Developing a reduced consumer-led lexicon to measure emotional response to beer

Carolina Chaya[a], Curtis Eaton[b], Louise Hewson[b], Rocío Fernández Vázquez[c], Virginia Fernández-Ruiz[d], Katherine A. Smart[e], Joanne Hort[b]*

[a] Department of Agricultural Economics, Statistics and Business Management, Technical University of Madrid, Spain

[b] Division of Food Sciences, The University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Leicestershire, LE12 5RD, United Kingdom

[c] Food Colour & Quality Laboratory, University of Seville, Spain

[d] Department of Nutrition and Bromatology II (Bromatology), Complutense University of Madrid, Spain

[e] SABMiller plc, Woking, United Kingdom

*Corresponding author Tel: +44 115 951 6222. Fax: +44 115 951 6142. Email: Joanne.Hort@nottingham.ac.uk

Abstract

Previous researchers have recently recommended and utilised consumer-led lexicons to measure emotional response. This study further advances this approach by 1) making the lexicon generation process more efficient by using consumer focus groups as opposed to individual consumer interviews and 2) decreasing the number of responses required from each consumer by reducing the lexicon to categories of similar terms. In response to 10 lager samples which were manipulated in order to control selected sensory properties, focus groups generated a lexicon of 44 emotion terms. This lexicon was reduced to 12 distinct emotion categories using linguistic checks and cluster analysis. Naïve beer consumers ($n = 113$) used these 12 emotion categories to rate their emotional response to the 10 samples. The reduced consumer-led lexicon was validated through its ability to discriminate across samples as well as show differences in emotional response between genders and age groups. The 12 emotion categories were found to discriminate well between samples, although a number of categories grouped samples similarly. However, differences in responses to otherwise comparable emotion categories were identified between genders and age groups, highlighting the importance of including all emotion categories so as to not over-reduce the lexicon and risk missing out on valuable emotion data.

Key words

Emotion, consumer-led lexicon, reduced form, beer, gender, age

Highlights

- A reduced consumer-led emotion lexicon was developed in response to beer samples

- The 12 emotion categories successfully discriminated between samples

- Emotional response differences across genders/age groups were also highlighted

1 Introduction

Emotion research in sensory and consumer science has gathered significant momentum over recent years. This is in no small part due to the increased reliance on emotional characteristics of products for a differential advantage in the modern marketplace where products are of similar quality and price (Schifferstein, Fenko, Desmet, Labbe, & Martin, 2013) and comparable in technical and performance properties (Churchill & Behan, 2010). As interest in this research area continues to grow, the need for effective emotion measurement methodologies is increasing.

To date, the majority of sensory and consumer emotion research has focussed on self-report measures which require the consumer to directly indicate their emotional response to the given stimulus. In verbal self-report, this often requires an emotional lexicon for the consumer to refer to. Such lexicons can be divided into two categories: pre-determined or consumer-led. A prominent example of a pre-determined emotion lexicon is EsSense Profile® ((King and Meiselman (2010); King, Meiselman, and Carr (2010); Ng, Chaya, and Hort (2013); Jaeger and Hedderley (2013); King, Meiselman and Carr (2013); Jaeger, Cardello, and Schutz (2013); Piqueras-Fiszman and Jaeger (2014a); Piqueras-Fiszman and Jaeger (2014b); Chaya, Pacoud, Ng, and Hort (2015)). With considerable consumer input, emotion terms derived from pre-existing affective questionnaires were narrowed down to a final questionnaire of 39 terms which can be applied to a range of foods and beverages. The effectiveness of EsSense Profile® for differentiating emotional response both between and within product categories was demonstrated by King and Meiselman (2010) using both qualitative (checklist) and quantitative (rating) approaches.

The major advantage of using pre-determined emotion lexicons like EsSense Profile® for researchers is that such lexicons are general and, as such, can be applied to any group of products without the initial outlay of developing a product-specific lexicon. However, some emotion terms may be of little or no relevance to certain product categories, causing an already lengthy form to be longer than necessary and perhaps even confusing respondents (Jaeger et al., 2013). Ng, Chaya, and Hort (2013) reported six such redundant EsSense Profile® terms in the emotional assessment of a set of blackcurrant squashes. More significantly, terms may be excluded that are characteristic of the emotional response to certain product categories. A number of such omissions were identified by Ng et al. (2013) for their range of blackcurrant squashes (e.g. comforted, curious, disappointed). King et al.

(2010) noted that the exclusion of characteristic terms can be ameliorated by modifying or expanding the pre-determined list. Of course, this is associated with additional effort and expense for the researcher, negating somewhat the advantage of employing a pre-determined lexicon.

The alternative to using a pre-determined lexicon is to develop a consumer-led lexicon (e.g. Thomson, Crocker, and Marketo (2010); Ng et al. (2013); Manzocco, Rumignani and Lagazio (2013); Spinelli, Masi, Dinnella, Zoboli, and Monteleone (2014)). In response to products of interest to the researcher, consumers generate an emotional lexicon in their own words. This approach incurs increased costs in both time and resources as compared with pre-determined lexicons but has the advantage of excluding irrelevant terms, thus shortening the form and also removing potential consumer confusion (Jaeger et al., 2013). In addition, relevant terms are less likely to be left out, thereby increasing the discrimination ability of the lexicon (Ng et al., 2013). Approaches for generating consumer-led emotion lexicons have yet to become established, presenting the opportunity to further improve on previously published methods.

Recently, Ng et al. (2013) generated and used a consumer-led emotion lexicon to discriminate between the emotional response to 11 commercial blackcurrant squash products. Twenty-nine consumers generated their own individual lexicons in one-to-one interviews. The consumers then used check-all-that-apply (CATA) on their own personal list of terms to indicate their emotional response to all 11 products. Synonyms were combined and any terms checked by fewer than five consumers were excluded, giving a final lexicon of 36 terms. This approach was found to differentiate between the products based on their emotional profiles. However, one-to-one interviews were labour-intensive and the researchers recommended that small focus groups of subjects would be more efficient with the added benefit of enabling group discussion for deeper probing of consumer language. In addition, it was proposed that a quantitative rate-all-that-apply (RATA) approach would open up more opportunities for statistical analysis compared to the qualitative CATA approach. The present study implemented these suggestions by, firstly, conducting group interviews to generate a consumer-led emotion lexicon in order to increase efficiency and promote discussion. Secondly, consumers used RATA as opposed to CATA to record the emotional response to increase the capability for statistical analysis of the data.

A disadvantage of many verbal self-report approaches is that they require consumers to make a large number of evaluations per sample (e.g. 39 in EsSense Profile®; 36 in Ng et al. (2013)), leading to potential consumer fatigue and boredom. Such a large number of emotion terms can also make statistical product comparisons unwieldy. In order to allow an easier and quicker test for the respondent, Porcherot et al. (2010) developed a shorter version of the Geneva Emotion and Odor Scale (GEOS; Chrea et al. (2009)) questionnaire with a reduced number of measurement scales. GEOS consists of 68 affective terms which were reduced to a set of 6 summary scales through factor analysis. Porcherot et al. (2010) advanced this approach by having participants rate a series of 3 representative terms for each of the 6 GEOS dimensions instead of rating the 36 terms individually for each sample (ScentMove™). In spite of the fact that half of the evaluations were required as compared to the original form, similar product information was obtained by the GEOS and ScentMove™ questionnaires. The present study also takes this approach of reducing the number of consumer responses. However, instead of using factor analysis to reduce the terms to a set of summary scales, similar terms were grouped into emotion categories using cluster analysis. The aims of grouping terms were to reduce potential consumer fatigue and boredom and to increase the ease of subsequent product comparisons.

The effectiveness of any emotion lexicon is assessed by its ability to discriminate between the emotional responses to samples of interest. This validation is of particular importance when assessing the effectiveness of a reduced form because there is the potential to lose important emotional information. Therefore, the present study refers to the ability of emotion categories to discriminate between the selected samples to validate the reduced emotion form. Previous studies have implicated sensory properties in driving emotional response for a range of product categories (chocolate in Thomson et al. (2010); blackcurrant squashes in Ng et al. (2013); beer in Sester, Dacremont, Deroy, and Valentin (2013); chocolate and hazelnut spreads in Spinelli et al. (2014); coffee in Bhumiratana, Adhikari, and Chambers (2014)). However, these studies made use of commercial products with no experimental control of sensory properties and many associated potential sources of variation. An understanding of the direct relationship between products' sensory properties and emotional response is vital for the proposed applications of emotion methodologies for product development (King & Meiselman, 2010). Therefore, the present study exerted a degree of experimental control by manipulating individual sensory properties of commercial lagers. In using samples representing a wide range of sensory properties of beer, the

groupings presented here were intended to cover the whole emotional space of this product category, with no redundancy. The success of the reduced consumer-led lexicon was judged on its ability to discriminate between the emotional responses elicited by each sample based on their differences in selected sensory properties.

As well as discriminating between samples, a successful emotional lexicon should permit other investigations around emotional response. Studying the emotional response to beer as a product category is of particular interest when considering differences between consumer groups. Males are more frequent beers consumers than females (Mintel (2013); Serra and Aranceta (2003)). There is also a decline in the frequency of beer consumption with age (Mintel (2013); Serra and Aranceta (2003)). Perhaps differences in emotional response to the sensory properties of beer can go some way towards explaining this. Therefore, a further source of validation for the reduced consumer led lexicon was to assess the effectiveness of this approach for investigating differences in emotional response between genders and age groups. In addition, emotions have previously been found to discriminate beyond liking (Ng et al., 2013) and this study was particularly focussed towards exploring if this increased discriminability of emotions over liking could be extended to differences between consumer groups. Familiarity has also been found to have an important bearing on consumer experience (Sester et al., 2013). It was anticipated that there may be a particular effect of familiarity between consumer groups in their reported emotional responses.

The main objectives of the study presented here were twofold: (1) to create an approach for the development of a product category-specific reduced consumer-led lexicon utilising group interviews and cluster analysis, and (2) to validate the use of a reduced consumer-led lexicon by evaluating its ability to (a) discriminate across a range of beer samples specifically designed to elicit specific sensory properties, and (b) reveal differences in emotional response across different consumer segments related to gender and age.

2 Material and methods

With reference to 10 lager samples that varied in selected sensory properties, small focus groups of consumers generated emotion terms to describe their responses to the 10 beers. These subjects then used this lexicon to rate their emotional response to the samples and their responses were subjected to linguistic checks and cluster analysis, allowing the grouping of similar terms into distinct emotion categories and forming a reduced lexicon.

