DEVELOPING AN EFFECTIVE APPROACH TO MEASURE EMOTIONAL RESPONSE TO THE SENSORY PROPERTIES OF BEER

CURTIS EATON, BSc (Hons)

Thesis submitted to The University of Nottingham

for the degree of Doctor of Philosophy

JULY 2015

For Chloe

Abstract

Emotion research in sensory and consumer science has gathered significant momentum over recent years and the development of effective emotion measurement methods is a priority in this rapidly growing area.

The aim of this research was to advance the use of consumer-led emotion lexicons by using focus groups to increase the efficiency of lexicon generation and by decreasing the number of consumer response categories. In parallel, the ability of the newly generated reduced lexicon to discriminate emotional response across different gender and age groups, and across sensorially distinct beer samples, was evaluated. The new approach was largely effective at discriminating across samples and revealed significant differences in emotional response between genders and between age groups.

The reduced lexicon was compared to a full lexicon to ascertain their relative efficacies. Whilst there were differences between the two form lengths, neither was convincingly more effective at sample discrimination than the other, although the full form better differentiated between age groups.

The reduced form was also applied to cross-cultural comparisons through the generation of a reduced product-specific consumer-led emotion lexicon in Spain. As in the UK, the approach discriminated well between samples and was able to differentiate between consumer groups. Comparing Spanish and UK responses, ratings of emotions associated with pleasure/pleasantness were similar but there were differences in the use of emotions associated with arousal/engagement/activation. This new methodology was therefore

demonstrated to be a valuable tool for investigating cross-cultural emotional response.

The approach developed in this thesis provides researchers with an enhanced consumer-led emotion methodology for use with food and beverages. As well as being relatively quick, the approach has been proven to differentiate between products and reveal differences concerning emotional response across different consumer groups and between cultures. These attributes make this emotional measurement approach extremely valuable to this young research area.

Publications and presentations

Publications

CHAYA, C. EATON, C., HEWSON, L. FERNÁNDEZ VÁZQUEZ, R., FERNÁNDEZ-RUIZ, V., SMART, K. A., & HORT, J. accepted. Developing a reduced consumer-led lexicon to measure emotional response to beer. Food Quality & Preference.

Oral presentations

EATON, C., CHAYA, C., SMART, K. A. & HORT, J. 2014. Emotional response to the sensory properties of beer. Young Scientists Symposium on Malting, Brewing and Distilling, Ghent, Belgium.

EATON, C., CHAYA, C., SMART, K. A. & HORT, J. 2014. Happiness from hoppiness? Comparing emotional response to sensory properties of beer between the UK and Spain. ASBC Annual Meeting, Chicago, IL, USA.

EATON, C., BEALIN-KELLY, F. & HORT, J. 2013. Measuring emotional response to food products. Symposium in Memory of Harry Nursten, Reading, UK.

Poster presentations

EATON, C., CHAYA, C., SMART, K. A. & HORT, J. 2014. Measuring emotion: the long and short of it. 6th European Conference on Sensory and Consumer research, Copenhagen, Denmark.

EATON, C., CHAYA, C., HEWSON, L., FERNÁNDEZ VÁZQUEZ, R., FERNÁNDEZ-RUIZ, V. SMART, K. A., BEALIN-KELLY, F., FENTON, A. & HORT, J. 2013. Happiness from hoppiness? Using consumer lexicons to compare emotional response between the UK and Spain. 10th Pangborn Sensory Science Symposium, Rio de Janeiro, Brazil.

Acknowledgements

None of this would have been possible without the belief of Professor Joanne Hort. You've provided me with an environment in which I have been able to thrive and I can't thank you enough.

Thank you to Dr. Carolina Chaya who has been a continued source of enthusiasm and constructive criticism throughout this project. Many thanks to the team in Madrid who contributed to Chapter 5 of this thesis and to Laura Lorido Contreras for the three months spent assisting me.

Many thanks to the beer panel: Neil, Graham, Kevin, Gary, Diane, Sue K., Eric, Sue S., Pete, and Rob whose patience, dedication, and good humour made this the part of this project particularly rewarding and enjoyable. Also, thanks to all the consumers that took part in this research.

Thanks to Dr. Louise Hewson and Helen Allen; Louise for always having time for "just one more question" and Helen for being there with a smile when it was needed most.

To SABMiller for sponsoring this research; Professor Katherine Smart for providing me with a valued industrial perspective on this research and Dr. Adam Fenton for initial guidance. Also thanks to Cara Technology for their generous support.

My time at Sutton Bonington has been touched by too many individuals to name here, but I extend my sincere thanks to all those that I've had the pleasure of working alongside, learning from, and befriending.

Preface

Sensory science is concerned with the evoking, measurement, analysis, and interpretation of product perception through the senses of sight, smell, touch, taste, and hearing (Stone and Sidel, 1993). It is a field that has seen rapid growth since its origins in the 1940s. The need to understand the relationships that products' sensory properties have with consumer perception and, importantly, behaviour continues to drive research in this ever growing area. A recently emerging trend has focussed on consumer emotional response to products. Traditionally, sensory practitioners have referred to hedonic measures to gain an understanding of product performance. However, in an increasingly competitive modern marketplace, hedonic response does not offer such a significant differential advantage as it once did. It is in its potential capacity for product differentiation that emotion research has garnered so much interest.

It is a fallacy to believe that humans are rational beings, with emotions consistently shown to be important drivers of decision-making. Damasio (2006) describes neurological patients with disorders of emotion and related defects in even simple decision-making, underlining the importance of emotions in everyday life. By understanding the relationships between sensory properties of products and emotional response, a window is opened to consumer behaviour. This is notoriously difficult to predict, as evidenced by the high rate of failures for newly launched products. By leveraging emotional information about products, there is an expected commercial advantage of being able to identify niches, steer product development, differentiate

products, and align sensory with branding and marketing to create a synergistic relationship.

As a relatively young area of research in sensory and consumer science, methods to measure emotional response have yet to become established, affording researchers the opportunity to further develop existing methods. The aim of the present study was to produce an effective approach to permit the investigation of the relationships between sensory properties and consumer emotional response. The product of study in this thesis was beer as it is complex and varied in its sensory properties and has been shown to be an emotive product category (Chaya et al., 2015). Furthermore, it is a product with very different consumption habits across consumer groups, offering an opportunity to research the potential role emotions may play in this.

Thesis structure

This thesis presents an improved reduced product-specific consumer-led reduced emotion lexicon methodology and considers its effectiveness through its ability to discriminate between samples with different sensory properties, its ability to differentiate between the responses of different consumer groups, its relative advantages and disadvantages as compared with a full emotion form, and its application to cross-cultural emotion research.

The thesis is organised into six chapters. Chapter 1 gives a general introduction to emotions and their measurement in sensory and consumer science. The current understanding of the relationship between emotional response and sensory properties is also described. Finally, as the product used in this research is beer, the brewing process is outlined, with particular

attention given to the origins of key sensory properties. Chapter 2 describes the selection of sensorially-distinct beer samples with specific sensory properties. Together with the final data collection, an overview of panel recruitment and training necessary for this project is presented. Chapters 3, 4 and 5 describe the measurement of emotional response to the samples characterised in Chapter 2. Chapter 3 details the development of a reduced product-specific consumer-defined emotion lexicon and explores its efficacy. Chapter 4 then compares the use of a reduced emotional lexicon to a full lexicon. Chapter 5 describes the creation of a reduced product-specific consumer-defined emotion lexicon in Spain using the approach developed in Chapter 3. Firstly, the effectiveness of the approach in its application to Spanish consumers is determined. Then, emotional responses to the samples are then compared cross-culturally between the UK and Spain using the countries' respective reduced lexicons. Finally, Chapter 6 provides an overview of major findings, suggestions for future work, and general conclusions.

List of abbreviations

ANOVA	Analysis of variance
ANS	Autonomic nervous system
CATA	Check all that apply
CNS	Central nervous system
DMS	Dimethyl sulphide
ECG	Electrocardiogram
EEG	Electroencephalography
EMG	Electromyography
F	Full form
FACS	Facial action coding system
fMRI	Functional magnetic resonance imaging
GEOS	Geneva Emotion and Odor Scale
HSD	Honest significant difference
I-PANAS-SF	International Positive and Negative Affect Schedule Short Form
MAACL	Multiple Affect Adjective Check List
MAACL-R	Multiple Affect Adjective Check List Revised
MEG	Magnetoencephalography
MFA	Multiple factor analysis
NES	Neuro-endocrine system
PANAS	Positive and Negative Affect Schedule
PANAS-X	Positive and Negative Affect Schedule Expanded Form
PC	Principal component
PCA	Principal components analysis
POMS	Profile of Mood States
PrEmo	Product Emotion Measurement Instrument
R	Reduced form
RATA	Rate all that apply
SNS	Somatic nervous system
UK	United Kingdom
UniGEOS	Universal Geneva Emotion and Odor Scale

Table of contents

Abstract iii

Publications and presentations v

Acknowledgements vi

Preface vii

Thesis structure viii

List of abbreviations x

Table of contents xi

Chapter 1: General Introduction 1

1.1 Defining and classifying emotions 1

1.1.1 What is (and what isn't) an emotion? 1

1.1.2 Classifying emotions 3

1.2 Why measure emotional response to sensory-driven products? 5

1.2.1 Beyond liking measures 5

1.2.2 Potential differences in emotional response between consumer segments 6

1.2.2.1 Gender 7

1.2.2.2 Age 8

1.2.2.3 Culture 9

1.3 Measuring emotional response 9

1.3.1 Self- report measures 10

1.3.1.1 Verbal self-report measures 11

1.3.1.2 Non-verbal self-report measures 14

1.3.1.3 An important note about the cognitive component of emotion 15

1.3.2 Implicit reactions 16

1.3.3 Physiological reactions 17

1.3.4 Functional brain imaging 19

1.3.5 Concluding emotion measurement approaches 20

1.4 Current understanding of the relationship between sensory properties and emotional response 21

1.5 Origins of the sensory properties of beer 23

1.5.1 Introduction to the brewing process 23

1.5.2 Malting 25

1.5.3 Wort production 25

1.5.4 Fermentation 26

1.5.5 Maturation, finishing and packaging 27

1.6 Research aim and objectives 28

Chapter 2: Selecting and characterising samples 31

2.1 Introduction 31

2.1.1 Objective 31

2.2 Selection of beer samples for sensory analysis 32

2.3 Sensory panel recruitment 36

2.3.1 Pre-screening 37

2.3.2 Screening 38

2.3.2.1 Detection of stimuli 38

2.3.2.2 Discrimination between stimuli 40

2.3.2.3 Recognition and description of stimuli 43

2.3.2.4 Screening results 46

2.3.3 Post-screening 48

2.3.3.1 Detection of stimuli....48
2.3.3.2 Discrimination between stimuli....49
2.3.3.3 Recognition and description of stimuli....52
2.3.3.4 Post-screening results....54
2.4 Panel training....54
2.5.1 Materials and method....58
2.5.2 Profiling results....59
2.6 Conclusions....61
Chapter 3: Developing an approach for the measurement of emotional response to the sensory properties of beer....62
3.1 Introduction....62
3.1.1 Selecting an emotional measurement method....62
3.1.2 Verbal self-report....62
3.1.3 Objectives....65
3.2 Lexicon development....65
3.2.1 Subjects....66
3.2.2 Procedure....67
3.2.3 Grouping of terms into emotion categories....69
3.3 Measuring consumer emotional response using the reduced emotion lexicon....73
3.3.1 Subjects....73
3.3.2 Procedure....74
3.3.3 Data analysis....76
3.4 Results....77
3.4.1 Emotion space....77

3.4.2 Discrimination ability of emotion categories between samples 80

3.4.3 Consumer group effects 88

3.4.3.1 Gender 89

3.4.3.2 Age group 89

3.5 Discussion 91

3.6 Conclusions 100

Chapter 4: Comparing the effectiveness of full and reduced emotion lexicons 103

4.1 Introduction 103

4.1.1 Reduced emotion forms in psychology 103

4.1.2 Reduced emotion forms in sensory and consumer science 104

4.1.3 Objectives 105

4.2 Materials and method 106

4.2.1 Procedure 106

4.2.2 Data analysis 107

4.3 Results 108

4.3.1 Emotion space 108

4.3.2 Comparing the discrimination abilities of the full and reduced forms 111

4.3.3 Consumer group effects 130

4.3.3.1 Gender 130

4.3.3.2 Age group 131

4.4 Discussion 135

4.5 Conclusions 141

Chapter 5: Cross-cultural comparison of emotional response using reduced emotion lexicons 144

5.1 Introduction 144

5.1.1 Universality of emotions 144

5.1.2 Cultural differences in emotional expression 145

5.1.3 Reduced lexicons as cross-cultural emotion comparison tools 147

5.1.4 Objectives 148

5.2 Materials and method 149

5.2.1 Samples 149

5.2.2 Lexicon development 150

5.2.2.1 Subjects 151

5.2.2.2 Procedure 151

5.2.2.3 Grouping of terms into emotion categories 151

5.2.3 Measuring consumer emotional response using the Spanish reduced emotion lexicon 155

5.2.3.1 Subjects 155

5.2.3.1 Procedure 157

5.2.4 Data analysis 157

5.2.4.1 Spanish emotional response 157

5.2.4.2 Cross-cultural comparisons 158

5.3 Measuring Spanish consumer emotional response using the reduced emotion lexicon 159

5.3.1 Emotion space 159

5.3.2 Discrimination ability of emotion categories between samples 161

5.3.3 Consumer group effects 169

5.3.3.1 Gender 169

5.3.3.2 Age group 172

5.4 Discussion of the results obtained using the Spanish reduced product-specific consumer-led emotion lexicon 174

5.5 Results of cross-cultural comparisons 183

5.5.1 Comparing the grouping of samples using emotion categories across cultures 183

5.5.2 The association between emotion categories and sensory properties of beer across cultures 184

5.5.3 The relative abilities of each culture's reduced lexicon to discriminate across consumer segments 192

5.6 Discussion of cross-cultural comparisons 193

5.7 Conclusions 200

Chapter 6: General Discussion 202

6.1 Developing an approach for the measurement of emotional response 202

6.1.1 Beyond self-report 204

6.2 Emotional response to the sensory properties of beer 205

6.2.1 Beyond fixed intensities of sensory properties 206

6.2.2 Beyond single sensory properties 207

6.2.3 Beyond beer 208

6.3 Consumer groups 210

6.3.1 Gender 210

6.3.2 Age group 211

6.3.3 Beyond extrinsic grouping of consumers 211

6.3.4 Cross-cultural comparisons 212

6.3.4.1 Beyond the UK and Spain....213

6.3 Conclusions....214
References....215

Appendices....227

Appendix A: Pre-screening questionnaire....227
Appendix B: Scaling exercises (redrawn from Meilgaard et al. (2007))....231
Appendix C: Panel sample assessment protocol....232

Chapter 1: General Introduction

This chapter is organised into a number of sections. The first addresses the complex issue of defining and classifying emotions. In the next section, the purpose of measuring emotional response is discussed, before the measurement of emotion is described in the section that follows. The current understanding of the relationship between sensory properties and emotional response is then explored. As beer is the product of study in this thesis, the origins of key sensory properties in the brewing process are subsequently outlined. The chapter closes with the aim and objectives of the research presented in this thesis.

1.1 Defining and classifying emotions

1.1.1 What is (and what isn't) an emotion?

The study of emotions has attracted many noted minds throughout history including Hippocrates, Descartes and Darwin. Nevertheless, it is relatively straightforward for even the layperson to list a number of emotions, name situations where they may feel these emotions, and even understand when they share an emotion with somebody else. However, defining emotion proves to be a problematic task for both the man on the street and great thinkers alike. This difficulty owes no small part to the fact that emotion seemingly cannot be defined as unitary concept. This has been well demonstrated by the wide variation in definitions given by scientists (Izard, 2010). Inevitably, this introduces the potential for significant confusion in the study of emotions unless researchers clearly iterate how they are defining emotion for the purposes of their research.

This thesis adapts the approach of Scherer (2005) who outlined the key constituents of an emotion (Table 1.1). According to this definition, emotions have a rapid onset, are high in intensity, and of short duration, impacting an organism's behaviour through the synchronised response to relevant events. In this way, distinctions can be made between different aspects of emotion, as well as distinguishing emotion from other affective phenomena, for example:

- Preferences are relatively stable evaluative judgements of stimulus liking/disliking, independent of currents needs or goals. They have low behavioural impact apart from approach or avoidance
- Attitudes are relatively enduring beliefs and predispositions towards specific objects or persons
- Moods are lower in intensity and longer in duration than emotions and may emerge without specific cause
- Affect disposition describes the tendency of a person to experience certain moods more frequently or to be prone to react with certain types of emotions (care must be taken because certain terms, e.g. 'anxious', can be used to describe both an emotion and an affect disposition)

Table 1.1 The key constituents of emotions as outlined by Scherer (2005).

