Food Quality and Preference 27 (2013) 108–119

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

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

"All-In-One Test" (AI1): A rapid and easily applicable approach to consumer product testing

Davide Giacalone, Wender L.P. Bredie, Michael Bom Frøst *

Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

ARTICLE INFO

Article history:
Received 6 October 2011
Received in revised form 28 September 2012
Accepted 29 September 2012
Available online 12 October 2012

Keywords:
Consumer research
Sensory perception
Check-All-That-Apply
L-PLSR
Preference mapping
Beer

ABSTRACT

Methods to collect sensory profiles relying directly on consumers' perceptions are increasingly employed. A consumer test method simultaneously collecting information about the consumer background, appropriateness rating for specific sensory properties, hedonic ratings and a sensory profile by the Check-All-That-Apply (CATA) technique is reported. In this exploratory "All-In-One Test" (AI1), subjects ($N = 160$) filled out a questionnaire with demographic and psychographic variables, and appropriateness ratings for specific sensory descriptors of beer. Subsequently, subjects gave hedonic ratings for six beers and assessed the presence of specific sensory properties by CATA. The dataset was analyzed by L-shaped Partial Least Square Regression (L-PLSR) to link product information (X), liking ratings (Y) and consumer background information (Z). The CATA technique effectively discriminated between all beers and identified the underlying sensory dimensions. Consumer psychographics – particularly previous knowledge and interest in beer – explained liking better than demographics. Appropriateness correlated well with liking for some sensory properties, whereas others showed large discrepancy between appropriateness and actual hedonic response. Overall, the AI1 test provided interpretable results concerning consumer perception (sensory/hedonic) of the beers, and revealed relations with consumers' background information. Initial results with AI1 test show that it is an efficient and versatile approach for exploratory product testing with consumers.

1. Introduction

1.1. Consumer data

Methods to collect sensory information directly from consumers (i.e. individuals without formal training in sensory evaluation) seem to be of great importance in sensory research. In the present context, we define *consumer data* as information about food products collected "with more or less natural conditions, with appropriate samples and with sufficient representative *untrained* individuals" (Schutz, 1999) (italics added), be they hedonic or analytic in nature. Hedonic tests – *viz.*, asking consumers to quantify their degree of liking for a given product – are typical test in which untrained individuals are involved (Lawless & Heymann, 2010; Schutz, 1999). According to a traditional view on sensory evaluation, these are also the *only* tests where untrained individuals should be involved, whereas trained assessors are suitable for analytical sensory tests (Lawless & Heymann, 2010; Stone & Sidel, 1993).

Nevertheless, a broader involvement of consumers is advocated by many as a way to enhance *external validity*, i.e. the ability of a sensory test to predict actual marketplace behavior.

Sensory analysis has traditionally played a prominent role in quality control of food products (e.g. to identify deviations from a prescribed quality standard, determine shelf-life of food products, etc.), generally focusing on understanding the links between physicochemical characteristics of the products and sensory responses (van Trijp & Schifferstein, 1995). This type of sensory analysis generally employs a small sample of often well-trained panelists, maintains good control over testing situation (i.e. mostly laboratory based research), and a clear separation of sensory and hedonic measurements (Moskowitz & Chandler, 1979).

Sensory analysis and methods have gradually expanded in the area of food product development and consumer preferences (Köster, 1981; McBride, 1990; van Trijp & Schifferstein, 1995), causing a shift in terms of scientific focus: from studying the food, to studying *both* the food *and* the consumer (Meiselman, 1993). By doing so, it has moved towards marketing, a domain with a rather different approach to product testing, based on the use of large samples of untrained consumers, little or no control over testing situation and conjoint measurement of sensory and hedonic ratings (Moskowitz & Chandler, 1979). Marketing approaches to sensory analyses would stress external rather than internal validity (van Trijp & Schifferstein, 1995). Schutz (1988) suggested four major elements (or dimensions) that influence the external validity of food acceptance studies: type of respondents, type of test

* Corresponding author. Tel.: +45 3533 3207; fax: +45 3533 3509.
E-mail address: mbf@life.ku.dk (M.B. Frøst).

0950-3293/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2012.09.011

environment, type of measurement and type of test stimuli. These dimensions can also be applied to product perception methodology (van Trijp & Schifferstein, 1995). Concerning the type of respondent, a small panel of trained subjects is the preferred choice from an internal validity perspective, because by effect of training their use of terminology achieves consistency and their response are then traceable to objective product characteristics. The product oriented approach to sensory testing is based on the assumption that panelists are interchangeable (sensory acuity and performance are the only things that are monitored). On the contrary, market oriented research maintains that people differ fundamentally in their perception (Wansink, 2003), and thus require larger sample to ensure that data are projectable and that the peculiarities of specific segments can be observed. Moreover, since the vast majority of consumers in the market place do not have specific skills in terms of sensory evaluation, there is a high risk that consumers and trained panel could perceive the products differently (Ares, Deliza, Barreiro, Gimenez, & Gámbaro, 2010; van Trijp & Schifferstein, 1995). Therefore a trained sensory panel could describe the product differently from consumers, and/or take into account variables that are irrelevant to the end-consumer. On this line of reasoning, consumer data based on simple sensory concepts would ensure a better understanding of how people perceive the sensory characteristics of food products (Ares, Deliza et al., 2010). Indeed, with regards to the external validity, trained judges yield the lowest level, whereas random consumers the highest (Schutz, 1988; van Trijp & Schifferstein, 1995). Concerning the type of test environment, a product oriented approach would emphasize a highly controlled environment, designed to minimize biases (via e.g. climate and noise control, detailed instructions, palate cleansers and so on). Although this makes perfectly sense from an internal validity perspective, such procedures are highly artificial when compared to how real consumers interact with food. Indeed, the criterion of external validity would prefer the test circumstance to be as natural as possible (Köster, 1981; Meiselman, 1993, 2008; Schutz, 1988; van Trijp & Schifferstein, 1995). Concerning the type of measurement, external validity would suggest that sensory and hedonic measurements be corroborated by other form of measurement such as appropriateness by use (Schutz, 1988), expected sensory properties (Moskowitz & Chandler, 1979), and possibly non cognitive responses such as observation of actual behavior (Dijksterhuis, 2006). Concerning the type of stimuli, the closer the item tested is to the version seen by the consumer, the higher the external validity (Schutz, 1988).

Nevertheless, these prescriptions towards what could be called consumer oriented sensory evaluation have been difficult to implement and have encountered much opposition and generated a debate over the value that can be given to consumer data.

Attribute meaning is one of the primary concerns. When trained panels are used, consistency is achieved during training, when subjects agree on the meaning of the sensory attributes and on the use of scale. Moreover, during training the specific sensation can be linked up to a specific standard reference (*viz.*, a concrete stimulus with known properties), enhancing the degree to which sensory response can be linked to physical product characteristics (though in practice verbal description is more frequent – see Murray, Delahunty, & Baxter, 2001). Consumers, for the most part, are able to express their opinion on more superficial (i.e. non-technical and idiosyncratic attributes), which limits the actionability of this data for product guidance (Muñoz, 2003). Discriminative ability is another controversial point. It has been demonstrated that sensory discrimination ability increases to a specific stimulus or attribute domain (Gibson, 1969), and indeed trained panels are usually able to detect very small differences in the products, whereas consumers allegedly do not (Hough, 1998). However, discriminative ability is mostly relevant when the panel is used for quality control, or other tasks where the goal is to maintain product integrity, in which cases trained panels appear the most appropriate option. When consumers are used for sensory tasks, the purpose is often to obtain developmental guidance, i.e. getting a direct feedback on sensory characteristics, and not detection of subtle differences (Moskowitz, 1998).

Nevertheless, the (prescribed) involvement of consumers continues hitherto to be a bone of contention among researchers, limiting the market orientation of sensory studies (Meiselman, 1994; van Trijp & Schifferstein, 1995).

More recently, this issue was again debated in this journal almost a decade ago after the publication of ''Measuring consumers' response to food products'' by Garber, Hyatt, and Starr (2003), which led to a lively discussion about the way sensory tests and food-related consumer research should be conducted (Food Quality and Preference, Volume 14, Issue 1).

Change advocates argued for a series of measures to increase the realism of sensory tests and improve the external validity of the results obtained. Such measures included, among others: use of real consumers in taste tests; performing the experiment in a natural or naturalistic setting (for a review, see Meiselman, 1993, 2008); identifying target consumer segments; and including certain elements of the marketing mix (e.g. brand) (Garber et al., 2003). Critical commentators dismissed these suggestions for various reasons. Cardello (2003) argued that, while Garber's suggestions would benefit product-focused research (*viz.*, applied research), sensory science should be concerned with basic underlying mechanism for acceptance and rejection of food. Other critiques concerned the challenges of performing traditional sensory tests with untrained subjects, the uncertainty connected to performing experiments in uncontrolled settings, and the necessity to avoid confusing overlaps between the realm of sensory science and that of marketing (Buck, 2003; Cardello, 2003; Civille & Heylmun, 2003).