This reduced lexicon was then used by over 100 beer consumers to rate the same 10 samples and this data was used to validate the effectiveness of the reduced consumer-led lexicon in discriminating emotional responses between different sensory properties of beer.

2.1 Samples

Ten lager samples were included in this study (Table 1). Two samples were 'control' commercial beer samples, one of which was a commercial non-alcoholic beer. The other eight samples were based upon the control samples and were each manipulated in a single sensory property. Several sensory properties were chosen to represent key characteristic properties of beer (e.g. bitterness, hoppiness, etc.) whilst others reflected off-flavours and/or hypothesised drivers of emotional response (e.g. isoamyl acetate, dimethyl sulphide (DMS)). The commercial beers were modified using ethanol (Merck Chemicals Ltd, UK), dextrose (Myprotein, UK), specific Aroxa flavour capsules (Cara Technology, UK), or controlled decarbonation. All 10 samples had been evaluated by the University of Nottingham's trained expert beer panel who had rated each sample for the 8 sensory properties of interest (data not shown). The assessments revealed significant differences between the spiked and control samples, indicating that the samples differed in their sensory properties, at least for a trained panel. From these results, it was anticipated that subjects in subsequent studies would perceive the differences in relevant sensory properties across samples.

Samples were prepared by adding the relevant materials to samples and 10ml decanted into transparent closed screw cap universal containers 2-4 hours before assessment by consumers. This was with the exception of the Low CO_2 samples which was decarbonated by leaving open and refrigerated ($4\pm1°C$) for 3 hours before re-sealing. Low CO_2 samples were then decanted just prior to consumer assessment. Products were presented blind (labeled with three digit random codes) at $4\pm1°C$. Unsalted crackers (Carrefour, Spain) and mineral water (Fuente Liviana, Spain) were provided as palate cleansers.

2.2 Lexicon development

Focus groups of consumers generated an emotional lexicon using their own words to describe their emotional responses to the 10 samples. These subjects subsequently used this lexicon to rate all 10 samples and this data was submitted to cluster analysis and linguistic

Table 1. Ten beer samples and their related treatments[1].

	Sample	Treatment
1	Control	Commercial lager
2	Hoppy	0.75mg Aroxa kettle hop extract/litre commercial lager
3	Light struck	0.3µg Aroxa 3-methyl-2-butene-1-thiol/litre commercial lager
4	Isoamyl acetate	10.5mg Aroxa isoamyl acetate/litre commercial lager
5	DMS	0.9mg Aroxa dimethyl sulphide capsules/litre commercial lager
6	Bitter	25mg Aroxa iso-α-acids/litre commercial lager
7	Sweet	25g dextrose/litre commercial lager
8	Low CO_2	Commercial lager decarbonated to ~1.6 units
9	Non-alcohol control	Commercial non-alcohol lager
10	High alcohol	96% ethanol added to commercial non-alcohol lager (8% ABV)

[1]Treatment refers to the manipulation to the base beer to control the relevant sensory property.

checks in order to group similar terms into emotion categories. This section describes this process in detail.

2.2.1 Subjects

Seventeen reasonably articulate Spanish consumers (aged 18-60 years), who consumed beer at least once per month, took part in this study after signing consent forms in line with local ethical procedures. As women have previously been suggested to be more adept with emotional language (Fugate, Gouzoules, & Barrett, 2009), the majority of the subjects recruited were female (70%) to facilitate term generation although males were included to

ensure relevant terms from both genders were included. Participants were divided into three groups of 5-7 subjects and attended a total of four 90 minute-2 hour sessions.

2.2.2 Procedure for the development of a reduced consumer-led lexicon

Sessions 1, 2 and 3 concerned the explanation of 'emotion', warm-up exercises and the generation of emotion terms in response to the beer samples. At the start of the first session, consumers received a short explanation of the meaning of 'emotion' based on the following reasoning. According to King and Meiselman (2010), one can distinguish at least three different affective behaviours: (1) attitudes, which include an evaluation component, e.g. "I like beer"; (2) moods, which are more enduring, build up gradually, are more diffuse, and not focused on a referent, e.g. "I'm happy" (3) emotions, which are brief, intense, and focused on a referent, e.g. "This comment makes me feel angry". Therefore, whilst consumers could precede all three with "I feel..." (necessary for inclusion in the lexicon), they were encouraged to focus their elicitation on the third type of affective behaviour. In order to make this distinction clear, warm-up exercises were carried out with reference to pictures and prompt cards following Ng et al. (2013).

Drawing on the experience of previous emotion research (Ng, 2013), participants were presented with warm-up samples of the 2 'base' control beers before generating terms in session 1 and 2 in order to both contextualise the beer and aid participants in considering the differences in their emotional response between the presented samples (as opposed to their response to beer *per se*).

Triadic elicitation (Fransella, Bell, & Bannister, 2004) was used to generate terms. That is to say, participants were asked to assess triads of samples and describe 'in what way two samples were similar but different from the third in terms of the emotional response they elicited'. After performing this task individually, the participants shared and discussed their response with the group and a consensus between the members was reached. To ensure that each sample appeared at least once for each group, four triads were presented to each group (two samples appeared more than once for each group). The samples included in each triad were selected to be different for each group (and different samples appeared more than once for each group).

A total of 80 terms were generated by all groups in this elicitation phase. An initial reduction of terms was performed with a convenience subset of participants. They indicated

their perceived meaning of terms which were ambiguous in whether they described emotional experiences or the samples themselves (e.g. bland, unappealing). Where possible, synonymous terms were combined using a thesaurus (Microsoft Word 2007), resulting in a condensed list of 54 terms.

This list of terms was then used by the subjects to rate their emotional response to all 10 samples during the fourth and final session. The question asked, translated into English, was 'Please taste the sample and rate the intensity of the following evoked feeling'. Each of the 54 terms was associated with a 150mm line scale, anchored at 15 mm from the line ends from 'very low' to 'very high' and responses were recorded using Fizz Forms (Biosystèmes, Couternon, France). Subjects were also asked 'How much do you like this sample?' and 'Is this sample familiar to you?' using similar line scales anchored from very little to very much. These responses were subsequently expressed as a percentage distance along the line scale, i.e. 0-100%. Emotions were presented in a randomised order (emotion list order has been found to affect consumer responses (King, Meiselman, & Carr, 2013). Having rated all 10 samples and assuming subjects had become familiar with the lexicon, they were also asked to rate each term for relevance in describing emotions elicited by beer as a product category *per se* using the question 'In your opinion, are the following terms relevant to describe how do you feel while consuming a beer?', with a line scale anchored from absolutely irrelevant to absolutely relevant. This element of the questionnaire was included to provide additional data to help in the elimination of redundant terms which may have been generated in efforts to appease the researcher in their persistent requests for more terms to describe participants' emotional experiences of the samples. Again, consumers used 150mm line scales but this time anchored from 'not relevant at all' to 'extremely relevant'. This session was carried out in an air conditioned room (21±1ºC), under Northern Hemisphere daylight lighting.

2.2.3 Grouping of terms into emotion categories

Ten terms with a mean 'relevance' score of less than 33% (i.e. less than one third of the scale) were excluded as being evaluated as not very relevant to beer. These included relieved, distressed, tired, sickly, infantile, afraid, suspicious, embittered, sad, and empty.

The next stage was to group similar terms into emotion categories. A number of multivariate statistical techniques (factor analysis, principal components analysis, hierarchical cluster

analysis) were applied to the mean ratings of samples for the remaining 44 terms (Table 2) in order to ascertain the relative effectiveness of each technique for grouping terms which produced similar patterns of data. Cluster analysis was deemed to do this in the most practical way for this research because there is some degree of control at the hands of the researcher in deciding how many clusters are appropriate. Only the results of the cluster analysis approach are presented in this paper for brevity. Terms were segmented using Euclidean distances and Ward's criterion of aggregation (XLSTAT Version 2009.6.03, Addinsoft, USA). The coefficient, Cronbach's α, was calculated in order to assess internal consistency of clusters.

Table 2. Final lexicon of 44 terms translated into English (with original Spanish terms shown in italics).

Agreeable	*Conforme*	Enjoyment	*Divertido*	Nostalgic	*Nostálgico*
Annoyed	*Contrariado*	Excited	*Emocionado*	Objectionable	*Indeseable*
Appetised	*Apetecible*	Festive	*Festivo*	Pleasant	*Placentero*
Authentic	*Auténtico*	Fresh	*Fresco*	Positive	*Positivo*
Bad	*Mal*	Friendly	*Amistoso*	Relaxed	*Relajado/tranquilo*
Bored	*Aburrido*	Happy/	*Alegre/chispeante/*	Repulsed	*Reacio/rechazo/*
Cheated	*Engañado*	cheerful	*contento*		*repulsión*
Classic	*Clásico*	Indifferent	*Indiferente*	Satisfied	*Satisfecho*
Curious	*Curioso*	Intense	*Intenso*	Shocked	*Sorprendido/*
Different	*Diferente*	Lacking in	*Inapetente*		*inesperado*
Disappointed	*Decepcionado*	appetite		Strong/powerful	*Fuerte/potente*
Disgusted	*Asqueado/*	Lively	*Animado*	Traditional	*Tradicional*
	disgustado	Mild	*Ligero/suave/ flojo*	Uncomfortable	*Incómodo*
Disillusioned/	*Desilusionado/*	Natural	*Natural*	Unmotivated	*Desmotivado*
disenchanted	*desencantado*	Negative	*Negativo*	Unpleasant	*Desagradable*
Dissatisfied	*Insatisfecho*	Nice	*Agradable*	Unusual	*Extrañado/raro/*
Eager	*Ansioso*	Normal	*Esperado/normal*		*atípico*

To aid the grouping of terms, all participants were asked to identify the meaning of ambiguous generated terms by indicating their interpretation of the word through the use of a thesaurus. This was found by the researchers to be particularly relevant for the terms 'emocionado' (which could be interpreted as excited or moved) and 'ansioso' (which could be interpreted with positive connotations as eager/desirous or with negative connotations as anxious). Their responses were subsequently very useful when defining emotion categories.

Most consumers associated 'emocionado' with excitement and 'ansioso' with eagerness/desire.

