrained assessors for the second set (Set 3-b). So it seems that the performance of both groups of assessors is globally similar. Nevertheless this performance is lower than what we could expect.

2.5.3.3. Variability. A Student t -test performed on R_V coefficients computed between the individual matrices of each assessor and the rest of her/his group for trained and untrained assessors showed a significant difference in consensus between trained and untrained assessors [$t(29)=2.18$; $p < .05$]. This shows that trained assessors were more consensual ($R_V = 0.38 \pm 0.03$) than the untrained assessors ($R_V = 0.33 \pm 0.07$) on their categorization.

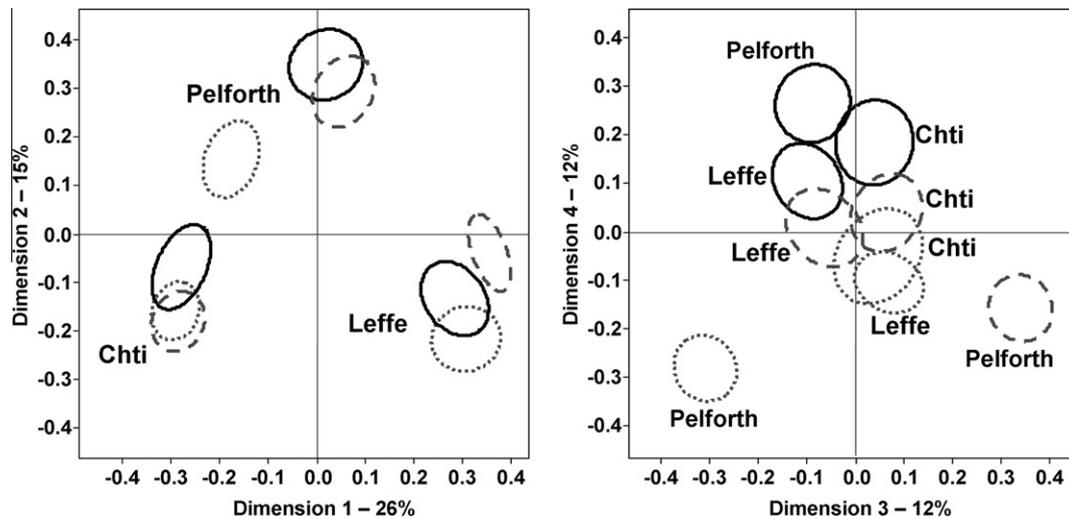


Fig. 8. Four dimensional compromise maps of the products with 95% confidence ellipsoids computed by bootstrap on trained assessors and untrained assessors' data (68 assessors). The solid black lines represent the dark beers, the dashed gray lines represent the amber beers and the dotted gray lines represent the blond beers. (from Lelièvre et al., 2009).

Fig. 8 shows the four-dimensional compromise map. The confidence ellipsoids represent the variability of the results over the 68 assessors. We observe that the assessors' categorizations are stable because of the small size of ellipsoids.

2.5.4. Discussion

To conclude the sorting task is a robust tool. Our result replicate and confirms those of Falahee and MacRae (1997), Cartier et al. (2006) and Lawless and Glatter (1990) who observed similar results in repeated sorting tasks on different products. Moreover the assessors' agreement is relatively high especially for trained assessors. However the duplicated samples were not always classified in the same group but this result could be explained by a carry-over effect. As the order of presentation was random and different for all assessors the taste of a beer could have been "contaminated" by a strong beer tasted before for some assessors. Perhaps we would not observe similar results with products less persistent than beer.

2.6. Experiment 5: How many beers can be categorized with a sorting task?

A recurrent question when using sorting task is: Is there a limit to the number of products that can be evaluated? Besides the fact that some products cannot be tasted in large number because of their alcohol content or their high persistence, the question of how many products to be sorted is important from a memory point of view. As already mentioned, because of the necessity to compare products, performing a sorting task certainly involves short term memory. Short term memory is known to have a limited capacity: the number of items or chunks that can be retained at once, also called memory span, is estimated to be around seven plus or minus two (Miller, 1969). As a consequence when the numbers of products to sort exceed assessors' memory span they have to taste several times the products and the risk of interference increases. And so we can conjecture that there is an optimal number of beers to sort. We tested this hypothesis by examining the maximum number of beers that can be categorized in a sorting task.

