

# Sensory drivers of perceived situational appropriateness in unbranded foods and beverages: Towards a deeper understanding

Davide Giacalone<sup>a,\*</sup>, Sara R. Jaeger<sup>b</sup>

<sup>a</sup> University of Southern Denmark, Odense, Denmark

<sup>b</sup> The New Zealand Institute for Plant & Food Research Ltd., Auckland, New Zealand

## ARTICLE INFO

### Keywords:

Context  
Use situations  
Appropriateness  
'Item-by-use'  
Sensory differences

## ABSTRACT

Measures of product performance that effectively predict food and beverage choices are highly sought after. A simple method to add value to hedonic data is that of item-by-use (IBU) appropriateness, where consumers are presented with a list of possible consumption situations and asked to indicate how well a product fits them. A persistent misconception surrounding this approach is that it is relevant for discriminating between different products, but not between variants within the same product category, which is often the focus of sensory and consumer studies. To provide a deeper understanding of the sensory underpinnings of appropriateness, the present work presents experimental evidence from six consumer studies (116–210 consumers per study) employing unbranded product variants from the same category. Products were successfully discriminated based on IBU appropriateness in all studies, even when sensory variation was unidimensional and controlled (such as a basic lemonade recipe varying in sugar content). While product differentiation based on the sensory profiles was greater than differentiation based on appropriateness, the results clearly show that sensory variation, in and of itself, is sufficient to elicit differences in perceived appropriateness. As expected, the degree of inter-product differences in appropriateness was approximately linearly related to the degree of differences in sensory profiles. Finally, while some sensory properties independently affected perceived appropriateness, the magnitude (and in some cases the direction) of the effects often depended on the level of product liking.

## 1. Introduction

### 1.1. Motivation for the research

Hedonic responses to foods and beverages have traditionally been regarded as the key product performance indicator in central location tests (CLT) (Cardello, 2020). Yet, empirical evidence indicates that strong sensory/hedonic performance may not correlate with marketplace success. This is due to the fact that consumers' decisions to buy or consume particular products do not depend solely on the intrinsic quality of the product, but also on the anticipated usage situations (Marshall, 1993), among other factors. Accordingly, situational appropriateness, here defined as *perceived fit between products and use situations*, is considered as an important cognitive-contextual measure of consumer attitudes towards foods and beverages (Schutz, 1994; Giacalone, 2019; Jaeger & Porcherot, 2017).

The 'item-by-use' (IBU) approach in which consumers evaluate products across a range of relevant usage situations (Schutz, 1988,

1994) is popular for measurement of situational appropriateness, and previous research has shown that appropriateness (IBU) ratings are associated with expected liking (Cardello et al., 2000; Jaeger et al., 2019), willingness to pay (Elzerman et al., 2015), frequency of consumption (Sosa, Martínez, Arruiz, Hough, & Mucci, 2005) and emotional responses (Piqueras-Fiszman & Jaeger, 2019). When co-elicited with hedonic ratings for tasted products, samples that are similarly liked often differ in appropriateness (Cardello & Schutz, 1996; Lähteenmäki & Tuorila, 1995, 1997; Jaeger, Lee, et al., 2019, 2019b) indicating that appropriateness enables product differentiation beyond liking. It has also been possible to predict consumer choices from IBU ratings (Giacalone & Jaeger, 2019a, 2019b).

The use of situational appropriateness evaluations in central location tests (CLTs) has appeal, but uptake has been slow, possibly due to lack of methodological research (Jaeger et al., 2019). Early reports of a positive correlation between appropriateness and liking may have led some researchers to question the added value of including appropriateness evaluations beyond collecting hedonic ratings from consumers.

\* Corresponding author. Dept. of Technology and Innovation, Campusvej 55, DK-5230, Odense M, Denmark.

E-mail address: [dg@iti.sdu.dk](mailto:dg@iti.sdu.dk) (D. Giacalone).

<https://doi.org/10.1016/j.appet.2021.105589>

Received 17 February 2021; Received in revised form 30 May 2021; Accepted 2 July 2021

Available online 9 July 2021

0195-6663/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

However, the relationship between appropriateness and liking is believed to be more complex than initially thought. By and large, this is due to non-linearity where low appropriateness more likely couples with disliking, as opposed to the decoupling that has been observed for liked products. This renders appropriateness measures particularly relevant for discriminating between products that have high hedonic value and are similarly liked (Jaeger et al., 2019; Giacalone & Jaeger, 2019b).

As argued by Giacalone and Jaeger (2019a), the magnitude of difference in perceived appropriateness is related to how similar/different products in the sample set are to each other, and the difference will generally be higher when extrinsic factors (e.g., brand name and packaging) are included and/or products from different categories are compared. Notably, some studies have reported inter-product differences in appropriateness within a range of difference that is typical of CLTs. For example, Lähteenmäki and Tuorila (1995, 1997) showed, respectively, that appropriateness ratings differed on the basis of sensory differences in commercial brands of vanilla ice-creams, and between blueberry-raspberry juices varying in sweetener content. Cardello and Schutz (1996) successfully discriminated between chocolate bars differing in storage time and conditions, and other studies with similar findings for both foods and beverages include Geertsens et al. (2016), Hersleth et al. (2011), Mejhlholm and Martens (2006), and Stolzenbach et al. (2016). This suggests that appropriateness is useful for differentiating product variants within the same category, even when evaluations are obtained in blind (i.e., without any extrinsic information).

### 1.2. Research questions

What this body of research suggests is, in turn, that if relationships between sensory attributes and situational appropriateness can be established, the sensory profile of a product can likely be modified to achieve an increase in appropriateness for a target situation(s) envisioned during product development. To achieve this, a deeper understanding of the sensory underpinnings of appropriateness is needed, in particular drawing on experimental results on unbranded products within the same product category. The present research is situated in this knowledge gap.

In particular, we sought to confirm that sensory attributes in and of themselves influence appropriateness, and, further, to understand to what degree of sensory differences between products is needed to elicit a difference in terms of appropriateness. The first research question (RQ1) was: Can sensory variation within the same product category determine differences in situational appropriateness? And if so, which degree of sensory differences is necessary to elicit differences in appropriateness?

Another important question is to determine whether relationships between sensory attributes and appropriateness are independent from liking; in other words, whether the same relationships hold across individuals who differ in their preference for a specific sensory attribute. The literature points to relative stability of appropriateness data compared to hedonic evaluations, which is thought to be due to the cultural influence of appropriateness which makes associations between products and contexts likely to be widely held within the same consumer population (Giacalone, 2019; Marshall, 1993; Schutz, 1988). Lähteenmäki and Tuorila (1997) reported that appropriateness ratings obtained in response to tasted products are highly correlated to liking, but only when appropriateness is above a certain level, whereas “inappropriateness seems to be a culturally determined concept which cannot be invalidated by liking” (p. 90), a finding also replicated by Jaeger et al. (2019). More recently, Giacalone and Jaeger (2019b) found that appropriateness and liking tend to diverge most in products which are characterized by 1) relatively high sensory acceptability and 2) small sensory differences, two conditions that are typical in CLT tests (say, when comparing a product against its immediate competitors). On this basis, it could be expected that consumers with different preference patterns would not largely differ in the way they evaluate products as (in)appropriate for specific usage situations. At the same time, some

differences are likely to exist given that consumers form associations between products and contexts through exposure – accordingly, product familiarity has repeatedly been found to drive appropriateness (Giacalone et al., 2015; Giacalone, 2018; Giacalone & Jaeger, 2016, 2019b).

Drawing on the above, we sought to evaluate the moderating effect of liking on the relationship between sensory properties and appropriateness. Specifically, the second research question (RQ2) was: Are sensory drivers of appropriateness moderated by liking? In other words, if relationships between sensory attributes and usage situations can be identified, do they differ or hold across consumers segments with different preference patterns?