Event focus	An emotion is triggered by an internal or external stimulus
Appraisal driven	The stimulus and its consequences must be relevant to the major concerns of the organism
Response synchronisation	The prepared response results in the coordinated mobilisation of resources
Rapidity of change	Response changes rapidly to track rapid changes in stimulus and associated appraisals
Behavioural impact	Emotions have a strong impact on behaviour (particularly communication), often interrupting ongoing behaviour
Intensity	The intensity of emotions is relatively high because of their importance in behavioural adaptation
Duration	Due to their high intensity, the duration of emotions is short so as to not tax the resources of the organism

1.1.2 Classifying emotions

Many attempts have been made at classifying emotions into groups that have different functions (e.g. whilst fear of threat and happiness to see a loved one are both emotions, they are clearly different to one another). Ekman and Friesen (1971) suggested that there are six basic emotions: happiness, sadness, fear, surprise, anger, and disgust. To explain the wealth of emotions we feel beyond the basic emotions, distinctions have been made between lower-order and higher-order emotions. Lower-order emotions (so-called Type 1 emotions (Rossiter and Bellman, 2005)) occur automatically and do not require cognitive input. These are analogous to the basic emotions posited by Ekman and Friesen (1971). Cognitive appraisal and conscious labelling leads to higher-order emotions (so-called Type 2 emotions (Rossiter and Bellman, 2005)). For example, cognitive processes might give rise to Type 2 emotions like ‘fulfilment’ or ‘contentment’ arising from the Type 1 emotion happiness.

However, it has been proposed that basic emotion theories are unable to adequately explain empirical observations and that these are better accounted for by circumplex models of affect (Posner et al., 2005). Circumplex models state that each emotion represents a point on a continuum that varies along two or more fundamental axes. For example, the model of Larsen and Diener (1992) organises affect into two bipolar dimensions (Figure 1.1), the first related to pleasantness (unpleasant-pleasant) and the second to activation (low activation-high activation). Earlier models included similar dimensions (misery-pleasure and sleep-arousal, Russell (1980); unpleasantness-pleasantness, engagement-disengagement, Watson and Tellegen (1985)).

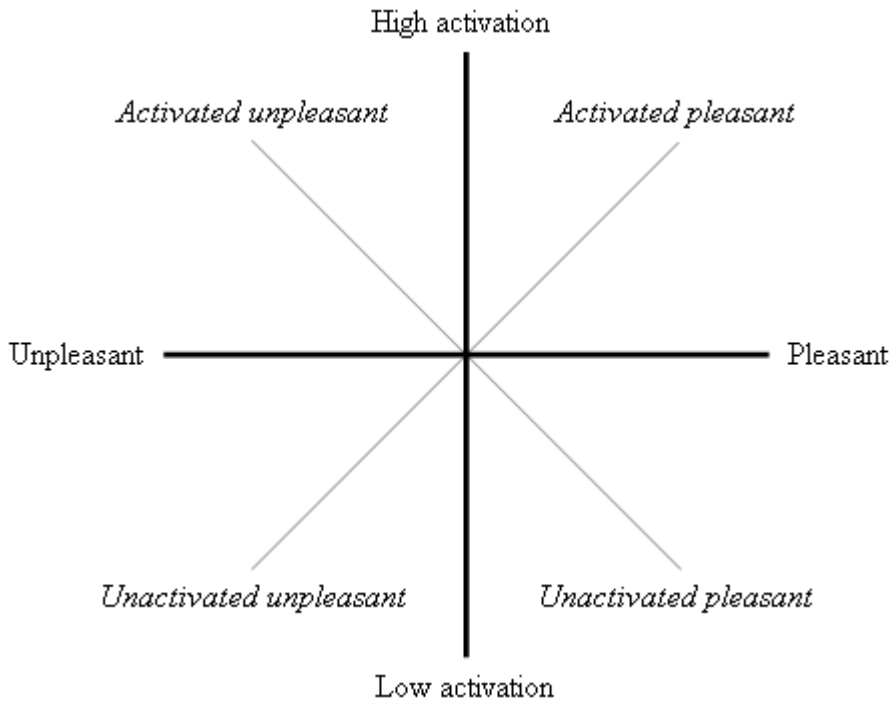


Figure 1.1 The circumplex model of emotion of Larsen and Diener (1992).

1.2 Why measure emotional response to sensory-driven products?

Measuring emotional response has the potential to provide more in depth understanding concerning the consumer's relationship with sensory-driven products such as food and beverages. This will be described in this section.

1.2.1 Beyond liking measures

Most new brands and products do not give the predicted return of investment and are withdrawn from the market (Thomson and Crocker, 2015). These failures represent a significant cost in terms of product development, marketing, etc. This situation has become increasingly problematic given that differentiation of modern products is more and more difficult as they are now more similar in terms of their technical or performance properties (Churchill and Behan, 2010). Traditionally, hedonic measures have played an important role in corporate research (Moskowitz et al., 2012). However, it is clear from the failure rates of new products that liking is not necessarily a reliable predictor of success. By measuring emotional response, it has been shown that similarly liked products can be differentiated in a number of reported emotions (Ng et al., 2013). Therefore, the emotional quality of products is becoming important to maintain a differential advantage (Schifferstein et al., 2013).

A further advantage of measuring emotions in addition to liking is that emotions allow for a shared lexicon between sensory and marketing, thereby strengthening the link between the two historically distinct fields (King and Meiselman, 2010). This cooperation has the potential to be particularly fruitful as marketing already has a slightly firmer foothold in emotions research than sensory and consumer science (see Poels and Dewitte (2006)). By aligning the

emotional response to the sensory properties of a product with its branding and marketing, ‘consonance’ is achieved, thereby reinforcing the brand message (Thomson et al. (2010); Thomson and Crocker (2015)).

1.2.2 Potential differences in emotional response between consumer segments

Emotional response may offer a partial explanation with regards to the reasons for consumption differences observed between consumer groups. In sensory and consumer science, there is precious little published work investigating differences in emotional response between consumer segments (e.g. genders, age groups, socioeconomic classes, etc.) to confirm or refute this suggestion. The literature concerning consumer liking highlights differences between consumer segments (e.g. King et al. (2012) found distinctions between consumer groups that liked certain wine styles; Cooke and Wardle (2005) found effects of gender for liking of various food products in children) and it is suggested that a similar effect could be observed for emotions.

Beer is the focus of study throughout this thesis and is a pertinent product for the investigation of the differences between consumer segments because the relative consumption of different products between groups of consumers is quite marked in the alcoholic beverage industry. A better understanding of the differences in emotional response to beer between consumers would allow the brewing industry to gain a clearer picture of the roles emotions may play. If, indeed, emotions do make a significant contribution to these differences, it may be possible to develop products with sensory properties that elicit selected emotions in order to appeal to particular groups.

1.2.2.1 Gender

It has been reported in the UK (Mintel, 2013) and Spain (Serra and Aranceta, 2003) that women are less frequent drinkers of beer than men. Although a number of factors could contribute to this consumption difference (e.g. genetic variation in taste sensitivity (Tepper et al., 2009); papillae density (Bartoshuk et al., 1994); trigeminal sensation (Komiyama et al., 2009); carbonation (Bere et al., 2008); brand (Guinard et al., 2000)), it may be the case that differences in emotional response to beer may play at least a partial role. Indeed, it is perhaps to be expected as differences in emotionality have been reported between genders. Women have been shown to be stereotyped as more emotional than men (Fabes and Martin (1991); Plant et al. (2000); Timmers et al. (2003)) and research generally supports this stereotype, with females exceeding males in reported emotionality and emotion expressivity Allen and Haccoun (1976); Gross and John (1995)). This gender difference has been explained as having an evolutionary basis, particularly in that emotions play an important role in child-rearing (Babchuk et al., 1985). However, King et al. (2010) reported that reports of emotion were dependent on the product category being assessed and discussed gender roles (Fischer (1993); Grossman and Wood (1993)) as perhaps playing a role. As beer is viewed as a relatively masculine beverage (Landrine et al., 1988), this could give very different emotional responses between genders as a result of gender roles.

It is clear that there are many potential sources of emotional variation across genders in response to products, making this a potentially fruitful area of research and one that effective emotion measurement approaches should be able to probe.

1.2.2.2 Age

Another interesting avenue for emotions research is the investigation of the relationship between emotional response and age. In particular relation to beer, there is a decline in consumption with age (Mintel (2013); Serra and Aranceta (2003)). It is known that taste and odour discrimination decline with age (Kaneda et al., 2000) and this could have an important impact on emotional response. Perhaps the key sensory properties of beer which drive positive emotional response show deterioration with age. More likely is a change in lifestyle and attitudes to beer with age, particularly relative to other alcoholic beverages, which decreases consumption and possibly affects emotional response. The literature suggests that researching the relationships between emotion and age could be difficult because there is a reported reduction in emotional expressivity in older adults (Gross et al., 1997), meaning that, even when the intensity of emotions are high, this may not be reflected in reports of the intensity of the emotional experience. Rather than being an evolutionary quirk, this has been interpreted as increased emotion regulation with age. There is also a reported trend for adults to experience more positive affect and less negative affect with age (Mroczek and Kolarz, 1998), so an increase in pleasant emotions may be seen in spite of decreasing consumption of beer. Again, this does not seem to be evolutionary but instead linked to sociodemographic variables associated with older people. It is important then that emotion measurement approaches are able to effectively explore the relationship between age and emotional response in consumers.

1.2.2.3 Culture

Emotional response could also inform about cross-cultural differences between consumers. It is now commonplace for products to compete at a multinational level. As a result, there is an increasing need for consumer research to be conducted across markets (Thomson and Crocker, 2013). Sensory and consumer science has already begun to explore the relationships between emotion and culture in aromas (Ferdenzi et al. (2011); Ferdenzi et al. (2013)) and beverages (van Zyl and Meiselman, 2015). Findings so far suggest that there are similarities across cultures, but also differences that could be significant for multinational products. Therefore, effective emotion measurement approaches should be able to reveal differences between cultures.

1.3 Measuring emotional response

According to the component process model, the elicitation of an emotion results in changes in the states of all or most of the five organismic subsystems (Table 1.2; Scherer (2005)).

Table 1.2 Relationships between organismic subsystems and the functions and components of emotion. Adapted from Scherer (2005).

Emotion function	Organismic subsystem (and major substrata)	Emotion component
Evaluation of objects and events	Information processing (CNS)	Cognitive component (appraisal)
System regulation	Support (CNS, NES, ANS)	Neurophysiological component (bodily symptoms)
Preparation and direction of action	Executive (CNS)	Motivational component (action tendencies)
Communication of reaction and behavioural intention	Action (SNS)	Motor expression component (facial and vocal expression)
Monitoring of internal state and organism–environment interaction	Monitor (CNS)	Subjective feeling component (emotional experience)

CNS = central nervous system; NES = neuro-endocrine system; ANS = autonomic nervous system; SNS = somatic nervous system

Due to the multiple processes that comprise emotions, it is only by assessment of all component changes that a comprehensive measure of an emotion can be obtained (Scherer, 2005). However, most measures of emotion only access one of these subsystems. Current methods can be categorised into four main approaches: (1) self-report (verbal and non-verbal); (2) implicit reactions; (3) physiological reactions; and (4) functional brain imaging. The following sections describe and exemplify these approaches whilst detailing the emotion components that are addressed by each.

1.3.1 Self- report measures

Self-report requires consumers to consciously select their emotional response to a stimulus. Conscious emotions are emotions of which the person is aware and has access to. Using self-report, rich data about complex and specific emotions from the appraisal and subjective feeling components of the

component process model (Scherer, 2005) can be obtained (e.g. nostalgia, adventurousness).

1.3.1.1 Verbal self-report measures

Psychological research into emotions has historically concerned mood and anxiety disorders (e.g. depression, obsessive compulsive disorder, etc.) and the associated emotion questionnaires reflect this. Profile of Mood States (POMS; (McNair et al., 1971)), Multiple Affect Adjective Check List (MAACL; Zuckerman (1960)), and Positive and Negative Affect Scale (PANAS; Watson et al. (1988)) represent three of the most applied clinical and academic questionnaires. POMS measures emotion on six dimensions: anger–hostility, vigour–activity, tension–anxiety, depression–dejection, fatigue–inertia, and confusion–bewilderment. Note that all dimensions are concerned with negative emotion, reflecting the underlying clinical nature of this questionnaire. MAACL has three negative scales (anxiety, depression, hostility) but also two positive scales (sensation seeking and positive affect). PANAS measures just two dimensions, positive affect and negative affect and it is claimed that the questionnaire provides independent measures of each.

The differences between these questionnaires in their measurements of positive and negative affect are significant when considering their application to sensory and consumer science. This is because Desmet and Schifferstein (2008) revealed a so-called "hedonic asymmetry" in that participants who were asked to describe food experience showed a positive bias. That is to say, consumers used largely positive as opposed to negative words to describe their emotional experience of food. This comes as no surprise as food consumption

is a fundamentally positive experience (Gibson, 2006). Although attempts have been made to utilise the established psychological questionnaires (e.g. Kuesten et al. (2014) used PANAS), approaches have been developed which utilise emotion language more appropriate for use by consumers rather than for the intended clinical use of the previously described questionnaires.

The EsSense Profile has proved a popular choice for verbal-self report of emotions in its short history (King and Meiselman (2010); King et al. (2010); Ng, Chaya, and Hort (2013); Jaeger and Hedderley (2013); King et al. (2010); Jaeger et al. (2013); Piqueras-Fiszman and Jaeger (2014a); Piqueras-Fiszman and Jaeger (2014b); Chaya et al. (2015)). EsSense Profile is a general use pre-determined emotion questionnaire to test food with consumers in person or on the internet in a consumer context. Importantly, it is aimed at product category users and product users who typically like the product (King and Meiselman, 2010). EsSense Profile includes a basic lexicon of 39 emotions which can be modified for specific product categories. Mainly positive emotions are measured, consistent with the aforementioned hedonic symmetry associated with emotional response to food. Data collection may include checklists (check all that apply; CATA) or visual analogue scales, the latter of which is more detailed when comparing small product differences (King et al., 2010). The ballot asks how respondents "feel" while evaluating a stimulus (according to Robinson and Clore (2002), self-reports of current emotional experiences are likely to be more valid than self-reports of emotion made after the event).

An alternative to pre-determined lexicons is to instead make use of a product-specific emotion lexicon. One example of the application of such an approach

is the development of the Geneva Emotion and Odor Scale (GEOS) by Chrea et al. (2009), which was generated in response to odours. GEOS consists of 36 affective terms which were reduced with participant input from a much larger pool of words derived from the literature. Participants' responses to a number of representative odours were then subjected to factor analysis to give a set of six summary scales, each of which included similar terms. Therefore, the ratings of the 36 terms were able to inform about the responses of participants to a set of odours in the defined factorial structure of emotional response to odours. In order to allow an easier and quicker test for the respondent. Porcherot et al. (2010) developed a shorter version of the GEOS questionnaire with a reduced number of measurement scales. Participants rated a series of three representative terms for each of the six GEOS dimensions instead of rating the 36 terms individually for each sample (ScentMove). Despite the fact that only 50% of the evaluations were required by each consumer as compared to the original GEOS questionnaire, similar product information was obtained by GEOS and ScentMove.

At present, verbal self-report represents the most active area of emotion research in sensory and consumer science. Whilst some methods have gained an early foothold (e.g. EsSense Profile), the area is still at the development stage, with many modified (e.g. GEOS and ScentMove) and new approaches being published all the time with the aim to create more effective emotional measurement methods.

1.3.1.2 Non-verbal self-report measures

Non-verbal self-report measures avoid the necessity for language, allowing for potential cross-cultural application (although validation is required that cultures agree on the emotions).

One example is the Product Emotion Measurement Instrument (PrEmo; Desmet et al. (2000)), in which basic emotions (7 positive and 7 negative) are represented non-verbally as expressive cartoon animations with dynamic facial, bodily and vocal expressions. During the self-running procedure, consumers are presented with stills of the 14 emotion animations which are activated by clicking on them. It is the respondent's task to rate to what extent they feel that particular emotion to a stimulus on a 3-point scale ("I do feel the emotion"; "to some extent I feel the emotion"; and "I do not feel the emotion") that appears by the animation.

Another example is Mood Portraits (Churchill and Behan, 2010). Respondents are asked to sample the stimulus and experience the mood (emotion using the definition of Scherer (2005) used throughout this thesis) evoked and then select up to five pictures (e.g. mother and child laughing and playing, a woman deeply relaxed) from a visual library of pictures (screened for its emotional content by consumers) that evoked the same or similar mood.

Overall, non-verbal self-report appears to be an effective but relatively little researched approach to measuring emotion. This is perhaps due to one of two reasons. First, a richer response is potentially obtained from verbal self-report in that language is able to be more specific. Secondly, it may simply be that sensory and consumer science practitioners feel less comfortable with the

unfamiliar approach of non-verbal self-report. This same factor may also influence the lack of research relative to verbal self-report in the following approaches.

1.3.1.3 An important note about the cognitive component of emotion

Whilst cognition forms a component of emotion according to Scherer (2005), it is important to note that an emotion may not need to be elicited at all for a person to be able to give a response about what type of emotion might be associated with a stimulus. For example, if a person were to be asked how they would feel if they won the lottery, the emotions associated with winning the lottery (and changes within all the relevant organismic subsystems) do not need to be directly experienced for that person to give a response like "excited" or "happy".