2.6.1. Procedure

We used data from Sets 1 and 3. For Set 1 trained and untrained assessors had to sort first 15 beers out of 20 from Set 1 and then

the entire set of 20 beers. For Set 3, trained assessors had to sort 24 beers presented either as two separate sets of 12 beers each (Sets 3-a and 3-b) or a single set of 24 beers (Sets 3-a + and 3-b). To evaluate whether increasing the number of beers to be sorted at once affected assessors' performance, we compared the results obtained with (a) 15 versus 20 beers and (b) 12 versus 24 beers.

2.6.2. Data analysis

Because we want to specifically explore the structure of the groups obtained from the sorting task, we decided to analyze the result with additive trees which are particularly suited for this aim (Abdi, 1990).

2.6.3. Results

2.6.3.1. Comparison of 15 beers versus 20 beers: Set 1. Fig. 9 presents the additive trees with 15 beers and 20 beers for trained and untrained assessors. A strong structure is observed with 15 beers and 20 beers for trained and untrained assessors (as indicated by a large proportion of explained variance of 87.9%, 88.4%, 83.7% and 83.3% respectively for the four additive trees). We can just notice a slight decrease of variance between sorts with 15 beers and those with 20 beers. Moreover we observe roughly a similar structure for trained and untrained assessors. Globally, on the four additive trees we identify similar groups of beers: stout beers (Extra Stout, Pony Stout, Louwaeghe's, Leroy Stout), alcohol-free beers (Kronenbourg sans alcool, Celta, Tourtel, Buckler) and Gueuze beers (Saint-Louis, Linderman's, Chapeau, Mort Subite). So a sorting task could be carried out with 20 beers without losing efficiency.

2.6.3.2. Comparison of sorts of 12 beers versus 24 beers: Set 3. Fig. 10 presents the additive trees of trained assessors obtained after analyzing the data of three sorting tasks: two with two sets of 12 beers (Sets 3-a and 3-b) and one with the complete set of 24 beers (Set 3-a plus Set 3-b). For the two sets of 12 beers, a strong structure of classification was observed (78.3% and 83.5% of variance). For Set 3-a we can identify four groups of beers: one group with the two repeated Tuborg beers, one group with the three Leffe beers, one group with the Jenlain beer, the "Tradition anglaise," the Loburg and one of the two George Killian's beers and the last group with the two repeated Pelforth beers and the other George Killian's beer. For Set 3-b the additive tree shows also four groups of beers: one group with one of the two repeated Ur Pils beers, the Adelscott