Eight clusters of terms were identified by the initial cluster analysis (Table 3) and the associated dendrogram is presented in Figure 1. However, it was perceived by the authors that some of the clusters were unclear and could potentially cause confusion, an observation supported by a low Cronbach's α associated with many clusters. To reduce confusion and increase internal consistency, a number of clusters were modified by the authors. For example, Cluster 1 (mild, bored, and indifferent) had a very low Cronbach's α of just 0.44. Based on the authors' discussions with participants, particularly at the elicitation phase, mild was deemed to be distinct in meaning from bored/indifferent. Therefore, Cluster 1 was split into two categories: 'Mildness' (including just mild) and 'Indifference' (including bored and indifferent). Cluster 2 was judged to include too many terms and as a result was split into two new categories: 'Pleasure' (including positive, pleasant, relaxed, satisfied, etc.) and 'Classic' (including authentic, natural, traditional, etc.). A particular reason for this distinction was that 'Classic' contained terms that could be deemed as more 'abstract' than emotional. Consumers felt very strongly during the elicitation phase that such terms should be included in the lexicon because they were important for differentiating their feelings towards the samples presented. Finally, Cluster 8 (Cronbach's α = 0.34) was split into two categories: 'Excited' and 'Nostalgic' because these terms were used to describe different emotions by the participants who had generated them.

Table 3. Cluster analysis of 44 terms grouped into 8 clusters with associated Cronbach's αs (translated into English).

| Cluster 1 (Cronbach's α = 0.44)	Cluster 2 (Cronbach's α = 0.97)	Cluster 3 (Cronbach's α = 0.85)	Cluster 4	Cluster 5 (Cronbach's α = 0.97)	Cluster 6 (Cronbach's α = 0.89)	Cluster 7 (Cronbach's α = 0.85)	Cluster 8 (Cronbach's α = 0.34)
Bored Indifferent Mild	Nice Friendly Appetised Authentic Classic Agreeable Normal Fresh Natural Pleasant Positive Relaxed Satisfied Traditional	Happy/ cheerful Lively Curious Enjoyment Festive	Eager	Disgusted Annoyed Disappointed Unpleasant Unmotivated Cheated Uncomfortable Objectionable Bad Negative Repulsed Shocked	Disillusioned/ disenchanted Unusual Lacking in appetite Dissatisfied	Strong/ powerful Intense Different	Excited Nostalgic
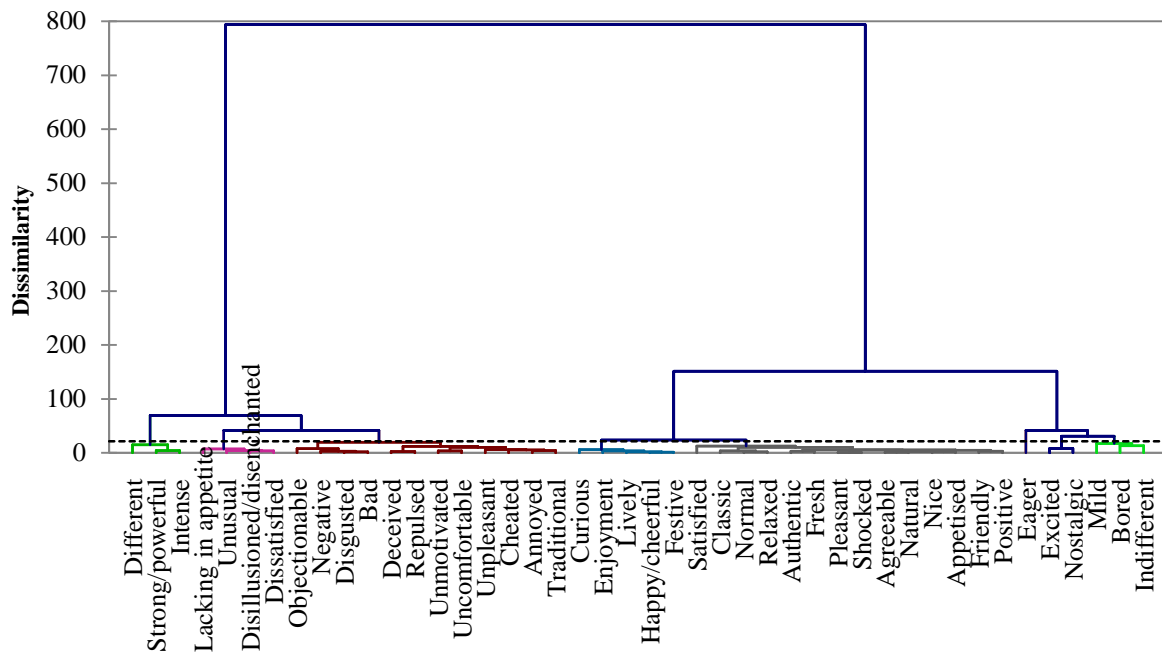
Figure 1. Dendrogram of the 8 solution cluster analysis based on mean scores of focus group subjects (n = 17) rating xx emotion terms across 10 beer samples

The revised grouping resulted in a total of 12 clearly defined emotion categories (Table 4). For each modification, the new Cronbach's α was higher than calculated from the initial cluster analysis results. For 11 of the 12 emotion categories, Cronbach's α indicated adequate internal consistency (Cronbach's α>0.8; Streiner (2003)). The exception to this was the category 'Indifferent'. As the category contains just two terms, a low Cronbach's α is to be expected as the coefficient is greatly affected by the number of items (Streiner, 2003).

It is important to further discuss two emotion categories in particular: Intensity and Mildness. The terms belonging to these emotion categories were not used by consumers to describe their general judgement of emotions. In relation to the previous literature, the Intensity emotion category is similar to "energetic" in Spinelli et al. (2014) or the "Energy" dimension in Ferdenzi et al. (2011) (based on the terms refreshed, revitalised, and energetic in Singapore and, clean, revitalized, rejuvenated, stimulated, energetic in Liverpool). The emotion category Mildness was associated with the opposite, i.e. a lack of a feeling of vigour or energy. To summarise this point, when responding using the Intensity or Mildness emotion categories, consumers were not referring to their overall emotional intensity (or lack thereof) but instead specific qualities of emotion. The terms listed within both categories helped the consumer to identify the expected nature of feelings.

2.3 Consumer emotional response

In the second part of the study, a large number of naïve consumers rated their emotional response to the 10 beer samples using the 12 emotion categories.

2.3.1 Subjects

One hundred and thirteen subjects (52% male) who consumed beer at least once per month took part in this study. Approximately two-thirds were aged 18-34 (68%), with the remaining third aged 35+. Consumers were weighted towards the younger age group because beer consumption peaks before consumers reach 35 years of age (Mintel (2013); Serra and Aranceta (2003)).

2.3.2 Procedure

When rating their emotional responses, consumers were presented with 12 continuous line scales. Each scale was associated with an emotion category. Emotion categories were presented as a list of terms that belonged to each category. After tasting the sample, consumers were instructed to read all of the terms associated with each emotion category and to rate (on a 150mm continuous line scale anchored at 5mm from the line ends from 'very low' to 'very high') the overall intensity of their feeling of the underlying emotion that the words described. The order of emotion categories was randomised for each consumer, although this order remained consistent within each consumer for every sample. Once the 12 emotion categories were rated, consumers then rated the sample for both liking and familiarity on 2 additional 150mm line scales. The inclusion of liking allowed a comparison between traditional hedonic measures and emotional response in order to observe if emotions discriminate beyond liking as has been found previously (Ng et al., 2013). Familiarity was included in the form to add supplementary data for the purposes of interpretation of consumer response. It has been found that familiarity has an important bearing on consumer experience (Sester et al., 2013) and it was anticipated that there may be a particular effect of familiarity between consumer groups in their reported emotional responses.

A 'dummy' sample was presented in the first position for all consumers to overcome first-order effects (Macfie, Bratchell, Greenhoff, and Vallis (1989); Lawless and Heymann (2010); Dorado, Picard, Pérez-Hugalde and Chaya (2014)) and familiarise consumers with the task. This data was subsequently discarded. This first 'dummy' sample was always the

1 **Table 4.** Final grouping of 44 terms into 12 clusters with associated Cronbach's αs (translated into English).

Mildness	Indifference (Cronbach's α = 0.55)	Pleasure (Cronbach's α = 0.97)	Classic (Cronbach's α= 0.93)	Fun (Cronbach's α = 0.85)	Desire	Disgust/ Negative feelings (Cronbach's α = 0.97)	Disillusionment (Cronbach's α = 0.89)	Disappointment (Cronbach's α = 0.87)	Intensity (Cronbach's α = 0.85)	Nostalgia	Excitement
Mild	Bored Indifferent	Nice Friendly Appetised Agreeable Fresh Pleasant Positive Relaxed Satisfied	Authentic Classic Natural Normal Traditional	Happy/ cheerful Lively Curious Enjoyment Festive	Eager	Disgusted Annoyed Unpleasant Unmotivated Cheated Objectionable Bad Negative Repulsed Shocked	Disillusioned/ disenchanted Unusual Lacking in appetite	Disappointed Uncomfortable Dissatisfied	Strong/ powerful Intense Different	Nostalgic	Excited

Control sample, although it must be noted that data for this sample was also collected in the main data collection phase. The presentation order of the 10 samples that followed the 'dummy' sample was randomly assigned for each consumer. Samples were assessed in an air conditioned room (21±1ºC) under Northern Hemisphere daylight lighting. Unsalted crackers (Carrefour, Spain) and mineral water (Fuente Liviana, Spain) were provided as palate cleansers between sample assessments. After evaluating the first set of five 10ml samples, consumers were given a 20 minute break after which they evaluated the remaining samples. Availability of assessors meant that samples needed to be assessed in one session which also limited ethically the amount of sample that could be consumed.

2.3.3 Data analysis

Multivariate analysis was performed using principal components analysis (PCA) on mean ratings of every emotion category for each sample and in order to map out the emotional space of the ten samples (XLSTAT, Version 2009.6.03). The liking and familiarity data were included in the PCA as supplementary variables to determine their relationships with the emotional data. Mixed model analysis of variance (ANOVA) was carried out for each emotion category as well as liking and familiarity with sample as a fixed factor and subject as a random factor (SPSS Statistics 22, IBM, USA). Tukey's HSD was applied where significant effects of sample were found in order to ascertain how each emotion category discriminated between samples. Further ANOVAs were carried out with fixed effects of sample, gender and age group for each emotion category. Interactions between sample*gender and sample*age group were explored in order to investigate the effects of individual sensory properties on consumer group ratings of emotion categories.

3 Results

Group interviews and cluster analysis were successfully utilised to develop a reduced consumer-led lexicon which was applied to a set of 10 controlled beer samples. The following section demonstrates the validity of the lexicon through its creation of a discriminating emotional space across samples and between genders and age groups.