Thomson et al. (2010) described this in terms of emotional 'conceptualisations'. Conceptualisations are constructs created in the mind which allow the interpretation, understanding and assignment of meaning to our experiences. The authors put forward that most emotional measurement tools access emotional conceptualisations as opposed to emotional 'consequences' (which would be described by the component process model (Scherer, 2005) as the cognitive component of an elicited emotion).

The implication of this is that emotional responses measured by self-report may or may not be capturing experienced emotion. Instead, respondents may instead be reporting emotional conceptualisations associated with the stimulus.

1.3.2 Implicit reactions

In his 1872 book The Expression of the Emotions in Man and Animals, Charles Darwin wrote that "...the young and the old of widely different races... express the same state of mind by the same movements" (Darwin, 1872). This observation forms the basis of implicit measures of emotion where each emotion is associated with a particular pattern of expression (Ekman, 1994). The measurement of these patterns of expression reflects the communication of reaction and behavioural intention function of the component process model (Scherer, 2005). This pattern is particularly evident on the face and it is claimed that, in some cases, facial expression provides a means of communication of emotions that is even more effective than verbal expression (Etcoff and Magee, 1992). Facial expressions can be measured in a number of ways:

- Facial Action Coding System (FACS; Ekman and Friesen (1978)) – Requiring approximately 100 hours of self-instruction to encode the movements of individual facial muscles and categorise expressed emotion.

- Electromyography (EMG) – measuring the electrical activity associated with the activation of specific facial muscle groups

    - Hu et al. (1999) showed that negative hedonic sensations were associated with higher EMG activity in the levator labii muscle region, whilst positive hedonic sensations were associated with lower EMG activity in this same muscle region.

- Software packages – using a camera to recognise different facial expressions (but only 7 basic emotions, 1 positive - happiness)
    - Danner et al. (2014) applied FaceReader (Noldus Information Technology, Wageningen, Netherlands), to the measurement of emotional response to orange juices, successfully discriminating between samples.

However, the methods are restricted in the range of emotions (FACS to a lesser extent but this is significantly more labour intense than the other methods) and it is also difficult to measure simultaneous emotions. Furthermore, mainly negative emotions are measured by FaceReader, whereas people overwhelmingly use positive rather than negative words when describing food (Desmet and Schifferstein, 2008). Perhaps the most important consideration in a food and beverage sampling context is that motor artefacts, caused by eating and drinking, can be easily misinterpreted as reported by Danner et al. (2014) when using FaceReader to measure emotional response to orange juice. These artefacts are only expected to increase with solid samples.

1.3.3 Physiological reactions

Physiological reactions result from changes in the autonomic nervous system (ANS), central nervous system (CNS) and neuro-endocrine system (NES) that accompany emotions and represents the neurophysiological component of the component process model (Scherer, 2005). As with implicit reactions, this approach is language independent and, as a result, can be implemented cross-culturally. There are a number of physiological parameters that can be measured to indicate emotional state:

- Skin parameters (potential, resistance, blood flow, temperature)
- Cardio-respiratory parameters (electrocardiography (ECG), instantaneous respiratory frequency, instantaneous heart rate)

It has also been identified that pupil size increases when people are exposed to positive and negative stimuli (Partala and Surakka, 2003).

Rousmans et al. (2000) simultaneously and continuously measured the emotional reactivity associated with sweet, sour, bitter, and salty tastes using five ANS parameters: skin potential, skin blood flow, skin temperature, skin resistance, and instantaneous heart rate. The innate-accepted sweet taste induced the weakest ANS responses whereas the unpleasant tastes (salty, sour and bitter) induced stronger ANS responses, with the innate-rejected bitter taste the strongest.

Physiological methods have been shown to able to go beyond mere emotional reactivity as studies have implicated patterns of ANS response associated with distinct emotions (see Kreibig (2010) for a review). In addition, it has also been shown the physiological measures are able to differentiate between the responses of men and women. Robin et al. (2003) used the same five ANS parameters as Rousmans et al. (2000) to study the influence of gender on primary tastes. A similar distribution of basic emotions was associated with each primary taste for men and women for sweet, bitter and control solutions but differences were observed for salt and sour solutions.

1.3.4 Functional brain imaging

Functional brain imaging methods measure changes in brain activity associated the motivational changes produced by appraisal results in the component process model (Scherer, 2005), in particular action tendencies with associated neural signatures in their respective motor command circuits.

Electroencephalography (EEG) is a lightweight, compact, and portable system for measuring the synchronous firing of large populations of neurons (therefore the method has low spatial acuity). The frequency and amplitude of waves are related to different brain activities and can be linked to emotion. For example, higher regularity of the rhythm of frontal alpha-waves is associated with positive emotions. Using EEG, Kaneda et al. (2011) found that the aromas of essential oil extracted from Saaz hops as well as ester aromas exhibited a significant relaxing effect (lowering of arousal) on subjects. The method is susceptible to movement artefacts, meaning the alpha-waves of many subjects need to be measured in order to attain statistical power.

Magnetoencephalography (MEG) offers higher spatial resolution than EEG as it measures magnetic, as opposed to electric, fields which are less distorted by living tissue. The combination of high spatial and temporal resolution is ideal for the study of emotion as the timing of activation of specific brain sites proves important, as observed by Leon-Carrion et al. (2006). However, the method is expensive as it requires a very low noise environment or magnetic shielding.

Any modern discussion of functional brain imaging would be incomplete without mentioning functional magnetic resonance imaging (fMRI). Interest in

the application of fMRI to study emotion in neuroscience has exploded over the last 20 years (see Phan et al. (2002) for an early review). In a food and beverage context, fMRI has practical issues as it is difficult to present subjects with samples due to the participants being recumbent in a confined space. However, it is only a matter of time until the method is applied to better understand the relationship between neural systems linked to emotion and sensory-driven products.

Taking functional brain imaging as a whole, a great deal of progress has certainly been made in identifying the neural substrates of emotion. Nevertheless, this type of research is still relatively new and there is a lot of ground to cover. Some neural associations with basic emotions may be possible to identify at present but results are inconsistent (Barrett, 2006).

1.3.5 Concluding emotion measurement approaches

It is clear that the richest emotional response data is obtained using self-report because every emotion for which there is a word can be measured. As a relatively young area of research, methods have yet to become established, with current methods regularly being refined and new methods proposed. Although the literature thus far has mainly focussed on verbal self-report, non-verbal measures also offer advantages, particularly in cross-cultural application. However, any self-report method, by its very nature, is only able to measure conscious responses, whereas our reactions to food and beverages may be largely unconscious (Thomson et al., 2010). That is to say, components of emotion beyond awareness and access may be important when considering consumers' responses to products. Indeed, unconscious elements

of emotion have been shown to drive behaviour (e.g. Berridge and Winkielman (2003)). Implicit and physiological methods access unconscious components of emotional response, although they are limited in the parameters that can be measured. Functional brain imaging has the potential to be a powerful tool in the measurement of emotional response, although, at present, it is difficult to interpret the data in a meaningful way with regard to a consumer's emotional experience of a product.

1.4 Current understanding of the relationship between sensory properties and emotional response

The distinction between a stimulus and its evoked sensory property (or properties) is an important one. Take the example of colour vision. Light with a wavelength of 620-750nm (stimulus) appears red to those with normal trichromatic vision (sensory property). However, there is nothing inherently 'red' about that range of wavelengths. Indeed, those with red-green colour blindness perceive red from wavelengths of light that would be described as green by the majority of the population. Through receptor transduction and subsequent neural processing, a stimulus is converted into a meaningful representation to the organism perceiving it (e.g. red fruit stands out against the green foliage background). Therefore, there is nothing 'sweet' about the stimulus sucrose until it interacts with and depolarises a taste receptor cell, sending the information to the relevant brain regions through the brainstem and is perceived by the organism. Thus, it is the interaction between an organism's sensory system and the stimulus that imbues the stimulus with a 'sensory property'. In terms of food and beverages, a product does not possess sensory properties unless it is perceived by a consumer.

Some sensory properties are associated with innate emotional responses. Rosenstein and Oster (1988) showed that neonates experience positive affect in response to sweet solutions and negative affect to bitter solutions. This has an evolutionary basis in that it is advantageous to the survival of neonates to be motivated to consume their mother's sweet milk and to avoid the consumption of potentially toxic bitter compounds. Studies have also explored beyond general valence. Using solutions representative of four of the five primary tastes (bitter, salty, sour, and sweet), Robin et al. (2003) measured six basic emotions: happiness, surprise, sadness, fear, disgust, and anger. Sweetness was associated with happiness and surprise, whereas the bitter solution was associated with anger and disgust. Salty and sour solutions were associated to some degree with all six emotions, possibly reflecting more variable taste associations and evolutionary functions.

The previous evidence shows that different taste qualities elicit different emotional responses but it has also been observed that other types of sensory property give rise to different emotional responses. In commercial chocolate, Thomson et al. (2010) showed that vanilla flavour and brown appearance were associated with the term 'sensual', whilst creamy mouthfeel and sweetness were associated with fun, comforting and easy-going. Ng et al. (2013) identified a link between normal added sugar blackcurrant squash drinks and positive emotions. However, this was not the case for niche added sugar samples, suggesting that the emotions were being driven by other sensory properties. These flavours present in commercial products are less likely to have a strong evolutionary basis for their associations with particular emotions and, more likely, these associations are learnt through experience. In fact,

experience has an important bearing on basic tastes; for example, exposure to bitter-tasting compounds has been shown to lead to increased hedonic ratings (Stein et al., 2003), explaining the large consumption of bitter products such as coffee or cruciferous vegetables.

To the author's knowledge, no research has been published linking the manipulation of a product's individual sensory properties with emotional response. This thesis describes such an approach through the experimental control of sensory properties of beer. This is a particularly interesting product to study because of the range and complexity of sensory characteristics across modalities (taste, aroma, mouthfeel), in spite of very few raw materials.

1.5 Origins of the sensory properties of beer

This section describes the process of converting beer's raw materials into the final product and how each stage in the process has an effect on the product's many and varied sensory properties.

1.5.1 Introduction to the brewing process

The main ingredients from which most beers are brewed are malted barley, water, yeast, and hops. The processes that convert these raw materials into beer can be broadly grouped into malting, wort production, and fermentation. Maturation, finishing, and packaging are the final steps in the process. A simplified representation of the main stages of this process is shown in Figure 1.2 for reference.

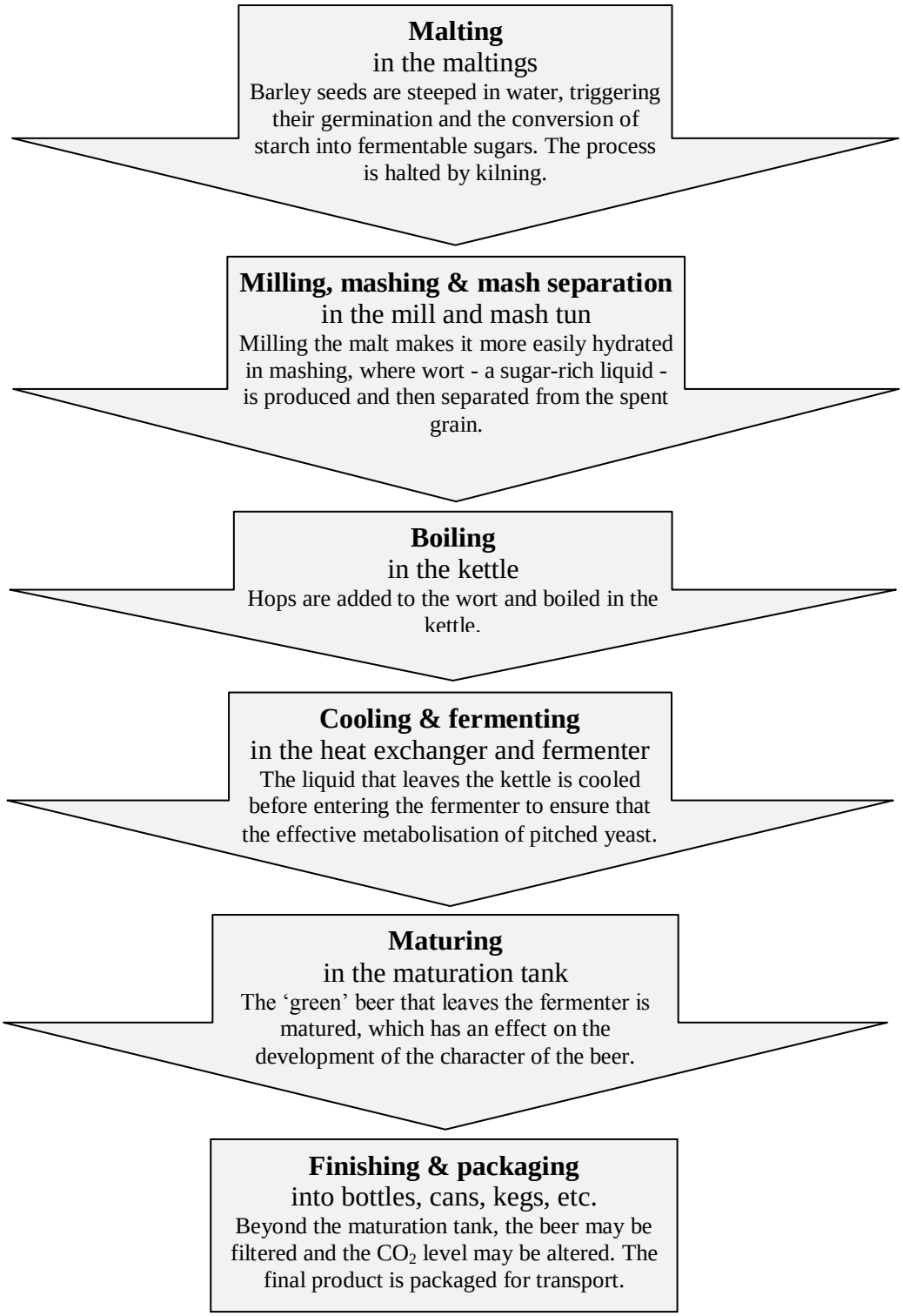


Figure 1.2 A simplified flow diagram of the main stages of the brewing process.

1.5.2 Malting

When discussing barley in a brewing context, it is the plant's grain that is being referred to. During malting, the kernels' embryos are hydrated by steeping the seeds in water. This activates hormones which trigger the production and secretion of enzymes, starting the partial breakdown of the starchy endosperm. This starch is the source of fermentable sugars in brewing.

Before excessive tissues associated with germination are produced, the process is stopped by heating the grain in a process called kilning. The aim is to reduce the moisture level in the grain, thereby halting the metabolism of the barley and stabilising the product. Lager-style beers are generally kilned to lower temperatures than ales. High kilning temperatures result in a darker colour product with more complex flavours. This process is the origin of the malty characteristics associated with beer. Dimethyl sulphide (DMS) can be produced from malt-derived precursors at this stage, although it can also be caused by bacterial infection during fermentation. DMS is considered a characteristic property of some beers at a relatively low concentration but can also be an off-flavour (Bamforth, 2009).

1.5.3 Wort production

First, the malt is milled so as to be easily hydrated. This is important for the activation of enzymes and solvation of substrates during the mashing stage, where the malt is mixed with water under controlled conditions (e.g. time, temperature, pH) in the mash tun to start the hydrolysis process. The resulting sugar solution is called wort, which is separated from the spent grains in the mash separation stage.

From here, the wort enters the kettle for hopping. Hops are perennial climbing plants. The important components of hops for the brewing process are located in the lupulin glands of the cones. The use of whole hop cones is now rare and most commonly, pelletized hops are used although resin and oil extracts can also be used.

It is in the kettle that hops are added. Hops contain resins (the most important of which are α-acids) that are extracted in the wort boil and isomerized into more soluble forms (iso-α-acids). These acids are perceived to have bitter sensory characteristics. Hops also possess a complex mixture of essential oils which provide the wide range of different hoppy characters associated with beer (e.g. floral, fruity).

The timing of the addition of hops during the boil is important. Hops added at the beginning of a boil will lose virtually all their oils through evaporation, therefore contributing bitterness but hardly any flavour character. As a result, it is commonplace for a proportion of hops to be held back for addition during the final few minutes of the boil in order to contribute flavour. This is known as late hopping. An alternative and more traditional procedure for imparting hop-derived flavours is to ‘dry hop’ the product by adding hop cones to the cask.

1.5.4 Fermentation

Fermentation is primarily concerned with the conversion of carbohydrates (approximately 70% of which are fermentable sugars: maltose, glucose, fructose, sucrose, and maltotriose; Gibson (2006)) into alcohol, although it is also about producing a subtle mix of flavours. Esters, such as isoamyl acetate

which imparts an ‘artificial banana’ flavour to the beer, are typical of flavours produced during fermentation. Before yeast can be pitched for fermentation, the liquid must first be cooled to such a temperature that yeast can metabolise effectively. This temperature is different depending on the genus/species of a yeast strain and the intended characteristics of the final product. Brewing yeasts are divided into two categories: Saccharomyces cerevisiae (associated with the brewing of ales) and Saccharomyces pastorianus (associated with the brewing of lagers). Ale yeast strains ferment at a higher temperature (typically 18-22°C) than lager yeast strains (typically 6-15°C) (Bamforth, 2009). Any unfermented sugars contribute to the sweetness of the final product.