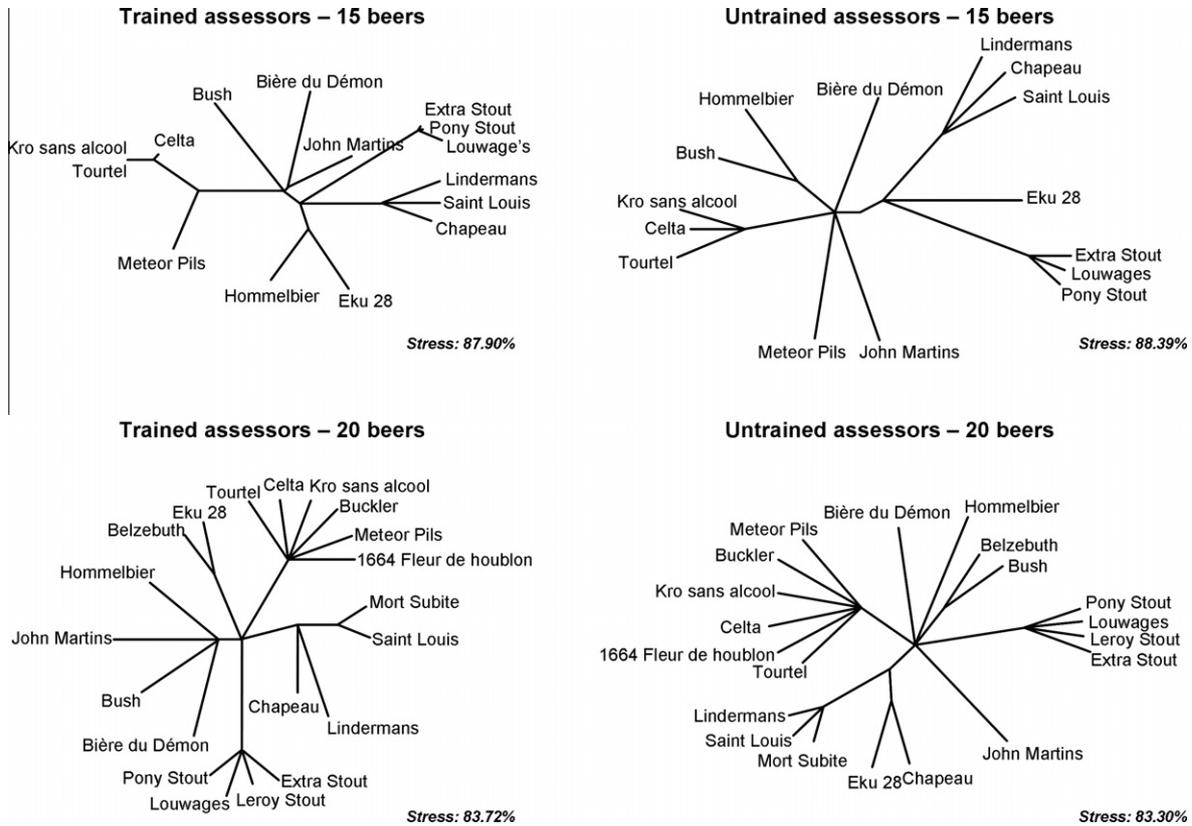


Fig. 9. Additive trees with 15 beers and with 20 beers for trained assessors and for untrained assessors.

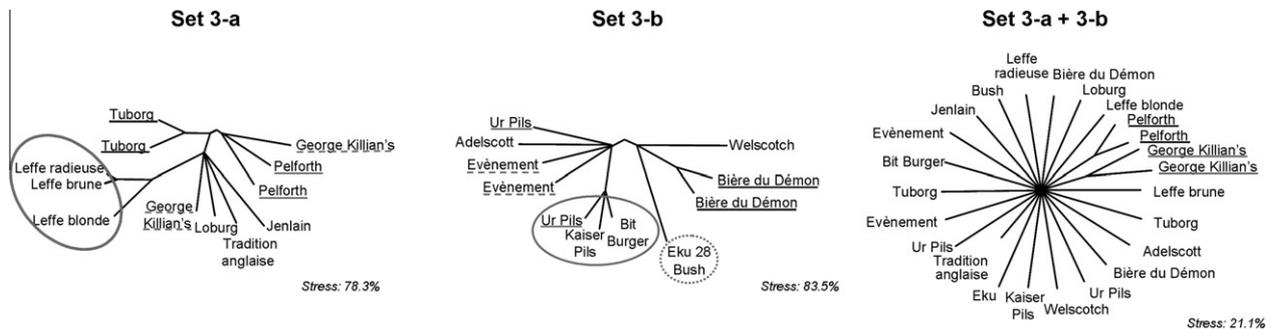


Fig. 10. Additive trees for trained assessors with set 3-a, set 3-b and set 3-a + set 3-b.

and the two repeated “Evenement” beers, one group with the second Ur Pils, the Bit Burger and the Kaiser Pils, one group with the Eku 28 and the Bush and the last group with the two repeated Demon and the Welscotch beers. However for the complete set of 24 beers, a weak structure is observed (21.1% of variance) with almost as many groups as beers: just the two repeated Pelforth and George Killian’s beers are grouped together. So it seems that 24 beers are too many beers for performing a sorting task.

2.6.4. Discussion

Two phenomena might explain the decrease in efficiency with the larger number of beers. The first phenomenon is directly linked to the product itself: beer is a complex product which contains alcohol and persistent bitter compounds. These characteristics do not facilitate beer tasting and a large number of beer samples could lead to sensory fatigue as well as to a diminution of assessors’ acuteness and attention. The second phenomenon is linked to the assessors and more precisely to their short term memory abilities. When an assessor begins the task, he or she tastes one

beer and tries to memorize it. Then the assessor tastes a second beer and compares it to the first one to decide if they belong to the same group and so on until the last beer is evaluated. So the larger the number of beer samples in a sorting task, the greater the short term memory load. These memory problems in sorting tasks have previously been highlighted in an original experiment (Patris, Gufoni, Chollet, & Valentin, 2007) evaluating the strategies used by participants during the task. In this experiment, verbal data collected after the sorting task were analyzed in conjunction with behavioral indicators. Results showed that trained and untrained participants expressed difficulties to memorize the beer samples during the task. These difficulties are likely to increase when beers have similar tastes. Thus the number of beers which can be used in sorting task depends also on the resemblance between beers.