1.2. Consumer-based sensory evaluation

A pragmatic look at the field suggests that several intermediate form of sensory analysis exist. Many practitioners today advocate, under certain conditions and for specific purposes, the use of consumers for analytical tests. Moreover, sensory profiles given by consumers have also been shown to be discriminative and repeatable (Albert, Varela, Salvador, Hough, & Fiszman, 2011; Husson, Le Dien, & Pagès, 2001; Moskowitz, 1996; Moussaoui & Varela, 2010; Worch, Lê, & Punter, 2010).

Indeed, especially in the last years new methods to collect sensory data relying directly on consumers' perception of specific products are increasingly employed.

Methodological advancements brought forth a series of profiling techniques suitable for working with untrained subjects, such as Free-Choice profiling (Williams & Langron, 1984), Projective Mapping (Risvik, McEwan, Colwill, Rogers, & Lyon, 1994), Flash Profiling (Dairou & Sieffermann, 2002), Sorting (Abdi, Valentin, Chollet, & Chrea, 2007) and Check-All-That-Apply (CATA, Adams, Williams, Lancaster, & Foley, 2007). Especially CATA has recently gained popularity due to its high rapidity and ease of use. This method consists in having respondents tick all the words they deem appropriate to describe a food product, and it was introduced to the sensory community at the 7th Pangborn Sensory Science Symposium (Adams et al., 2007). Since then, it has then been already applied with promising results in several studies (Adams et al., 2007; Ares, Barreiro, Deliza, Gimenez, & Gámbaro, 2010; Ares, Deliza et al., 2010; Puyares, Ares, & Carrau, 2010) who showed the discriminative ability of CATA as a method and its use in e.g. consumer-led food product development.

Such data consists of dichotomous response (presence = 1; absence = 0) in correspondence to each attributes, and several multivariate statistical techniques have been applied for exploration and representation of this data. The usability of CATA with consumers as a profiling method was documented in a few studies: Dooley, Lee and Meullenet (2010), in a study on vanilla ice cream, compared the product space obtained by CATA with consumers and that obtained by trained assessors with the Spectrum® method (Sensory Spectrum Inc., Chatham, NJ, USA), and found the two spaces to be in very close agreement. The same result was obtained when CATA was compared with Quantitative Descriptive Analysis (Ares, Barreiro et al., 2010). Other studies further validated the method by showing that it provides similar results to other more commonly used profiling techniques used with consumers, such as projective mapping/Napping® and intensity scaling (Ares, Barreiro et al., 2010; Ares, Deliza et al., 2010; Ares, Varela, Rado, & Gimenez 2011; Reinbach, Giacalone, Machado, Bredie, & Frøst, 2012). Additionally, an increasing number of scholars have successfully used consumer-based CATA for preference mapping (Dooley et al., 2010; Parente, Ares, & Manzoni, 2010; Parente, Manzoni, & Ares, 2011; Piqueras-Fiszman, Ares, Alcaide-Marzal, & Diego-Más, 2011).

Theoretically, the disadvantage of this type of data compared to more conventional techniques (specifically those based on multidimensional scaling of several attributes) is that it is based on relatively impoverished dichotomized data, masking relative differences between specific attributes and, in general, making its application unviable for product sets with small products differences. However, provided that sufficient sensory differences exist, it has been shown that CATA has a discriminatory power analogous to intensity ratings and Napping (Reinbach et al., 2012).

(Possible) Drawbacks of CATA seem to be compensated from the main advantages that the method delivers: firstly, it is a very fast and spontaneous method, minimizing the amount of time and cognitive effort that is asked of a subject, and thus is regarded as a very appropriate method to use with naïve consumers. An important corollary to this observation is the fact that CATA questions have been reported by the method proponents to be easier to understand and to have a smaller halo effect on perceived liking than JAR or intensity scales (Adams et al., 2007, see also Ares, Barreiro et al., 2010), yet it is still just a claim. Secondly, the rationale for having consumers performing a sensory task is generally not to obtain a very accurate sensory characterization of the product, but rather to obtain insights into how consumers understand the product from a sensory point of view, and how sensory characteristics may structure their preference patterns.

Several users found CATA to be a simple and valid tool for gathering information about food products based on consumer perception (Ares, Barreiro et al., 2010; Ares, Deliza et al., 2010; Puyares et al., 2010;, both in terms of complementarities with other methods and as a standalone method when e.g. a sensory trained panel is not available (Ares, Barreiro et al., 2010).

1.3. AI1 approach

Building on these premises, the aim of this study is to further expand the possibilities of combining different domains of consumer product perceptions simultaneously in one test method, which we coined the "All-In-One Test" (AI1).

This test combines several existing techniques for collection of consumer data into a unique experimental approach, and aims at simultaneously collect information about consumer background, appropriateness for specific sensory properties, hedonic ratings of food products and a sensory profile by Check-All-That-Apply (CATA) technique – using consumers as sole data source.

This paper reports a first application of the AI1 test in an exploratory study on six commercial Danish beers with a convenience sample of consumers ($n = 160$).

The following sections illustrate the general experimental design, as well as a data analytical procedure to efficiently handle the generated data.

2. Materials and methods

2.1. General design

The overall study design is shown in Fig. 1. The experiment consisted of two consecutive steps: a questionnaire part, where consumers provided information about specific demographic and psychographic variables, as well as appropriateness ratings for a series of flavors and sensory descriptors for an ideal beer; a sensory part, where consumers were presented with six beer samples (in a monadic sequence), for which they had to score perceived liking and provide a sensory profile of the beer by answering to a series of CATA questions.

2.2. Location

The study was conducted in the Danish city of Aarhus within the 2010 "Aarhus Week Festival", a large and renowned cultural event in Denmark. The experiment took place in two consecutive days in one of the event facilities.

2.3. Subjects

A convenience sample of subjects was used: they were essentially passers-by taking part in the larger city event. In total, 160 subjects (68 women and 92 men) took part in the study. Their age spanned unevenly from 18 to over 75 years old, with a prevalence of younger subjects (age interval 18–25, $n = 48$; 26–35, $n = 21$, 36–45: $n = 17$, 46–55, $n = 27$; 56–65, $n = 31$; 66–75, $n = 13$; Over 75, $n = 3$).

2.4. The beers

Six commercially available beers were used in the study (Table 1). They were chosen to represent different beer styles available in the Danish beer market. The beers were stored in a refrigerator prior to serving (5 °C), and they were served in clear glasses, blind labeled with a three digit code.

2.5. Experimental procedures

The test included two consecutive steps: a conceptual part (questionnaire) and a beer tasting.

2.5.1. Step 1: Computer-based questionnaire

After agreeing to take part, participants were asked to complete a computer-based questionnaire designed to collect two kinds of information: (1) demo- and psychographic data, and (2) appropriateness ratings in relation to a series of flavors and sensory descriptors for an "ideal" beer. The questionnaire was designed and conducted with the software Katim v.1.0. (Laugesen, 2007).

Demographic information collected were gender, age, education level, monthly income, number of persons in the household, number of beers consumed per month. Psychographic questions concerned either product specific attitudes (interest in beer, self-reported beer knowledge) or general attitudes towards food (e.g. general interest in food, willingness to try novel food) and were collected on a 7-points Likert scale.

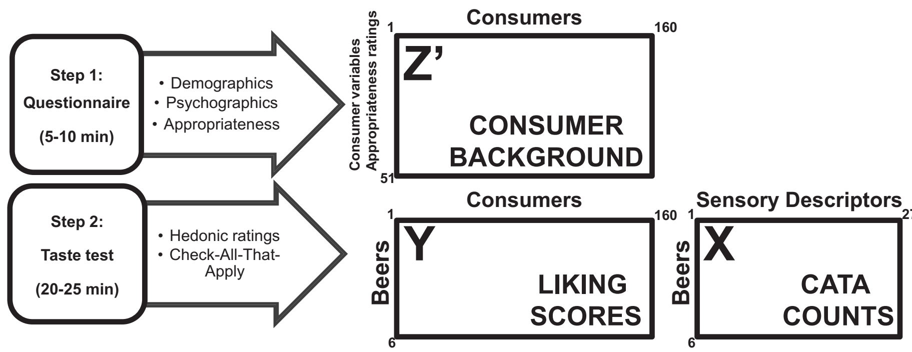

Fig. 1. Experimental procedures and resulting data structure.

Table 1
Description of the beers.