The metabolism of yeast produces by-products. Some are desirable and characteristic of most beers. For example, carbon dioxide contributes to the perception of ‘tingliness’ and ‘bubbliness’ in the final product. Others are largely undesirable, such as diacetyl (Bamforth, 2009) which contributes a ‘buttery’ sensory characteristic to the final product. Yeast is able to ‘mop up’ the diacetyl again, converting it into compounds which do not have the same intense aromas. Another undesirable flavour produced during fermentation is acetaldehyde, a precursor of ethanol (Bamforth, 2009). To reduce levels of both diacetyl and acetaldehyde, fermentation time must be increased, which adds to the cost of brewing.

1.5.5 Maturation, finishing and packaging

The ‘green’ beer must be cooled in order to stabilise the product. It is then matured in order to, among other things, develop the flavour of the beer. After

this period of storage, the beer may be clarified using filters. Finally, the level of CO_2 can be altered before packaging.

Packaging is an important consideration for stability of the product during storage before reaching the consumer. For example, the exposure of beer to certain wavelengths of light can lead to photolysis of iso-α-acids and the production of 3-methyl-2-butene-1-thiol, commonly called 'light struck' by brewers (Stephenson and Bamforth, 2002) and described as having 'skunky' or 'freshly brewed coffee' aroma. This undesired characteristic can be avoided by not using clear bottles and instead using green or, better still, brown bottles. The coloured glass filters out the wavelengths of light that cause the photolysis. Alternatively, beers that use clear packaging (for brand or aesthetic purposes) can make use of tetrahydroiso-α-acids as bittering agents for the reduced susceptibility to photolysis (Briggs et al., 2004).

1.6 Research aim and objectives

The overarching aim of this research was to build upon previous studies to produce an effective emotion measurement approach which allowed the investigation of the relationships between sensory properties of beer and consumer emotional response. In order to achieve this, a number of key objectives were established, as listed below:

1. To develop and validate sensorially-distinct beer samples with specific sensory attributes for use in subsequent studies. This is addressed in Chapter 2.
2. (a) To create an approach for the development of a reduced product-specific consumer-led lexicon; and (b) to apply the reduced consumer-

led lexicon to (i) discriminate across a range of beer samples specifically designed to elicit specific sensory properties, and (ii) reveal differences in emotional response across different consumer segments related to gender and age. This is described and discussed in Chapter 3.

3. To compare the relative efficacy of the reduced emotion lexicon against the full lexicon upon which it was based by assessing (a) the relative discriminability between samples by each approach, and (b) the ability of the reduced form emotion categories to differentiate between the responses of different consumer groups as compared to the full form terms. Further objectives were (c) to explore the potential effect of form reduction on halo dumping, and (d) to assess the effectiveness of the modified cluster analysis for grouping terms that elicit similar patterns of response to one another as well as to the reduced form emotion category to which they belonged. This is the subject of Chapter 4.
4. (a) To create a Spanish reduced consumer-led emotion lexicon for use in (i) discriminating across a range of beer samples designed to elicit specific sensory properties, and (ii) revealing differences in emotional response across different consumer segments; and (b) to compare and contrast emotional response to the selected sensory properties of beer between UK and Spanish beer consumers by exploring the similarities/differences in (i) how emotion categories were associated with specific sensory properties of beer across the two cultures, and (ii)

the extent to which the respective lexicons discriminated across consumer segments. This forms the basis of Chapter 5.

Chapter 2: Selecting and characterising samples

2.1 Introduction

Beer is a product category ripe for research into the association between sensory properties and emotional response due to the broad and diverse range of sensory properties associated with products belonging to this category. Furthermore, it has previously been demonstrated that beer is an emotive product with differential responses dependent on the product (Chaya et al., 2015). This thesis focuses on lager-style beer and its associated sensory properties because this style dominates beer sales in the UK and Spanish markets (Euromonitor, 2014a, Euromonitor, 2014b) which are studied in this thesis. This chapter describes the characterisation of a number of samples which were designed to be controlled in chosen sensory properties. It was important that the sample manipulations led to changes in the perception of the relevant sensory properties because these samples would subsequently be used to measure the relationship between sensory properties and emotional response. Therefore, it was necessary to recruit and train a beer-specific sensory panel and a significant part of this chapter describes this lengthy but important process. Through statistical analysis of the panel's responses, it is shown that the 'base' and 'spiked' samples significantly differed in the pertinent sensory properties. As a result, the effects of the differences in sensory properties between samples could be investigated in subsequent studies using these samples.

2.1.1 Objective

The main objective of the work presented in this chapter was to develop sensorially-distinct beer samples with specific sensory properties. This was in

order that they could be used in subsequent studies as a basis for the investigation of the relationship between sensory properties of beer and consumer emotional response.

2.2 Selection of beer samples for sensory analysis

The fourteen selected sensory properties (Table 2.1) were chosen in order to reflect a wide range of variation in beer that can result from malting (e.g. maltiness, DMS), wort production (e.g. hoppiness, bitterness), fermentation (e.g. alcoholic flavour, isoamyl acetate, etc), and storage (light struck). Some properties were characteristic of beer (e.g. Bitterness, hoppiness, etc.) whilst others were anecdotal drivers of emotional response (e.g. off-flavours like DMS or acetaldehyde). The sensory properties also span a range of modalities (i.e. taste, aroma, flavour, mouthfeel).

Table 2.1 Sensory properties of beer and the treatments of the 'manipulated' samples.

Sensory property	Control sample	Manipulated sample	Intermediate sample (for evaluation of panel discriminability)
Tingliness	Commercial lager decarbonated and recarbonated to ~1.6 volumes (2psi at 4.4°C)	Commercial lager decarbonated and recarbonated to ~4 volumes of CO_2 (26psi at 4.4°C)	Commercial lager decarbonated and recarbonated to ~2.5 volumes of CO_2 (13psi at 4.4°C)
Bubbliness			
Alcoholic flavour	Commercial non-alcohol lager	8% ethanol added to commercial non-alcohol lager	4.7% ethanol added to non-alcohol commercial lager
Body			
Warming			
Sweetness	Commercial lager	25g dextrose/litre commercial lager	15g dextrose/litre commercial lager
Acetaldehyde		45mg Aroxa acetaldehyde/ litre commercial lager	30mg Aroxa acetaldehyde/ litre commercial lager
Bitterness		25mg Aroxa iso-α-acids/litre commercial lager	15mg Aroxa iso-α-acids/litre commercial lager
Isoamyl acetate		10.5mg Aroxa isoamyl acetate/litre commercial lager	7mg Aroxa isoamyl acetate/litre commercial lager
DMS		600µg Aroxa dimethyl sulphide/litre commercial lager	450µg Aroxa DMS/litre commercial lager
Hoppiness		750µg Aroxa kettle hop extract/litre commercial lager	500µg Aroxa kettle hop extract/litre commercial lager
Maltiness		720µg Aroxa 2-acetyl pyridine/litre commercial lager	540µg 2-acetyl pyridine/litre commercial lager
Diacetyl		390µg Aroxa diacetyl/litre commercial lager	312µg Aroxa diacetyl/litre commercial lager
Light struck		300ng Aroxa 3-methyl-2-butene-1-thiol/litre commercial lager	100ng Aroxa light struck/litre commercial lager

Differences in sensory properties were achieved by manipulating base beers (Table 2.2 gives the brand profiles of the two base beers). Two base beers were selected, a normal strength commercial lager (chosen due to its general

low levels of sensory characteristics making it easier to 'spike') and a non-alcohol commercial lager. The non-alcohol lager was chosen to explore the effect of alcohol content. (this could not so easily be done with the normal strength commercial lager because perceptible increases in sensory properties associated with alcohol could only be achieved when the alcohol volume extended beyond a typical range for commercial lagers). By 'spiking' these base beers (Table 2.1), individual sensory properties could be altered, offering greater experimental control than using a range of commercial lagers. Some of the manipulated attributes were associated with just one major compound (e.g. the chemical compound acetaldehyde contributes a sensory property frequently labelled as 'green apples'). There are a few examples in this thesis of sensory properties that can be dependent on the contribution of a number of compounds, namely, sweetness, bitterness, maltiness, and hoppiness. For example, sweetness is not associated with just one compound but a number of unfermented sugars like maltose, glucose, fructose, and sucrose. In such cases, the relevant flavour standards developed by Aroxa (Cara Technology, Leatherhead, UK) were referred to. This is with the exception of sweetness where dextrose (Myprotein, UK) was used in preference to sucralose which is the flavour standard supplied for sweetness by Aroxa. This was the result of informal preliminary tests wherein it was found that dextrose contributed a quality of sweetness more typical of beer than sucralose (or, indeed, glucose, fructose, and sucrose). Maltose offered a comparably beer-typical sweetness to dextrose but was not selected for this project to avoid potential confusion between attributes due to its close association with maltiness.

Table 2.2 Brand profiles of the two base beers used for manipulation throughout this thesis.

	Commercial lager	**Non-alcohol commercial lager**
Real extract	3.7 Plato	5.6 Plato
Alcohol by volume	4.7%	0.5%
International Bittering Units	11mg/l	22mg/l
CO_2	2.7 volumes	2.6 volumes
European Brewery Convention colour units	6.4	7

Levels of each attribute (Table 2.1) were chosen to reflect typical variation within lagers but to also be easily discriminated by consumers when used in subsequent emotions research. In some instances this was easily achieved (e.g. CO_2 volume and alcohol volume) but in other instances, a considerable amount of sensory panel work was required to define the necessary level of attribute. For the Aroxa compounds, a level nine times greater than the detection threshold in beer as identified by Aroxa was used as the starting point. In some cases this was increased to ensure perceptible differences between base and manipulated samples. The most extreme example of this was the DMS sample; the level of DMS in this sample was 12-20 times greater the detection threshold in beer. This level may not, however, be wholly accurate as the Aroxa preparation instructions state that the capsule should be added to the beer and presented immediately. For this research, both sensory and consumer, this was not practical due to resource limitations, meaning that samples were prepared up to three hours in advance. This had a definite effect on the perception of individual sensory properties in that they were far less detectable. As such, the preparation times of samples were carefully controlled.

The base and manipulated samples needed to be characterised in terms of their sensory profiles in order to establish that differences existed in the selected sensory properties. This was because the samples would be used as a basis to explore the relationship between individual sensory properties of beer and emotional response. Consequently, it was necessary to recruit and train a beer-specific sensory panel.

2.3 Sensory panel recruitment

The purpose of a sensory panel is to provide objective data on perception. This is achieved by training a group of assessors to become accurate and reliable in their measurements, whilst also tending to the panel average. A process of recruitment and screening was required in order to identify individuals with the potential to become precise measurement instruments for sensory evaluation. The key characteristics sought were:

- Sensory acuity (ability to detect, recognise and discriminate between stimuli)
- Descriptive and communication skills (ability to describe character and communicate intensity)
- Interpersonal skills (ability to co-operate as a member of a group)

Other important factors that were considered were health, general availability, and motivation. As it is undesirable for a panel to consist of less than 10 assessors (ISO 8586-1:1993), the aim was to recruit between 12-15 panellists to allow for drop-outs. Classified advertisements were placed in local newspapers and posters were put up in the surrounding community inviting

interested individuals to request an application form by email. The recruitment process was divided into three stages: (1) pre-screening, (2) screening, and (3) post-screening.

2.3.1 Pre-screening

Respondents to the advertisements were sent a questionnaire (Appendix A) by email. This was used to obtain background information about the candidates, including interest and motivation, attitudes to foods, knowledge and aptitude, health, ability to communicate, availability, and personality characteristics (see ISO 8586-1: 1993 for more detail). Information on whether applicants currently or ever have smoked was also recorded, although candidates were not excluded on this basis.

Together with the questionnaire, scaling exercises (Appendix B; redrawn from Meilgaard et al. (2007)) were sent to applicants in order to assess their ability to use scales. Five geometric shapes were presented and the task of candidates was to mark on a continuous line scale the proportion of the shape that was shaded (those invited to screening completed a further set of five scaling exercises in order to ensure that applicants were able scale accurately under controlled conditions).

Of the 109 applicants who submitted questionnaires and scaling exercises, 69 were selected to attend screening sessions at The University of Nottingham's Sensory Science Centre based on general availability (e.g. those in full time employment were excluded), health (e.g. those with hypertension were excluded), and descriptive ability (e.g. those that provided single word or single sentence responses to descriptive questions were excluded).

2.3.2 Screening

In total, 43 candidates completed the two screening sessions (exceeding the minimum recommended 40 recruited individuals needed in order to obtain a final panel of 10 selected assessors according to ISO 8586-1: 1993). A number of invited applicants did not attend the first session and others lacked interest after the first session. This was part of the reason that more than one session was conducted as it reveals those with a genuine interest and motivation to be panellists. The 43 candidates participated in a number of activities in order to assess their ability to detect, recognise, discriminate, and describe stimuli in addition to their communication skills. All tests took into account the intended application of the panel for the assessment of beer and made use of relevant materials. The activities were created to be deliberately challenging in order to avoid ceiling effects where all candidates perform well. In this way, it was easier to differentiate between potential panellists, although this had the effect of generating low scores on some tasks, even where performance is good relative to other candidates.

2.3.2.1 Detection of stimuli

It was important that the selected panellists did not have any impairment in colour vision because the panel was anticipated to continue their participation beyond this project and colour may play an important part in future studies. Although all candidates invited to screening indicated in the pre-screening questionnaire that they were not colour blind, it was necessary to check as this vision deficiency sometimes remains undiagnosed. Candidates completed the short version of the Ishihara test for colour blindness (Ishihara, 1972) in which six plates were presented. The task was to identify the numbers present in the

individual plates. According to the responses given, it could be determined if an impairment was present. No candidates scored below 100%, meaning all had normal colour vision.

Hypogeusia (reduced ability to taste) is relatively common in the general population (5% according to Welge-Lüssen et al. (2011)), so it was important to screen out any candidates who were unable to detect bitter and sweet tastes as these represent two significant sensory properties of beer. In order to assess candidates' ability to detect and recognise bitter and sweet compounds, applicants took part in absolute discrimination tests for taste (adapted from ISO 8586-1: 1993). Six samples were presented and it was explained that these samples may taste bitter, salty, sour, sweet, or of nothing at all (i.e. water). The candidates' task was to identify the taste, if any, of each sample by ticking the appropriate box on the associated Fizz Forms (Biosystèmes, Couternon, France) sheet. It was made clear that each sample would not contain more than one tastant. Two samples were tasteless (i.e. water), two were sweet (dextrose (Myprotein, UK) at 10g/l and 15g/l water), and two were bitter (Aroxa iso-α-acids at 3mg/l and 4mg/l water, Cara Technology, Leatherhead, UK). ISO 8586-1: 1993 recommends that successful panellists score 100% in this type of task. However, none of the 43 candidates achieved this (Activity 1 in Table 2.7). As observed by Welge-Lüssen et al. (2011), further investigation of the data showed that there were many instances of applicants confusing tastes, particularly bitter and sour. Therefore, candidates participated in further taste detection testing with samples at higher concentrations post-screening (see section 2.3.3.1) in order to better assess absolute discrimination.

2.3.2.2 Discrimination between stimuli

In the first of the two screening sessions, ranking tests were used to assess candidates' ability to discriminate the relative intensity of various attributes that vary between lager-style beers: colour intensity, sweetness, bitterness, hoppiness, and carbonation. Candidates were presented with three intensities of each attribute in water (see Table 2.3) and asked to rank them from lowest to highest intensity on a Fizz Forms printout. Successful candidates were expected to rank the lowest and highest intensity samples of each attribute correctly, with particularly sensitive candidates ranking the intermediate intensity sample correctly. For each ranking test, 1 point was given for every correct ranking, with half a point given where the intermediate sample was ranked in the wrong order with either the lowest or highest sample. This gave a maximum score of 5 points (100%; see Activity 2a in Table 2.7). Any other ranking received 0 points. During post-screening (see section 2.3.3.1), candidates carried out a very similar activity except in beer instead of water (and with adjusted levels of each attribute).

Table 2.3 Attributes and their levels in the ranking activity.

Attribute	Low (1)	Intermediate (2)	High (3)
Colour intensity	⅓ water, ⅔ commercial lager	Commercial lager	30µl red food colouring in 100ml commercial lager
Sweetness	10g dextrose/l commercial lager	15g dextrose/l commercial	20g dextrose/l commercial lager
Bitterness	3mg Aroxa iso-α-acids/l commercial lager	4mg Aroxa iso-α-acids/l commercial lager	5mg Aroxa iso-α-acids/l commercial lager
Hoppiness	125µg Aroxa kettle hop extract/l commercial lager	187.5µg Aroxa kettle hop extract/l commercial lager	250µg Aroxa kettle hop extract/l commercial lager
Carbonation	Commercial lager at ~4.2 volumes of CO_2	Commercial lager at ~4.5 volumes of CO_2	Commercial lager at ~4.8 volumes of CO_2

In the second session, rank rating tests were used. The same five attributes were assessed but this time four samples were presented within each attribute as opposed to three (see Table 2.4). Three of these four samples were identical to those presented in session 1 in order to allow for a test of replication and learning. Again, candidates were required to rank the samples in order from lowest to highest on a Fizz Forms sheet. Successful candidates were expected to be able to rank all of these attributes correctly (1 point) or, if not, only confuse samples that neighbour one another in their concentrations (half point). There was a maximum score of five points (100%; Activity 4 in Table 2.7). It was also expected that experience with the sample would improve performance in the replication of the three samples from session 1 (Activity 2b in Table 2.7) and this was scored in the same way as Activity 2a. Indeed, a number of candidates showed improved performance from session 1, demonstrating an ability to learn specific attributes with exposure, which is an important skill for panellists.