3.1 Emotional space

The PCA enabled the visualisation of the emotional space for the samples tested. The first 2 principal components accounted for 95.34% of the data variance (Figure 2a). PC1 (72.65%)

was highly positively correlated with emotion categories Disgust, Disillusionment and Disappointment and negatively correlated with Pleasure, Classic, Fun, Desire, Nostalgia and Excitement. Liking and familiarity were not active in the PCA but were highly and negatively correlated with PC1. Mildness and Indifference were positively correlated with PC2 (22.69%) whilst Intensity was negatively correlated. This emotional space was consistent with circumplex models of emotion (Russell (1980), Watson and Tellegen (1985), Larsen and Diener (1992)) with PC1 associated with pleasure/pleasantness and PC2 related to engagement/activation. Such an emotional space has been previously observed in other food and beverage categories (Chrea et al. (2009); Porcherot et al. (2010); Ng et al. (2013); Chaya et al. (2015)). It is worth also noting that the pairs of terms Mild and Indifference and also Nostalgia and Excitement are located quite closely on the PCA indicating that the authors splitting of these terms for the final emotion categories may have been unnecessary as the pattern of response in the mean data appears to be similar. However, while this is true in general terms for the PCA, which is in fact a descriptive method on the mean data, results on mixed ANOVA and Tukey's HSD groupings based on
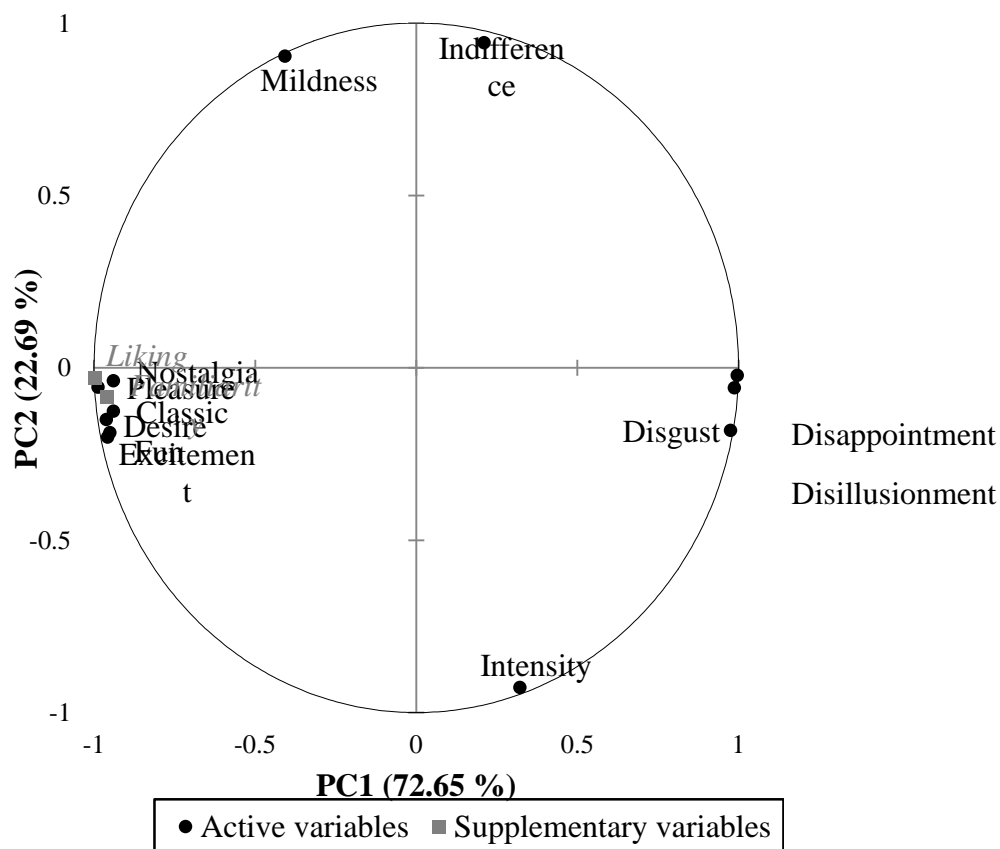

Figure 2a. PCA correlation circle of the emotion categories (PC2 vs PC1).1 1Based on 12 emotion categories means (n = 113) across 10 beer samples.

individual consumers' data show different conclusions. Although Nostalgia and Excitement categories discriminate similarly the beer samples, different groupings were found between Mildness and Indifference: e.g., Control is different from Isoamyl acetate for Indifference but not for Mildness.

When plotting the sample positions in the emotional space (Figure 2b), it was observed that the Non-alcohol control and High alcohol samples were projected highly positively onto PC1 (with Hoppy also loading less highly in this direction), therefore associated with displeasure/unpleasantness. In contrast, the Control was highly negatively correlated with PC1 (with Low CO_2 and Light struck samples also projected somewhat in this direction)
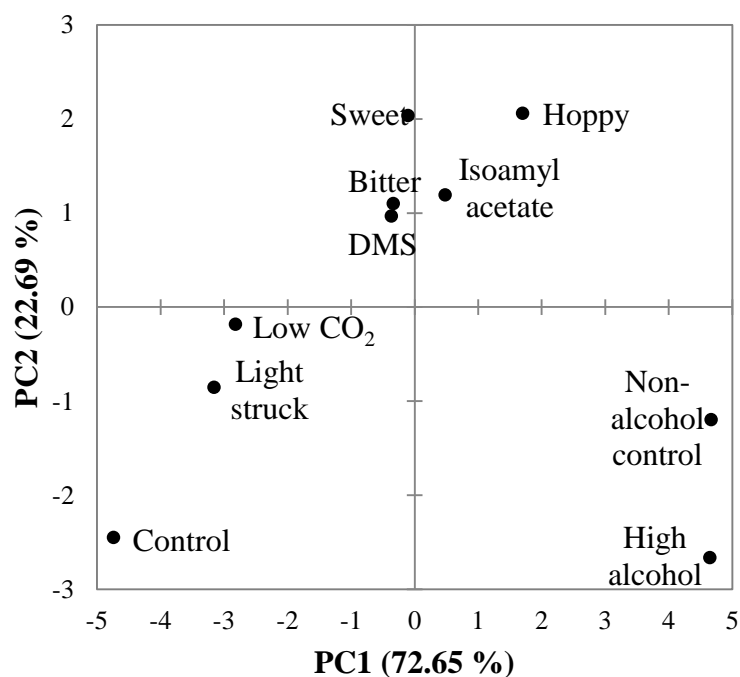

Figure 2b. PCA product plot (PC2 vs PC1).1 1Based on 12 emotion categories means (n = 113) across 10 beer samples.

and, as such, was associated with pleasure/pleasantness. Both the Control sample and the High alcohol sample were projected negatively onto PC2, showing an association with engagement/activation. In contrast, Hoppy and Sweet (and to a lesser extent Bitter, DMS, and Isoamyl acetate) samples were highly positively correlated with PC2 and unengaging emotions.

3.2 Discrimination ability of emotion categories between samples

Mixed ANOVA indicated that all emotion categories discriminated between samples (Table 5) and the subsequent post hoc analyses identifying differences between samples are shown in Table 6. Comparisons between these post hoc groupings highlighted patterns of sample groupings which, unsurprisingly, related to how emotion categories and samples loaded onto the two dimensions identified by the PCA. This offered a useful guide for comparing and contrasting the discrimination ability of individual emotion categories. The following highlights some of these key differences.

The three emotion categories identified as unpleasant by PC1 (Disappointment, Disgust, and Disillusionment; Figure 2a) overlapped considerably in their sample groupings (Table 6). The Non-alcohol control and High alcohol samples were rated significantly higher for these emotions than most others, but were not significantly different to one other. This is clearly shown on PC1 of the PCA (Figure 2b). The Hoppy sample loaded in the same direction on PC1 as these two samples but not as highly and hence was not so unpleasant. Nevertheless, significant differences were found between the Hoppy sample and the most pleasant samples. The Control and Light struck samples were rated significantly lower in unpleasant emotions than the Hoppy sample. In addition, the Low CO_2 sample was rated significantly lower in Disillusionment than the Hoppy sample. The isoamyl acetate sample was rated

Table 5. *p*-values for main effects of sample, gender, and age group, and interactions between sample*gender and sample*age group for each emotion category (and liking and familiarity).

Emotion category	Sample	Gender	Sample* Gender	Age group	Sample* Age group
1 Mildness	**<0.001**	**0.007**	0.841	**0.002**	0.085
2 Indifference	**<0.001**	**0.001**	0.306	**0.004**	0.903
3 Pleasure	**<0.001**	0.406	**0.030**	**<0.001**	0.284
4 Classic	**<0.001**	**0.004**	0.322	0.475	**0.028**
5 Fun	**<0.001**	0.416	0.137	**0.001**	0.198
6 Desire	**<0.001**	**0.032**	0.258	**0.012**	0.251
7 Disgust	**<0.001**	0.795	**0.016**	**0.008**	0.109
8 Disillusionment	**<0.001**	0.180	0.207	**<0.001**	0.140
9 Disappointment	**<0.001**	**0.034**	0.063	**<0.001**	**0.041**
10 Intensity	**<0.001**	**0.044**	0.636	**0.005**	0.129
11 Nostalgia	**<0.001**	**<0.001**	0.599	0.257	**0.017**
12 Excitement	**<0.001**	0.186	0.659	**0.001**	0.131
Liking	**<0.001**	0.529	0.134	**0.003**	**0.031**
Familiarity	**<0.001**	**0.006**	0.220	**0.003**	**0.002**

Emboldened *p*-values represent statistical significance ($p < 0.05$)

significantly higher than the Control in all 3 unpleasant emotion categories, whilst just Disappointment and Disillusionment showed Bitter and Sweet samples to receive significantly higher ratings than the Control. Disillusionment also revealed that DMS was rated as significantly more disillusioning than the Control. The Low CO_2 and Light struck samples were located in the same quadrant of the PCA as the Control and were the only samples found to not differ significantly to the Control for all 3 unpleasant emotion categories.

Several emotion categories loaded highly negatively on PC1 (Pleasure, Fun, Desire, Excitement, Classic, and Nostalgia; Figure 2a) and grouped samples similarly when considering post hoc tests (Table 6). As with the unpleasant emotion categories, none of the pleasant emotion categories differentiated between Non-alcohol control and High alcohol samples. The Control was rated significantly higher for Pleasure, Fun, Desire, Classic and Excitement than all other samples except Light struck and Low CO_2 (and also Bitter for Classic). This was evident in the PCA where only Light struck and Low CO_2 samples were located in the same quadrant as the Control. Other smaller differences between samples were shown by Pleasure, Fun, Desire, Classic, and Excitement. For example, Fun did not discriminate between the High alcohol and Bitter samples but Pleasure, Desire, Classic, and Excitement did. Desire did not discriminate between High alcohol and Isoamyl acetate samples but the other three emotion categories did. Nostalgia was markedly less discriminating than the other five pleasant emotion categories, indicating that adjusting the Control's sensory properties did not show any significant effect. Spinelli et al. (2014) identified that Italian consumers struggled in applying the term equivalent to 'nostalgic' to food and perhaps this was also the case here. However, consumers did discriminate between samples that were particularly opposed on PC1 (e.g. the High alcohol sample was rated significantly lower for Nostalgia than the Control, Light struck and Low CO_2 samples).