## 2. Materials and methods

### 2.1. Empirical overview

To address the two RQs, we present results from six consumer studies focusing on a diverse set of product categories, two sets of beverages – lemonades and various fruit/berry juices – and four snack foods – chips, nuts, chocolate, and popcorn (Table 1). Inclusion of several datasets helped to generalize beyond individual product categories and – linked to RQ1 – to span variation with regards to inter-product differences in sensory and IBU profiles. All studies focused on product evaluations of unbranded samples within the same product category. The latter was with a view to greater relevance for typical R&D applications and CLT studies.

### 2.2. Participants

Sample size, age and gender composition of the participants in each study are shown in Table 1. Participants in Studies 1–3 and 5–6 lived in Auckland (New Zealand) and were recruited by a marketing research provider according to specified criteria that besides age and gender quota included willingness to eat the focal product categories. The research was covered by a general approval from the human ethics committee at the New Zealand Institute for Plant & Food Research (PFR). All participants gave informed written consent and were compensated in cash. Participants in Study 4 were recruited among customers of a popular food market in Copenhagen (Denmark) and were recruited based on interest and availability. The study did not require formal ethical approval by the Danish National Committee on Health Research Ethics. Consumers in this study participated on a voluntary basis and received no compensation.

### 2.3. Samples, sensory terms and use situations

Table 1 contains summary information about samples, sensory terms and use situations. The number of samples used in each study varied from three to eight, representing a typical range in CLTs. Information on the product categories and the specific samples used in the study are given in Table 1. Test products in Studies 2, 3, 5 and 6 were commercially available in New Zealand, while Study 4 used beverage prototypes developed by author DG and Danish colleagues as part of a previous product-focused investigation (Geertsens et al., 2016).

In keeping with the aim of linking variation in appropriateness to sensory variation alone, all samples were presented unbranded (i.e., in blind) in disposable plastic containers labelled with a 3-digit code. Samples were served at room temperature in all studies, except for Study 4, where they were served at ~6 °C. Serving size, determined through pilot work, allowed for two or three sips or bites of each sample.

For each product category, lists of relevant sensory and use situations terms were developed based on pilot work. The number of sensory terms varied from a minimum of 8 to a maximum of 22 (Table 1). Priority was given to term relevance for the focal product categories but, wherever possible, the number of sensory and IBU terms was identical or very close in order to eliminate possible biases due to differences in the

**Table 1**  
Overview of studies and experimental conditions.

Study	Country	Product category	Test samples	N Consumers	Age range (y. o.) and gender (% female)	N samples	Experimental variables <sup>a</sup>	Experimental design and versioning
1	New Zealand	Lemonade with added sugar	Identical lemonades varying in sugar content (range: from 0 g/L to 120 g/L added sucrose)	139	20-66; 51%	6 (1 repeated)	H + S (8) + IBU (10)	Incomplete within-Ss design V1 (n = 70): samples S1, S2, S3 V2 (n = 69): samples S1, S4, S5
2	New Zealand	Salted snacks	Potato chips with tzatziki flavour, puffed wheat snacks with cheese flavouring, chickpea chips, potato chips with salt and vinegar flavour, wholegrain chips with honey and mustard flavour, and salted potato chips	116	20-66; 50%	6	H + S (22) + IBU (22)	Full within-Ss design, but different order of CATA questions V1 (n = 60): H + S + IBU V2 (n = 56): H + IBU + S
3	New Zealand	Nuts	Honey roasted peanuts, tamari roasted almonds, roasted and salted macadamia, walnut pieces, brazil nuts and salt and pepper cashew nuts	170	21-67; 0%	8 (1 repeated)	H + S (18) + IBU (15)	Full between-Ss design V2 (n = 67): H + IBU V3 (n = 103): H + S
4	Denmark	Juices	Juice from apple, orange, beetroot, rosehip, sea-buckthorn berries, and combinations hereof	200	18-60; 77%	7	H + S (13) <sup>b</sup> + IBU (8)	Full within-Ss design
5 <sup>c</sup>	New Zealand	Chocolate	Milk chocolate samples different price/quality brackets (budget, average, artisan) and with varying textures	210	20-66; 51%	4	H + S (8) + IBU (12)	Full between-Ss design V1 (n = 104): H + S V2 (n = 106): H + IBU
6 <sup>c</sup>	New Zealand	Popcorn	Popcorn samples flavoured to taste sweet and salty, buttery, and salty	204	20-66; 50%	3	H + S (10) + IBU (10)	Full within-Ss design, but different order of CATA questions V1 (n = 104): H + S + IBU V2 (n = 100): H + IBU + S

<sup>a</sup> H = hedonic (9-pt scale); S = Sensory (CATA); IBU = item-by-use (CATA). The numbers in brackets indicate number of CATA terms.

<sup>b</sup> Only sensory terms were included. The original ballot for this study also included some terms of more hedonic nature (e.g., complex, familiar, typical Danish) which were excluded in this paper. For the complete lists of attributes please refer to [Geertsens et al., 2016](#).

<sup>c</sup> A substantial number of consumers (67%) participated in both Study 5 and Study 6.

number of terms presented to the consumers. The full list of sensory and IBU terms in each study is given in the appendix (Table S1).

## 2.4. Procedures

Studies 1–3 and 5–6 took place in standard sensory booths under white lighting and regulated temperature (20–22 °C) and air flow. Study 4 took place in a large and popular food market in central Copenhagen, Denmark (<http://torvehallernekbh.dk/>) where a stand and a small test area was available to the researchers.

In all studies, products were presented monadically. The serving order was randomized between participants following a William's Latin square design to minimize first order and carry-over effects. Crackers and still mineral water were provided for rinsing of the palate in between samples (not enforced).

For each sample, consumers were first asked to report their liking using a 9-pt hedonic scale (1 = dislike extremely to 9 = like extremely; [Peryam & Pilgrim, 1957](#)), then consumers proceeded to complete two check-all-that-apply (CATA, [Ares & Jaeger, 2015](#)) questionnaires with sensory and IBU data respectively. Within the two individual CATA questionnaires, the order in which the attributes appeared in the ballot was randomized between consumers to reduce response biases associated with the term position ([Ares & Jaeger, 2013](#)).

There were minor methodological differences between the studies (dictated by concurrent investigations in the CLT sessions), pertaining to sample presentation and question order (see the last column of Table 1). Study 1 used an incomplete design, whereas in all other studies consumers evaluated all samples. Studies 1, 2, and 4 employed a within-subject design where consumers evaluated each product monadically

and rated their liking and completed two CATA questions focusing on sensory attributes and IBU. Studies 3 and 5 employed a between-subjects design where consumers evaluated samples using either a sensory or IBU CATA question. Regarding question order, the hedonic question was always the first, followed by the CATA questions. Four studies obtained the sensory CATA responses first, followed by the IBU CATA, whereas in two studies (2 and 6) half of the consumer completed the task in the opposite order (first IBU, then sensory).

## 2.5. Data analysis

The data were analyzed in R ([R Core Team, 2017](#)). Statistical significance was set at  $\alpha = 5\%$ .

### 2.5.1. Preliminary analyses

Preliminary analyses were performed to ascertain that the samples differed with respect to key variables of the studies, i.e., liking, sensory and IBU terms and would therefore be relevant to addressing the RQs. Analysis of variance (ANOVA) was performed with samples and consumers as a fixed and random sources of variation, respectively. Cochran's Q test was used to evaluate differences in sensory and IBU data. In studies that used an incomplete design, chi-squared goodness of fit test was used to the same end. Where a significant sample effect on a CATA term, the overall test was supplemented by pairwise comparisons between samples. *P* values from these comparisons were corrected for multiple testing in accordance with [Benjamini and Hochberg \(1995\)](#).

### 2.5.2. Analyses pertaining to RQ1

To address RQ1, we used both overall (i.e., across all terms in the

CATA questions) and term-specific analyses. When IBU data are obtained using rating scales, Cardello and Schutz (1996) proposed a measure of overall (dis)similarity in appropriateness between pairs of products obtained using a standardized Euclidean distance formula. In the same vein, we used the chi-squared distances, which is an Euclidean distance measure applicable to counts. From the contingency table (products  $\times$  CATA terms), the chi-squared distance between two products is defined as follows:

$$d_{ij} = \sqrt{\sum_k \frac{(x_{ik} - x_{jk})^2}{c_k}}$$

where  $d_{ij}$  is the distance between product  $i$  and product  $j$ ,  $x_{ik}$  and  $x_{jk}$  are the products' row profiles for term  $k$  (row profiles are obtained by taking each row point and dividing by its margin, i.e., the sum of all row points), and  $c_k$  is the average row profile for term  $k$ .