Name (Style)	Brewery	Description
Hvede (Eng. *Wheat*) I.P.A. *(India Pale Ale)*	Indslev Bryghus	Top-fermented beer. It is an unconventional wheat beer influenced by English pale ale. Light golden in color and more hoppy than normal light wheat beers. Its aromatic taste is characterized by a combination of different kind of hops (Hallertauer Perle, Amarillo, Cascade and Hallertauer Saphir). It has a soft sweet touch and a long bitter aftertaste. Alcohol by volume (ABV): 5.5%
New York Lager *(American Lager)*	Nørrebro Bryghus	Bottom-fermented beer, powerful, rich and dark golden in color. It has an intense malt character with caramel notes, pronounced bitterness and a flowery aroma of North American Cascade hops. ABV: 5.4%
Ravnsborg Rød (Eng. *Red*) *(Red Ale)*	Nørrebro Bryghus	Top-fermented beer, it is a variant of the classic British Amber or Red Ale. It is characterized by its russet color and a light malt character, mixed with an intense fruity and aromatic scent. The taste presents also a spicy touch given by Amarillo hops. ABV: 5.5%
River Beer *(Pilsner)*	Grauballe Bryghus	Bottom-fermented beer, it is a pilsner characterized by a stronger, hoppier flavor than usual lagers. The taste is very influenced by two varieties of hops: Saaz and First Gold. ABV: 4.6%
Brown Ale *(Brown Ale)*	Thisted Bryghus	Top-fermented beer, it is inspired by the English brown ales. This beer is fermented with three different kinds of yeast, giving it a slightly refined and spicy aroma. Its flavor is rich and slightly sweet, with a note of fresh bitterness. ABV: 6.9%
Porse Bock *(Bock)*	Thisted Bryghus	Bottom-fermented beer, it is a quite complex lager beer. It is dark golden in color and is characterized by a distinct bitterness, achieved through the combination of hops and bog myrtle (*Myrica gale*, "porse" in Danish). ABV: 5.8%

Appropriateness ratings for 27 sensory descriptors in an "ideal" beer was given, on a 7-points scale (e.g. "how appropriate do you think it is for a beer to be *bitter*?"); the two semantic anchors were "1 = not at all appropriate" and "7 = extremely appropriate". Since some of the descriptors referred to specific ingredients, some of them unusual for beer, a visual aid portraying each ingredient was provided. The 27 descriptors used for the study are listed in Table 2. They were selected on the basis of the guidelines provided by The Danish Beer Academy, an organization active in promoting knowledge and awareness about beers.

In the sensory science field, the concept of appropriateness is mostly used in relation to appropriateness by use of a food product, to add info to a taste test. In the present context, rating appropriateness was a pure cognitive task inasmuch as we asked consumers to rate the perceive appropriateness of a series of sensory descriptors for an "ideal" beer. This type of data might be useful for knowing how consumers conceptualize expected sensory properties of a product, based on prior consumption experiences, advertising, etc. Moreover, this kind of data can also be related to hedonic response and sensory information – as we do in this paper – to explore whether people in fact like what they find appropriate on a conceptual level.

A full version of the questionnaire is available in the online version (see Appendix).

2.5.2. Step 2: Hedonic and sensory test

After completion of the questionnaire, subjects proceeded to the tasting.

Each consumer received six beer samples in a monadic sequence, with a balanced serving order (MacFie, Bratchell, Greenhoff, & Vallis, 1989) to minimize systematic carry-over and position effects. For mouth rinsing, water was provided after each sample.

Table 2
Sensory descriptors used for conceptual appropriateness ratings and Check-All-That-Apply task, and examples provided in the visual aid.

English name	Examples provided
Flowers	Geranium, rose, elder flower and others
Beans	Chocolate beans and coffee beans
Intense berries	Sea-buckthorn, elderberry, rose hip, aronia
Caramel	None (self-explanatory)
Nuts	Walnuts, hazelnuts, almonds, pine seeds
Savoury spices	Cinnamon, cloves, cardamom etc.
Dessert spices	Vanilla and Honey
Regional spices	Juniper and bog myrtle
Herbs	Anise, rosemary, laurel etc.
Citrus fruit	Lemon, orange, mandarin etc.
Berries	Strawberry, raspberry, black currant etc.
Fruit	Apple, peer, banana, etc.
Dried fruit	Rosins, dried plums, dried figs
Liquor	Porto, sherry, grappa
Bitter	None (self-explanatory)
Sparkling	None (self-explanatory)
Refreshing	None (self-explanatory)
Fruity	None (self-explanatory)
Aromatic	None (self-explanatory)
Pungent	None (self-explanatory)
Smoked	None (self-explanatory)
Foamy	None (self-explanatory)
Still	None (self-explanatory)
Sour	None (self-explanatory)
Sweet	None (self-explanatory)
Warming	None (self-explanatory)
Vinous	None (self-explanatory)

Each respondent received six evaluation sheets (one for each sample) where they were asked to evaluate the perceived liking on a 7-point hedonic scale. Subsequently, each respondent had to answer a series of CATA questions with 27 flavor nuances and sensory attributes which they found appropriate for describing the beer sample. The CATA descriptors were exactly the same as the one used in the preliminary questionnaire (see Table 2). This approach made possible a comparison between expected sensory properties and actual sensory experience.

Subjects used the same evaluation sheet both to rate liking and to fill out the CATA questions, on a sample-by-sample basis. A sample of the evaluation sheet used in the study is available in the Appendix.

2.6. Data analysis

2.6.1. Univariate analyses

Analyses of variance were performed on the appropriateness ratings (considering descriptors as fixed and consumers as random factor) and on the hedonic ratings (samples fixed factor, consumers random factor). Post-hoc Tukey's test was performed to uncover the honestly significant differences between samples and descriptors. Differences were considered significant when $p \leqslant 0.05$.

ANOVA and post hoc testing were carried out in R, Version 2.11.1 (R Development Core Team, 2010).

2.6.2. Multivariate analyses

Subsequently, the collected data were organized into three data matrices: *X*, *Y* and *Z* (Fig. 1).

The *X* matrix contained the sensory data gathered from consumers, and was built as a frequency table by counting how many time the consumers had marked each of the descriptors for each of the beer samples. The first data block thus had six rows representing each beer, and the significant sensory descriptors as columns. Matrix *Y* contained consumer liking data, and had the six beers as rows, and 160 consumers as columns. Matrix *Z* contained the consumer background information collected by the computerized questionnaire, i.e. demographic, psychographic and appropriateness data. Demographic variables were rendered as category data (1/0), whereas the other variables (psychographic and appropriateness) were semi-continuous.

The resulting three blocks data structure was analyzed by L-shaped Partial Least Square Regression (Martens et al., 2005). As in classical PLS regression, the goal is to model latent variables (components) representing the common variance between two matrices (Martens & Martens, 2001). L-PLSR is a three block extension in which two matrices (*X* and *Z*) share no matrix size dimension, but are instead connected via *Y*, with which they each share one of the dimensions (the consumer background matrix, *Z*, is transposed into *Z*′, in order to have the same number of columns as *Y*). A few underlying components are extracted from *X* and *Z*′, and their interactions are used for bilinear modeling of *Y*, as well as for *X* and *Z* (Martens et al., 2005).

Prior to the L-PLSR analysis, three separate full cross validated PLSR analyses were performed to evaluate the validated variance and the significance of the variables via Martens' uncertainty test (Martens & Martens, 2000). This procedure is suggested by the method developers to overcome current limitations of the cross-validation method of L-PLSR.

The first model dealt with the sensory part only, i.e. the results from CATA questions.

ANOVA Partial Least Square Regression (A-PLSR) was carried out to uncover significant differences between the samples existed, based on consumers' evaluation of the product by CATA (*X* = indicator variables 1/0 for the six samples, *Y* = sensory variables by consumers), according to the procedure described by Martens and Martens (2001). The model was calculated using individual CATA scores (instead of summed CATA counts, as in the L-PLSR setting). Cross-validation was used, with individual consumers ($N = 160$) as cross-validation segments. The experiment was carried out with a convenience sample of products, not a designed set of products. Thus it was not sensible to use the samples for cross-validation, as properties of one sample cannot be predicted from others. Martens' uncertainty test (a jack-knife based elimination of noisy variables) was used as a method to assess the significance of inter-products differences (differences were considered significant when $p \leqslant 0.05$). The use of A-PLSR as an alternative to classical factorial ANOVA is discussed in Martens, Høy, Westad, Folkenberg, and Martens (2001).

The second preliminary PLSR model was obtained by interchanging the two matrices of the first model (*X* = sensory variables by consumers, *Y* = indicator variables 1/0 for the six samples), into a so-called Discriminant Partial Least Square Regression (D-PLSR) (Martens & Martens, 2001). This model was used to reveal the specific sensory descriptors responsible for the sensory differences, again via jack-knifed based estimated significance of the sensory attributes (Martens & Martens, 2000).

The third preliminary PLSR model was performed with consumer background information as predictor matrix and liking scores as response matrix (with regards to matrices names in Fig. 1, this was a PLS Regression of *Z* on *Y*′), to reveal what consumer background variables significantly contributed to explaining the variation in consumer liking. Again, the model was validated with full cross-validation, using consumers as cross-validation segments, and using Martens' uncertainty test as part of the analysis.