Turning attention to emotion categories correlated with PC2 (Figure 2a), low activation/engagement/ Mildness and Indifference were rated significantly higher compared to the Control for the Sweet, DMS, Hoppy, and Bitter samples (and also Isoamyl acetate in Indifference; Table 6). These samples can be seen to load positively (low activation/engagement) on PC2 and oppose the highly negatively loading (high activation/engagement) Control sample (Figure 2b). Intensity showed an inverse correlation with Mildness and Indifference as would be expected because it loads in the opposite

Table 6. Mean scores (%) for the 12 emotion categories (and liking and familiarity) across 10 samples.

Emotion category	Sample									
	Control	Hoppy	Light struck	Isoamyl acetate	DMS	Bitter	Sweet	Low CO_2	Non-alcohol control	High alcohol
1 Mildness	40.6 BC	53.4 DE	45.1 CD	50.0 CDE	52.4 DE	52.1 DE	57.4 E	47.9 CDE	33.2 AB	29.8 A
2 Indifference	30.0 A	43.8 C	34.9 ABC	44.0 C	42.1 BC	40.5 BC	41.5 BC	37.6 ABC	36.7 ABC	32.8 AB
3 Pleasure	59.0 E	40.9 ABC	52.7 DE	44.9 CD	44.1 BCD	44.3 CD	45.8 CD	51.6 DE	31.6 A	34.7 AB
4 Classic	52.7 E	31.7 AB	51.6 E	35.4 B	40.5 BCD	45.4 CDE	37.2 BC	49.0 DE	32.5 AB	26.3 A
5 Fun	54.5 D	35.9 AB	49.0 CD	43.3 BC	41.4 BC	39.6 AB	42.4 BC	49.1 CD	30.7 A	35.8 AB
6 Desire	52.2 D	33.8 AB	43.3 CD	38.8 ABC	40.0 BC	40.9 BC	39.5 BC	47.4 CD	29.6 A	32.5 AB
7 Disgust	26.7 A	40.6 B	28.7 A	40.4 B	37.3 AB	36.7 AB	34.8 AB	30.3 AB	51.7 C	55.6 C
8 Disillusionment	28.0 A	43.4 CDE	31.7 AB	41.6 BCD	39.1 BC	39.3 BC	40.5 BC	32.1 AB	51.6 DE	53.3 E
9 Disappointment	27.9 A	43.8 CD	30.6 AB	40.3 BC	36.9 ABC	39.2 BC	39.4 BC	35.2 ABC	53.2 D	52.5 D
10 Intensity	54.8 D	37.5 AB	48.5 CD	43.4 BC	45.6 BC	40.0 ABC	33.0 A	45.1 BC	57.5 D	66.8 E
11 Nostalgia	38.0 C	32.1 ABC	38.2 C	30.9 ABC	33.6 ABC	33.2 ABC	30.6 ABC	36.9 BC	28.9 AB	26.0 A
12 Excitement	51.9 D	34.6 AB	46.1 CD	40.3 BC	42.0 BC	39.3 BC	39.4 BC	45.2 CD	30.1 A	34.7 AB
Liking	60.5 E	40.1 AB	53.3 CDE	44.1 BC	46.4 BCD	46.4 BCD	45.3 BCD	54.4 DE	30.7 A	32.6 A
Familiarity	55.4 F	29.0 B	49.4 EF	33.0 BC	38.8 CD	43.9 DE	33.6 BC	49.9 EF	26.2 AB	18.8 A

ABCDEF: Letters within the same row indicate post hoc groupings by Tukey's HSD ($p < 0.05$)

direction on PC2 (high activation/engagement; Figure 2a). Therefore, Sweet, DMS, Hoppy, Bitter, and Isoamyl acetate samples were rated significantly lower than Control for Intensity. The Low CO_2 sample also received significantly lower ratings of Intensity compared to the Control. In addition, Intensity was able to discriminate between the Non-alcohol control and High alcohol samples with the high alcohol been rated higher for Intensity. These findings demonstrate that Intensity was a particularly discriminating emotion category as it was the only emotion category to demonstrate a change in emotional response associated with decreased and increased alcohol content.

Drawing together results from across emotion categories, the reduced consumer-led lexicon revealed an individual emotional profile for almost all samples. The Control sample (Figure 3a) scored very low in unpleasant emotions (Disgust, Disillusionment, Disappointment) and relatively high in a number of pleasant emotions (e.g. Fun, Excitement). However, exaggerating other sensory properties, for example hoppiness, was shown to generally increase ratings of negative emotion categories and decrease ratings of positive emotion categories (Figure 3a). Just the Light struck sample was shown to have no significant emotional effects as compared to the Control (Figure 3a) although the samples were
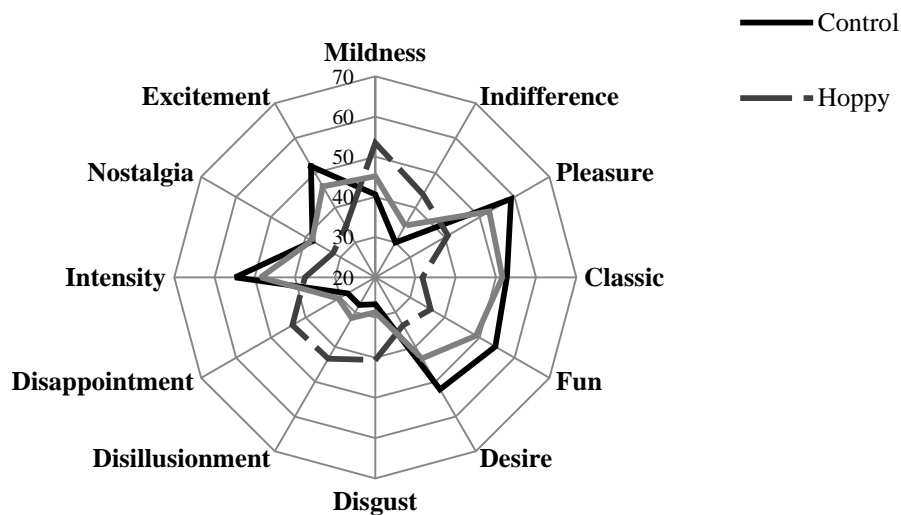

Figure 3a. Spider plot showing mean scores of all 12 emotion categories for Control, Hoppy, and Light struck samples. Emboldened emotions denote significantly different ratings between Control and Hoppy samples according to Tukey's HSD (p < 0.05). There were no significant differences between Control and Light struck samples.

sensorially different (see section 2.1). In a couple of further instances, only Intensity was able to discriminate between samples. For example, the Non-alcohol control and High alcohol samples were rated non-significantly different for 11 emotion categories but were significantly different in Intensity (Figure 3b).

3.3 Consumer group effects

The next section further validates the reduced consumer-led emotion lexicon by showing how it was able to reveal differences in emotional response across consumer segments (gender and age group).

3.3.1 Gender

A significant main effect of gender was found for emotion categories Classic, Desire, Disappointment, Indifference, Intensity, Mildness, and Nostalgia (Table 5). In all except Intensity, women gave significantly lower ratings than men. This result supports what was found by King and Meiselman (2010). These authors found that on average females rate emotion intensities stronger than males: however, this pattern is product specific and is reversed for some products. Interestingly, such gender differences were not apparent in liking. However, significantly lower ratings of familiarity were also given by women, which may have had some bearing on their emotion ratings (particularly Nostalgia and Classic).

There were significant interactions between sample and gender for emotion categories Disgust (Figure 4a) and Pleasure (Figure 4b). The key gender differences were driven by the Non-alcohol control and High alcohol samples. Although generally similar ratings were obtained for Disgust, simple main effects of gender on sample rating showed that the High alcohol sample was rated significantly higher in this emotion by women than men. Pleasure was also rated similarly for most samples between genders, except for the Non-alcohol control and High alcohol samples which females rated significantly lower in Pleasure than males according to simple main effects analyses.

3.3.2 Age group

A significant effect of age group was found for emotion categories Desire, Disappointment, Disgust, Disillusionment, Excitement, Fun, Indifference, Intensity, Mildness, and Pleasure,
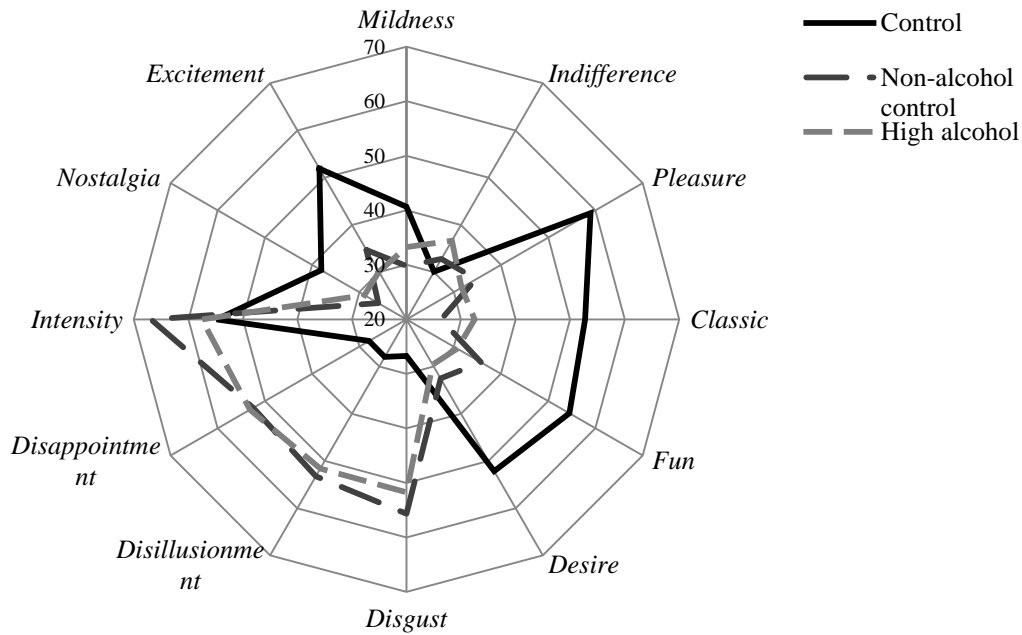
Figure 3b. Spider plot showing mean scores of all 12 emotion categories for Non-alcohol control and High alcohol samples. Emboldened emotions denote significantly differentratings between Non-alcohol control and High alcohol samples according to Tukey's HSD (p < 0.05). Italicised emotions denote significantly different ratings between High alcohol and Control samples according to Tukey's HSD. Underlined emotions denote significantly different ratings between Non-alcohol control and Control samples according to Tukey's HSD.

as well as liking and familiarity (Table 5). On the whole, the 35+ age group evaluated the neutral, unpleasant, and low engagement emotion categories (Disappointment, Disgust, Disillusionment, Indifference, Mildness) higher than the 18-34 group. Conversely, ratings for pleasant and high engagement emotion categories (Desire, Excitement, Pleasure, Intensity) as well as liking and familiarity were higher for the younger age group.