To conclude, the efficiency of the sorting task decreases as the number of beers increases and it seems that 20 beers might be the maximum number of beers that can be efficiently sorted at once. The efficiency of the sorting task also seems to decrease

when the number of beers is too small: sorting task performed with 8 beers gave poor results compared to a similar task carried out with 12 beers (Nava Guerra, Chollet, Gufoni, Patris, & Valentin, 2004). More generally we can advise to carry out a sorting task using 9–20 beers with an optimum number being around 12 beers (depending of the type of beers, their similarities and the degree of alcohol). But what to do with more than 20 beers to sort? Several solutions could be considered. The first one is the use of incomplete block. But in this case, as when we use incomplete block in consumer test, more assessors are needed. Unfortunately to obtain relevant results we rapidly need a large number of assessors. Another possibility is to split the large set of beers in several smaller sets and to add in each smaller set the same beer (called prototype) and to compare to this prototype. In this case the choice of the prototype is a crucial step.

2.7. Experiment 6: Do instructions on the number of groups matter?

The instructions for the sorting are quite vague and in general do not specify the number of groups to make. Yet, because some authors have found that experts make more groups than novices (Augustin & Leder, 2006; Chatard-Pannetier, Brauer, Chambres, & Niedenthal, 2002; Chi, Feltovich & Glaser, 1981) providing an indication on the number of groups might help novices to sort the beers more precisely.

2.7.1. Procedure

We compared the results obtained without and with instructions on the number of groups to make on Set 1 (15 beers) for trained and untrained assessors. In this set, because we have a *priori* 5 family of beers (hoppy beers, alcoholic beers, Stout beers,

Gueuze beers, and alcohol-free beers), the instructions were to make 5 groups of beers.

2.7.2. Data analysis

Results were analyzed using additive trees.

2.7.3. Results

Fig. 11 presents the additive trees in the conditions without (top of Fig. 11) and with instructions (bottom of Fig. 11) for trained and untrained assessors. A strong structure of the tree is observed without and with instructions for trained and untrained assessors (87.9%, 88.4%, 90.2%, and 87.8% of variance respectively for the four additive trees). We can notice that providing instructions did not increase the variance. Moreover similar structures are observed for both groups of assessors. Globally we identify similar groups of beers on the four additive trees: Stout beers (Extra Stout, Pony Stout, Louwages), free alcohol beers (Kronenbourg sans alcool, Celta, Tourtel) and Gueuze beers (Saint-Louis, Linderman's, Chapeau).

2.7.4. Discussion

We can conclude that for this set of beers, there is no impact of the instructions on the results of the sorting task. Providing the number of groups to make does not seem to change the structure of classification of both trained and untrained assessors. In other words, spontaneous categorization (without instruction) is similar to oriented categorization (with instructions on the number of groups). This means that assessors spontaneously used the expected groups of beers. The absence of difference observed in this experimentation might be due to the strong structure of the data in five groups. So the constraint of making five groups corresponded to the natural mental representations of the assessors. Results might have been quite different if we had asked assessors to