All PLSR analyses (incl. L-PLSR) were run with non-standardized variables in The Unscrambler X, Version 10.0.1 (CAMO ASA, Norway).

3. Results and discussion

3.1. Appropriateness ratings

Table 3 shows results for the mean appropriateness ratings given by the consumers to the list of 27 descriptors, sorted from most appropriate to least appropriate. Analysis of variance revealed significant differences in the perceived appropriateness of the descriptors ($F_{(26,4134)} = 32.1$, $p < 0.0001$). The descriptors "refreshing", "foamy", "sparkling" and "bitter" were rated as the four most appropriate for a beer, whereas "vinous" "still", "dried fruit" and "sour" were rated as the least appropriate. All the four most appropriate sensory descriptors are generally associated with pale lagers (Daems & Delvaux, 1997; Mejlholm & Martens, 2006), by far the most brewed and consumed beer type in Denmark (Statistics Denmark. Sales of alcohol, 2011). This could indicate that exposure and familiarity may breed perceived appropriateness in consumers' conceptualization of specific sensory properties.

3.2. Hedonic ratings

The mean overall liking scores (on a 7-points hedonic scale) of the six beers ranged from 3.9 (Wheat I.P.A.) to 4.9 (Brown Ale), as shown in Table 4. Analysis of variance performed on mean hedonic ratings showed that the differences were not large, but were significant ($F_{(5,795)} = 9.1$; $p < 0.0001$). At an overall level the beers appeared to be almost similarly liked. Subsequent multivariate analysis (see below) showed that differences existed and corresponded to specific consumer segments. Interestingly, the two beers that received the highest overall scores were two ales, a beer type with an established presence in the Danish market. This is consistent with the findings of Mejlholm and Martens (2006), even

Table 3
Mean appropriateness ratings and standard deviation for the sensory descriptors. Appropriateness ratings were collected on a 7 points scale. Different letters indicate significantly different descriptors (Tukey $p < 0.05$).

Sensory descriptor	Mean appropriateness (±S.D.)
Refreshing[a]	5.7 ± 1.3
Foamy[a,b]	5.4 ± 1.4
Sparkling[b,c]	45 ± 1.6
Bitter[b,c]	4.9 ± 1.4
Aromatic[c,d]	4.7 ± 1.3
Regional spices[c,d,e]	4.6 ± 1.5
Citrus fruit[c,d,e]	4.5 ± 1.6
Caramel[c,d,e,f]	4.3 ± 1.7
Nuts[c,d,e,f]	4.3 ± 1.4
Dessert spices[c,d,e,f]	4.3 ± 1.5
Fruity[c,d,e,f]	4.3 ± 1.4
Pungent[c,d,e,f]	4.3 ± 1.4
Beans[d,e,f,g]	4.2 ± 1.8
Flowers[d,e,f,g,h]	4.2 ± 1.7
Savoury spices[d,e,f,g,h,i]	4.1 ± 1.6
Sweet[d,e,f,g,h,i]	4.1 ± 1.3
Liquor[d,e,f,g,h,i,j]	4.0 ± 1.6
Herbs[e,f,g,h,i,j]	4.0 ± 1.5
Berries[e,f,g,h,i,j]	3.9 ± 1.7
Warming[f,g,h,i,j]	3.9 ± 1.6
Smoked[f,g,h,i,j]	3.8 ± 1.8
Intense berries[f,g,h,i,j]	3.7 ± 1.6
Fruit[g,h,i,j]	3.7 ± 1.6
Vinous[g,h,i,j]	3.7 ± 1.4
Still[h,i,j]	3.6 ± 1.4
Dried fruit[j,k]	3.4 ± 1.5
Sour[k]	2.9 ± 1.4

Table 4
Consumer liking for the beer samples: scores across all, male and female consumers (a 7-points hedonic scale was used). Beers are sorted by most liked to least liked. Different letters indicate significantly different beers (in overall values, Tukey $p < 0.05$).

Samples	Average Hedonic scores (±S.D.)		
	Overall ($n = 160$)	Male ($n = 92$)	Female ($n = 68$)
Brown Ale[a]	4.9 ± 1.7	4.8 ± 1.6	5.0 ± 1.9
Ravnsborg Red[a]	4.7 ± 1.7	4.7 ± 1.7	4.6 ± 1.7
New York Lager[b]	4.2 ± 1.6	4.0 ± 1.6	4.5 ± 1.6
River Beer[b]	4.1 ± 1.5	4.1 ± 1.5	4.0 ± 1.5
Porse Bock[b]	4.1 ± 1.6	4.2 ± 1.7	3.8 ± 1.5
Wheat IPA[b]	3.9 ± 2	4.1 ± 1.9	3.7 ± 2.1

though for this study the beers were not selected to vary systematically over styles and market presence.

3.3. Sensory differences characterized by Check-All-That-Apply

Consumers checked between 0 and 12 descriptors to describe each sample (average per sample: 4; standard deviation: 2). The most liked sample, Brown Ale, was the one for which consumers used the highest number of terms to describe (Table 5), and was the only sample to vary significantly from the other beers in this regard ($p < 0.001$).

Results from ANOVA-PLSR model (two components retained, X explained variance: 9%, Y: 5.8%) revealed that although the explained validated variance was very low, all products were perceived as significantly different from each other ($p \leqslant 0.05$). D-PLSR analysis (two components retained, X explained variance: 14%, Y: 7%) showed that 17 of 27 descriptors varied significantly ($p \leqslant 0.05$) between products: *flowers*, *fruity*, *dessert spices*, *sweet*, *savory spices*, *citrus fruit*, *refreshing*, *sparkling*, *sour*, *liquor*, *aromatic*, *caramel*, *beans*, *nuts*, *warming*, *dried fruits* and *pungent*. The 17 significant descriptors were kept for further analysis (L-PLSR).

These results also indicate the suitability of these two types of PLSR (A-PLSR and D-PLSR) for handling category data such as those CATA deliver, allowing evaluation of both significant product differences and significance of individual descriptors. As argued elsewhere (Frøst & Giacalone, 2011), the low explained variance is a result of a huge (960×27) and sparse data matrix (over 75% of 0s), which inevitably lead to a low level of explained variance in PLS models. However, the low variance did not per se constituted a problem, since the goals of these two models was to use their diagnostics, to ascertain that there are significant differences between the samples (A-PLSR), and which descriptors differ across samples (D-PLSR). This approach differ from the usual data analysis of CATA data, which is done on the frequency table (with the counts), and there the procedure is to consider the explained variance and interpret the results. Drawing a parallel with classical descriptive analysis, this approach corresponds to performing a PLS on sensory data from a trained panel unfolded by judges, replicates and samples vs. a PCA on the same data averaged over panelists and replicates. In the first model you can derive more information than the sensory variation, but clearly in the second case the explained variance will be much higher. This is implicitly shown, later in the paper, from the high explained variance of the X matrix in the L-PLSR.

3.4. Consumer background variables and relationship to liking

A third PLSR analysis was carried out using consumer background information as predictor matrix and liking scores as response. The cross-validation revealed that two components were optimal (Z explained variance: 33%, Y': 12% – matrices' names refer to the letter coding of Fig. 1).

The correlation loadings plot (Fig. 2) shows that the first component is mostly describing psychographic variables (e.g. interest in beer, beer knowledge, number of beer types consumed per month), whereas the second component seem to describe mainly the variation in perceived appropriateness. Only variables with a significant effect on liking were retained for further analysis.

3.5. Relationship between sensory characteristics, background consumer information and liking ratings: overall data structure by L-PLSR

The overall structure between the variables was provided by L-PLSR analysis. The correlation loadings plot (Fig. 3) summarizes the results visualizing the systematic covariation between the three data matrices, giving an overview of consumer liking explained by both consumer background and sensory characteristics.

Inspection of the plot shows that consumer liking was quite heterogeneous, as consumers (indicated as dots) are scattered in all directions. Restricting the analysis to sensory variables with a significant effect on liking differences showed more clearly that most of the consumers are in the directions of Ravnsborg Red and Brown Ale (the plot is not interpreted here as it would have caused a loss of important sensory variation between the products).

With regard to the perceived sensory differences, the first PLSR component mainly distinguished between dark ales (Brown Ale and Ravnsborg Red) and pale ales/lagers, whereas the second component further discriminated the latter into two clusters: a pilsner cluster (River Beer and Porse Bock) and a fruity cluster (Wheat I.P.A. and New York Lager). Sensory descriptors positively correlated to the ale clusters were, among others, "sweet", "caramel", "aromatic", "warming", "beans" (i.e. coffee and chocolate), "nuts" and "liquor". The two most liked sample, Ravnsborg Red and Brown Ale were also perceived as the sweetest samples. It is not unusual that in a group of products the sweeter ones receive the highest liking ratings (Frøst, 2006); this finding is interesting in

Table 5
Frequency table showing the occurrence of each sensory descriptor for each of the sample checked by the consumers during the CATA task (with the exclusion of 10 non significant variables, this table constituted the *X* matrix in the L-PLSR analysis).