For each study we computed a scaled distance  $D$ , defined as  $D = \sum d_{ij}/n$ , with  $n$  denoting the total number of product pairs. The same procedure was applied to both appropriateness and sensory data (in the remainder of the paper,  $D^A$  is used to denote scaled distances in IBU appropriateness in each dataset, and  $D^S$  to denote scaled distances pertaining to sensory attributes) to assess the degree of differences within each dataset. Additionally, Pearson's correlation coefficient computed on the distances between individual product pairs (denoted  $d_{ij}^A$  and  $d_{ij}^S$ ) was used to quantify the relationship between the two distance measures at an overall level.

To further explore sample differentiation on the basis of chi-squared distances, classical correspondence analysis (Greenacre, 2017) was applied on the contingency tables for, respectively, sensory and IBU data. The stability of sample configuration was established by bootstrap resampling and confidence ellipses containing 95% of the bootstrapped models were used to visually infer sample separation in the multivariate space.

At the level of individual CATA terms, logistic regression was used to uncover the relationships between specific sensory attributes and usage situations in the four datasets. This analysis is described in the next section as it is central to both RQs.

### 2.5.3. Analyses pertaining to RQ2

For RQ2, we used logistic regression models where selection of sensory (1/0) terms, hedonic scores (9-pt), and their interaction were used as predictors, and selection of IBU terms (1/0) as response variables. The inclusion of both sensory attributes and hedonic scores in the models enabled us to simultaneously assess the independent effect of sensory terms and liking, as well as possible moderation effects (through the interaction terms), thereby addressing both RQ1 and RQ2. Odds ratios (O.R.) from individual regression model were used to estimate the probability of a usage situation to be selected as appropriate given the applicability of each sensory term (for guidance on O.R. interpretation see Hailpern & Visintainer, 2003). Note that the logistic regression analysis requires all sets of data to be obtained from the same consumers; therefore, Studies 3 and 5 were excluded as sensory and IBU data were obtained from different consumers in those studies.

Additionally, Multiple Factor Analysis (MFA, Escoufier & Pages, 1994) was used to compare multivariate configurations from the sensory, IBU, and liking scores, to assess the correspondence between the perceptual maps obtained by these three sets of data. The data blocks for MFA consisted of contingency tables crossing products and attributes for the two CATA datasets, and a table crossing products (rows) and individual consumers (columns) containing the liking scores. Only sensory and IBU terms which significantly discriminated between products according to Cochran's Q test were included in this analysis. Qualitative (visual) assessment was supplemented by computation of RV coefficients (Robert & Escoufier, 1976) in order to quantify the degree of agreement between product configurations, considering the first two

MFA dimensions. Study 1 was excluded from the MFA analysis as it used an incomplete design.

## 3. Results

### 3.1. Preliminary analyses

In all studies, products significantly varied in the three focal dimensions: liking, sensory profiles, and appropriateness. Table 2 presents a summary of these analysis (detailed results for individual CATA terms for each study are given in Table S1 in the Supplementary Materials).

The range of product liking means varied between 1.3 points (Study 6, popcorn) and 4.3 points (Study 4, juice) on a 9-pt scale and was generally on the positive side of the hedonic scale (Table 2). Study 4 (juice) had a slightly larger extension into the negative part of the hedonic scale: this was, however, due to a single disliked sample (mean = 2.6), without which the range would be very similar to the other studies (the next least liked sample in that study had a mean of 5.0).

Collectively the preliminary analyses confirmed the suitability of the datasets to addressing the RQs, as they showed that the test samples 1) differed with respect to key variables in all studies (liking, sensory and IBU terms), 2) had generally high acceptability (thus ensuring relevance for commercial product testing), and 3) differentially varied with respect to sensory differences.

### 3.2. Is sensory variation related to differences in appropriateness? (RQ1)

RQ1 explored whether and to which degree appropriateness is related to sensory differences within a single product category. With respect to sensory and appropriateness CATA terms, significant differences between the samples were present in all studies, however their extent varied from one study to another (Table 2). Study 1 (lemonade) showed the least amount of difference in both sensory and appropriateness data, which was expected due to the small degree of sample differences in this study. In particular, less than 1% of the comparisons based on IBU terms were significant (Table 2); the only attribute where a significant difference was found was "For health conscious people", again consistently with the nature of differences of this study which featured lemonades with increasing sugar content, with the unsweetened sample being significantly more appropriate for that situational use. By contrast, differences between the samples in both sensory and appropriateness terms were more numerous in the other five studies, where the differences between products were comparatively larger.

Fig. 1 shows CA plots from each study showing the perceptual maps based on sensory data (top) and the corresponding one from IBU data (bottom). Product names and sensory/IBU terms are omitted from the plots to improve legibility, but Figs. S7–S12 in the Supplementary Material show each individual plot with full details.

The amount of explained variance in the first two CA dimensions was  $\geq 64\%$  and  $\geq 83\%$  for models based on sensory and appropriateness, respectively, suggesting a clear structure in both types of datasets. Confidence ellipses from bootstrap resampling are superimposed to the product points, from which sample separation, in a multivariate sense, can be inferred (i.e., if two ellipses do not overlap, the products can be considered significantly different). Also note that in order to enable an easier comparison, all plots in Fig. 1 have the same axis value range.

It is easy to see that the products are better separated from a sensory point of view than they are in terms of appropriateness, as 1) products are more spread out on the CA dimensions and 2) there are fewer overlapping confidence ellipses (Fig. 1). This was the case in all studies and fits with the univariate analyses reported in Table 2. Nevertheless, all studies showed successful sample separation in both sensory and, crucially for RQ1, appropriateness data. Moreover, when interpreting the CA plots qualitatively, most studies showed a clear correspondence between the sensory and the appropriateness product spaces, as well as good face validity with respect to interpretation. For example, in Study 1



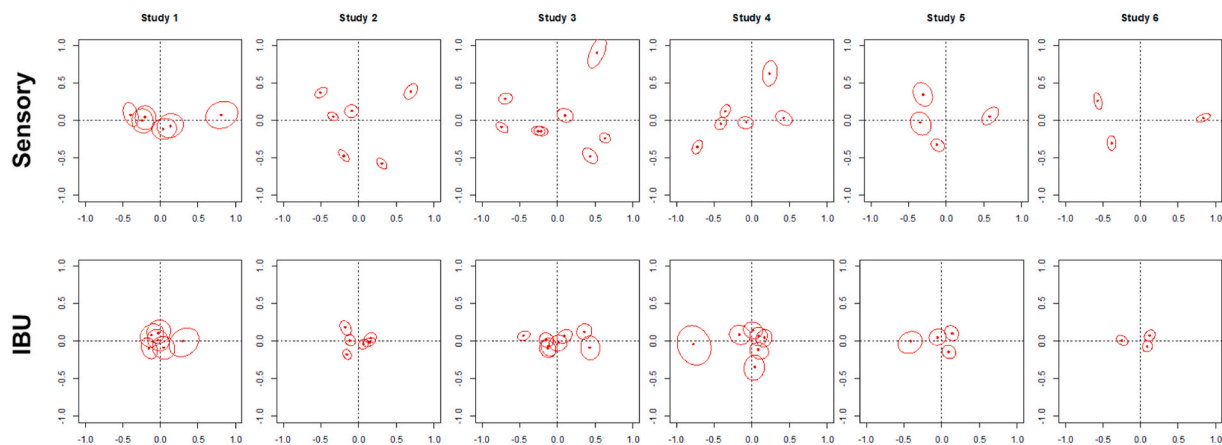
**Table 2**

Summary of inter-sample differences in key variables. The columns  $D^S$  and  $D^A$  report scaled chi-squared differences between products pertaining to sensory and IBU, respectively. Chi-squared distances between individual product pairs for all studies are reported in the Supplementary Materials (Figs. S1–S6).