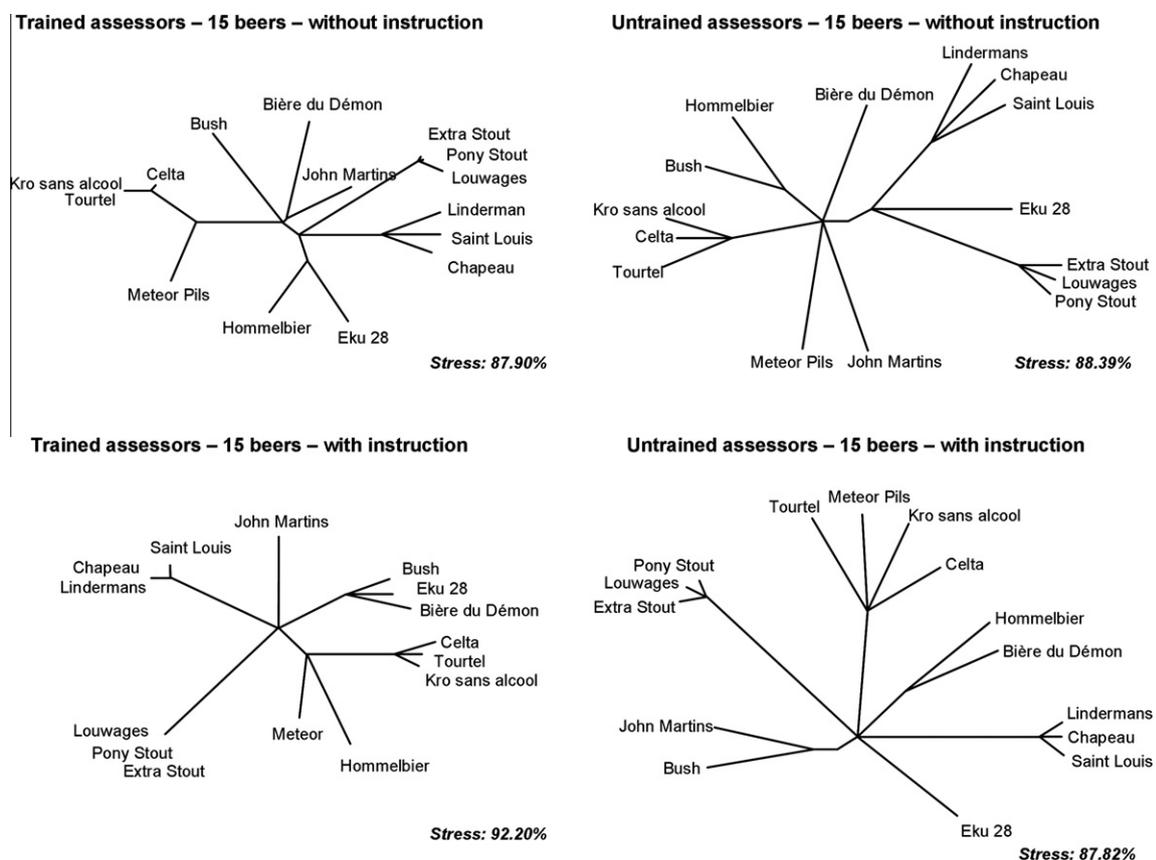


Fig. 11. Additive trees without instruction and with instructions for trained and untrained assessors.

separate the beers in three groups. This interpretation is confirmed by the results of Lelièvre (2010) who observed different results when assessors were asked to perform a free sorting task and a binary sorting task with one group of Trappist beers and one group of non-Trappist beers. In the free sorting task, trained and untrained participants tended to split the Trappist and non-Trappist beers whereas no such separation was observed in the binary sorting task. These results suggest that, in the free sorting task, assessors relied mostly on perceptual clues whereas in the binary sorting they relied on the “concept” of Trappist which seemed to be somewhat fuzzy for these assessors.

3. Conclusion

The review of the results from different sorting experiments suggests answers to several questions about the methodology of the sorting task—in particular a sorting task could be used with about 20 novice assessors, is a robust tool, is efficient with no more than 20 beers and provides groups similar to those obtained from a profile. However the sorting task can, in some cases, provide descriptions difficult to interpret and not as precise as those obtained from a standard profile.

From a practical point of view, it is important to remember that sorting is much less time-consuming than the conventional profile. The sorting task needs more assessors (at least 20) than the profile does (10–15). However the sorting task can be performed with trained as well as untrained assessors. In the sorting task the samples have to be present in the same time and the sample number is limited compared to the profile. As in the profile the assessors' performance could be evaluated in term of repeatability, discrimination and agreement; but in the sorting task this estimation is global and not detailed attribute by attribute. Moreover authors using the sorting task generally report that this is a rapid method for obtaining perceptual maps of a large set of products. However even if the principle of the task is easy to understand, the task itself is not so easy: Indeed assessors report some memory problems and have difficulties to decide upon the specific criteria to use to make the groups (Patris et al., 2007). Finally it seems that the sorting provides qualitative information whereas the profile provides quantitative information.