Table 2.4 Attributes and their levels in the rank rating activity.

Attribute	Low (1)	Intermediate (2)	Intermediate (3)	High (4)
Colour intensity	⅓ water, ⅔ commercial lager	Commercial lager	20µl red food colouring in 100ml commercial lager	30µl red food colouring in 100ml commercial lager
Sweetness	10g dextrose/l commercial lager	12.5g dextrose/l commercial lager	15g dextrose/l commercial lager	20g dextrose/l commercial lager
Bitterness	3mg Aroxa iso-α-acids/l commercial lager	4mg Aroxa iso-α-acids/l commercial lager	4.5mg Aroxa iso-α-acids/l commercial lager	5mg Aroxa iso-α-acids/l commercial lager
Hoppiness	125µg Aroxa kettle hop extract/l commercial lager	187.5µg Aroxa kettle hop extract/l commercial lager	218.75µg Aroxa kettle hop extract/l commercial lager	250µg Aroxa kettle hop extract/l commercial lager
Carbonation	Commercial lager at ~4.2 volumes of CO_2	Commercial lager at ~4.4 volumes of CO_2	Commercial lager at ~4.6 volumes of CO_2	Commercial lager at ~4.8 volumes of CO_2

After the ranking had been completed for each attribute, candidates were told to rate each of the samples on a scale with '1' and '10' as anchors. The sample ranked as the lowest intensity was automatically assigned a value on the scale of '1', whilst the sample ranked as the highest intensity was automatically assigned a value on the scale of '10'. Candidates then rated the relative magnitude of the two intermediates in relation to these. The purpose of the scaling aspect of this task was to introduce candidates to scaling sensory properties (using the completed shading exercises to illustrate how to do this). It also provided supplementary data (not shown) as candidates who confused samples on the ranking would be expected to rate them very similarly. If this

was not the case, then the candidate's suitability as a panellist would require extra consideration.

Discrimination between samples that varied in alcohol concentration was also assessed. In order to give candidates experience of a wider range of sensory approaches, triangle tests were used instead of ranking. Three triangle tests were completed and each included two of three possible ethanol concentrations: 0.49% (commercial low alcohol lager), 2.4% (ethanol (Merck Chemicals, Nottingham UK) added to commercial low alcohol lager), and 4.7% (ethanol added to commercial low alcohol lager). In each triangle test, two of the samples were identical and one was different. It was the task of candidates to identify the 'odd one out' and to indicate how the 'odd one out' was different to the other two (i.e. higher or lower in alcohol concentration). Successful candidates were expected to be able to discriminate between the samples correctly successfully indicate how the odd one out was different. One point was allocated for a correct sample identified as the 'odd one out', with a further point given if the direction of this difference was correctly identified, giving a maximum of 6 points (100%, Activity 3 in Table 2.7). Further evaluation of candidates' ability to discriminate alcohol concentration was made in post-screening (see section 2.3.3.2).

2.3.2.3 Recognition and description of stimuli

In order to assess candidates' ability to recognise and describe unknown stimuli, an odour recognition and description task was completed in which a number of aroma compounds were presented. Some aromas were characteristic of beer and some were not. A direct method of sample

preparation was employed by applying a few drops of the relevant compound to a cotton wool ball which was sealed in a small glass bottle. Candidates were instructed to accurately identify the aroma or if they could not, describe the aroma as precisely as they could on the provided sheet. The scoring system used by Hollowood (2002) was adapted for use in the assessment of responses (Table 2.5). Although a satisfactory success level is dependent on the materials, ISO 8586-1: 1993 recommends that candidates must score above 65% on these tests. In fact, only a few candidates scored higher than 65% (Activity 5 in Table 2.7), reflecting difficulty in assigning descriptors to the chosen aroma compounds.

Table 2.5 Example identifications and descriptions of compounds in the odour identification screening test. This list is not exhaustive but a summary of acceptable responses given by candidates. Table adapted from Hollowood (2002).

Compound	One point	Half point
Isoamyl acetate	Banana Pear drops	Fruity Sweet
Diacetyl	Butterscotch Butter	
Benzaldehyde	Marzipan Almonds	Sweet
Cis-3-hexen-1-ol	Green	
Dimethyl sulphide	Cabbage Sulphurous	Beetroot Aparagus Corn Seafood
Trans-cinnemaldehyde	Cinnamon	
Water	Nothing	

A similar recognition and description task was also carried out with flavours (i.e. samples were consumed) in order to determine how accurately candidates could identify and describe flavours presented individually in water (Table 2.6). All of the flavours selected were specifically related to beer. Flavour

recognition and description (Activity 6 in Table 2.7) drew out a wider range of performance between candidates than aroma recognition description, with a couple scoring 100% and three as low as 25%.

Table 2.6 Example identifications and descriptions of compounds in the flavour identification screening test. This list is not exhaustive but a summary of acceptable responses given by candidates.

Compound	One point	Half point
Isoamyl acetate	Banana Pear drops	Fruity Sweet
Diacetyl	Butterscotch Butter	
Acetaldehyde	(Green) apple Metallic	Sour
2-acetyl pyridine	Malty Biscuity	Horlicks
Dimethyl sulphide	Cabbage Sulphurous	Beetroot Aparagus Corn Seafood
Water	Water Nothing	

To assess descriptive ability specific to lager-style beer, candidates were presented with two commercially available lagers and were asked to objectively describe the sensory attributes of the samples in as much detail possible. The comments sheet was divided into appearance, aroma, flavour, texture, and aftertaste, each of which was explained to candidates (although it was not necessary that they stick to this format). The activity was limited to ten minutes. Responses were assessed in terms of their detail, relevance and objectivity (results not shown).

In the second session, candidates were provided with reminder samples of one of the commercial lagers as well as their description of the lager. After a few minutes of prompting their memories and adding anything they feel would be

relevant, the candidates participated in a group discussion lead by the panel leader where the key aromas, flavour, textures and aftertastes were discussed. Observations were made about how individuals interacted within the group. This included contribution to the discussion, ability to develop others' contributions, ability to communicate own experience of beer, etc.

A separate group discussion was also carried out in which a more general topic was discussed. Candidates were asked to individually make a note of a few meals which they felt were best accompanied by beer and by wine. They were told to consider if there are any particular types of each beverage that go well with certain meals (e.g. red wine with beef; stout with shepherd's pie). This allowed for further assessment of candidates' ability to communicate as well as consideration of their interest, motivation, and personality characteristics.

2.3.2.4 Screening results

Table 2.7 shows the performance of each of the 43 candidates that completed both screening sessions across activities. In deciding who to invite to join the panel, it was important to consider performance across all activities, meaning poor performance in one or two tasks was weighed against good performance in others (for example, candidate 11 performed averagely or above in most tasks except aroma recognition (Activity 5)). Supplementary data (not shown here) from pre-screening, scaling activities (shaded shapes task and sample intensity ratings), group discussion, and sample description were used where necessary to better inform the selection process. Ranking performance was generally quite good (Activities 2a and 2b), although performance was

Table 2.7 Screening results.

		Activity						
		1	**2a**	**2b**	**3**	**4**	**5**	**6**
Candidate	1	16.7%	70%	80%	66.7%	40%	66.7%	41.7%
	2	**33.3%**	**60%**	**60%**	**66.7%**	**60%**	**70.8%**	**50%**
	3	16.7%	60%	60%	0%	20%	25%	41.7%
	4	33.3%	70%	70%	0%	40%	58.3%	50%
	5	**66.7%**	**70%**	**60%**	**16.7%**	**40%**	**33.3%**	**66.7%**
	6	**33.3%**	**80%**	**90%**	**50%**	**60%**	**66.7%**	**41.7%**
	7	83.3%	70%	60%	33.3%	40%	50%	50%
	8	**33.3%**	**90%**	**80%**	**66.7%**	**40%**	**50%**	**50%**
	9	**33.3%**	**80%**	**80%**	**33.3%**	**70%**	**50%**	**33.3%**
	10	**50%**	**80%**	**70%**	**16.7%**	**40%**	**54.2%**	**41.7%**
	11	**83.3%**	**90%**	**70%**	**66.7%**	**40%**	**37.5%**	**50%**
	12	**83.3%**	**90%**	**60%**	**66.7%**	**60%**	**50%**	**50%**
	13	**50%**	**80%**	**80%**	**0%**	**50%**	**54.2%**	**41.7%**
	14	**50%**	**60%**	**60%**	**33.3%**	**40%**	**58.3%**	**66.7%**
	15	**50%**	**70%**	**80%**	**50%**	**50%**	**41.7%**	**33.3%**
	16	50%	90%	60%	16.7%	50%	25%	25%
	17	33.3%	70%	60%	16.7%	30%	37.5%	50%
	18	**33.3%**	**70%**	**70%**	**100%**	**50%**	**37.5%**	**37.5%**
	19	50%	70%	70%	33.3%	0%	33.3%	29.2%
	20	**66.7%**	**80%**	**100%**	**33.3%**	**60%**	**33.3%**	**50%**
	21	**33.3%**	**80%**	**60%**	**83.3%**	**40%**	**37.5%**	**50%**
	22	33.3%	70%	70%	0%	40%	41.7%	50%
	23	50%	90%	50%	16.7%	40%	33.3%	70.8%
	24	**83.3%**	**70%**	**70%**	**33.3%**	**20%**	**50%**	**33.3%**
	25	**33.3%**	**90%**	**70%**	**33.3%**	**40%**	**33.3%**	**50%**
	26	**83.3%**	**80%**	**80%**	**0%**	**40%**	**58.3%**	**58.3%**
	27	**50%**	**60%**	**70%**	**33.3%**	**40%**	**58.3%**	**83.3%**
	28	66.7%	50%	70%	33.3%	20%	33.3%	41.7%
	29	**50%**	**90%**	**80%**	**66.7%**	**20%**	**66.7%**	**100%**
	30	33.3%	70%	80%	0%	40%	50%	75%
	31	50%	30%	50%	33.3%	20%	66.7%	58.3%
	32	50%	80%	60%	0%	40%	45.8%	33.3%
	33	66.7%	40%	70%	50%	40%	16.7%	25%
	34	33.3%	50%	80%	0%	40%	41.7%	66.7%
	35	33.3%	60%	30%	33.3%	20%	16.7%	25%
	36	33.3%	70%	90%	66.7%	40%	41.7%	45.8%
	37	**66.7%**	**50%**	**80%**	**50%**	**40%**	**66.7%**	**58.3%**
	38	**66.7%**	**50%**	**70%**	**50%**	**40%**	**33.3%**	**100%**
	39	**33.3%**	**70%**	**80%**	**33.3%**	**60%**	**75.0%**	**45.8%**
	40	33.3%	70%	60%	33.3%	40%	25%	50%
	41	50%	90%	70%	33.3%	50%	41.7%	37.5%
	42	16.7%	40%	50%	33.3%	40%	50%	33.3%
	43	50%	60%	40%	0%	20%	33.3%	41.7%

Emboldened candidates were invited to post-screening

impaired when including an additional sample (Activity 4). It was clear that most candidates found the absolute discrimination task (Activity 1), ethanol triangle tests (Activity 3) and aroma (Activity 5) and flavour (Activity 6) recognition tasks particularly difficult. On this basis, modified tasks were included in post-screening in order to make a more informed decision about who to include in the final panel.

2.3.3 Post-screening

Twenty-two candidates were invited back the Sensory Science Centre for two paid sessions before the final panellists were selected. As well as providing an extra opportunity to assess the abilities of candidates, post-screening sessions were added to provide candidates with a more representative experience of the type of work that they would be carrying out as panellists. This allowed them to better inform their decision about committing to the panel.

The methodologies and materials at this stage reflected the work the panel would be carrying out, providing a more accurate indication of the candidates' potential. The general availability of candidates was compared at this stage and the days and times for regular future panel training sessions set. Seventeen candidates were able to attend the chosen days and times.

2.3.3.1 Detection of stimuli

At the screening stage, it was frequently found that candidates confused bitter with sour in the absolute discrimination of taste and, as a result, low scores were generally obtained in this task. This likely reflected a failure to recognise the taste upon detection and not bitter specific hypogeusia. Therefore, in the post-screening sessions, candidates were presented with three reference

samples: bitter (Aroxa iso-α-acids at 10mg/l water), sour (0.6ml lactic acid (Sigma-Aldrich, UK)/litre water), and sweet (30g dextrose/litre water) in order to familiarise themselves with the taste attributes. Following this, candidates were set the task of matching each of nine samples to one of the references. These samples varied in terms of their concentration of taste compounds (bitterness: 5mg, 7.5mg, 10mg Aroxa iso-α-acids/litre water; sourness: 0.2ml, 0.4ml, 0.6ml lactic acid/litre water; sweetness: 10g, 20g, 30g dextrose/litre water). The concentrations of bitter and sweet compounds were higher in post-screening than in screening to ensure that specific ageusias would be picked up. As with all post-screening tasks, data was acquired using Fizz computer software. Due to the higher levels of attributes, 100% correct responses were expected from successful candidates in this task. Fourteen of the seventeen candidates gave 100% correct responses, with the remaining three only erring on either the low bitter or low sour samples. Correct responses for the higher concentrations suggest that these candidates were not aguesic, although they may be less sensitive than other candidates.

2.3.3.2 Discrimination between stimuli

Many candidates had difficulties in discriminating between samples of varying hoppiness or carbonation in water during screening. The opportunity was taken in post-screening to further assess candidates' group descriptive ability. Samples of beer that were high in hoppiness and carbonation were discussed and the groups generated several relevant descriptors. Following this, hoppiness and carbonation were ranked. In contrast to screening, the attributes were present in beer as opposed to water to make it more relevant to the tasks panellists would be expected to perform. In addition, the range of sensory

properties in samples was increased in line with the range typical of commercially available lager (see Table 2.8). As in screening, accurate ranking was expected from successful candidates.

Table 2.8 Attributes and their levels in the post-screening ranking activities.

Attribute	Low (1)	Intermediate (2)	High (3)
Hoppiness	Commercial lager	375µg Aroxa kettle hop extract/litre commercial lager	750µg Aroxa kettle hop extract /litre commercial lager
Carbonation	Commercial lager at ~1.6 volumes of CO_2	Commercial lager at ~2.7 volumes of CO_2	Commercial lager at ~4 volumes of CO_2

It was also common for candidates to be poor at discriminating between alcohol contents during screening. As a result, the range of concentrations was increased for post-screening to reflect the range of concentrations in commercial lager (0.47% low alcohol lager; 4.7% low alcohol lager with added ethanol; 8% low alcohol lager with added ethanol). Candidates discussed a sample high in alcohol and generated a number of descriptive terms. The task was changed from triangle tests in screening to ranking in order to keep in line with the other tests the panel were completing during post-screening. Given the large differences, it was anticipated that the final panellists would be able to rank these samples correctly.

Candidates identified and described flavours in water during screening. Their task during post-screening was to rank samples of beer that varied in these flavour attributes. This ensured that they were able to discriminate between varying intensities of these stimuli in the complex matrix of lager. The range of samples within an attribute reflected the range in commercially available

lagers (see Table 2.9). Before carrying out the ranking, candidates were given samples high in each of the attributes and discussed a number of relevant descriptors. Consistent performance across all attributes was expected from successful candidates.

Table 2.9 Flavour attributes and their levels in the post-screening ranking activities.

Attribute	Low (1)	Intermediate (2)	High (3)
Isoamyl acetate	Commercial lager	5.25mg Aroxa isoamyl acetate/litre commercial lager	10.5mg Aroxa isoamyl acetate/litre commercial lager
Diacetyl		260µg Aroxa diacetyl/litre commercial lager	320µg Aroxa diacetyl/litre commercial lager
Acetaldehyde		22.5mg Aroxa acetaldehyde/litre commercial lager	45mg Aroxa acetaldehyde/litre commercial lager
2-acetyl-pyridine		540µg Aroxa 2-acetyl-pyridine/litre commercial lager	1.08mg Aroxa 2-acetyl-pyridine/litre commercial lager
Dimethyl sulphide		450µg Aroxa DMS /litre commercial lager	900µg Aroxa DMS /litre commercial lager

The discussion of attributes and descriptor generation before ranking hoppiness and carbonation, alcohol volume, and the five flavours allowed for candidate interactions to be assessed in a realistic setting. It was possible to observe how each candidate contributed to the generation of terms, both in putting forward their own ideas and building on the contributions of others. Ranking was scored in the same way as in the screening sessions with 1 point given for a correct ranking and half a point given if the intermediate sample was mixed up with either the lowest or highest sample, giving a maximum score of eight points (100%; Activity 7 in Table 2.11). Although no

particularly high scores were attained, a closer examination of the data (not shown) revealed that there were far fewer instances of no points being awarded for a ranking compared to screening. This was particularly impressive given the complexity of the sample (i.e. commercial beer) compared to in screening (i.e. water).