Study	Liking range <sup>a</sup> and test statistic for sample effect	Percentage (%) of significant pairwise sample comparisons based on sensory terms <sup>b</sup>	$D^S$	Percentage (%) of significant pairwise sample comparisons based on IBU terms <sup>b</sup>	$D^A$
1 (Lemonades)	4.6–6.4, $F_{(4,274)} = 17.4$ , $p < 0.001$	27	0.33	0.8	0.06
2 (Salted snacks)	5.6–6.5, $F_{(5,575)} = 5.3$ , $p < 0.001$	64	0.99	26	0.07
3 (Nuts)	5.7–7.4, $F_{(7,1183)} = 33.6$ , $p < 0.001$	49	1.16	19	0.16
4 (Juices)	2.6–6.9, $F_{(6,1357)} = 103.6$ , $p < 0.001$	66	1.11	52	0.25
5 (Chocolate)	4.6–7.4, $F_{(3,627)} = 149.8$ , $p < 0.001$	77	0.42	61	0.12
6 (Popcorn)	6.0–7.3, $F_{(2,406)} = 48.2$ , $p < 0.001$	87	0.88	60	0.06

<sup>a</sup> Range of average liking scores on 9-point scale where 1 = ‘dislike extremely’ and 9 = ‘like extremely.’

<sup>b</sup> Based on Cochran’s Q test or chi-squared goodness-of-fit depending on whether studies employed, respectively, a complete or incomplete design. For detailed results refer to Table S1 in the appendix.



**Fig. 1.** CA plots showing perceptual maps obtained from sensory (top) and IBU (bottom) data for each study. Each plot shows the first two dimensions of the respective CA models, with superimposed 95% confidence ellipses obtained by bootstrap resampling around the samples (if two ellipses do not overlap, the products can be considered significantly different). For visual clarity an ease of comparison, product names and attributes have been removed, and all plots have the same axis dimensions. Detailed plots for individual studies are given in the Supplementary Materials (Figs. S7–S12).

(lemonade) the samples were separated in by sweetness from a sensory point of view, and for appropriateness “for health conscious people” in the IBU data, with the unsweetened sample being the most frequently associated to this term compared to those to which sugar was added (see Fig. S7 in Supplementary Material).

Sample separation based on IBU data appeared better in the other studies (than in Study 1) where the sensory differences were larger and more multidimensional. This suggests that with increasing sensory differences the products were also perceived as more diverse in terms of appropriateness. To quantitatively verify this interpretation, we calculated the correlation between chi-squared distances between each product pair based on sensory data ( $d_{ij}^S$ ) against the corresponding distance based on appropriateness ( $d_{ij}^A$ ). Overall, across all possible product pairs, there was a linear relationship between the magnitude of sensory difference and the magnitude of appropriateness difference, but the strength of this relationship was moderate ( $r_{(86)} = 0.44$ ,  $p < 0.001$ ), as also indicated by the presence of several outliers (cf. Fig. S13 in the supplementary material which plots the chi-squared distance between each product pair based on sensory data ( $d_{ij}^S$ ) against the corresponding distance based on appropriateness ( $d_{ij}^A$ )).

Accordingly, between-study differences were large. In particular, a strong linear relationships between perceptual differences in sensory and appropriateness was observed in Studies 1 and 6 (respectively:  $r_{(13)}$

$= 0.91$ ,  $p < 0.001$ , and  $r_{(1)} = 0.99$ ,  $p = 0.06$ ), whereas the other four studies showed weak to moderate correlations (Study 2:  $r_{(13)} = 0.21$ ,  $p = 0.455$ ; Study 3:  $r_{(26)} = 0.37$ ,  $p = 0.049$ ; Study 4:  $r_{(19)} = 0.39$ ,  $p = 0.076$ , Study 5:  $r_{(4)} = 0.12$ ,  $p = 0.818$ ).

### 3.3. Moderation of perceived appropriateness by liking (RQ2)

#### 3.3.1. Logistic regression

RQ2 focused on whether the effect of sensory attributes on appropriateness is moderated by liking. To this end, we consider the results of the logistic regression models linking selection of each IBU term to the selection of each sensory term, the level of liking, and their interactions. These results are first reported in a summarized form in Table 3, which shows that liking independently affected IBU term selection in a majority of cases (from 60% to over 90% of models). Whenever a main effect of liking was found, the direction of the effect was always positive, meaning that, everything else being equal, higher liking always made it more likely that a consumer would select an IBU term. Sensory terms independently affected the selection of appropriateness terms less frequently – from a minimum of 6% of logistic regression models in Study 6 to a maximum of 22.5% of models in Study 1, meaning that a minority of sensory terms were responsible for differences in appropriateness. Similar percentages of models were also found to contain a significant interaction with liking (Table 3), indicating that the effect of

**Table 3**

Overview of logistic regression results modelling the likelihood of selecting each IBU term using liking scores, individual sensory term, and their interaction as the independent variables. Studies 3 and 5 are excluded from this analysis as sensory and IBU data were obtained from different consumers in those studies.

Study	Models with a significant effect of liking (%)	Models with a significant effect of sensory terms (%)	Models with a significant 2-way interaction (%)
1 (Lemonades)	58.7	22.5	27.5
2 (Salted snacks)	90.7	8.0	6.0
4 (Juices)	97.0	12.5	9.0
6 (Popcorn)	60.0	6.0	7.0

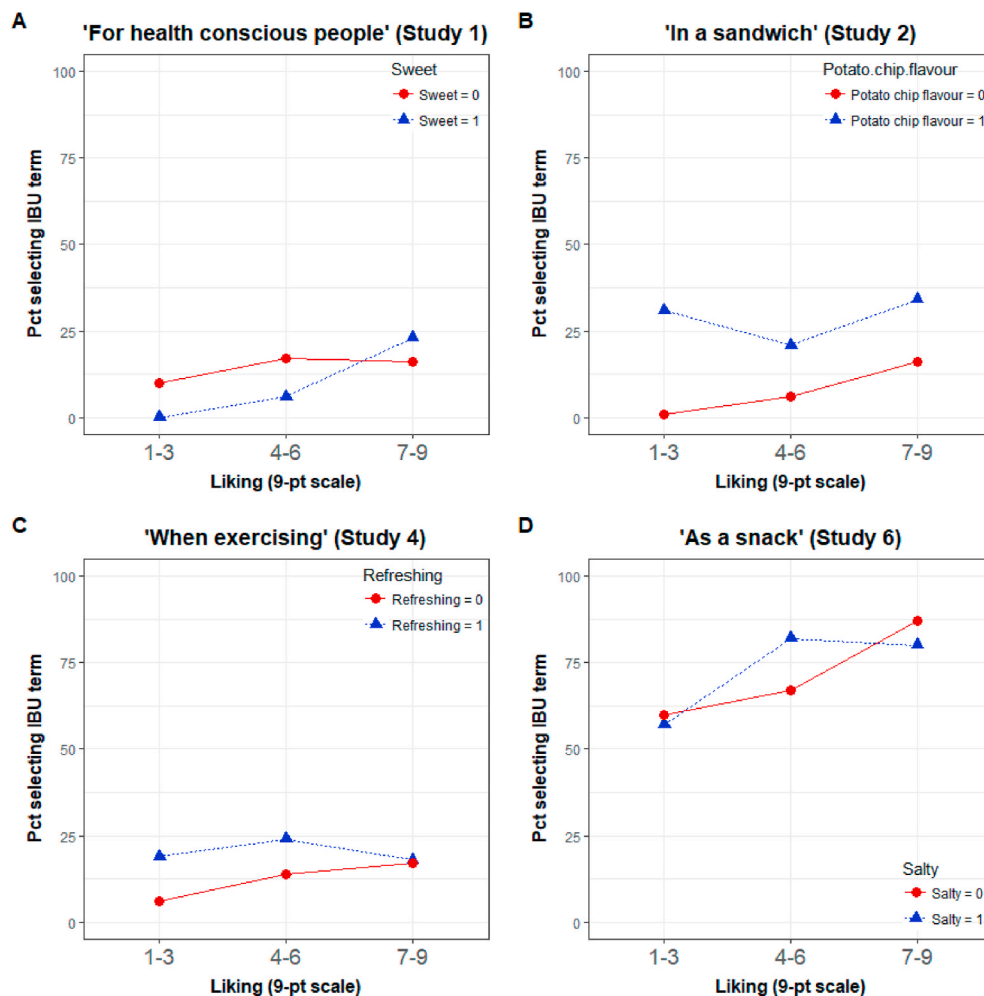
sensory attributes on appropriateness depends on the level of liking.