Significant interactions were found between sample and age group for Classic (Figure 5a), Disappointment (Figure 5b) and Nostalgia (Figure 5c) (also liking and familiarity). On the whole, simple main effects analysis showed non-significant effects of age group on ratings of Classic. However, significant differences were found for Non-alcohol control and High alcohol samples which were rated higher in Classic by 35+ year old consumers than the younger age group. In contrast, the Bitter sample was rated significantly higher in Classic by18-34 year old consumers. With regard to the Disappointment emotion category, simple main effects analysis showed that four samples (Control, Bitter, Sweet, Low CO_2) were rated significantly higher by 35+ year old consumers. As would be expected, the inverse
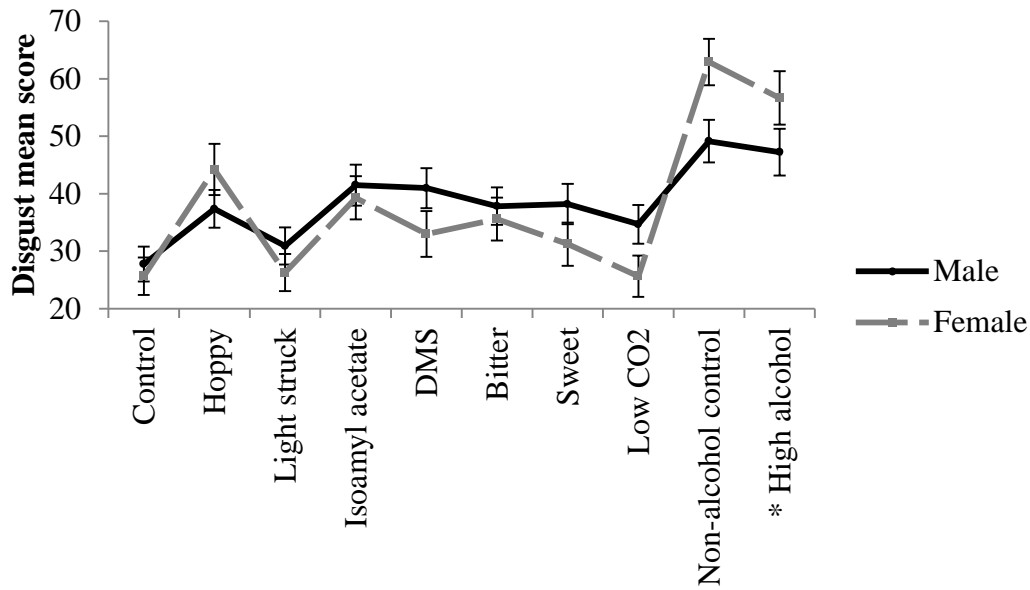
Figure 4a. Mean ratings (and SEM) of Disgust for each sample by gender. *denotes a significant difference (p < 0.05) in the rating of disgust between gender groups.
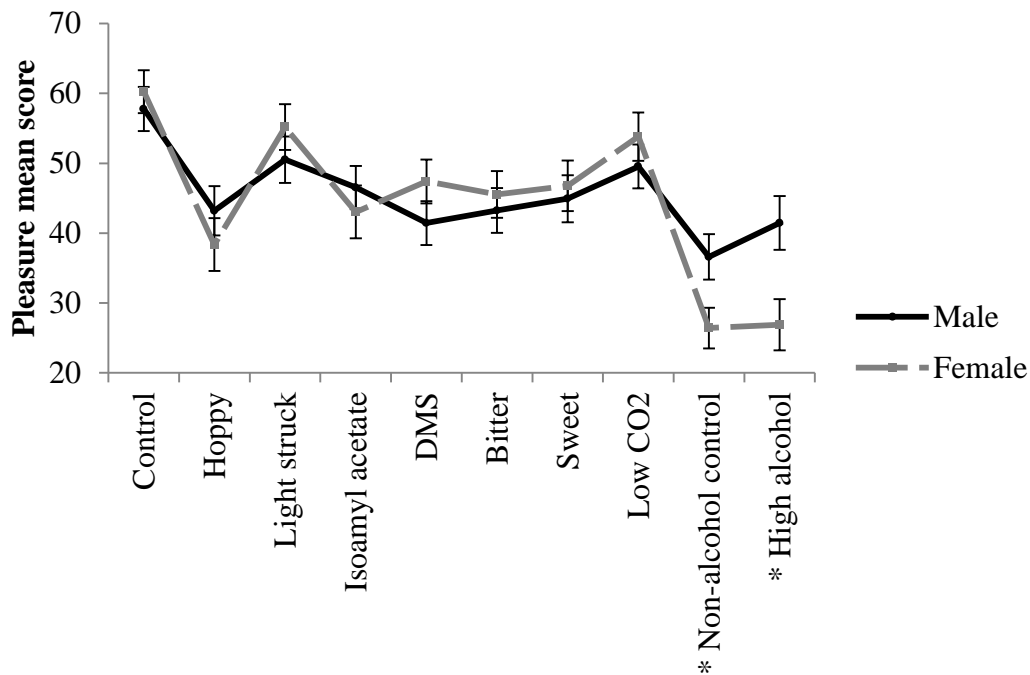

Figure 4b. Mean ratings (and SEM) of Pleasure for each sample by gender. ⁄ denotes a significant difference (p < 0.05) in the rating of Pleasure between gender groups.

was revealed for these four samples in liking (i.e. Control, Bitter, Sweet, and Low CO_2 were rated significantly lower in liking by 35+ year old than 18-34 year old consumers) and, interestingly, these same four samples were also rated significantly lower by the older age group in familiarity. However, differences in familiarity were not related to differences in

Nostalgia which was not rated significantly differently between age groups for most samples according to simple main effects analysis. The High alcohol sample was rated significantly higher in Nostalgia by the 35+ year old consumers and, conversely, the Light struck sample was rated significantly higher by the 18-34 year old consumers.
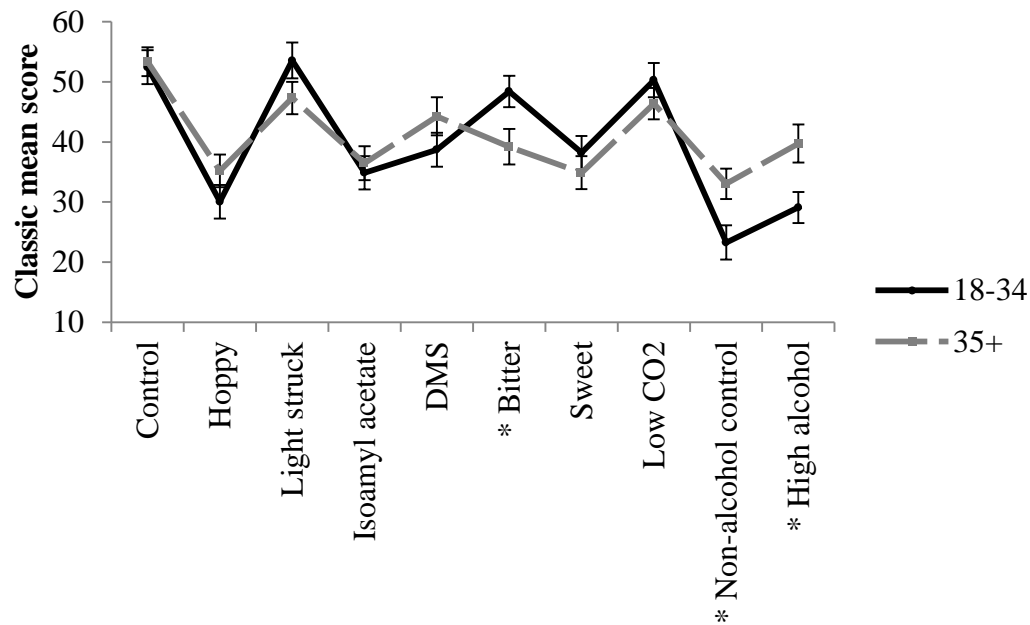

Figure 5a. Mean ratings (and SEM) of Classic for each sample by age group. *denotes a significant difference (p < 0.05) in the rating of Classic between age groups.
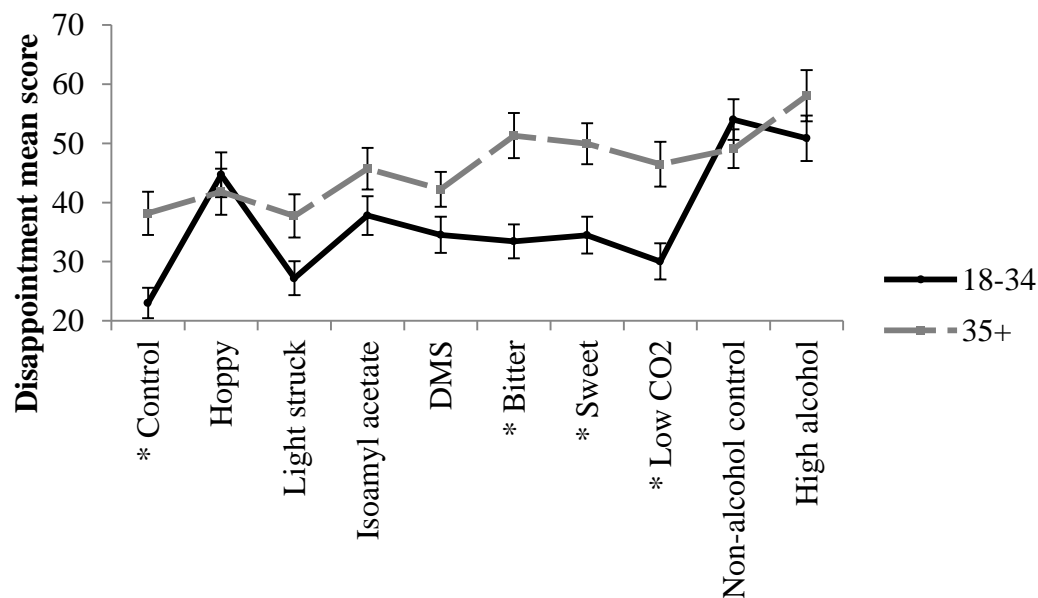