	Flowers	Beans	Intense berries	Caramel	Nuts	Savory spices	Dessert spices	Reg. spices	Herbs
Wheat IPA	45	9	17	9	10	28	24	16	26
NY Lager	46	11	17	16	9	22	20	26	22
Ravnsborg Red	17	25	15	36	25	19	19	20	23
River Beer	22	13	15	4	14	14	9	20	13
Brown Ale	4	55	14	52	26	14	13	13	20
Porse Bock	27	2	21	7	12	6	16	26	24
Total	161	115	99	124	96	103	101	121	128
Range	42	53	7	48	17	22	15	13	13

	Citrus fruit	Berries	Fruit	Dried fruit	Liquor	Bitter	Sparkling	Refreshing	Fruity
Wheat IPA	30	9	21	6	2	57	24	55	43
NY Lager	30	5	15	6	10	69	17	38	45
Ravnsborg Red	12	6	12	14	9	71	24	51	27
River Beer	29	4	18	4	11	66	38	54	24
Brown Ale	10	6	12	21	26	64	20	25	22
Porse Bock	34	5	8	7	9	68	39	57	22
Total	145	35	86	58	67	395	162	280	183
Range	24	5	13	17	24	14	22	32	23

	Aromatic	Spicy	Still	Smoked	Foamy	Sour	Sweet	Warming	Vinous
Wheat IPA	29	22	14	25	26	23	25	6	8
NY Lager	38	20	18	24	28	17	22	9	8
Ravnsborg Red	52	17	24	28	16	25	20	12	5
River Beer	23	16	23	7	31	39	13	3	5
Brown Ale	58	18	21	38	27	19	44	26	12
Porse Bock	20	20	17	13	26	36	13	7	6
Total	220	113	117	135	154	159	137	63	44
Range	38	6	10	31	15	22	31	23	7

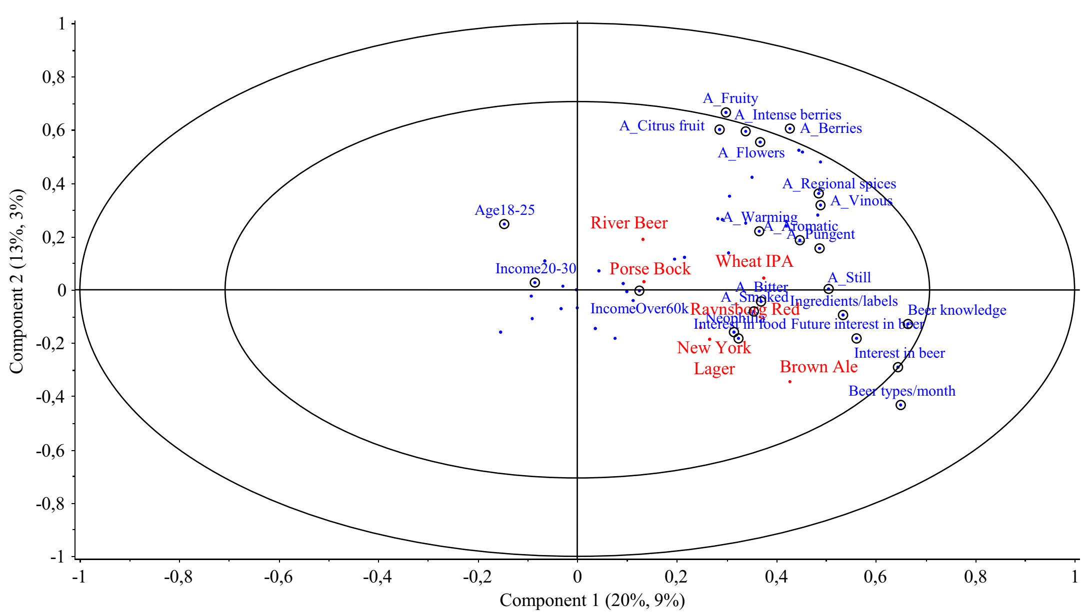

Fig. 2. Correlation loadings plot (Predictor matrix: *Z'*, Response matrix: *Y*), Component 1 vs. Component 2. Only significant variables are labeled. The inner and outer ellipses represent 50% and 100% of explained variance respectively.

the light of the well established trend, especially in Europe, to brew increasingly sweet beers (Jentsch, 2007).

River Beer and Porse Bock, in the lower left quadrant, are defined by the descriptors "sparkling", "sour" and "refreshing". The last pale ale/lager cluster contains New York Lager and Wheat IPA, correlated well with the descriptors "fruity" and "flowers". These descriptors match the beers style as well as the commercial description provided by the breweries.

Consumer psychographics explained liking better than demographics. Combined, the first two components segmented consumers by beer knowledge and interest. The big cluster of variables – beer knowledge, beer current interest, beer future interest, number of beer types consumed a month, attention to ingredients and labels – all positively loaded on the first component, indicate that knowledgeable consumers had a preference for ales, whereas less experienced consumers preferred lagers.

Demographic variables generally failed to predict consumer variation in liking, with the partial exception of age. The youngest consumer segment (18–25) was found to be significantly different from the mean, with their preference in the direction of sensory

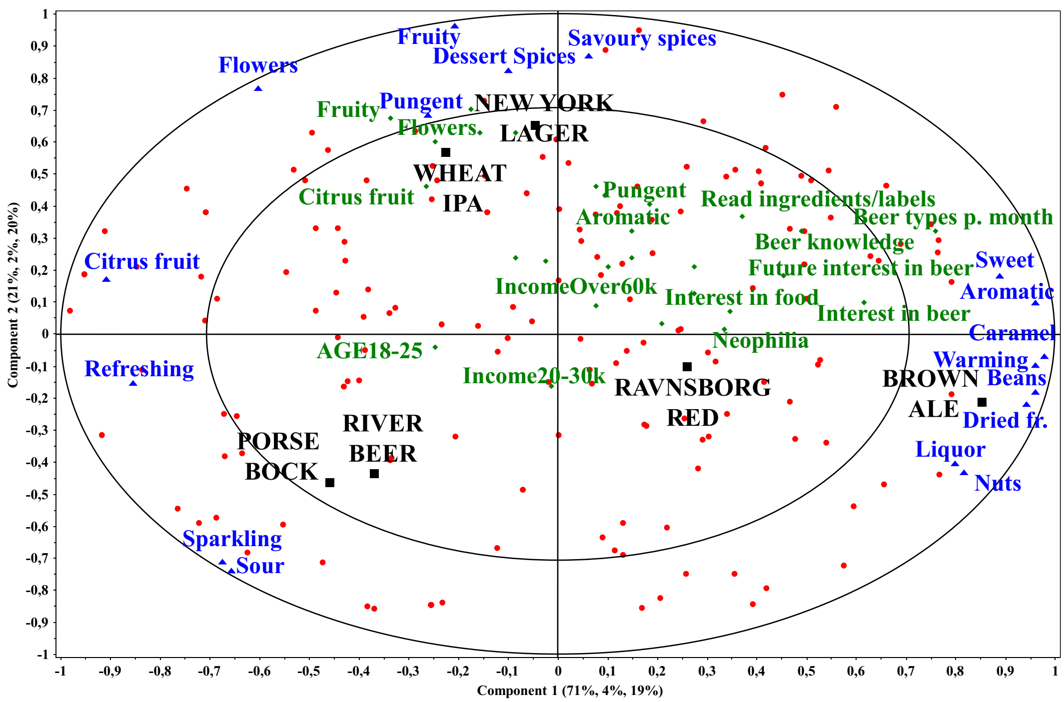

Fig. 3. L-PLSR correlation loadings plot showing consumer preferences (●) for the six beers (■) in relation to both sensory characteristics (▲) and consumer background information (♦). Only the first two components are shown here (*X* explained variance: 92%, *Y*: 6%, *Z*: 39%). The inner and outer ellipses represent 50% and 100% of explained variance respectively. Labels for *Z* variables of minor interest have been omitted.

descriptors "citrus fruit" and "refreshing". Further, they were all negatively correlated to all psychographic variables, indicating that youngest consumers are less knowledgeable when it comes to beer. This result seem to fit with the industrial trend, particularly by large breweries, to market new soft-drink like beers with these sensory characteristics, specifically targeted for young consumers as alternative to sales leaders in the ready to drink (RTD) or flavored alcoholic beverages category .