To understand these interactions more in depth, consider Fig. 2 which shows four examples of links between sensory and IBU terms: (a) “Sweet” and “For health conscious people” (from Study 1), (b) “Potato chip flavor” and “In a sandwich” (from Study 2), (c) “Refreshing” and “When exercising” (from Study 4), (d) “Salty” and “As a snack” (from Study 6). These examples were chosen as they all show a significant main effect and a significant interaction (Table S2 in the Supplementary Materials provide full results with main effect and interactions for each individual CATA term in each study). For each plot, the percentage of consumers who checked that specific IBU term is plotted (on the y-axis) for different levels of liking (x axis). These percentages of term selection are shown separately for consumers who did/did not check the predictor sensory term (the red and blue lines). In other words – taking Fig. 2a as example – the red line shows the percentage of consumers who selected “For health conscious people” out of those who checked “Sweet” to describe the sample, whereas the blue line shows the percentage of consumers who selected “For health conscious people” out of those who

did NOT checked “Sweet” to describe the sample. Note that the liking scores are grouped into a low (1–3), medium (4–6), and high (7–9) level in Fig. 2, both for simplicity and because there otherwise would have been too few data points in each point of the hedonic scale (particularly at the extremes) for the results to be meaningful.

In general, all four plots in Fig. 2 show that IBU term selection increases with liking, consistent with the finding that the main effect of liking in the logistic regressions is always positive. However, they also show that sensory properties affect IBU term selection independently, and also that the difference may depend on level of liking.

Fig. 2a shows the effect of selecting the term “sweet” on the likelihood of selecting the term “for health-conscious people” (Study 1). The main effect for that term, reported in Table S2, indicate that sweetness had a negative effect on selection frequency of that use context. In particular, the odd ratio associated to the main effect of “sweet” is close to 0: everything else being equal, consumers who found the term “sweet” applicable to a lemonade were extremely unlikely to then select the IBU term “for health conscious people”. However, Table S2 also shows a

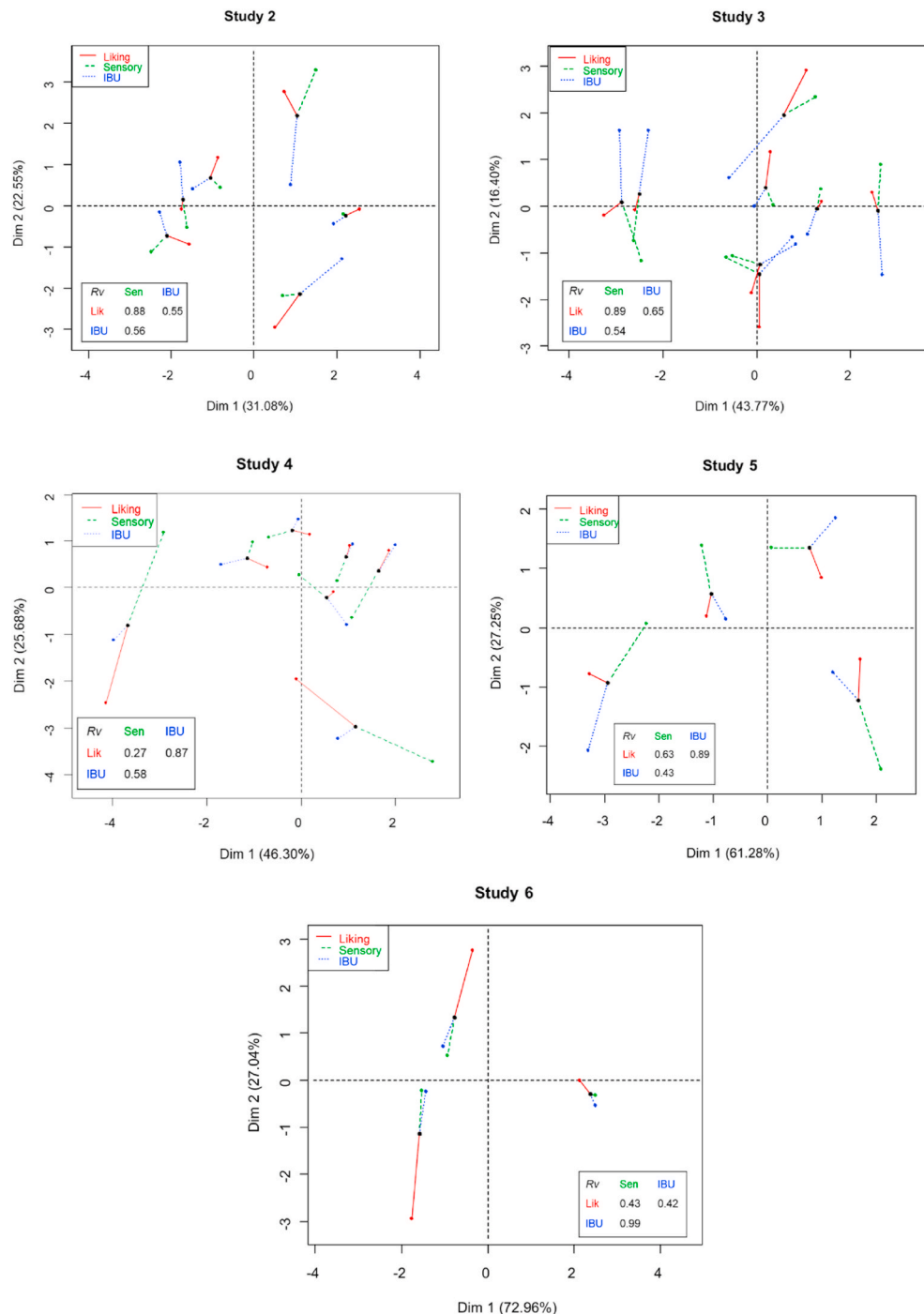


**Fig. 2.** Examples of interactions between sensory and liking in driving the selection of specific IBU terms (a: ‘For health conscious people’, b) ‘In a sandwich’, c) ‘When exercising’, d) ‘As a snack’). In each example, the percentage of consumers (y axis) who checked the IBU term for each level of liking (x axis) is plotted. The red circled (●) and blue triangle lines (▲) separately show the selection frequency of the IBU term (e.g. ‘For health conscious people’ in Fig. 2a) for consumers who, respectively, did and did not check the predictor sensory term (e.g. ‘Sweet’ in Fig. 2a) to describe the samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

significant interaction with liking, and the odd ratio for the interaction term is positive (i.e.,  $>1$ ). Accordingly, Fig. 2a shows that while for low and medium levels of liking, consumers who perceived the samples as sweet almost never think it is appropriate ‘for health-conscious people’. This changes as liking increases we can see that sweetness become unrelated to the probability of term selection and actually for high liking level (7–9) the two line are very close, meaning that even consumers who found the product sweet could consider it appropriate for health conscious people, if they liked it very much.