Figure 5b. Mean ratings (and SEM) of Disappointment for each sample by age group. * denotes a significant difference (p < 0.05) in the rating of disappointment between age groups.
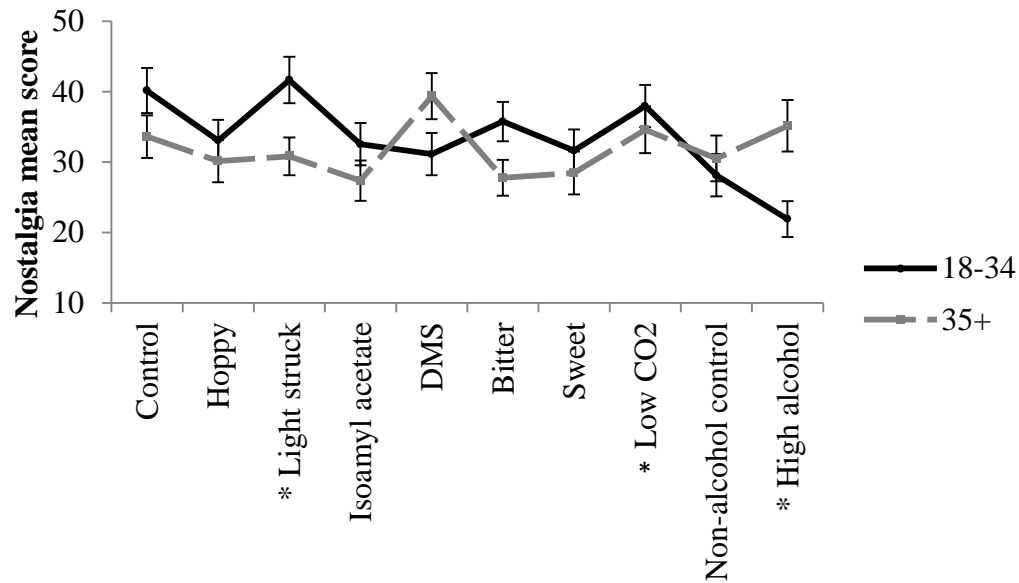
Figure 5c. Mean ratings (and SEM) of Nostalgia for each sample by age group. * denotes a significant difference (p < 0.05) in the rating of Nostalgia between age groups.

4 Discussion

One of the main objectives of this study was to create an approach for the generation of a product category-specific reduced consumer-led lexicon. Firstly, it was found that the process was accelerated by the use of small focus groups as opposed to the one-to-one interviews employed by Ng et al. (2013). Notwithstanding this large saving in time, the final lexicon was comparable to other published emotional lexicons in the number of emotion terms generated (e.g. Ng et al. (2013), King and Meiselman (2010)). Cluster analysis proved a useful tool for grouping terms into emotion categories of similar terms but required subtle modifications to reduce overlap and potential confusion between categories. With reference to each emotion category's internal validity (i.e. Cronbach's α) and by using linguistic checks, the final 12 emotion categories were clearly defined and still allowed for a breadth of emotions to be reported by consumers when assessing samples.

A second objective of this study was to validate the reduced consumer-led lexicon by assessing its ability to discriminate between samples with varying sensory properties in their elicited emotional responses. Furthermore, the ability of the reduced lexicon to reveal differences between consumer groups in their emotional responses to the samples was also considered in validating the approach. The 2-dimensional structure of emotional space revealed by PCA was consistent with circumplex models of emotion (Russell (1980),

Watson and Tellegen (1985), Larsen and Diener (1992)) and was in line with previous sensory findings using both long (Chrea et al. (2009); Ng et al. (2013); Chaya et al. (2015)) and short (Porcherot et al., 2010) emotion forms. This emotional space provided a useful guide for comparing the discriminability of emotion categories between samples because categories co-located in the emotional space grouped samples similarly.

Post hoc groupings of samples for the emotion categories Disgust, Disillusionment and Disappointment (which loaded highly positively on PC1 and were associated with unpleasantness/displeasure) showed only small differences in their discrimination between samples, underlining the close correlation between these emotions. Of the three, Disillusionment was the most discriminating between samples. However, the other two unpleasant emotion categories identified differences between consumer groups where Disillusionment did not. Disgust revealed an interaction between gender and sample, with the High alcohol sample rated as more disgusting by women. This could be related to the finding that women are more sensitive to the alcohol burn associated with ethanol (Duffy, Peterson, & Bartoshuk, 2004). Women have also been reliably shown to score higher than men on disgust sensitivity scales (Haidt, McCauley, and Rozin (1994); Davey (1994); Curtis, Aunger, and Rabie (2004); Olatunji, Arrindell, and Lohr (2005); Tybur, Lieberman, and Griskevicius (2009)) so, though there were no significant differences between genders in liking, a higher disgust sensitivity may have contributed to the higher ratings of the emotion by women for the particularly disgusting High alcohol sample. All three emotion categories showed a tendency for 35+ year old consumers to rate the samples higher than 18-34 year old consumers but Disappointment implicated four samples in particular (Control, Bitter, Low CO_2, Sweet) in driving this difference between age groups. The high ratings of Disappointment for these four samples by the older consumer group appeared to be closely linked to their ratings of liking as the same four samples were found to be rated significantly lower in liking by 35+ year old than 18-34 year old consumers.

Pleasure, Fun, Desire, Excitement, Classic, and Nostalgia were negatively correlated with PC1 and were associated with pleasantness/pleasure. These emotion categories revealed many similarities in their groupings of samples with only a few subtle differences of samples. However, consumer group comparisons highlighted much larger differences between emotion categories. Nostalgia, though not as discriminating as the other pleasant emotion categories when considering just the main effect of sample, was able to draw out

interactions between age group and sample, with older consumers assigning higher ratings to the High alcohol sample and lower ratings to the Light struck sample than their younger counterparts, in spite of no significant differences between age groups for those samples in familiarity. In fact, this was the only case of a difference between the Light struck and Control samples throughout this study which is surprising given the fact that Light struck aroma is considered undesirable by the brewing industry (Stephenson & Bamforth, 2002). For Classic, the Bitter sample was rated higher by 18-34s whereas the Non-alcohol control and High alcohol samples received lower ratings from 18-34 year olds. Both Nostalgia and Classic were rated lower overall by females than males. This could be linked to a lack of familiarity with the sensory properties of the selected samples as familiarity was also rated significantly lower by women. Interactions between gender and sample were found for Pleasure with females giving lower ratings to Non-alcohol control and High alcohol samples than men. Such gender differences were not found for liking.

A number of sensory properties (Bitter, Sweet, Hoppy, Isoamyl acetate, DMS) showed similar patterns when comparing with the Control in that they were scored lower for pleasant emotions and higher for unpleasant emotion categories. This is likely to be in no small part due to the fact that, in this study, optimised commercial products were modified so any change was detrimental and consequently affected consumers' emotional responses. Indeed, no modification in this study was found to significantly increase consumer ratings of pleasant emotion categories or significantly decrease scores for unpleasant emotion categories. The similarity in response between these sensory properties is particularly interesting as some are characteristic attributes of beer (bitterness, sweetness, hoppiness) whilst others are more commonly accepted as off-flavours (DMS, isoamyl acetate) although at low concentrations can also be characteristic of some beers. For the emotion categories related to pleasantness, just Classic was able to demonstrate a significant difference in rating between these samples (Bitter was rated significantly higher than both Isoamyl acetate and Hoppy).

Intensity (high activation/engagement) and Mildness and Indifference (low activation/engagement) loaded very highly in opposite directions on PC2 and grouped samples comparably but very differently to emotion categories highly correlated with PC1. Mildness and Indifference showed similar sample groupings with subtle differences (e.g. increasing Isoamyl acetate significantly increased Indifference but not Mildness). Intensity had greater discrimination ability than its two opposing emotion categories and, in fact, was

the only emotion category of the 12 to successfully discriminate between the Control and Low CO_2 samples as well as between the Non-alcohol control and High alcohol samples. This was unexpected as it was anticipated that there would be differentiation between the Control and Low CO_2 samples and between the Non-alcohol control and High alcohol samples in emotion categories correlated with pleasure/pleasantness based on the findings of Chaya et al. (2015) who reported that increased carbonation or body associated with increased alcohol content elicited more pleasant emotions. The finding that sweetness was associated with less engaging emotions agreed with previous findings (Chaya et al., 2015). There were no interaction effects between samples and gender or age group for any of these three emotion categories correlated with the activation/engagement dimension. In addition, 35+ year old consumers gave higher ratings for Mildness and Indifference than those aged 18-34, whereas Intensity was rated higher by the younger consumer group. Taken together, these results suggest that the emotion categories correlated with activation/engagement were important for discriminating between samples.

A surprising finding across emotion categories was that, where main effects of gender were found, women generally gave lower ratings than men (the exception was the emotion category Intensity). This was unexpected as women are stereotyped as more emotional than men (Fabes and Martin (1991); Plant, Hyde, Keltner, and Devine (2000); Timmers, Fischer, and Manstead (2003)). Research appears to bear this out with females exceeding males in reported emotionality and emotion expressivity (Allen and Haccoun (1976); Gross and John (1995)). Gender roles have been discussed as playing an important role in emotions (Fischer (1993); Grossman and Wood (1993)) and gender role characteristic have indeed been found to moderate the relationship between gender and emotion expressivity (Kring & Gordon, 1998). This could offer a partial explanation for the lower emotional ratings by women as beer has been historically, and continues to be, viewed as a relatively masculine beverage. This fact, therefore, may be influencing both male and female reports of emotion. However, it is interesting that Intensity is rated higher by women than men in general and the reasons for this could be explored further. The literature also reveals a trend for adults to experience more positive affect and less negative affect with age (Mroczek & Kolarz, 1998), though this appears not to be the case in response to this sample set, with higher ratings given for unpleasant emotion categories by the older group and lower ratings assigned to pleasant emotion categories by the 35+ age group. Of particular interest was the fact that Classic and Nostalgia showed significant interactions between sample and age group with some samples

rated higher by the younger consumer and some by the older consumers. Nostalgia has been described as referring to a preference for objects that were more common when one was younger (Holbrook & Schindler, 1991) and, obviously, these objects (beers in the case of this study) will likely have been different for the two age groups, perhaps explaining the differences seen in both nostalgia and classic between the two age groups in the study.