Unlike previous studies, gender was not found to have a significant effect on liking (cfr. Fig. 2) and was therefore not included in the L-PLSR analysis. Accordingly, mean liking ratings (Table 3) showed quite a good agreement between men and women with regards to their preferences. Interestingly, however, Table 3 also shows that the preference of women consumers spanned a larger range than their male counterparts (3.7–5 for women vs. 4–4.8 for men), suggesting that, while men like beer in general, women exhibit more variability. Subsequent multiple t-tests revealed that the two gender groups differed significantly ($p = 0.02$) only in relation to their liking of one specific sample, New York Lager, which was most characterized by the attributes "fruity" and "flowers" and was significantly better liked by the female consumers.

A good correlation between the conceptual appropriateness ratings and the preference for sensory characteristics was found for certain descriptors, such as "fruity", "flowers" and "citrus fruit". Others, e.g. "aromatic" and "pungent", showed large discrepancy between appropriateness and actual hedonic response. Due to much residual variance, results do not clearly indicate whether a perceived appropriateness for a sensory attribute correlated well with liking. It is suggested that the usability of this type of appropriateness ratings could improve by restricting the product set to a specific product type (e.g. lagers only), and possibly with frequent users of the product, that have formed clear sensory expectations for the product category.

4. General discussion

In the previous sections we have presented an exploratory approach to product testing with consumers, the AI1 test. The AI1 test is not a single method in itself, but rather a series of complementary operating procedures to collect, analyze and represent data, that can be used to collect consumer response early in the product development life-cycle rapidly and inexpensively.

As suggested, this approach is advantageous in terms of external validity for three main reasons. First, because it allows consumer segmentation on the basis of demographics and psychographic variables, so that the results can be linked to specific consumer segments. Secondly, because it derives a sensory profile of the products directly on consumer perception. Thirdly, because it is particularly suitable for working with consumers in naturalistic settings, a factor that has been identified as important for producing ecologically valid results (Meiselman, 2008) and that ensures that subjects are from a comparable pool as those about whom we would like to generalize.

The suggested analytical path, based on ad hoc PLSR models, provides diagnostic tools to ensure the statistical significance of the results obtained (incl. within samples differences and removal of noisy variables). L-PLSR provides a graphically efficient tool for exploring relations between preferences, perceived sensory attributes and consumers' characteristics. Moreover, as an exploration technique L-PLSR is advantageous because regressors can be included in a model to explain a response that is characterized by both the sensory characteristics and the characteristics of the consumers, leading to greater insights and giving "double" validity to the extracted latent components (Martens & Tschudi, 2010; Plaehn & Lundahl, 2006).

The inclusion of appropriateness measurements in the AI1 test can provide researchers with consumer expectations concerning the characteristics of the product, and how these relate to the actual sensory experience. This is quite important for product developers because the difference between expected and experienced product performance is a key determinant of customer satisfaction and re-purchase (see e.g. Churchill & Surprenant, 1982).

It is important to carefully consider the gains and losses of using consumers as unique data source compared to a more conventional approach, e.g. performing descriptive analysis with a trained panel and collecting only liking ratings and background information from consumers in separate sessions. The advantage of the AI1 test are that it provides an overview of the main product issues and differences with relevance for the consumer in a fast and inexpensive way. It may have a higher degree of external validity than traditional preference mapping approaches, as all data, both descriptive and hedonic originate from consumers, but this requires further testing.

Because of its relative inexpensiveness and easy implementation to real-life situations (food festivals, trade fair, other local events), it could thus applicable by SMEs in the food sector (which may have limited or no budget for R&D) in need to gather a wide range of consumer impressions about their products. For larger companies, the AI1 approach can be useful *inter alia* as a fast screening tool to incorporate the "voice of the consumer" in early stages of product development (e.g. getting feedbacks of initial formulations of a new products), supporting knowledge-based design of further, more focused, studies.

These characteristics makes it particularly suitable for exploratory studies, e.g. to identify windows of opportunity (e.g. preference segmentation) by involving consumers in early stages of new product development, thus increasing the market orientation of sensory research (Grunert et al., 2008; Stewart-Knox & Mitchell, 2003; van Kleef, van Trijp, & Luning, 2005). Further, the AI1 test appears to be best suitable for product sets where the sensory differences are rather large (such as the case presented in this paper). Conversely, it seems to be less suitable for product sets with small sensory differences. These authors think this aspect is secondary to the primary goal of AI1 test, which is not that of providing a very accurate profile of products, but rather to show what attributes consumers recognize in the product, and to directly connect the attributes to preferences.

A classical feature of sensory profile given by consumers is that they are simple and provide large features (Husson et al., 2001), and that consumers data may be not immediately actionable since their response are typically limited to non-technical terms (we elaborate on this issue in the next section). To enhance their actionability, consumer data can be interpreted based on *a priori* knowledge of the samples by the experimenter and/or the producer. Moreover, consumer data can be related to e.g. physico-chemical characteristics of the product, and/or data from a trained panel, to avoid misleading indications. A number of statistical regression techniques can be employed to link data from different domains. PLSR itself is classically employed to establish to establish weighted linear combinations between variables of different nature, e.g. predict a product's sensory profile from its chemical composition (Martens & Martens, 2001; Martens, Tenenhaus, Esposito Vinzi, & Martens, 2007; Martens & Tschudi, 2010). Other available techniques include Moskowitz's reverse engineering approach (Moskowitz, 1994), metric multidimensional scaling (Cardello et al., 1982), and multiple factor analysis (Escofier & Pagès, 1994). Additionally, if instrumental or chemical measurements are available on the same samples, they could be used as X matrix in the L-PLSR, as in earlier examples from the literature (e.g. Lengard & Kermit, 2006; Mejlholm & Martens, 2006; Thybo, Kühn, & Martens, 2003).

At last, the AI1 test is a versatile approach that can be customized according to specific needs. For instance, one might want to restrict the experimental design to a specific product category (e.g. only lagers) or specific consumer segments (e.g. only product users). The L-PLSR analysis can be also restricted to specific variables to highlight different aspects (e.g. effect of demographic variables, effect of psychographics, relationship between appropriateness and sensory variables, product profiling etc.). Furthermore, consumer hedonic variables other than liking that are known to be important for food product experience, such as perceived complexity, novelty, etc. (see e.g. Lévy, MacRae, & Köster, 2006), could be included in the study and used as Y matrix.

4.1. Further developments and research perspectives

The AI1 test presents some potential issues that need further exploration.

The first one is the data produced will generally have a lot of noise and variability, as was the case in the presented study. This is a well known problem when using of consumers for descriptive analysis, exacerbated in the AI1 test due to it being performed under very limited control. However, the incidence of this problem appears quite limited for two reasons: (a) first, because the suggested multivariate methods provide a powerful diagnostic tool for assessing the significance of the results obtained; and then (b) because it makes it easy to involve a large amount of consumers, rapidly reaching a number –100 or more consumers – beyond which results are not particularly influenced by the base size (Moskowitz, 1997; Worch et al., 2010).

Secondly, careful consideration should be given to attribute selection when preparing CATA questionnaires. In the presented study, consumers selected on average only four out of 27 attributes. Consumers are known to minimize their efforts when answering questionnaire, and many will tend to simply aim at providing a satisfactory answer (Krosnick, 1991). In this context, a list with 27 descriptors must have proven a substantial cognitive effort. Further, it is noteworthy that the many of the most frequently chosen descriptors (e.g. "bitter", "fruity", "refreshing") were placed somewhat at the top of the list in the evaluation sheet (cfr. Appendix). This suggests that the attribute order might have had an influence in the choice of attribute, and that in future applications it would be wise to randomize attribute orders when designing CATA questionnaire. However, it is also apparent that, regardless of their position, simpler and less equivocal terms were consistently more used by consumers. This was the case of e.g. "sour", "sweet" and "citrus fruit" that were very frequently used although being among the last in the attribute list. Conversely, more complex categorical attributes such as "intense berries" or "nuts" were less frequently mentioned, despite being on top of the list. This concurs with previous claims that simpler terms (i.e. understandable without the use of references) are usually dominant when untrained subjects are used in descriptive tasks, as it was observed both for beer evaluation (Chollet & Valentin, 2007; Clapperton & Piggott, 1979) as well for other food products (for a discussion, see van Trijp & Schifferstein, 1995). Consumers seemed to prefer integrated terms (*viz.*, which combines several product attributes into one term), such as "refreshing" and "fruity", over more specific terms (e.g. "vinous", "dried fruit"), which may need to be related to e.g. a descriptive profile to be translated and broken down into more actionable attributes.