Similarly, the other three plots show instances where sensory

properties affect IBU term selection both independently and depending of level on level of liking. For example, in Fig. 2b we see that consumers who checked the sensory term “potato chips flavor” selected more often the IBU term “in a sandwich” and their selection frequency was unrelated to the level of liking. By contrast, for consumers who did not check that sensory term, frequency of selection increased linearly with the level of liking (Fig. 2b). A very similar situation is shown in Fig. 2c, where consumers who perceived the juices “refreshing” were more likely to select the IBU term “when exercising” regardless of how much they liked the samples. For consumers who did not check “refreshing”,



**Fig. 3.** MFA plot (first two dimensions) showing sample similarities and differences in terms of liking, sensory profiles, and IBU appropriateness for Studies 2, 3, 4, and 6. In addition to overall sample positioning, the plots show the partial configurations obtained from each dataset separately superimposed to the consensus points. The boxes in the bottom left corners report RV coefficients between these configurations. Study 1 is excluded from this analysis as it used an incomplete block design.

frequency of selection of the IBU term again followed liking in a linear fashion. Finally, in Fig. 2d we see a very similar trend between the sensory term “Salty” and the IBU term “As a snack”, except that in this case the main difference related to sensory term is usage is for moderate liking level, whereas for low and high level of liking the two consumer groups select the IBU term at a similar rate.

### 3.3.2. Multiple Factor Analysis

To further assess the relative importance of sensory attributes on perceived appropriateness (RQ2), Fig. 3 shows MFA plots with product maps obtained considering these three groups of variables (sensory, liking, and appropriateness) in Studies 2–6 (Study 1 is omitted as this analysis requires a full experimental design). Each plot also shows partial configurations obtained from each dataset separately, and RV coefficients between these configurations.

Fig. 3 supports the indication of a significant variation between studies. Two studies - Studies 4 and 5 - showed a poor correspondence between sensory attributes and appropriateness, and a high degree of correspondence between liking and appropriateness. Note that the partial product points from liking (in red in Fig. 3) are identical to those one would obtain from an internal preference map using principal component analysis. Therefore, it appears that liking was the main driver of appropriateness in these studies and that these two datasets provided similar information about products. By contrast, in Studies 2 and 3 liking and sensory data were highly correlated with each other, and to a lesser degree with IBU, suggesting that IBU provided additional information in terms of product characterization in these studies. Finally, in Study 6, product configurations from sensory and IBU were nearly identical, and clearly different from liking, consistent with the correlation analysis for this dataset reported in the previous section.

## 4. Discussion

The overall aim of this research was to investigate whether unbranded food and beverage products from the same product category could be differentiated on the basis of perceived appropriateness (RQ1), and to evaluate the degree to which sensory variation and liking contributed to drive appropriateness evaluations (RQ2).

### 4.1. Linking sensory variation to degree of appropriateness

With respect to RQ1, the key finding was that sensory variation among products from the same product category can produce differences in appropriateness. This conclusion was supported by univariate analyses (Cochran's Q test) as well as visual inspection of CA plots showing successful sample separation in all studies in both sensory - which was expected - but also IBU data. Additionally, results from analyses of overall product differences using chi-squared differences showed that the magnitude of sensory differences was linearly related to the magnitude of difference in appropriateness. Intuitively, this makes good sense, as it means that the more different two products are from a sensory point of view, the more likely is that consumers will associate them to different use situations.

However, evidence of non-linearity and large study-to-study differences with respect to the strength of the relationship also emerged. Considering the results in light of other analyses (Table 3, Fig. S13 and Table S2 in Supplementary Materials) suggest that the reason for the moderate correlation may reside in the fact that differences in appropriateness are generally related to one or a few specific sensory attributes (e.g., sweetness in Study 1), whereas chi-square distances consider overall differences across all attributes, thereby “diluting” the effect. It may possibly reflect differences in versatility (how many different uses exist for a product) associated to specific product categories. For example, Study 4 (fruit juice) showed average product separation with respect to sensory attributes but the highest degree of product differentiation with regard to appropriateness (Table 2 and Fig. S10 in

Supplementary Materials), consistent with that fact that beverages are generally consumed in a wider range of occasions than solid food items (McCrickerd et al., 2014; Mueller Loose & Jaeger, 2012).

Interestingly, multivariate findings of perceptual maps (Figs. S7 – S12 in Supplementary Materials) revealed that the sensory product spaces did not always have a one-to-one relationship with the appropriateness spaces; put differently, the sensory attributes that most discriminated between the samples were not always the most important drivers of appropriateness. For instance, in Study 5 (chocolate) the largest sensory differences between the chocolate samples related to texture, yet taste (sweetness) was more important for important for explaining differences in appropriateness. A parallel can be drawn here with research on (external) preference mapping, where sensory attributes most represented in the first component are not necessarily the ones most predictive of consumer preference (MacFie, 2007).

Another finding with face validity was that sensory differences between the products were consistently larger than differences in appropriateness. This could be expected given that consumers evaluated products in blind conditions. Giacalone and Jaeger (2019a) showed that variation in IBU appropriateness is closely linked to the heterogeneity in the product set, meaning that product differentiation based on appropriateness would have likely been much larger had consumers evaluated the same products in presence of brand and packaging elements, and/or if they had evaluated foods belonging to different product categories. Moreover, it should be noted that this difference in discriminative ability could have been accentuated by the strategy for generating CATA terms, which was slightly different for sensory and IBU terms. For sensory terms, the criterion was to select terms that discriminated between samples, reflecting established practices in sensory evaluation. This was more loosely enforced when generating IBU terms, where the focus was more on terms that were appropriate (or not) for the category as a whole.

### 4.2. Accounting for liking when examining the role of sensory variation on appropriateness

RQ2 sought to disentangle the effect of sensory attributes from that of liking. The results first and foremost showed that liking in and of itself always increased the likelihood that a product would be evaluated as appropriateness. This was expected as it is consistent with multiple previous reports of a positive relationship between liking and appropriateness (e.g., Lähteenmäki & Tuorila, 1995, 1997; Jaeger, Roigard, et al., 2019, 2021). Nevertheless, independent associations between sensory attributes and appropriateness for specific use situations were also established. These associations were, as expected, specific to the product category and dependent on the specific products included in the studies, and in many cases appeared to have good interpretability. For example, “salty” popcorns were perceived as better for snacking, a “refreshing” fruit juice was considered more appropriate for consumption after exercise, and “sweet” lemonades were negatively associated with health orientation (Fig. 2).

Furthermore, logistic regression results (Table 3 and Table S2 in Supplementary Material) revealed that, in almost all cases where sensory attributes had a significant effect on appropriateness, the magnitude and direction of the effect was dependent on liking. More precisely, the probability that a sensory attribute increases (or decreases) appropriateness for a consumption situation depends on the level of product liking. Accordingly, the general indication seems to be that the more consumers like the product, the less individual sensory attributes matter in and of themselves, to the point that even strong negative associations (e.g., sweetness and health) and positive associations (e.g., saltiness and snacks) may be overridden at very high levels of liking. This result appears quite consequential from an applied perspective, as it means that it may be difficult to disentangle such sensory-specific effects from that of liking, especially if hedonic differences between products are large (by contrast, all studies showed instances of pairs of sample where average



liking was similar but sensory and IBU profiles were very different). Thus, controlling for level of liking in the data analysis or by comparing products that do not differ in liking seems to be necessary to correctly interpret any sensory specific effect on appropriateness.

#### 4.3. Implications for appropriateness measurement in CLT tests

In general, the inclusion of appropriateness measures for product differentiation has been deemed most useful when comparing products characterized by small or no difference in liking (Giacalone & Jaeger, 2019b; Jaeger et al., 2019b, 2018). The present results further support this recommendation, as do the multivariate analyses. Specifically, the highest degree of configurational similarity between liking and IBU was achieved in Study 4, which had the largest span in liking means between the products and also extended the most into the disliking dimension; by contrast in the two studies (2 and 6) where products were closest to one another in terms of liking (Table 2), these two types of data clearly provided different information on the products.

Still, all studies provided examples where product rankings in terms of liking and appropriateness differed, and of less liked products outperforming others in terms of appropriateness for specific situations, such as the case of the very disliked prickly/sour sea-buckthorn juice in Study 4 that was nonetheless strongly associated to use in cocktails. In general, the multivariate analyses clearly indicate that the inclusion of all these three sets of variables – liking, sensory and appropriateness – is necessary for a complete product characterization as they provide different and complementary information, a point often highlighted in similarly motivated work (e.g., Jaeger et al., 2017; Spinelli et al., 2019).