In relationship with this conclusion the sorting task might be a useful tool for selecting products before another test such as a profile or a consumer test. For example if we have 10 new formulations and want to organize a consumer test just with three products we can use the sorting task to select the three products the more different or the more similar according to the objectives of the study. The sorting task could also be used in control quality, for example, in order to obtain an estimation of the variation of sensory characteristics according to the age of products or to the different batches. We can even imagine that the sorting task could replace a series of triangular tests. Moreover the free sorting task could be an appropriate method to determine the general characteristics of a product from a given family when we know *a priori* the relevant sensory characteristics of different members from this family. Indeed, based on the proximity structure of the members, we can deduce the membership of the studied products and thus derive its sensory characteristics. And finally the sorting task could also help marketing research by providing map in which products are compared to their competitors.

Acknowledgment

This work was supported by the Institut Supérieur d'Agriculture, CESA, the region of Nord-Pas-de-Calais, Qualtech and Danone. The authors would also like to gratefully thank the anonymous review-

ers who for their helpful comments on a previous version of this paper.

References

- Abdi, H. (1990). Additive-tree representations. *Lecture Notes in Biomathematics*, 84, 43–59.
- Abdi, H. (2007a). The R_V coefficient and the congruence coefficient. In N. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 849–853). Thousand Oaks (CA): Sage.
- Abdi, H. (2007b). Metric multidimensional scaling. In N. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 849–853). Thousand Oaks (CA): Sage.
- Abdi, H. (2010). Congruence coefficient, R_V coefficient, and mantel coefficient. In N. J. Salkind, D. M. Dougherty, & B. Frey (Eds.), *Encyclopedia of Research Design*. Thousand Oaks (CA): Sage. pp. 222–229.
- Abdi, H., Valentin, D., O'Toole, A.J., & Edelman, B. (2005). DISTATIS: The analysis of multiple distance matrices. *Proceedings of the IEEE Computer Society: International Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA, pp. 42–47.
- Abdi, H., & Valentin, D. (2007a). STATIS. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage. pp. 955–962.
- Abdi, H., & Valentin, D. (2007b). Multiple factor analysis. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage. pp. 657–663.
- 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, 627–640.
- Augustin, M. D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychology Science*, 48, 135–156.
- Avancini de Almeida, T. C., Cubero, E., & O'Mahony, M. (1999). Same-different discrimination tests with interstimulus delays up to one day. *Journal of Sensory Studies*, 14, 1–18.
- Bárceñas, P., Pérez Elortondo, F. J., & Albu, M. (2003). Comparison of free choice profiling, direct similarity measurements and hedonic data for ewes' milk cheeses sensory evaluation. *International Dairy Journal*, 13, 67–77.
- Bitnes, J., Rødbotten, M., Lea, P., Ueland, Ø., & Martens, M. (2007). Effect of product knowledge on profiling performance comparing various sensory laboratories. *Journal of Sensory Studies*, 22, 66–80.
- Blancher, G., Chollet, S., Kesteloot, R., Nguyen Hoang, D., Cuvelier, G., & Sieffermann, J.-M. (2007). French and Vietnamese: How do they describe texture characteristics of the same food? A case study with jellies. *Food Quality and Preference*, 18, 560–575.
- Cadoret, M., Lê, S., & Pagès, J. (2009). A factorial approach for sorting task data (FAST). *Food Quality and Preference*, 20, 410–417.
- Cartier, R., Rytz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., et al. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, 17, 562–571.
- Chatard-Pannetier, A., Brauer, M., Chambres, P., & Niedenthal, P. (2002). Représentation, catégorisation et évaluation: Différence entre experts et novices dans le domaine des meubles d'antiquités. *L'Année Psychologique*, 102, 423–448.
- Chauhan, J., & Harper, R. (1986). Descriptive profiling versus direct similarity assessments of soft drinks. *Journal of Food Technology*, 21, 175–187.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152.
- Chollet, C., & Valentin, D. (2001). Impact of training on beer flavor perception and description: Are trained and untrained assessors really different? *Journal of Sensory Studies*, 16, 601–618.
- Chollet, S., & Valentin, D. (2006). Impact of training on beer flavour perception. *Cerevisia, Belgian Journal of Brewing and Biotechnology*, 31, 189–195.
- Chollet, S., Valentin, D., & Abdi, H. (2005). Do trained assessors generalize their knowledge to new stimuli? *Food Quality and Preference*, 16, 13–23.
- Civille, G. V., & Lawless, H. T. (1986). The importance of language in describing perceptions. *Journal of Sensory Studies*, 1, 203–215.
- Dairou, V., & Sieffermann, J.-M. (2002). A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *Journal of Food Science*, 67, 826–834.
- Delarue, J., & Sieffermann, J.-M. (2004). Sensory mapping using Flash profile comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15, 383–392.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Boca Raton: Chapman & Hall.
- Escofier, B., & Pagès, J. (1990). Multiple factor analysis. *Computational Statistics & Data Analysis*, 18, 121–140.
- Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751–760.
- Escoufier, Y. (1980). L'analyse conjointe de plusieurs matrices de données. In M. Jolivet (Ed.), *Biométrie et Temps* (pp. 59–76). Paris: Société Française de Biométrie.
- Falahee, M., & MacRae, A. W. (1997). Perceptual variation among drinking waters: The reliability of sorting and ranking data for multidimensional scaling. *Food Quality and Preference*, 8, 389–394.
- Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Profiling free sorting verbalization tasks with naive assessors: An alternative to descriptive mappings. *Food Quality and Preference*, 15, 781–791.

- Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A., & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference*, 17, 604–614.
- Gower, J. C. (1971). Statistical methods of comparing different multivariate analysis of the same data. In F. R. Hodson, D. Kendall, & P. Tautou (Eds.), *Mathematics in the Archeological and Historical Sciences* (pp. 138–149). Edinburgh: University Press.
- Guerra, B., Méndez, J., Taboada, M., & Fernandez-Albalat, M. P. (1999). Training of a panel for the evaluation of cheese flavor. *Milchwissenschaft*, 54(4), 203–205.
- Guinard, J. X., & Noble, A. C. (1986). Proposition d'une terminologie pour une description analytique de l'arôme des vins. *Science de l'Aliment*, 6, 657–662.
- Guy, C., Piggott, J. R., & Marie, S. (1989). Consumer profiling of Scotch whisky. *Food Quality and Preference*, 1, 673–693.
- Hughson, A. L., & Boakes, R. A. (2002). The knowing nose: The role of knowledge in wine expertise. *Food Quality and Preference*, 13, 463–472.
- Kemp, S. E., Hollowood, T., & Hort, J. (2009). *Sensory Evaluation. A Practical Handbook*. Oxford: Wiley–Blackwell.
- Kennedy, J., & Heymann, H. (2009). Projective mapping and descriptive analysis of milk and dark chocolates. *Journal of Sensory Studies*, 24, 220–233.
- Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (STATIS) method. *Computational Statistics & Data Analysis*, 18, 97–119.
- Lawless, H. T., & Glatler, S. (1990). Consistency of multidimensional scaling models derived from odor sorting. *Journal of Sensory Studies*, 5, 217–230.
- Lawless, H. T., Sheng, N., & Knoops, S. S. C. P. (1995). Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference*, 6, 91–98.
- Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Lelièvre, M. (2010). *De la première gorgée de bière au concept: Le même processus pour tous? Effet de l'expertise sur la catégorisation et la représentation des bières*. Dijon: Thèse de l'université de Bourgogne.
- Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2008). What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 19, 697–703.
- Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2009). Beer trained and untrained assessors rely more on vision than on taste when they categorize beers. *Chemosensory Perception*, 2, 143–153.
- Lim, J., & Lawless, H. T. (2005). Qualitative differences of divalent salts: Multidimensional scaling and cluster analysis. *Chemical Senses*, 30, 719–726.
- MacRae, A. W., Rawcliffe, T., Howgate, P., & Geelhoed, E. N. (1992). Patterns of odour similarity among carbonyls and their mixtures. *Chemical Senses*, 17, 119–125.
- Marshall, R. J., & Kirby, S. P. J. (1988). Sensory measurement of food texture by free-choice profiling. *Journal of Sensory Studies*, 3, 63–80.
- Meilgaard, M., Civille, G. V., & Carr, B. T. (1999). *Sensory Evaluation Techniques* (3rd Edition). Boca Raton: CRC Press.
- Meilgaard, M. C., Dalglish, C. E., & Clapperton, J. F. (1979). Beer flavour terminology. *Journal of the American Society of Brewing Chemists*, 37, 47–52.
- Miller, G. A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 6, 169–191.
- Munoz, A. M., & Civille, G. V. (1992). The spectrum descriptive analysis method. In R. C. Hootman (Ed.), *ASTM Manual Series MNL 13, Manual on Descriptive Analysis Testing*. PA: West Conshohocken.
- Nava Guerra, M.T., Chollet, S., Gufoni, V., Patris, B., & Valentin, D. (2004). *Chemosensorial Expertise and Categorization: The Case of Beers*. European Chemoreception Research Organisation, Dijon, France.
- Oreskovich, D. C., Klein, B. P., & Sutherland, J. W. (1991). Procustes Analysis and its applications to Free-Choice and other sensory profiling. In H. T. Lawless & B. P. Klein (Eds.), *Sensory Science Theory and Application in Foods* (pp. 353–393). New York: Marcel Dekker.