Thirdly, CATA may present some limitations compared to rating attributes. The analysis conducted showed that consumers were able to discriminate the products based on 17 out of the 27 attributes, whose significance was estimated by jack-knife extension of cross-validation in our D-PLSR analysis. Had we used a classical intensity rating approach, we should conclude that the remaining 10 variables only represented unstructured variation (i.e. noise), due to e.g. subjects' inconsistency in the use of the scale, poor understanding of the attribute, etc. However, the use of nominal data (as in CATA) requires a more prudent interpretation. Take for example the attribute "bitter": it failed to discriminate between the products, even though it was the most frequently mentioned attribute (cfr. Table 5) and it is a basic taste whose meaning can be taken to be fairly understood even by naïve consumers. Is it a noisy variable? Probably not. Hop bitter substances are primary flavor constituents of almost all beer (Meilgaard, 1982). It is not surprising that bitter was the attribute which was so often mentioned for all of the beers. However, from a product perspective, the six beers employed in this study differed considerably in their degree of bitterness, measured by alpha acids content. It is possible that this difference would have been detected with a scaling approach: thus the lack of systematic variation in perceived bitterness, thus, could indicate more a limitation of the method than an actual lack of difference between beers.

The last methodological issue concerns the use of AI1 test as a tool for investigating consumer preference. There is a potential halo effect which has sometimes been observed when consumers perform analytical tasks together with hedonic ratings (Earthy, MacFie, & Hedderley, 1997). Overall, this issue is underinvestigated and the existing results are complex and inconsistent. Popper, Rosenstock, Schraidt, and Kroll (2004) found that different types of attribute questions affected liking response to different degrees (and in some cases not all). Other studies reported no such effects, suggesting that concurrent (sensory/hedonic) rating procedures may be a valid method of collection of this type of sensory data (Mela, 1989; Vickers, Christensen, Fahrenholtz, & Gengler, 1993). However, recent findings again suggest that the presence of analytical attributes might bias liking ratings, specifically in the form of lower ratings (Prescott, Lee, & Kim, 2011). It is noteworthy that all previous studies concerned the use of scale attributes (e.g. JAR), and not specifically Check-All-That-Apply methodology. This is an important difference because CATA questions require significantly lower cognitive effort from the respondent compare to scale ratings (Rasinski, Mingay, & Bradburn, 1994), being more suitable to encompass consumers' natural tendency to minimize efforts in their response strategy (Krosnick, 1991). As stated earlier, CATA proponents reported that this method had a smaller effect on liking ratings than JAR or intensity scales (Adams et al., 2007; Ares, Barreiro et al., 2010), a claim that still lacks formal investigation. Further, the questions' order (first preference, then sensory or vice versa) is also known to be important, with liking ratings being lower when asked after analytical questions, which has been interpreted as implying that consumers become more critical only *after* they have been asked about specific sensory properties (Earthy et al., 1997). This suggests that asking hedonic ratings first – as we have done in our AI1 test – could limit the effects on liking (Popper et al., 2004). Nevertheless, an important continuation of the present research would be digging deeper into such systematic bias, and also into its specular risk that liking rating could distort the descriptive perception to the point that it could produce biased sensory ratings (see e.g. Moskowitz, 1999).

5. Conclusions

An integrated approach for gathering consumer product data has been proposed and explored. This approach, that we coin the AI1 test, can be considered as a fast and easily applicable tool for sensory and consumer scientists to support explorative stages in consumer-oriented food product development. In only one experimental setting with consumers, it provides insights in three domains (1) demographic and psychographic composition of the subject group; (2) consumer-based sensory profiles by CATA and (3) hedonic perception for the tested products, and demonstrates relationships between the different domains with the use of L-PLSR.

In the presented case, the AI1 test provided a description of consumer perception (sensory and hedonic) of the beers, and revealed the relationship with consumers' background. The CATA technique successfully discriminated the eight beers and identified the underlying sensory dimensions. By "successfully" we mean that statistically significant differences were obtained among the test samples, a highly relevant criterion for the success of a descriptive method. However, we recognize that it will always be dependent on the set of samples in the product set. With regards to preference mapping, consumer psychographics – particularly previous knowledge and interest in beer – explained liking better than demographics, confirming that this type of measurements in consumer studies provide an important supplement to demographics (Solomon, Bamossy, Askegaard, & Hogg, 2010). The stated appropriateness ratings for sensory properties in beer provided additional information concerning how consumers conceptualize sensory expectations to beer. However, appropriateness correlated well with hedonic for some sensory properties only, whereas others showed large discrepancy between appropriateness and actual hedonic response. This demonstrates that consumers are not aware of all sensory aspects in their actual hedonic perception pattern.

At a general level, the AI1 test formalizes a trend to expand the dominion of consumer data which has characterized our field in the last years. Of course, researchers have to be very careful to preserve the specificity of sensory science from pure marketing research, as well as the reliability of data collected from consumers. However, recent advancement in consumer-friendly sensory methods – including CATA – and the statistical methods for the analyses of relationships between different data domains should trigger a higher degree of consumer involvement in the descriptive parts of product tests.

Although more systematic studies are necessary, we see approaches such as the AI1 test, as an opportunity for growth of the sensory science field to better support product developers. The proposed combination of existing methods is particularly suitable for quantitative applications where external validity is a key issue. In the presented case, all data has external validity, as they are collected under informal naturalistic settings, using only consumer responses i.e. actual consumers' perception in a context that approaches a natural consumption situation.

Acknowledgments

This work was supported by the Danish Agency for Science, Technology and Innovation (through the consortium *Danish Microbrew – Product innovation and quality*) and by the Danish Ministry of Economic and Business Affairs (through the project *Local Foods in Denmark*). Additional support was provided by the Faculty of Science, University of Copenhagen. The authors wish to thank Bodil E. Pallesen, Mathias Andersen, Helene C. Reinbach, Letícia Machado Ribeiro-Vermeulen and Annemarie Olsen for help with conducting the experiment.

Davide Giacalone was supported with the Rick Bell Memorial Travel Scholarship for presenting this study at the 9th Pangborn Sensory Science Symposium.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodqual.2012.09.011.

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*, 627–640.

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In *7th Pangborn sensory science symposium*, 12–16 August 2007, Minneapolis, MN, USA.

Albert, A., Varela, P., Salvador, A., Hough, G., & Fiszman, S. (2011). Overcoming the issues in the sensory description of hot served food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. *Food Quality and Preference, 22*, 463–473.

Ares, G., Barreiro, C., Deliza, R., Gimenez, A., & Gámbaro, A. (2010). Application of a check-all-that-apply question to the development of chocolate milk desserts. *Journal of Sensory Studies, 25*, 67–86.

Ares, G., Deliza, R., Barreiro, C., Gimenez, A., & Gámbaro, A. (2010). Comparison of two sensory profiling techniques based on consumer perception. *Food Quality and Preference, 21*, 417–426.

Ares, G., Varela, P., Rado, G., & Gimenez, A. (2011). Identifying ideal products using three different consumer profiling methodologies. Comparison with external preference mapping. *Food Quality and Preference, 22*, 581–591.

Buck, D. (2003). Solicited commentary to Garber et al.: Measuring consumer response to food products. *Food Quality and Preference, 14*, 37–38.

Cardello, A. V. (2003). Ideographic sensory testing vs. nomothetic sensory research for marketing guidance: Comments on Garber et al. *Food Quality and Preference, 14*, 27–30.

Cardello, A. V., Maller, O., Kapsalis, J. G., Segars, R. A., Sawyer, F. M., Murphy, C., et al. (1982). Perception of texture by trained and consumer panelists. *Journal of Food Science, 47*, 1186–1197.

Chollet, S., & Valentin, D. (2007). Impact of training on beer flavor perception and description: Are trained and untrained subjects really different? *Journal of Sensory Studies, 16*, 601–618.

Churchill, G. A., & Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of Marketing Research, 19*, 491–504.

Civille, G. V., & Heylmun, J. (2003). Commentary on Garber et al. paper for Food Quality and Preference journal. *Food Quality and Preference, 14*, 31–32.

Clapperton, J. F., & Piggott, J. R. (1979). Flavour characterization by trained and untrained assessors. *Journal of the Institute of Brewing, 85*, 275–277.

Daems, V., & Delvaux, F. (1997). Multivariate analysis of descriptive sensory data on 40 commercial beers. *Food Quality and Preference, 8*, 373–380.

Dairou, J., & Sieffermann, J. (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.

Dijksterhuis, G. B. (2006). Diversity in methods and methods' merits. In *Invited keynote at "A sense of diversity", second European conference on sensory and consumer science of food and beverages*, 26–29 September 2006, The Hague, The Netherlands.

Dooley, L., Lee, Y., & Meullenet, J. (2010). The application of check-all-that-apply (CATA) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. *Food Quality and Preference, 21*, 394–401.

Earthy, P. J., MacFie, H. J. H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central location trials. *Journal of Sensory Studies, 12*, 215–237.

Escofier, B., & Pagès, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis, 18*, 121–140.

Frøst, M. B. (2006). Liking and exposure: First, second and tenth time around. *Physiology & Behavior, 89*, 47–52.

Frøst, M. B., & Giacalone, D. (2011). Using PLS-regression for verification of product differences and important variables in a consumer sensory profiles on beer obtained by a check-all-that-apply (CATA) technique. In *9th Pangborn sensory science symposium*, 4–8 September 2011, Toronto, Canada.