2.3.3.3 Recognition and description of stimuli

Aromas were presented to candidates during screening with the aim of identifying candidates that could describe and identify well. Aromas were also presented in post-screening, but instead with the aim of identifying candidates that could not detect compounds likely to be used as attributes (as well as some extra compounds that candidates struggled with during screening). The task was to match the aromas to descriptors. Eight aromas were presented and there were ten groups of descriptors (see Table 2.10). This task had the effect of producing a wide range of results across candidates (Activity 8 in Table 2.11).

Table 2.10 Descriptors of compounds in the odour matching post-screening test.

Compound	Descriptors	
Isoamyl acetate	Banana	Pear drops
Diacetyl	Butter	Butterscotch
Benzaldehyde	Marzipan	Almonds
Cis-3-hexen-1-ol	Freshly cut grass	Hedge cuttings
Dimethyl sulphide	Cabbage	Tomato sauce
Trans-cinnamaldehyde	Cinnamon	
Acetaldehyde	Green apples	Emulsion paint
2-acetyl-pyridine	Malty	Biscuity
No associated compound	Minty	Herbal
No associated compound	Rose petals	Floral

Having never been exposed to the 'light struck' attribute, it was important to present this to candidates before they were selected for the panel to ensure that they were able to detect it. Therefore, candidates performed three duo-trio tasks in which they only smelled the samples. They were required to identify which sample was the same as a reference. The sample containing the light struck attribute was the different sample in each case (300ng Aroxa light struck/litre commercial lager). It was anticipated that candidates would correctly match to the reference in all three duo-trio tests. Sensory adaptation to the attribute is anecdotally relatively quick, which the candidates had been warned of. In spite of this, a fair number of candidates identified the odd one out in all duo-trio tests (Activity 9 in Table 2.11) Supplementary descriptive data was also generated at this stage as candidates were asked to describe how the odd one out was different to the other two samples.

Table 2.11 Post-screening results.

		Activity		
		7	**8**	**9**
Candidate	**2**	**68.8%**	**50%**	**33%**
	5	**50%**	**38%**	**67%**
	6	**50%**	**50%**	**33%**
	9	**62.5%**	**38%**	**100%**
	10	37.5%	63%	67%
	12	**56.3%**	**63%**	**100%**
	15	50%	25%	33%
	18	43.8%	50%	67%
	20	**43.8%**	**50%**	**100%**
	21	**56.3%**	**38%**	**67%**
	24	43.8%	25%	67%
	25	37.5%	25%	67%
	26	**68.8%**	**25%**	**67%**
	27	**68.8%**	**25%**	**100%**
	29	**62.5%**	**63%**	**100%**
	37	**43.8%**	**88%**	**33%**
	38	**62.5%**	**13%**	**100%**

Emboldened candidates were invited to join the panel

2.3.3.4 Post-screening results

Table 2.11 shows that performance was generally high in post-screening, particularly compared to screening, reflecting more refined tests and learning on the part of the candidates. Ranking (Activity 7) was found to be consistently more difficult for some compounds (e.g. 2-acetyl pyridine) across candidates. Aroma recognition (Activity 8) yielded very variable performance, whilst detection of the light struck compound in the duo-trio tests (Activity 9) was generally high. Performance across screening and post-screening was taken into account when making the final panel selection. Twelve candidates were invited to join the final panel based on their pre-screening, screening and post-screening results. Ten individuals accepted this invitation.

2.4 Panel training

Before and during training, panellists were instructed to be objective at all times, setting aside their likes and dislikes. They were also directed not to use perfumed cosmetics (including hand soaps) before sessions as well as avoiding tobacco or strong tastes for at least one hour before a session (as per ISO 8586-1:1993 guidelines). All sessions (on average, two consecutive days per week) took place at The University of Nottingham's Sensory Science Centre which is equipped with twelve sensory booths, two training rooms and a panel lounge.The initial phase of training involved introducing panellists to the relevant sensory properties in order that they were able to successfully recognise each. For example, when familiarising the panellists with maltiness, they were presented with the chemical 2-acetyl pyridine in isolation as an aroma, allowing the panellists to focus on the specific qualitative sensory property elicited by 2-acetyl pyridine. The panel were encouraged to apply

their own terms to the sensory property (e.g. biscuity, sacky) and this was facilitated by having panellists smell and taste various reference products (e.g. malt extract, biscuits and malt-based beverages). These descriptors ensured that, when assessing 'maltiness', panellists were not assessing maltiness in general, but instead the very specific sensory property elicited by 2-acetyl pyridine.

The next stage of training involved discriminating between samples with varying levels of each sensory property. This was initially carried out in the absence of other sensory properties by adding each attribute to water. Panellists ranked three samples from low to high intensity. When accurate and reliable performance was achieved across all panellists, the panel moved on to rank three samples of beer that varied in their levels of each sensory property. The terms that were generated for each sensory property were continuously fine-tuned as experience with each increased.

In building up the panel's experience with the sensory properties gradually, it was relatively straightforward to ascertain the origin of problems and solve them. For example, a panellist that was able to rank an attribute accurately in water but had difficulties doing the same in beer may have been confusing the attribute with other properties of the beer. If the panellist systematically ranked incorrectly (e.g. always ranks the samples in reverse rank order) then it may have been that interactions between the property of interest and other properties result in confusion. Panellists' ranking performance was immediately fed back to them in order to facilitate this process.

Once the panel was confident in ranking the properties, they were also required to rate samples. It was at this point the panellists were introduced to the 'anchor' samples (i.e. samples that represented the extremes of each sensory property's scale). All samples were to be rated relative to these anchors (treatments shown in Table 2.1). To aid the transition from ranking to rating, panellists were initially required to rank three samples, of which they were informed two were anchors (i.e. '1' and '10' on the scale). After ranking the three samples, the panel were only required to rate the intermediate sample relative to the two anchor samples. The data generated during each rating session was immediately fed back to the panel and individuals were able to observe how their ratings compared to others', whilst also re-sampling. Once consistent ratings were obtained both within and between panellists, the panel were required to rate samples monadically (i.e. without anchor samples). An individual protocol for the assessment of each sensory property was also continuously developed and agreed upon in order that all panellists were assessing the each attribute in the same way (e.g. keeping the sample in mouth for the same amount of time, making their assessment at the same point in time).

Once consistent and reliable ratings were obtained for each sensory property by all panellists, focus moved to the assessment of the final samples. This involved having the panel practise assessing more than one attribute per mouthful (required because of limitations in alcohol consumption as dictated by local ethical considerations). At this stage, the assessment protocol was discussed and finalised, with the major decisions about which sensory properties to assess in the same mouthful. With consultation of the panel, it

was decided that a maximum of three attributes were to be assessed per 10ml mouthful. In addition, an attempt was made to separate sensory properties from different modalities into different mouthfuls. Also, the properties that were adapted to quickest were assessed earlier than other properties. The panel refined this list over several assessments until a widely agreed assessment protocol for all 14 sensory properties was obtained (Appendix C).

Before final data collection, panel performance was assessed through a pilot study, wherein only the anchors of each sensory property were presented, although panellists were unaware of this. The purpose of this pilot study was to identify any problems in order that further training could be carried out. Results (Table 2.12) were used to inform subsequent panel training in order to strengthen performance on weak attributes. For example, this pilot showed that the low samples associated with bitterness, tingliness, and bubbliness were all rated too highly in this attribute (panellists were trained to rate these samples as a '1' on the scale). Similarly, DMS, acetaldehyde, and maltiness should have been rated much higher for the high samples (these should have been rated as '10'). By referring to individual panellists' generated data, it could be ascertained which individuals were driving the inaccurate ratings. Furthermore, a number of attributes were identified as being systematically confused by some panellists (for example, maltiness was relatively frequently confused for DMS; data not shown). Group sessions and sessions with small subsets of panellists were focussed on improving performance on the weak attributes identified in the pilot. Panellists' performance was checked in smaller scale validation before continuing to the final profiling of samples.2.5 Sample profiling

Table 2.12 Summary of pilot mean ratings for the low and high samples associated with each sensory property.

Sensory property	Low sample (1)	High sample (10)
Light struck	1.79	6.42
DMS	1.38	3.21
Isoamyl acetate	2.14	7.77
Diacetyl	1.50	7.99
Sweetness	2.93	7.62
Bitterness	3.90	6.55
Warming	1.59	7.31
Acetaldehyde	2.37	3.79
Alcoholic flavour	1.56	8.05
Body	1.76	6.31
Maltiness	1.94	4.22
Hoppiness	2.70	6.99
Bubblines	3.39	9.91
Tingliness	3.25	9.60

Once satisfied that performance on all attributes was high, data collection began.

2.5.1 Materials and method

Overall, 25 samples (Table 2.1) were assessed in duplicate. Each sensory property had three samples associated with it: low (bottom scale anchor, trained to be rated as '1'), high (top scale anchor, trained to be rated as '10') and intermediate (typically approximately two thirds the amount of the relevant compound was included as compared to the high sample). The inclusion of the intermediate sample allowed for the further assessment of the discriminability of the panel for each sensory property.

Panellists were presented with three 50ml samples served at 4±1°C per session and followed the agreed protocol (Appendix C). Samples were presented in clear screw top bottles. All samples were presented with a lid adapted for use

with the carbonator so that panellists could not identify which had been manually carbonated and which ones had not.

Panellists indicated the intensity of each of the 14 sensory properties using the relevant continuous line scale (labelled from low to high). For each sample, the panel firstly assessed the intensity of light struck aroma. They then went on to assess 2-3 sensory properties per 10ml mouthful of sample as described in the protocol. Each mouthful was associated with one page on the computer upon which a line scale appeared for each sensory property. There were a total of five mouthfuls per sample. Panellists palate cleansed with water (Evian, France) and crackers (Rakusen's, UK) between each mouthful, except before the final mouthful where just water used. This was because the panel had identified during training that the flavour associated with the crackers was very close to the malty characteristic they were to assess in the final mouthful for each sample. Panellists undertook a break of at least ten minutes between the assessment of each sample to avoid fatigue and carry-over effects.

2.5.2 Profiling results

Two-way analysis of variance (ANOVA) was performed on the ratings of each of the 14 sensory properties with sample and judge as fixed factors. There were significant sample effects for every sensory property ($p < 0.05$). Therefore, post hoc Tukey's HSD tests were applied to reveal which samples significantly differed from one another (Table 2.13).

The panel were able to discriminate between all three samples for six sensory properties: isoamyl acetate, sweetness, warming, alcoholic flavour, body, and bubbliness. However, no significant differences were found between the

Table 2.13 Summary of mean ratings for each sensory property to the associated samples.

Sensory property	Low sample (1)	Intermediate sample	High sample (10)
Light struck*†	1.00 (A)	3.66 (B)	3.76 (B)
DMS†	1.00 (A)	6.14 (B)	6.99 (B)
Isoamyl acetate	1.13 (A)	7.05 (B)	8.17 (C)
Diacetyl*†	1.19 (A)	6.21 (B)	6.59 (B)
Sweetness*†	1.59 (A)	5.84 (B)	8.11 (C)
Bitterness*†	1.64 (A)	8.29 (B)	8.47 (B)
Warming*	1.67 (A)	4.45 (B)	7.62 (C)
Acetaldehyde*	1.76 (A)	5.30 (B)	6.42 (B)
Alcoholic flavour*†	1.83 (A)	4.71 (B)	8.25 (C)
Body*†	2.07 (A)	4.86 (B)	7.28 (C)
Maltiness*	2.67 (A)	4.41 (A)	7.22 (B)
Hoppiness*†	3.21 (A)	6.26 (B)	7.32 (B)
Bubbliness*	3.52 (A)	6.33 (B)	8.38 (C)
Tingliness*	3.89 (A)	6.23 (AB)	8.05 (B)

* Significant judge effect ($p < 0.05$)
† Significant product*judge interaction ($p < 0.05$)
Letters in brackets represent the results of post hoc Tukey's HSD tests carried out due to a significant main effect of sample ($p < 0.05$), with the same letter indicating no significant differences and different letters representing significant differences

intermediate and high samples for seven sensory properties: DMS, hoppiness, tingliness, light struck, diacetyl, bitterness and acetaldehyde. In addition, no significant differences were found between the intermediate and low samples for two sensory properties: maltiness and tingliness. Table 2.13 shows that discriminability was poorer for some sensory properties than others. This was perhaps unsurprising given that significant effects of judge were found for the majority of sensory properties (sweetness, alcoholic flavour, body, maltiness, hoppiness, bubbliness, tingliness, light struck, diacetyl, bitterness, and acetaldehyde), meaning that there were differences in scale usage for ratings of these sensory properties between judges. Post hoc tests showed that there was no panellist in particular driving this across sensory properties. A number

of sensory properties (sweetness, alcoholic flavour, body, hoppiness, light struck, diacetyl, bitterness, and DMS) also showed interactions between sample and judge, meaning that there were significant differences in the ratings for some individual samples across judges. In spite of the variability in scoring between panellists, the panel was able to successfully discriminate between the low and high samples for every sensory property.

2.6 Conclusions

Through the recruitment, screening, training, and validation of a beer-specific sensory panel, it has been shown here that it was acceptable to use the low and high samples for subsequent consumer experiments as it can be concluded that a perceptible increase in chosen sensory properties was achieved through the modification of selected attributes. Therefore, these samples were able to be used in the studies that follow which explore the relationship between sensory properties and emotional response.

Chapter 3: Developing an approach for the measurement of emotional response to the sensory properties of beer

3.1 Introduction

The measurement of emotional response is a relatively young area of sensory science and, as such, methods have yet to become established. Therefore, the opportunity exists to further develop current approaches to create more effective methodologies. This chapter will discuss the merits of present methods before describing and applying a further improved emotion measurement approach.

3.1.1 Selecting an emotional measurement method

Self-report has been the most commonly used approach in the short history of measuring emotions in sensory and consumer science. This is no doubt partially due to the relative familiarity to the sensory practitioner. Self-report requires consumers to consciously indicate their emotional response to a stimulus. In this way, rich data about complex and specific emotions can be assessed (e.g. nostalgia, adventurousness) which would be difficult to measure by any other method. Self-report measures can be either verbal (i.e. through the use of language) or non-verbal (e.g. through the use of pictures or cartoons) and the majority of sensory research has employed the former (for examples, see section 1.3.1.1).

3.1.2 Verbal self-report

Verbal self-report lexicons can be divided into two categories: pre-determined and consumer-led. A prominent example of a pre-determined emotion lexicon is EsSense Profile (see section 1.3.1.1). 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 (King and Meiselman, 2010). The effectiveness of EsSense Profile for differentiating emotional response both between and within product categories was demonstrated by King and Meiselman (2010) using both check-all-that-apply (CATA) and data scaling approaches.

The major advantage for researchers of using pre-determined emotion lexicons like EsSense Profile is that they 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 et al. (2013) reported six such redundant EsSense Profile terms in the emotional assessment of blackcurrant squashes. More significantly, emotion terms may be excluded that are characteristic of 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. This is, of course, 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 product-specific consumer-led lexicon. 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 removing potential confusion (Jaeger et al., 2013) and not missing relevant terms, thereby increasing discrimination ability (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 responses to 11 commercial blackcurrant squash products. Each of 29 consumers generated their own lexicon in one-to-one interviews. The consumers then used 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. These suggestions were implemented in the approach described in this chapter to increase the eficiency of lexicon generation and increase the capability for statistical analysis.

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 unwieldly. In order to allow an easier and quicker test for the respondent, Porcherot et al. (2010) developed a reduced version of the GEOS questionnaire (Chrea et al., 2009); see section 1.3.1.1). Participants rated a series of three representative terms for each of the six GEOS dimensions instead of rating the 68 terms individually for each sample (ScentMove). Regardless of the fact that 50% of the evaluations were required as compared to the original form, comparable product information was obtained by the GEOS and ScentMove questionnaires. The present study also takes this approach of reducing the number of consumer responses, with the aims of minimising the potential for consumer fatigue and boredom due to a lower number of required consumer responses as well as increasing the ease of statistical product comparisons.

3.1.3 Objectives

The main objectives of the study presented in this chapter were twofold: (1) to create an approach for the development of a reduced product-specific consumer-led emotion lexicon; and (2) to use the reduced consumer-led lexicon to (a) discriminate across a range of beer samples designed to elicit specific sensory properties, and (b) reveal differences in emotional response across different consumer segments related to gender and age.

3.2 Lexicon development

Focus groups of consumers generated an emotional lexicon in their own words to describe their responses to the set of 14 sensorially-distinct samples

developed in Chapter 2. These subjects subsequently used the created lexicon to rate the 14 samples (Table 3.1) and this data was submitted to cluster analysis and linguistic checks in order to group similar terms into emotion categories. This section describes the process in detail.

Table 3.1 Fourteen beer samples and their treatments.