- Pagès, J. (2003). Recueil direct de distances sensorielles: Application à l'évaluation de dix vins blancs du Val de Loire. *Sciences des aliments*, 23, 679–688.
- 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*, 16, 642–649.
- Patris, B., Gufoni, V., Chollet, S., & Valentin, D. (2007). Impact of training on strategies to realize a beer sorting task: Behavioral and verbal assessments. In D. Valentin, D. Z. Nguyen, & L. Pelletier (Eds.), *New trends in sensory evaluation of food and non-food products*. Ho Chi Minh (Vietnam): Vietnam National University-Ho Chi Minh City Publishing House.
- Perrin, L., Symoneaux, R., Maitre, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory descriptive methods for use with the Napping[®] procedure: Case of ten wines from Loire Valley. *Food Quality and Preference*, 19, 1–11.
- Qannari, E. M., Cariou, V., Teillet, E., & Schlich, P. (2009). SORT-CC: A procedure for the statistical treatment of free sorting data. *Food Quality and Preference*, 21, 302–308.
- Rainey, B. A. (1986). Importance of reference standards in training panelists. *Journal of Sensory Studies*, 1, 149–154.
- Risvik, E., McEwan, J. A., Colwill, J. S., Rogers, R., & Lyon, D. H. (1994). Projective mapping: A tool for sensory analysis and consumer research. *Food Quality and Preference*, 5, 263–269.
- Risvik, E., McEwan, J. A., & Rødbotten, M. (1997). Evaluation of sensory profiling and projective mapping data. *Food Quality and Preference*, 8, 63–71.
- Saint-Eve, A., Paçi Kora, E., & Martin, N. (2004). Impact of the olfactory quality and chemical complexity of the flavouring agent on the texture of low fat stirred yogurts assessed by three different sensory methodologies. *Food Quality and Preference*, 15, 655–668.
- Santosa, M., Abdi, H., & Guinard, J.-X. (2010). A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality and Preference*, 21, 881–892.
- Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to Multidimensional Scaling*. New York: Academic Press.
- Shortreed, G.W., Rickards, P., Swan, J.S., & Burtles, S. (1979). *The Flavour Terminology of Scotch Whisky*. Brewer's Guardian, November, 2–6.
- Soufflet, I., Calonnier, M., & Dacremont, C. (2004). A comparison between industrial experts' and novices' haptic perception organization: A tool to identify descriptors of handle of fabrics. *Food Quality and Preference*, 15, 689–699.
- Stampanoni, C. R. (1994). The use of standardized flavor languages and quantitative flavor profiling technique for flavored dairy products. *Journal of Sensory Studies*, 9, 383–400.
- Stone, H., & Sidel, J. (1993). *Sensory Evaluation Practices*. San Diego: Academic Press, Inc..
- Stone, H., Sidel, J., Oliver, S., Woolsey, A., & Singleton, R. C. (1974). Sensory evaluation by quantitative descriptive analysis. *Food Technology*, 28, 24.
- Takane, Y. (1981). MDSORT: A special purpose multidimensional scaling program for sorting data. *Journal of Marketing Research*, 18, 480–481.
- Takane, Y. (1982). IDSORT: An individual differences multidimensional scaling for sorting data. *Behavior Research Methods and Instrumentation*, 14, 546.
- Tang, C., & Heymann, H. (1999). Multidimensional sorting, similarity scaling and free choice profiling of grape jellies. *Journal of Sensory Studies*, 17, 493–509.
- Valentin, D., Chollet, S., Béal, S., & Patris, B. (2007). Expertise and memory for beers and beer olfactory compounds. *Food Quality and Preference*, 18, 776–785.
- Williams, A. A., & Langron, S. P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture*, 35, 558–568.