Garber, L. L., Hyatt, E. M., & Starr, R. G. (2003). Measuring consumer response to food products. *Food Quality and Preference, 14*, 3–15.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton-CenturyCrofts.

Grunert, K. G., Jensen, B. B., Sonne, A., Brunsø, K., Byrne, D. V., Clausen, C., et al. (2008). User-oriented innovation in the food sector: Relevant streams of research and an agenda for future work. *Trends in Food Science & Technology, 19*, 590–602.

Hough, G. (1998). Experts versus consumers: A critique. *Journal of Sensory Studies, 12*, 147–154.

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.

Jentsch, M. (2007). Top-fermented beer specialties in focus. *Brauwelt International, 25*, 332–334.

Köster, E. P. (1981). Sensory evaluation in a "natural environment". In P. Schreier (Ed.), *Flavour'81* (pp. 93–100). Berlin, Germany: Walter de Gruyter.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213–236.

Laugesen, J. L. (2007). Katim v. 1.0. Matlab-based software for 2k-factorial block design for images, Sensory Science Group, Department of Food Science, University of Copenhagen, Denmark.

Lawless, H. T., & Heymann, H. (2010). Sensory evaluation of food. In *Principles and practices*. New York, NY: Springer.

Lengard, V., & Kermit, M. (2006). 3-Way and 3-block PLS regression in consumer preference analysis. *Food Quality and Preference, 17*, 234–242.

Lévy, C. M., MacRae, A., & Köster, E. P. (2006). Perceived stimulus complexity and food preference development. *Acta Psychologica, 123*, 394–413.

MacFie, H. J. H., Bratchell, N., Greenhoff, N. K., & Vallis, I. 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., Anderssen, E., Flatberg, A., Gidskehaug, L. H., Høy, M., Westad, F., et al. (2005). Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR. *Computational Statistics & Data Analysis, 48*, 103–123.

Martens, H., Høy, M., Westad, F., Folkenberg, D., & Martens, M. (2001). Analysis of designed experiments by stabilised PLS Regression and jack-knifing. *Chemometrics and Intelligent Laboratory Systems, 58*, 151–170.

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

Martens, H., & Martens, M. (2001). *Multivariate analysis of quality*. Chichester, England: John Wiley & Sons.

Martens, M., Tenenhaus, M., Esposito Vinzi, V., & Martens, H. (2007). The use of partial least squares methods in new food product development. In H. J. MacFie (Ed.), *Consumer-led food product development*. Cambridge, England: Woodhead Publishing.

Martens, M., & Tschudi, F. (2010). Multivariate psychophysics, multivariate data: Human senses and their measurement. *Biological Theory, 5*, 337–343.

McBride, R. L. (1990). Three generations of sensory evaluation. In R. L. McBride & H. J. MacFie (Eds.), *Psychological Basis of Sensory Evaluation*. London, England: Elsevier Applied Science.

Meilgaard, M. C. (1982). Prediction of flavor differences between beers from their chemical composition. *Journal of Agricultural and Food Chemistry, 30*, 1009–1017.

Meiselman, H. L. (1993). Critical evaluation of sensory techniques. *Food Quality and Preference, 4*, 33–40.

Meiselman, H. L. (1994). Bridging the gap between sensory evaluation and market research. *Trends in Food Science and Technology, 5*, 396–398.

Meiselman, H. L. (2008). Experiencing food products within a physical and social context. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product experience*. San Diego, CA: Elsevier.

Mejlholm, O., & Martens, M. (2006). Beer identity in Denmark. *Food Quality and Preference, 17*, 108–115.

Mela, D. J. (1989). A comparison of single and concurrent evaluations of sensory and hedonic attributes. *Journal of Food Science, 54*, 1098–1100.

Moskowitz, H. R. (1994). *Food concepts and products: Just in time development*. Trumbull, CT: Food and Nutrition Press.

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

Moskowitz, H. R. (1997). Base size in product testing: A psychophysical viewpoint and analysis. *Food Quality and Preference, 8*, 247–255.

Moskowitz, H. R. (1998). Consumers versus experts in the light of psychophysics: A reply to Hough. *Journal of Sensory Studies, 13*, 291–298.

Moskowitz, H. R. (1999). Improving the "actionability" of product tests: Understanding and using relations among liking, sensory, and directional attributes. *Canadian Journal of Marketing Research, 11*, 19–37.

Moskowitz, H. R., & Chandler, J. (1979). Notes on consumer oriented sensory evaluation. *Journal of Food Quality, 2*, 269–276.

Moussaoui, K. A., & Varela, P. (2010). Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference, 21*, 1088–1099.

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

Murray, J. M., Delahunty, C. M., & Baxter, I. A. (2001). Descriptive analysis: Past, present and future. *Food Research International, 34*, 461–471.

Parente, M. E., Ares, G., & Manzoni, A. V. (2010). Application of two consumers profiling techniques to cosmetic emulsions. *Journal of Sensory Studies, 25*, 685–705.

Parente, M. E., Manzoni, A. V., & Ares, G. (2011). External preference mapping of commercial antiaging creams based on consumers' responses to a check-all-that-apply question. *Journal of Sensory Studies, 26*, 158–166.

Piqueras-Fiszman, B., Ares, G., Alcaide-Marzal, J., & Diego-Más, J. A. (2011). Comparing older and younger users' perception of mobile phones and watches using CATA questions and preference mapping on the design characteristics. *Journal of Sensory Studies, 26*, 1–12.

Plaehn, D., & Lundahl, D. S. (2006). An L-PLS preference cluster analysis on French consumers hedonics to fresh tomatoes. *Food Quality and Preference, 17*, 243–256.

Popper, R., Rosenstock, W., Schraidt, M., & Kroll, B. J. (2004). The effect of attribute questions on overall liking ratings. *Food Quality and Preference, 15*, 853–858.

Prescott, J., Lee, S. M., & Kim, K. (2011). Analytic approaches to evaluation modify hedonic responses. *Food Quality and Preference, 22*, 391–393.

Puyares, V., Ares, G., & Carrau, F. (2010). Searching a specific bottle for Tannat wine using a check-all-that apply question and conjoint analysis. *Food Quality and Preference, 21*, 684–691.

R Development Core Team (2010). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rasinski, K. A., Mingay, D., & Bradburn, N. M. (1994). Do respondents really "mark all that apply" on self-administered questions? *Public Opinion Quarterly, 58*, 400–408.

Reinbach, H. C., Giacalone, D., Machado, L. R., Bredie, W. L. P., & Frøst, M. B. (2012). Comparison of three sensory profiling methods to study consumer perception: CATA, CATA with intensity ratings and Napping®. In *A Sense of Inspiration – 5th European conference on sensory and consumer research*, 9–12 September 2012, Bern, Switzerland.

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.

Schutz, H. G. (1999). Consumer data-sense and nonsense. *Food Quality and Preference, 10*, 245–251.

Schutz, H. G. (1988). Beyond preference: Appropriateness as a measure of contextual acceptance of food. In D. M. H. Thomson (Ed.), *Food acceptability*. New York: Elsevier.

Solomon, M. R., Bamossy, G., Askegaard, S., & Hogg, M. K. (2010). Consumer behaviour. In *A European perspective*. Harlow, England: Prentice Hall.

Statistics Denmark. *Sales of alcohol and tobacco subject to excises duties by type*. Available from http://statistikbanken.dk/ALKO4. (Retrieved 20 September 2011).

Stewart-Knox, B., & Mitchell, P. (2003). What separates the winners from the losers in new food product development? *Trends in Food Science & Technology, 14*, 58–64.

Stone, H., & Sidel, J. L. (1993). *Sensory evaluation practices* (2nd ed.). San Diego, CA: Academic Press.

Thybo, A. K., Kühn, B. F., & Martens, H. (2003). Explaining children's preferences for apples using instrumental, sensory and demographic/behavioural data. *Food Quality and Preference, 15*, 53–63.

van Kleef, E., van Trijp, H. C. M., & Luning, P. (2005). Consumer research in the early stages of new product development: A critical review of methods and techniques. *Food Quality and Preference, 16*, 181–201.

van Trijp, H. C. M., & Schifferstein, H. N. J. (1995). Sensory analysis in marketing practice: Comparison and integration. *Journal of Sensory Studies, 10*, 127–147.

Vickers, Z. M., Christensen, C. M., Fahrenholtz, S. K., & Gengler, I. M. (1993). Effect of questionnaire design and the number of samples tasted on hedonic ratings. *Journal of Sensory Studies, 8*, 189–200.

Wansink, B. (2003). Response to "Measuring consumer responses to food products". Sensory tests that predict consumer acceptance. *Food Quality and Preference, 14*, 23–26.

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.

Worch, T., Lê, S., & Punter, P. (2010). How reliable are the consumers? Comparison of sensory profiles from consumers and experts. *Food Quality and Preference, 21*, 309–318.