Sample		Treatment
1	Control	Commercial lager
2	Hoppy	0.75mg Aroxa kettle hop extract/litre commercial lager
3	Malty	1.08mg Aroxa 2-acetyl pyridine/litre commercial lager
4	Light struck	0.3µg Aroxa 3-methyl-2-butene-1-thiol/litre commercial lager
5	Isoamyl acetate	10.5mg Aroxa isoamyl acetate/litre commercial lager
6	Diacetyl	520µg Aroxa diacetyl/litre commercial lager
7	DMS	0.9mg Aroxa dimethyl sulphide/litre commercial lager
8	Acetaldehyde	45mg Aroxa acetaldehyde/litre commercial lager
9	Bitter	25mg Aroxa iso-α-acids/litre commercial lager
10	Sweet	25g dextrose/litre commercial lager
11	Low CO_2	Commercial lager decarbonated and recarbonated to ~1.6 units
12	High CO_2	Commercial lager decarbonated and recarbonated to ~4 units
13	Non-alcohol control	Commercial non-alcohol lager
14	High alcohol	96% ethanol added to commercial non-alcohol lager (8% ABV)

3.2.1 Subjects

Seventeen UK consumers (aged 18-65 years) who consumed beer at least once per month took part in this study after signing consent forms in line with local ethical considerations. Each participant was judged to be reasonably articulate

following a short discussion with the session leader about beer prior to being invited to the sessions. Most (71%) consumed beer at least once per week and less frequent beer drinkers were included to ensure a range of emotional responses. However, all consumed beer at least once per month. As women have previously been suggested to be more adept with emotional language (Fugate et al., 2009), more female subjects were recruited (65%) to facilitate the term generation process, although males were included to ensure relevant terms from both genders were represented. Participants were divided into three groups of between five and seven subjects and attended a total of three 90min-2h sessions.

3.2.2 Procedure

At the start of the first session, consumers received a short explanation of 'emotion' in the context of other affective behaviours (see section 1.1.1). In order to make this distinction clear, warm-up exercises were conducted in which subjects described their emotional response with reference to pictures and prompt cards (Ng et al., 2013).

Drawing on the experience of previous emotion research (Ng, 2013), participants were presented with warm-up samples of the two '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 et al., 2004) was used to generate terms. That is to say, participants were asked to assess triads of samples (selected from the

set of 14 samples) and describe 'in what way two samples were similar but different from the third in terms of your emotional response'. These differences were not necessarily experienced emotions but instead emotional associations with products (see section 1.3.1.3). After performing this task individually, the participants shared and discussed their response with the group and a consensus between the members was reached. Five triads were presented to each group (to ensure each of the 14 samples (Table 3.1) appeared at least once) in a randomised design. Two triads were presented in the first session and three in the second session. Each sample was 10ml in volume and served at 4±1°C.

The terms generated by all three groups in this elicitation phase resulted in a list of 100 emotion terms. An initial reduction was performed by asking the subjects to indicate any words which more accurately described their sensory perception as opposed to their emotional response (e.g. bland, unappealing). Where possible, synonymous remaining terms were combined using a thesaurus (Microsoft Word 2007). This resulted in a lexicon of 48 terms.

This lexicon was then used by the subjects during the third session to rate their emotional response to all 14 samples. Each of the 48 terms was associated with a 150mm line scale, anchored from 'not at all intense' to 'extremely intense' and responses were recorded using Fizz Forms (Biosystèmes, Couternon, France). These responses were subsequently expressed as a percentage distance along the line scale, i.e. 0-100%. Emotions were presented in a randomised order, as emotion list order has been found to affect consumer responses (King et al., 2013). Having rated all 14 samples and assuming

subjects had become familiar with the lexicon, participants were instructed to rate each term for relevance to describing emotions elicited by beer as a product category per se. This element of the questionnaire was included to provide additional data to assist in the elimination of redundant terms. 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. Participants were instructed to palate cleanse using unsalted crackers (Rakusen's, UK) and mineral water (Evian, France) before the assessment of each sample.

3.2.3 Grouping of terms into emotion categories

Five 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 (angry, annoyed, optimistic, reassured, regretful).

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 43 terms in order to ascertain the relative effectiveness of each technique for grouping terms which produced similar patterns of data. Cluster analysis was deemed by the researchers 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 thesis 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.

Eight clusters were identified (Table 3.2a) and slight modifications were made with reference to Cronbach's α in order to improve the distinction between groups of emotions. It was also necessary to compare the clusters analysis using 14 samples to a second cluster analysis using just a subset of 10 samples. This was because the subset of 10 samples was to be used for comparison of emotional response with Spanish consumers (Chapter 5).

Based on discussions with participants, the term 'underwhelmed' was moved from the group containing 'tame' and 'safe' to be in an emotion category alone because it was deemed to represent a different emotion. Indeed, there was an increase in Cronbach's α (+0.18) in the group containing 'tame' and 'safe' without 'underwhelmed' (though this was largely due to the small number of terms as Cronbach's α is greatly affected by the number of items). In addition, the term 'unpleasantly surprised' was removed from the group containing 'disgusted', 'horrible', 'repulsed/repelled', and 'unpleasant' to the cluster including 'disappointed' and 'dissatisfied' (Table 3.2b) because it was felt by the researchers to fit more neatly in this cluster (also, cluster analysis using just 10 samples placed 'unpleasantly surprised' in this cluster). This hardly affected the Cronbach's α associated with the original group of the term 'unpleasantly surprised' and slightly increased Cronbach's α for its new group (+0.02; although again it must be noted that Cronbach's α is greatly affected by the number of terms). The Cronbach's α across all 9 final emotion categories indicated adequate internal consistency (Cronbach's $\alpha > 0.8$;

Table 3.2a Cluster analysis of 43 terms grouped into 8 clusters with associated Cronbach's αs (adequate internal consistency > 0.8 (Streiner, 2003)).

Cluster	1	2	3	4	5	6	7	8
Cronbach's α	0.96	-	0.99	0.99	0.91	0.95	0.99	0.77
	Alarmed	Bored	Calm	Curious	Desirous	Disappointed	Disgusted	Tame
	Cheated		Comfortable	Enthusiastic	Nostalgic	Dissatisfied	Horrible	Underwhelmed
	Confused		Comforted	Excited	Relieved		Repulsed/	Safe
	Overwhelmed		Content	Fulfilled			repelled	
	Shocked		Enjoyment	Fun			Unpleasant	
	Strange/weird		Good	Impressed			Unpleasantly	
			Happy	Interested			surprised	
			Nice	Optimistic				
			Pleasant	Pleasantly				
			Pleased	surprised				
			Relaxed	Want				
			Satisfied	Warm				

Table 3.2b Final grouping of the 43 terms into 9 emotion categories with associated Cronbach's αs (adequate internal consistency > 0.8 (Streiner, 2003)).

Emotion category	1 Shock	2 Boredom	3 Contentment	4 Excitement	5 Nostalgia	6 Disconfirmation	7 Disgust	8 Tame/Safe	9 Underwhelmed
Cronbach's α	**0.96**	**-**	**0.99**	**0.99**	**0.91**	**0.97**	**0.99**	**0.95**	**-**
	(a) Alarmed	Bored	(a) Calm	(a) Curious	(a) Desirous	(a) Disappointed	(a) Disgusted	(a) Tame	Underwhelmed
	(b) Cheated		(b) Comfortable	(b) Enthusiastic	(b) Nostalgic	(b) Dissatisfied	(b) Horrible	(b) Safe	
	(c) Confused		(c) Comforted	(c) Excited	(c) Relieved	(c) Unpleasantly surprised	(c) Repulsed/ repelled		
	(d) Overwhelmed		(d) Content	(d) Fulfilled			(d) Unpleasant		
	(e) Shocked		(e) Enjoyment	(e) Fun					
	(f) Strange/weird		(f) Good	(f) Impressed					
			(g) Happy	(g) Interested					
			(h) Nice	(h) Optimistic					
			(i) Pleasant	(i) Pleasantly surprised					
			(j) Pleased	(j) Want					
			(k) Relaxed	(k) Warm					
			(l) Satisfied						

Streiner (2003)). The only other difference when comparing the cluster analyses using 14 and 10 samples was that the 10 sample version included an emotion category including just 'cheated' and 'overwhelmed'. However, the version generated from the cluster analysis using 14 samples was maintained as the two terms were deemed to have a similar meaning to other terms within that group. Each emotion category was assigned a title that summarised the terms belonging to a given category. This was purely to aid reporting of consumer responses; consumers would never see the titles assigned to each category, just the terms that belonged to them. 'Boredom', 'Underwhelmed', and 'Tame/Safe' were directly named for all the emotion terms that belonged to those categories, whilst the others drew upon a term that best encapsulated the meaning of the category.

3.3 Measuring consumer emotional response using the reduced emotion lexicon

In the second part of the study, 109 naïve consumers rated their emotional response to the 14 beer samples using the 9 emotion categories.

3.3.1 Subjects

One hundred and nine subjects (54% female) who consumed beer at least once per month took part in this study (82% consumed beer at least once a fortnight and 52% at least once per week). 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 turn 35 years old (Mintel, 2013).

It is important to note that this study formed part of a larger study into form length (the other half of this study is reported in Chapter 4). The 109 consumers that took part in this reduced form study also assessed the 14 samples using a full emotion form of 43 terms. Fifty-two consumers completed the reduced form first and 59 completed the full form first. Two-way between subjects ANOVA was conducted for each reduced form emotion category (as well as liking and familiarity) to reveal any effects of the order that the forms were presented (IBM SPSS Statistics 22). Boredom, Disgust, Nostalgia, Shock and Tame emotion categories were all rated significantly higher by those consumers that completed the reduced form second than those that completed the reduced form first ($p < 0.05$). However, there were no significant interactions between the order of forms and sample for any emotion category ($p > 0.05$). Just familiarity showed a significant interaction between form order and sample ($p < 0.05$). Simple main effects analysis showed that the Control and Low CO_2 samples were rated as more familiar by those that completed the reduced form first but the High alcohol sample was rated higher in familiarity by those that completed the reduced form second. Therefore, it was concluded that, though the order of presentation of the forms affected the magnitude of ratings given by consumers, there was little or no difference in the response to individual samples no matter the form order. This is with the exception of familiarity which might have been expected to change anyway due to consumers' prior experience with the sample set.

3.3.2 Procedure

Consumers attended two sessions held on different days in order that local ethical considerations of alcohol consumption were adhered to and so that

intoxication would have a limited effect on emotional response. Eight 10ml samples served at 4±1°C were assessed in the first session and seven were assessed in the second session. A 'dummy' sample (always the normal strength alcohol commercial lager Control) was presented in the first position to overcome first-order effects Dorado et al. (in preparation) and familiarise consumers with the task. This data was subsequently discarded. The presentation order of the remaining 14 samples was randomly assigned for each consumer. Samples (Table 3.1) were assessed in an air conditioned room (21±1ºC) room under Northern Hemisphere daylight lighting. Unsalted crackers (Rakusen's, UK) and mineral water (Evian, France) were provided as palate cleansers between sample assessments.

When rating their emotional response, consumers were presented with 9 continuous line scales. Each scale was associated with one emotion category. Emotion categories were presented as a horizontal list of terms that belonged to that category (e.g. the Disconfirmation emotion category was presented as 'Disappointed/Dissatisfied/Unpleasantly surprised'). Consumers were instructed to read all of the terms associated with each emotion category and to rate the overall intensity of their feeling of the underlying emotion that the words were describing on a 150mm continuous line scale anchored from 'not at all intense' to 'extremely intense'. The order of emotion groups was randomised between consumers, although the order remained consistent across samples for individual consumers. Once the 9 emotion categories were rated, consumers then scored the sample for liking and familiarity on two further 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.

3.3.3 Data analysis

Multivariate analysis was performed using principal components analysis (PCA) on mean ratings of emotion categories for the 14 samples in order to map the emotional space of the samples (XLSTAT, Version 2009.6.03). The liking and familiarity data were included in the PCA as supplementary variables (i.e. this data was not used to generate the PCA space but was mapped in the space generated by the ratings of emotion categories) 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 (SPSS Statistics 22, IBM, USA). Further ANOVAs were carried out with fixed effects of sample, gender and age group for each emotion category (and liking and familiarity). In this way, differences between consumer groups in overall ratings of emotion categories could be seen, as well as interactions between sample and gender and between sample and age group. Where significant interactions were found, simple main effects analyses were conducted in order to ascertain

which samples were rated significantly differently for a given emotion category (or liking or familiarity) between genders or between age groups (SPSS Statistics 22, IBM, USA).

3.4 Results

The following section shows the effectiveness of the use of group interviews and cluster analysis to develop a reduced product-specific consumer-led emotion lexicon through its creation of a discriminating emotional space across samples. The effectiveness of the approach for differentiating between genders and age groups in their emotional responses is then explored.

3.4.1 Emotion space

The PCA enabled the visualisation of the emotional space for the samples tested. The first two principal components accounted for 94.63% of the data variance (Figure 3.1a). PC1 (75.03%) was highly positively correlated with emotion categories Disconfirmation, Disgust, and Shock and highly negatively correlated with Nostalgia, Contentment, Tame/Safe, Excitement. Liking was not active in the PCA but was highly negatively correlated with PC1, as was familiarity. Underwhelmed and Boredom correlated highly positively with PC2 (19.61%). No emotion category was particularly negatively correlated with PC2; excitement showed the most association with this dimension but was more correlated with PC1. This emotional space was consistent with circumplex models of emotion (see section 1.1.2), with PC1 associated with pleasure/pleasantness and PC2 related to arousal/engagement/activation.

With the samples projected onto the first two principal components (Figure 3.1b), it was observed that the High alcohol sample was projected highly

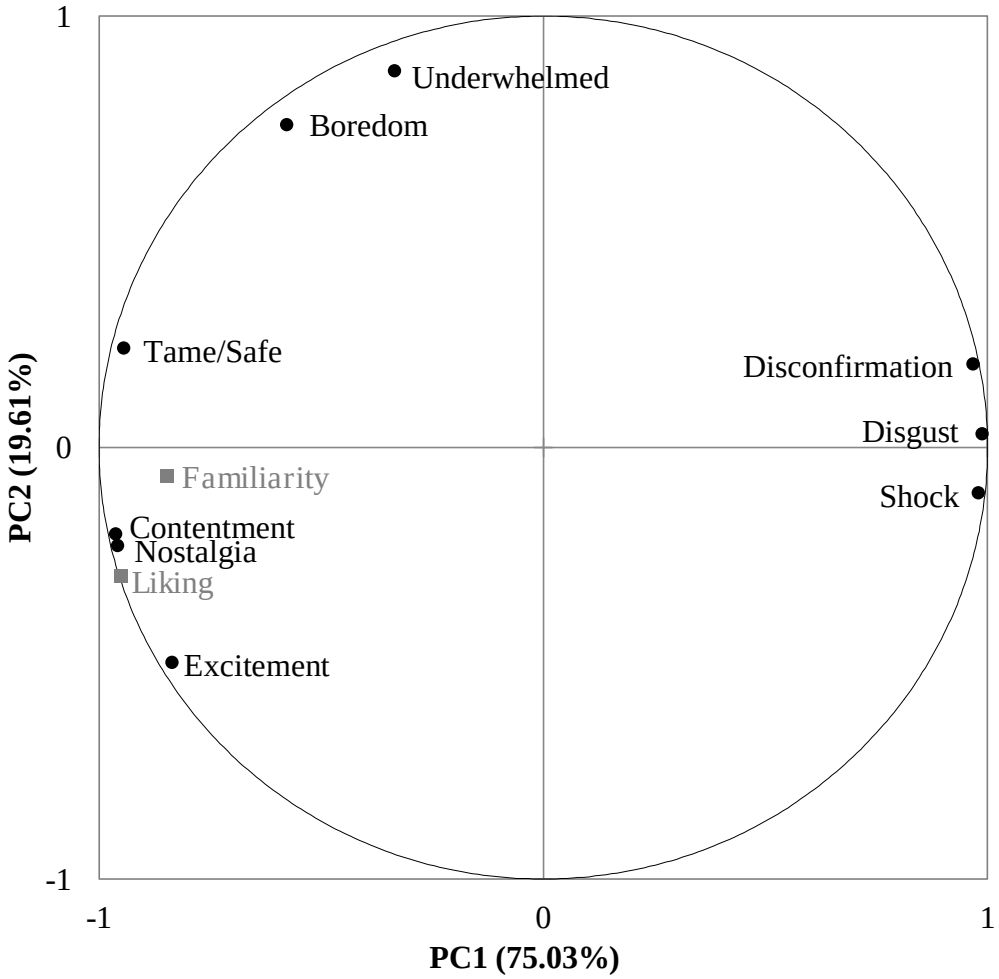


Figure 3.1a PCA correlation circle of the 9 emotion categories on PC1 and PC2 (liking and familiarity are included as supplementary variables).