Taken together, the results of this work show that differences in the sensory profiles of foods and beverages are associated to differences in perceived situational appropriateness. This supports growing calls for inclusion of appropriateness measures sensory and consumer science (Cardello & Schutz, 1996; Giacalone, 2019; Giacalone & Jaeger, 2019b; Jaeger & Porcherot, 2017) and extends their relevance to CLT studies employing variants within the same product category.

Successful product differentiation on the basis of appropriateness data was achieved in all studies. Our results are therefore consistent with previous reports (e.g., Cardello & Schutz, 1996; Hersleth et al., 2005; Lähteenmäki & Tuorila, 1997; Mejlholm & Martens, 2006) and contribute further evidence that appropriateness is a meaningful dimension on which unbranded product variants can be differentiated from a consumer perspective. However, it does not necessarily follow from our results that differences in appropriateness *always* exist or that one should *always* include appropriateness ratings in consumer tests. For instance, Kessler et al. (2019) used a similar methodology to this research in a study focusing on snack sausages, but did not find significant differences in appropriateness between the samples, likely because sausages do not have many different uses (consumers in that study pretty much only saw sausages appropriate as part of a main meal, and all test products performed similarly in that situation).

Neither does it follow that appropriateness evaluations need to be undertaken with the exact same methodological approach as in this paper. There may also be situations in which probing appropriateness for many possible usage contexts is unnecessary and it may instead be more relevant to focus on a single usage envisioned during product development, such as often seen in studies on meat replacers and other novel ingredients (Elzermann et al., 2011; Tan et al., 2016). Other times it may be sufficient to ask consumers to rate appropriateness “for many different occasions” to gauge the versatility of a product as exemplified in Cardello et al. (2016) and Jaeger et al. (2017). Clearly, these decisions should ultimately depend on specific considerations on the products and applications at hand. Pilot testing is strongly advised to elucidate whether differences in appropriateness are relevant to the focal samples (see Geertsen et al., 2016 for an example in a product development context).

#### 4.4. Limitations and suggestions for future research

In closing, we acknowledge limitations and suggest avenues for future research. First, all studies used products that were on, average, at or above the neutral point on the 9-pt hedonic scale. While this was part of the research strategy (i.e., to reflect the generally high eating quality of commercial food products) it leaves open the question of how the results would look like had the products spanned more into the disliking domain.

Even though several (six) products categories were studied, only two larger categories are covered (non-dairy soft drinks and snack products). Therefore, whether these findings generalize beyond product categories should be ascertained in future research. There is a clear assumption that this be the case, however, since situational appropriateness has been successfully employed for product differentiation in other product categories as well, such as milk and dairy (Jack, Piggott, 1994, & Patterson, 1994; Raats & Shepherd, 1991; Lähteenmäki & Tuorila, 2005), meat products (Elzermann et al., 2021; Hersleth et al., 2011), and alcoholic beverages (Schutz & Ortega, 1974; Mejlholm & Martens, 2006; Giacalone & Jaeger, 2016).

All results were based on the same methodological paradigm (IBU) for evaluation of appropriateness. Again, this was part of the research strategy in order to make results comparable. However, in the future it would be interesting to extend this research by using methods that provide deeper and more causal-like insights. For example, repertory grid interviews, which have also been used in appropriateness research (e.g., Jaeger et al., 2005; Spinelli et al., 2019a), could be used to elicit relevant use situations from consumers and have them choose between pairs or triads of products varying in key sensory attributes.

In this paper we have focused on aggregated/panel level analyses. While not impinging on the conclusions of this paper directly, one should keep in mind that individual differences almost certainly exist in the way consumers associate different sensory properties with different consumption situations, both between and within cultures (e.g., are salty or savory items appropriate for breakfast? Is pungency ever ok?). Depending on the sample size and/or on the specific research questions, segmentation approaches as shown in Jaeger, Roigard, et al. (2019) and Jaeger, Roigard, Ryan, Jin, and Giacalone (2021) may be more suitable for studying the sensory underpinnings of appropriateness in different consumer groups.

#### 5. Conclusions

With the aim of providing a deeper understanding of the sensory underpinnings of appropriateness, we presented data from six consumer studies employing product variants within the same product category. While products were better characterized based on the sensory data than on appropriateness, significant differences in IBU appropriateness between the products were consistently obtained in all studies, even when sensory variation was unidimensional and controlled. Crucially, since consumers evaluated the products in blind, differences in perceived appropriateness could be directly attributed to variation in sensory profiles. As expected, the degree of inter-product differences in appropriateness was approximately linearly related to the degree of differences in sensory profiles. Finally, while some sensory properties independently affected appropriateness, the intensity (and in some cases the direction) of the effects often depended on the level of liking for the product. Taken collectively, the results show that sensory variation, in and of itself is sufficient to elicit differences in perceived appropriateness of food and beverages. This calls for greater inclusion of appropriateness evaluation as measures of product performance in CLT evaluations, including those within a single product category. The results further suggested that the informational value of appropriateness data is maximized when comparing products that are similarly liked.

## Ethical statement

As concerns Studies 1–3 and 5–6, the research was covered by a general approval from the human ethics committee at the New Zealand Institute for Plant & Food Research (PFR). All participants gave informed written consent and were compensated in cash. Participants in Study 4 were recruited among customers of a popular food market in Copenhagen (Denmark) and were recruited based on interest and availability. The study did not require formal ethical approval by the Danish National Committee on Health Research Ethics. Consumers in this study participated on a voluntary basis and received no compensation.

The same information are reported in the paper, at section 2.2.

## Funding

Financial support for Studies 1–3 and 5–6 was received from two sources: 1) The New Zealand Ministry for Business, Innovation & Employment, and 2) The New Zealand Institute for Plant and Food Research Limited.

## CrediT author statement

**DG:** Conceptualization, Methodology, Formal analysis, Visualization, Writing – Original draft, Writing – Editing & Review. **SRJ:** Conceptualization, Methodology, Investigation, Writing – Original draft, Writing – Editing & Review.

## Declaration of competing interest

The authors declare no conflicts of interest.

## Acknowledgements

Staff at the Sensory & Consumer Science Team at Plant and Food Research are thanked for help in data collection and data management. Christina M. Roigard and Sok L. Chheang are also thanked for helpful comments to previous versions of this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.appet.2021.105589>.