These results have shown that, on the whole, the 12 emotion categories were able to discriminate across beer samples with varying sensory properties This work confirms the suggestions of previous authors that sensory properties act as a driver for emotional response (Thomson et al. (2010); Ng et al. (2013); Sester et al. (2013)). Nevertheless, a number of emotion categories grouped samples very similarly to one another (with just Intensity able to discriminate the Low CO_2 sample from the Control and also between the Non-alcohol control and High alcohol samples) no emotion category was able to discriminate between Light struck and Control samples. The consideration of consumer segments, however, revealed that some emotion categories were able to differentiate between the emotional responses of males and females and between younger and older age groups. For example, Nostalgia showed an interaction between age group and ratings of the previously discussed Light struck sample where no such interaction was found for the Control. Therefore, this study has been able to validate the use of a reduced consumer-led lexicon through the demonstration of its ability to discriminate between beer samples with varying sensory properties. However, the approach was of limited efficacy until consumer segments were considered, at which point the effectiveness of the 12 emotion categories for discriminating between samples and between the responses of consumer groups was shown.

5 Conclusions

It has been shown that the development of a reduced consumer-led lexicon offered a quicker test for consumers, whilst delivering emotional discriminability between a set of beer samples and revealing differences between consumer groups in emotional response to certain sensory properties of beer. Although a number of emotion categories appeared highly correlated, individual emotion categories were able to discriminate between emotional responses to samples as well as between the emotional responses of consumer groups. Hence, the categorisation of a full consumer-led emotional lexicon to a reduced lexicon presented here was successful.

Further work is required to determine the overall effectiveness of a reduced form as compared to a full lexicon. This is important to ascertain if we are "missing potentially valuable information" (King & Meiselman, 2010) by not including enough evaluations per product for consumers. However, indications suggest that a close relationship exists between full and reduced forms (Porcherot et al., 2010).

If found to be comparable to a full lexicon, this approach has the potential to open up possibilities for cross-cultural comparisons. Indeed, grouping similar terms in GEOS has already been shown to be useful for cross-cultural comparisons (Ferdenzi et al. (2011); Ferdenzi et al. (2013)). By following the described method, each culture of interest could generate its own reduced consumer-led lexicon. The main benefit of this is that groups of emotion terms would be compared as opposed to individual words, avoiding the problems associated with attempting direct one-to-one translations. This would facilitate understanding of the emotion concept that each culture is referring to, allowing comparisons between categories in each culture. In particular, use of multivariate analyses to use the established emotional space as a guide for cultural comparisons has the potential to prove a useful tool.

References

Allen, J. G., & Haccoun, D. M. (1976). Sex differences in emotionality - multidimensional approach. *Human Relations, 29*, 711-722.

Bhumiratana, N., Adhikari, K., & Chambers, E. (2014). The development of an emotion lexicon for the coffee drinking experience. *Food Research International, 61*, 83-92.

Chaya, C., Pacoud, J., Ng, M., & Hort, J. (2015). Measuring the emotional response to beer and the relative impact of sensory and packaging cues. *Journal of the American Society of Brewing Chemists*.

Chrea, C., Grandjean, D., Delplanque, S., Cayeux, I., Le Calve, B., Aymard, L., . . . Scherer, K. R. (2009). Mapping the Semantic Space for the Subjective Experience of Emotional Responses to Odors. *Chemical Senses, 34*, 49-62.

Churchill, A., & Behan, J. (2010). Comparison of methods used to study consumer emotions associated with fragrance. *Food Quality and Preference, 21*, 1108-1113.

Curtis, V., Aunger, R., & Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. *Proceedings of the Royal Society B-Biological Sciences, 271*, S131-S133.

Davey, G. C. L. (1994). Self-reported fear to common indigenous animals in an adult UK population - the role of disgust sensitivity. *British Journal of Psychology, 85*, 541-554.

Dorado, R., Picard, A., Pérez-Hugald, C., & Chaya, C. (2014). *First position effect on the emotional response of consumers: a case study in chocolate and beers* Sixth European Conference on Sensory and Consumer Research, Copenhagen, Denmark.

Duffy, V. B., Peterson, J. M., & Bartoshuk, L. M. (2004). Associations between taste genetics, oral sensation and alcohol intake. *Physiology & Behavior, 82*, 435-445.

Fabes, R. A., & Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin, 17*, 532-540.

Ferdenzi, C., Delplanque, S., Barbosa, P., Court, K., Guinard, J. X., Guo, T., . . . Grandjean, D. (2013). Affective semantic space of scents. Towards a universal scale to measure self-reported odor-related feelings. *Food Quality and Preference, 30*, 128-138.

Ferdenzi, C., Schirmer, A., Roberts, S. C., Delplanque, S., Porcherot, C., Cayeux, I., . . . Grandjean, D. (2011). Affective Dimensions of Odor Perception: A Comparison Between Swiss, British, and Singaporean Populations. *Emotion, 11*, 1168-1181.

Fischer, A. H. (1993). Sex differences in emotionality - fact or stereotype. *Feminism & Psychology, 3*, 303-318.

Fransella, F., Bell, R., & Bannister, D. (2004). *A Manual for Repertory Grid Technique* (2nd ed.). West Sussex: John Wiley & Sons Ltd.

Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2009). Separating production from perception: Perceiver-based explanations for sex differences in emotion. *Behavioral and Brain Sciences, 32*, 394-395.

Gross, J. J., & John, O. P. (1995). Facets of emotional expressivity - 3 self-report factors and their correlates. *Personality and Individual Differences, 19*, 555-568.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience - a social role interpretation. *Journal of Personality and Social Psychology, 65*, 1010-1022.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust - a scale sampling 7 domains of disgust elicitors. *Personality and Individual Differences, 16*, 701-713.

Holbrook, M. B., & Schindler, R. M. (1991). Echoes of the dear departed past - some work in progress on nostalgia. *Advances in Consumer Research, 18*, 330-333.

Jaeger, S. R., Cardello, A. V., & Schutz, H. G. (2013). Emotion questionnaires: A consumer-centric perspective. *Food Quality and Preference, 30*, 229-241.

Jaeger, S. R., & Hedderley, D. I. (2013). Impact of individual differences in emotional intensity and private body consciousness on EsSense Profile (R) responses. *Food Quality and Preference, 27*, 54-62.

King, S. C., & Meiselman, H. L. (2010). Development of a method to measure consumer emotions associated with foods. *Food Quality and Preference, 21*, 168-177.

King, S. C., Meiselman, H. L., & Carr, B. T. (2013). Measuring emotions associated with foods: Important elements of questionnaire and test design. *Food Quality and Preference, 28*, 8-16.

King, S. C., Meiselman, H. L., & Carr, T. B. (2010). Measuring emotions associated with foods in consumer testing. *Food Quality and Preference, 21*, 1114-1116.

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology, 74*, 686-703.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. *Review of Personality and Social Psychology, 13*, 25-59.

Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices*. New York: Springer.

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

Mintel. (2013). Beer - UK - December 2013. London: Mintel.

Mroczek, D. K., & Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology, 75*, 1333-1349.

Ng, M. (2013). *Emotional response: From sensory attributes to packaging and back again!* (PhD), The University of Nottingham.

Ng, M., Chaya, C., & Hort, J. (2013). Beyond liking: Comparing the measurement of emotional response using EsSense Profile and consumer defined check-all-that-apply methodologies. *Food Quality and Preference, 28*, 193-205.

Olatunji, B. O., Arrindell, W. A., & Lohr, J. M. (2005). Can the sex differences in disgust sensitivity account for the sex differences in blood-injection-injury fears? *Personality and Individual Differences, 39*, 61-71.

Piqueras-Fiszman, B., & Jaeger, S. R. (2014a). Emotion responses under evoked consumption contexts: A focus on the consumers' frequency of product consumption and the stability of responses. *Food Quality and Preference, 35*, 24-31.

Piqueras-Fiszman, B., & Jaeger, S. R. (2014b). The impact of evoked consumption contexts and appropriateness on emotion responses. *Food Quality and Preference, 32*, 277-288.

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly, 24*, 81-92.

Porcherot, C., Delplanque, S., Raviot-Derrien, S., Le Calve, B., Chrea, C., Gaudreau, N., & Cayeux, I. (2010). How do you feel when you smell this? Optimization of a verbal measurement of odor-elicited emotions. *Food Quality and Preference, 21*, 938-947.

Russell, J. A. (1980). A circumplex model of affecy. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Schifferstein, H. N. J., Fenko, A., Desmet, P. M. A., Labbe, D., & Martin, N. (2013). Influence of package design on the dynamics of multisensory and emotional food experience. *Food Quality and Preference, 27*, 18-25.

Serra, L. L., & Aranceta, J. (2003). La cerveza en la alimentación de los españoles: relación entre el consumo de cerveza y el consumo de energía y nutrientes, el índice de masa corporal y la actividad física en la población adulta española. Madrid: Centro de Información Cerveza y Salud.

Sester, C., Dacremont, C., Deroy, O., & Valentin, D. (2013). Investigating consumers' representations of beers through a free association task: A comparison between packaging and blind conditions. *Food Quality and Preference, 28*, 475-483.

Spinelli, S., Masi, C., Dinnella, C., Zoboli, G. P., & Monteleone, E. (2014). How does it make you feel? A new approach to measuring emotions in food product experience. *Food Quality and Preference, 37*, 109-122.

Stephenson, W. H., & Bamforth, C. W. (2002). The impact of lightstruck and stale character in beers on their perceived quality: A consumer study. *Journal of the Institute of Brewing, 108*, 406-409.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99-103.

Thomson, D. M. H., Crocker, C., & Marketo, C. G. (2010). Linking sensory characteristics to emotions: An example using dark chocolate. *Food Quality and Preference, 21*, 1117-1125.

Timmers, M., Fischer, A. H., & Manstead, A. S. R. (2003). Ability versus vulnerability: Beliefs about men's and women's emotional behaviour. *Cognition & Emotion, 17*, 41-63.

Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, Mating, and Morality: Individual Differences in Three Functional Domains of Disgust. *Journal of Personality and Social Psychology, 97*, 103-122.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219-235.