## References

- Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference*, 28, 141–153.
- Ares, G., & Jaeger, S. R. (2015). Check-all-that-apply (CATA) questions with consumers in practice: Experimental considerations and impact on outcome. In J. Delarue, B. Lawlor, & M. Rogeaux (Eds.), *Rapid sensory profiling techniques* (pp. 227–245). Amsterdam: Woodhead Publishing.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57, 289–300.
- Cardello, A. V. (2020). Human experience of eating and drinking: Perspectives on 50 Years of measurement progress. In H. Meiselman (Ed.), *Handbook of eating and drinking*. Cham: Springer.
- Cardello, A. V., Pineau, B., Paisley, A. G., Roigard, C. M., Chheang, S. L., Guo, L. F., & Jaeger, S. R. (2016). Cognitive and emotional differentiators for beer: An exploratory study focusing on “uniqueness”. *Food Quality and Preference*, 54, 23–38.
- Cardello, A. V., & Schutz, H. G. (1996). Food appropriateness measure as an adjunct to consumer preference/acceptability evaluation. *Food Quality and Preference*, 7, 239–249.
- Cardello, A. V., Schutz, H. G., Snow, C., & Leshner, L. L. (2000). Predictors of food acceptance, consumption and satisfaction in specific eating situations. *Food Quality and Preference*, 11, 201–216.
- Elzerman, J. E., Hoek, A. C., van Boekel, M. J., & Luning, P. A. (2015). Appropriateness, acceptance and sensory preferences based on visual information: A web-based survey on meat substitutes in a meal context. *Food Quality and Preference*, 42, 56–65.
- Elzerman, J. E., Keulemans, L., Sap, R., & Luning, P. A. (2021). Situational appropriateness of meat products, meat substitutes and meat alternatives as perceived by Dutch consumers. *Food Quality and Preference*, 88, 104108.
- Escofier, B., & Pages, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis*, 18, 121–140.
- Geertsen, J., Allesen-Holm, B. H., & Giacalone, D. (2016). Consumer-led development of novel sea-buckthorn based beverages. *Journal of Sensory Studies*, 31, 245–255.
- Giacalone, D. (2018). Product performance optimization. In G. Ares, & P. Varela (Eds.), *Methods in consumer research*, tome 1 pp. 159–185. Woodhead Publishing.
- Giacalone, D. (2019). Situational appropriateness in food-oriented consumer research: Concept, method, and applications. In H. Meiselman (Ed.), *Context: The effects of environment on product design and evaluation* (pp. 111–140). Amsterdam: Woodhead Publishing.
- Giacalone, D., Frost, M. B., Bredie, W. L., Pineau, B., Hunter, D. C., Paisley, A. G., Beresford, M., & Jaeger, S. R. (2015). Situational appropriateness of beer is influenced by product familiarity. *Food Quality and Preference*, 39, 16–27.
- Giacalone, D., & Jaeger, S. R. (2016). Better the devil you know? How product familiarity affects usage versatility of foods and beverages. *Journal of Economic Psychology*, 55, 120–138.
- Giacalone, D., & Jaeger, S. R. (2019a). Perceived situational appropriateness as a predictor of consumers' food and beverage choices. *Frontiers in Psychology*, 10, 1743.
- Giacalone, D., & Jaeger, S. R. (2019b). Consumer ratings of situational (‘item-by-use’) appropriateness predict food choice responses obtained in central location tests. *Food Quality and Preference*, 78, 103745.
- Greenacre, M. (2017). *Correspondence analysis in practice*. Boca Raton, FL: CRC Press.
- Hailpern, S. M., & Visintainer, P. F. (2003). Odds ratios and logistic regression: Further examples of their use and interpretation. *STATA Journal*, 3, 213–225.
- Hersleth, M., Lengard, V., Verbeke, W., Guerrero, L., & Næs, T. (2011). Consumers' acceptance of innovations in dry-cured ham: Impact of reduced salt content, prolonged aging time and new origin. *Food Quality and Preference*, 22, 31–41.
- Jack, F. R., Piggott, J. R., & Patterson, A. (1994). Use and appropriateness in cheese choice, and an evaluation of attributes influencing appropriateness. *Food Quality and Preference*, 5, 281–290.
- Jaeger, S. R., Cardello, A. V., Chheang, S. L., Beresford, M. K., Hedderley, D. I., & Pineau, B. (2017). Holistic and consumer-centric assessment of beer: A multi-measurement approach. *Food Research International*, 99, 287–297.
- Jaeger, S. R., Lee, P. Y., Jin, D., Chheang, S. L., Rojas-Rivas, E., & Ares, G. (2019a). The item-by-use (IBU) method for measuring perceived situational appropriateness: A methodological characterisation using CATA questions. *Food Quality and Preference*, 78, 103724.
- Jaeger, S. R., & Porcherot, C. (2017). Consumption context in consumer research: Methodological perspectives. *Current Opinion in Food Science*, 15, 30–37.
- Jaeger, S. R., Roigard, C. M., Le Blond, M., Hedderley, D. I., & Giacalone, D. (2019b). Perceived situational appropriateness for foods and beverages: Consumer segmentation and relationship with stated liking. *Food Quality and Preference*, 78, 103701.
- Jaeger, S. R., Roigard, C. M., Ryan, G., Jin, D., & Giacalone, D. (2021). Consumer segmentation based on perceived situational appropriateness: Partial replication and extension. *Food Quality and Preference*, 87, 104057.
- Jaeger, S. R., Rossiter, K. L., & Lau, K. (2005). Consumer perceptions of novel fruit and familiar fruit: A repertory grid application. *Journal of the Science of Food and Agriculture*, 85, 480–488.
- Jaeger, S. R., Spinelli, S., Ares, G., & Monteleone, E. (2018). Linking product-elicited emotional associations and sensory perceptions through a circumplex model based on valence and arousal: Five consumer studies. *Food Research International*, 109, 626–640.
- Kessler, F., Nielsen, M. B. R., Tøstesen, M., Duelund, L., Clausen, M. P., & Giacalone, D. (2019). Consumer perception of snack sausages enriched with umami-tasting meat protein hydrolysates. *Meat Science*, 150, 65–76.
- Lähteenmäki, L., & Tuorila, H. (1995). Consistency of liking and appropriateness ratings and their relation to consumption in a product test of ice cream. *Appetite*, 25, 189–198.
- Lähteenmäki, L., & Tuorila, H. (1997). Item-by-use appropriateness of drinks varying in sweetener and fat content. *Food Quality and Preference*, 8, 85–90.
- MacFie, H. (2007). Preference mapping and food product development. In H. MacFie (Ed.), *Consumer-led food product development* (pp. 551–592). Boca Raton, FL: Woodhead Publishing.
- Marshall, D. (1993). Appropriate meal occasions: Understanding conventions and exploring situational influences on food choice. *International Review of Retail Distribution & Consumer Research*, 3, 279–301.
- McCrickerd, K., Chambers, L., & Yeomans, M. R. (2014). Fluid or fuel? The context of consuming a beverage is important for satiety. *PLoS One*, 9, Article e100406.
- Mejlholm, O., & Martens, M. (2006). Beer identity in Denmark. *Food Quality and Preference*, 17, 108–115.
- Mueller Loose, S., & Jaeger, S. R. (2012). Factors that influence beverage choices at meal times. An application of the food choice kaleidoscope framework. *Appetite*, 59, 826–836.
- Peryam, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology*, 11, 9–14.
- Piqueras-Fiszman, B., & Jaeger, S. R. (2019). Evoked consumption context matters in food-related consumer affective research. In H. Meiselman (Ed.), *Context: The effects of environment on product design and evaluation* (pp. 545–563). Amsterdam: Woodhead Publishing.
- R Core Team. (2017). In R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Raats, M. M., & Shepherd, R. (1991). An evaluation of the use and perceived appropriateness of milk using the repertory grid method and the 'item by use' appropriateness method. *Food Quality and Preference*, 3, 89–100.
- Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 25, 257–265.
- Schutz, H. G. (1994). Appropriateness as a measure of the cognitive-contextual aspects of food acceptance. In D. M. H. Thomson (Ed.), *Measurement of food preferences* (pp. 25–50). London: Blackie Academic.
- Schutz, H. G., & Ortega, J. H. (1974). Consumer attitudes toward wine. *American Journal of Enology and Viticulture*, 25, 33–38.
- Sosa, M., Martínez, C., Arruiz, F., Hough, G., & Mucci, A. (2005). Degree of appropriateness and frequency of consumption of mayonnaise, ketchup, mustard and similar sauces in Argentina. *Food Quality and Preference*, 16, 667–674.
- Spinelli, S., Dinnella, C., Ares, G., Abbà, S., Zoboli, G. P., & Monteleone, E. (2019). Global Profile: Going beyond liking to better understand product experience. *Food Research International*, 121, 205–216.
- Spinelli, S., Dinnella, C., Ares, G., Abbà, S., Zoboli, G. P., & Monteleone, E. (2019a). Global Profile: Going beyond liking to better understand product experience. *Food Research International*, 121, 205–216.
- Spinelli, S., Monteleone, E., Ares, G., & Jaeger, S. R. (2019b). Sensory drivers of product-elicited emotions are moderated by liking: Insights from consumer segmentation. *Food Quality and Preference*, 78, 103725.
- Stolzenbach, S., Bredie, W. L., Christensen, R. H., & Byrne, D. V. (2016). Understanding liking in relation to sensory characteristics, consumer concept associations, arousal potential and “appropriateness for use” using apple juice as an application. *Journal of Sensory Studies*, 31, 135–142.
- Tan, H. S. G., Fischer, A. R. H., Van Trijp, H. C. M., & Stieger, M. (2016). Tasty but nasty? Exploring the role of sensory-liking and food appropriateness in the willingness to eat unusual novel foods like insects. *Food Quality and Preference*, 48, 293–302.