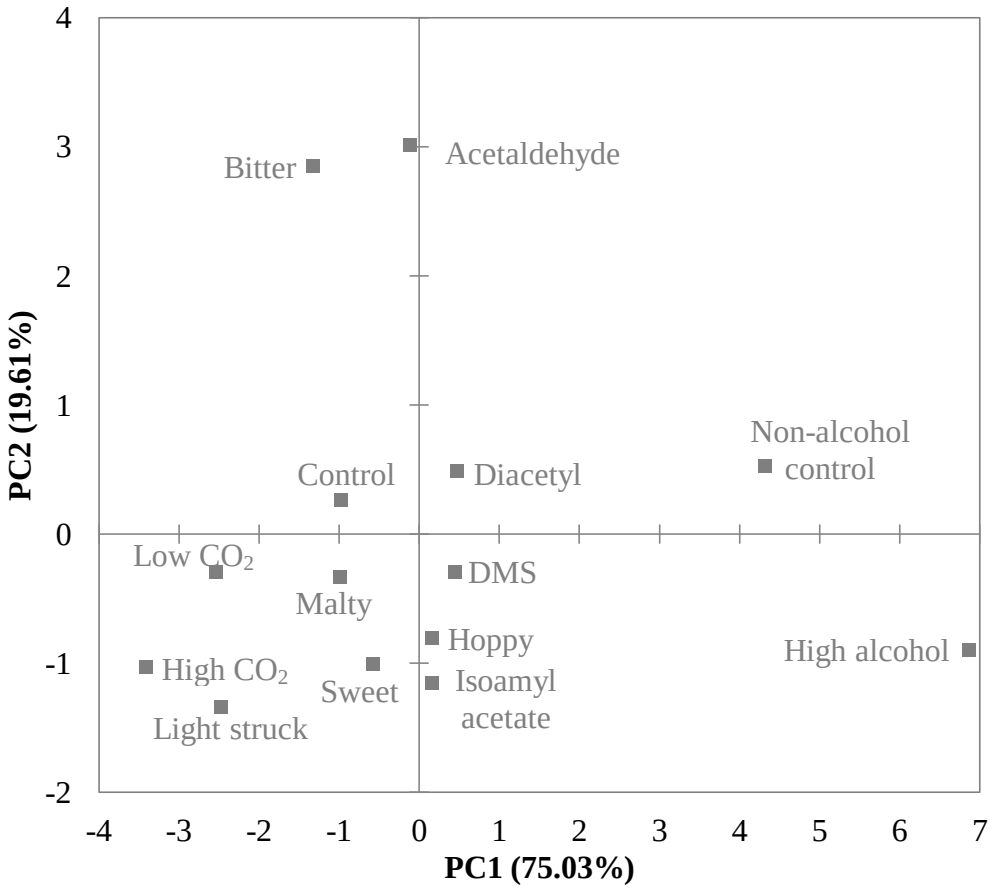


Figure 3.1b PCA product plot showing the projections of the 14 samples on PC1 and PC2.

positively onto PC1, therefore associated with displeasurable/unpleasant emotions. The Non-alcohol control also projected positively on PC1, but not as much relative to the High alcohol sample. In contrast, the Low CO_2, High CO_2 and Light struck samples were highly negatively correlated with PC1 and, as such, were associated with pleasurable/pleasant emotions. The Bitter, Control, and Malty samples were also projected somewhat negatively onto PC1, with Acetaldehyde, Diacetyl, DMS, Hoppy, Isoamyl acetate, and Sweet samples not loading particularly highly on this dimension. Bitter and Acetaldehyde samples projected highly positively on PC2, showing an association low arousal/engagement/activation emotions. A number of samples (Isoamyl acetate, Sweet, Hoppy, High alcohol, Light struck, High CO_2) were negatively associated with PC2, thereby eliciting emotions higher in arousal/engagement/activation. DMS and Diacetyl samples did not load particularly on either dimension (or on PC3 which, in any case, only accounted for 2.99% of the data variance).

3.4.2 Discrimination ability of emotion categories between samples

ANOVA showed that all 9 emotion categories (as well as liking and familiarity) significantly discriminated between the beer samples ($p < 0.05$; Table 3.3). Consequently, post hoc analyses ($p < 0.05$) were carried out for all emotion categories (and liking and familiarity) in order to reveal how each was able to discriminate between samples (Table 3.4). In this way, the reduced product-specific consumer-led emotion lexicon was shown to be effective through its ability to discriminate across the range of sensorially-distinct beer samples.

Table 3.3 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	Age	Sample* Gender	Sample* Age
1 Shock	**<0.001**	**0.002**	**0.013**	0.509	0.463
2 Boredom	**0.002**	**0.011**	**0.02**	0.198	0.185
3 Contentment	**<0.001**	0.1	**0.013**	0.398	0.101
4 Excitement	**<0.001**	**0.03**	**0.009**	0.543	**0.013**
5 Nostalgia	**<0.001**	**<0.001**	**0.003**	0.756	0.3
6 Disconfirmation	**<0.001**	**0.028**	0.524	0.215	0.118
7 Disgust	**<0.001**	0.78	**0.004**	0.722	0.459
8 Tame/Safe	**<0.001**	0.124	0.152	0.324	0.884
9 Underwhelmed	**0.007**	**0.005**	**<0.001**	0.52	0.848
Liking	**<0.001**	0.058	0.19	0.388	**<0.001**
Familiarity	**<0.001**	**0.001**	0.526	0.146	0.366

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

Comparisons of sample discriminations for each emotion category (and liking and familiarity) highlighted patterns of sample groupings related to how the emotion categories and samples loaded onto the two dimensions identified by PCA. These patterns offered a useful guide for comparing and contrasting the discrimination ability of individual emotion categories.

Table 3.4 Mean scores for the 9 emotion categories (and liking and familiarity) across the 14 samples.

Sample	Emotion category										
	1 Shock	2 Boredom	3 Conte-ntment	4 Excite-ment	5 Nostalgia	6 Discon-firmation	7 Disgust	8 Tame/ Safe	9 Under-whelmed	Liking	Famil-iarity
Control	26.3 ABC	31.1 AB	50.4 CD	42.3 BCDE	39.7 D	33.7 ABC	28.6 AB	44.1 CDE	36.7 AB	49.2 CDE	43.9 CDE
Hoppy	34.7 BC	30.0 AB	44.0 CD	43.0 CDE	32.5 BCD	34.3 ABC	28.7 AB	35.9 BC	32.5 AB	48.6 CDE	35.7 BC
Light struck	22.2 A	31.1 AB	41.7 BC	46.7 DE	29.3 ABC	29.0 AB	22.2 AB	45.5 CDE	31.2 A	56.1 DE	55.3 F
Isoamyl acetate	35.6 BC	28.2 AB	50.7 CD	44.5 CDE	33.9 BCD	36.3 ABC	29.5 AB	42.8 CDE	32.5 AB	48.5 CDE	34.8 BC
DMS	35.6 BC	27.7 AB	44.4 CD	40.1 ABCDE	34.0 BCD	37.7 BC	31.4 B	37.7 BCD	36.1 AB	46.0 CD	38.3 BCD
Malty	30.2 ABC	28.9 AB	45.9 CD	42.1 BCDE	34.0 BCD	34.2 ABC	25.9 AB	47.6 DE	35.4 AB	51.7 CDE	46.5 DEF
Diacetyl	36.9 CD	30.3 AB	44.1 CD	42.1 BCDE	32.1 BCD	40.5 BCD	32.7 BC	37.5 BCD	38.1 AB	44.1 BC	35.4 BC
Acetaldehyde	31.7 ABC	36.3 B	45.1 CD	35.7 ABC	32.0 BCD	40.9 CD	28.9 AB	39.2 CD	41.6 B	43.1 BC	36.0 BC
Bitter	25.1 AB	36.2 B	47.1 CD	38.1 ABCD	34.6 CD	38.3 BC	28.7 AB	41.5 CDE	42.8 B	45.9 CD	44.3 CDE
Sweet	34.5 BC	27.4 AB	53.5 D	46.0 CDE	38.8 CD	37.8 BC	28.9 AB	44.6 CDE	34.9 AB	50.6 CDE	32.3 B
Low CO_2	23.5 A	32.0 AB	47.4 CD	46.7 DE	35.7 CD	29.1 AB	21.9 AB	45.2 CDE	35.9 AB	53.2 CDE	51.3 EF
High CO_2	22.0 A	31.2 AB	54.3 D	49.1 E	41.3 D	25.7 A	20.3 A	49.4 E	34.0 AB	58.5 E	54.9 F
Non-alcohol control	46.9 D	29.9 AB	32.4 AB	32.1 AB	24.7 AB	51.4 DE	43.5 C	28.7 AB	32.2 AB	34.0 AB	34.6 BC
High alcohol	58.1 E	23.0 A	29.8 A	31.0 A	20.3 A	60.0 E	54.8 D	21.4 A	30.8 A	30.0 A	22.0 A

ABCDEF Letters within the same column indicate post hoc groupings by Tukey's HSD.
Different letters represent significant differences ($p < 0.05$) in rating between samples

The three emotion categories associated with unpleasantness according to the PCA (Disconfirmation, Disgust, and Shock; Figure 3.1a) discriminated between samples very similarly. The Non-alcohol control and High alcohol samples were highly positively projected onto PC1 (Figure 3.1b) and were found to be rated significantly higher in the unpleasant emotion categories than most samples (the only exceptions were that there was no significant difference in ratings between the Non-alcohol control and Diacetyl samples for all three categories and there was no significance difference between the Non-alcohol control and Acetaldehyde samples in ratings of Disconfirmation). The High alcohol sample showed a higher positive correlation with PC1 than the Non-alcohol control (Figure 3.1b) and, accordingly, the High alcohol sample was rated significantly more disgusting and shocking than the Non-alcohol control. None of the samples modified from commercial lager significantly differed from the Control upon which they were based for any of the three unpleasant emotion categories, although differences were observed between the individual manipulated samples. Again, this was related to sample positioning on PC1 (Figure 3.1b). For example, the Acetaldehyde sample (neutral on PC1) was rated significantly higher in Disconfirmation than the Light struck sample (highly negatively correlated with PC1). Also, the Diacetyl sample (relatively neutral on PC1) was rated higher in Shock than the Bitter sample (negatively correlated with PC1). There was no effect of changing the level of carbonation, with the Low CO_2 and High CO_2 samples showing no significant differences in any of the three unpleasant emotion categories.

The emotion categories Contentment, Excitement, Nostalgia, and Tame/Safe were associated with pleasantness according to the PCA (i.e. negatively correlated with PC1; Figure 3.1a) and were found to discriminate between samples similarly to one another. These pleasant emotion categories were unable to discriminate between the Non-alcohol control and High alcohol samples where unpleasant emotion categories were. In addition, Excitement and Nostalgia showed a reduced ability to discriminate Non-alcohol control and High alcohol samples from other samples as compared to other emotion categories associated with the pleasantness/pleasure dimension, with few significant differences observed (Table 3.4). Nevertheless, Nostalgia was the only emotion category, pleasant or unpleasant, to show a significant difference when comparing the Control with a sample modified from it, in that the Light struck sample received significantly lower ratings of Nostalgia than the Control. Differences between individual modified samples were also observed and this was somewhat related to the positioning of each sample on PC1 (Figure 3.1b). For example, the Malty sample (negatively projected onto PC1) was scored higher in Tame/Safe than the Hoppy sample (neutral on PC1). In addition, the Light struck sample (highly negatively correlated with PC1) was rated higher in Excitement than the Acetaldehyde sample (neutral on PC1). As with the unpleasant emotion categories, there were no significant differences between the Low CO_2 and High CO_2 samples, meaning there was no change in ratings of pleasant emotion categories associated with different carbonation levels.

As a supplementary variable, liking was also negatively correlated with PC1 (Figure 3.1a) and its discrimination between samples was similar to the

unpleasant and pleasant emotion categories. This was in so much as the Non-alcohol control and High alcohol samples were generally rated significantly differently to the other samples and scored lower in liking. In addition, liking was unable to differentiate between the Low CO_2 and High CO_2 samples, just like the emotion categories associated with pleasantness/pleasure. However, in contrast to two of the three unpleasant emotion categories, liking was unable to discriminate between the Non-alcohol control and High alcohol samples. Furthermore, the Control was not rated significantly differently in liking to any of the samples modified from it, whereas it was in Nostalgia. Just a couple of significant differences were found between the individual samples modified from the Control, with the Acetaldehyde and Diacetyl samples (neutral on PC1) rated significantly lower in liking than the Light struck sample (highly negatively projected onto PC1). Thus, the seven emotion categories associated with the pleasure/pleasantness emotion dimension were able to discriminate more effectively than liking in response to the 14 samples. The number of measures suggests that this is not just a spurious chance outcome and the emotional response genuinely provides more detailed consumer information than liking. As an aside, there was also a negative relationship between familiarity as a supplementary variable and PC1 (Figure 3.1a). Frequent differences between samples were seen in familiarity (e.g. the Control and Sweet samples significantly differed in familiarity where they did not for any emotion category) that did not seem to have much bearing on consumer discrimination between samples in emotional response when using the reduced product-specific consumer-led lexicon.

Boredom and Underwhelmed were associated with low arousal/engagement/activation (i.e. positively correlated with PC2; Figure 3.1a). Neither was particularly discriminating between samples according to post hoc tests. In both cases, the Bitter sample was rated higher (i.e. more boring and underwhelming) than the High alcohol sample. The Bitter sample was also rated higher in Underwhelmed than the Light struck sample. In addition, the Acetaldehyde sample received significantly higher ratings of Boredom than the High alcohol sample. Acetaldehyde and Bitter samples can be seen to oppose the High alcohol and Light struck samples on PC2 (Figure 3.1b). All other samples showed no significant differences in ratings for either of these low arousal/engagement/activation emotion categories. Excitement was the most highly negatively correlated emotion category with PC2 (Figure 3.1a) and therefore the most arousing/engaging/activating. However, Excitement was more correlated with PC1 and, indeed, post hoc tests showed that Excitement was more closely related in its sample discriminations to the pleasant emotion categories.

Taking the results from across emotion categories together, the samples based on the non-alcohol commercial lager (Non-alcohol control and High alcohol) were generally rated higher in unpleasant emotions and lower in pleasant emotion categories (and liking) than the samples based on the normal strength commercial lager (Figure 3.2a). Unpleasant emotion categories were able to discriminate between the Non-alcohol control and High alcohol samples, showing that an increase in the sensory properties associated with alcohol content increased ratings of unpleasant emotions (Figure 3.2a). Only Nostalgia was able to discriminate between the Control and any of the samples based

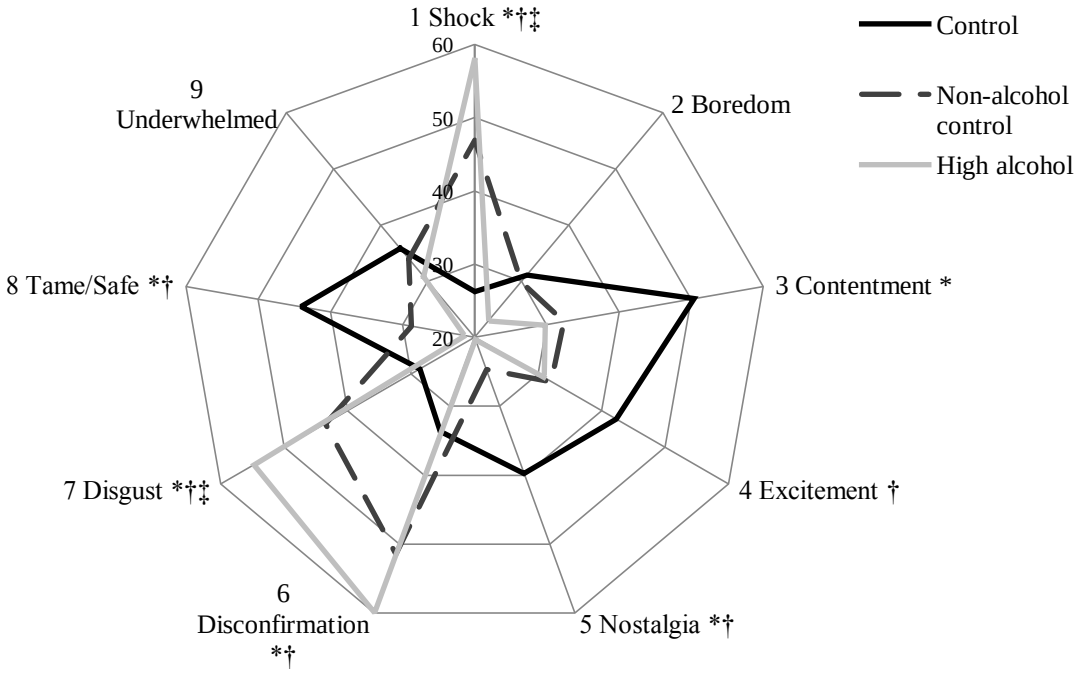


Figure 3.2a Spider plot showing mean score for all 9 emotion categories for Control, Non-alcohol control, and High alcohol samples. As per post hoc Tukey's HSD tests ($p < 0.05$), * denotes a significant difference between the Control and Non-alcohol control samples, † denotes a significant difference between the Control and High alcohol samples, and ‡ denotes a significant difference between the Non-alcohol control and High alcohol samples.

upon it, with the Light struck sample receiving significantly lower scores (Figure 3.2b). However, there were differences found between individual samples based on the Control in unpleasant and pleasant emotion categories. In particular, the Light struck sample, which was highly negatively correlated with PC1, was found to be rated significantly differently to a number of samples across emotion categories (see the example of differences in ratings of Excitement and Disconfirmation between the Acetaldehyde and Light stuck samples in Figure 3.2b), although there were differences between other samples as well. The unengaging Boredom and Underwhelmed emotion categories were relatively undiscriminating between the samples. None of the

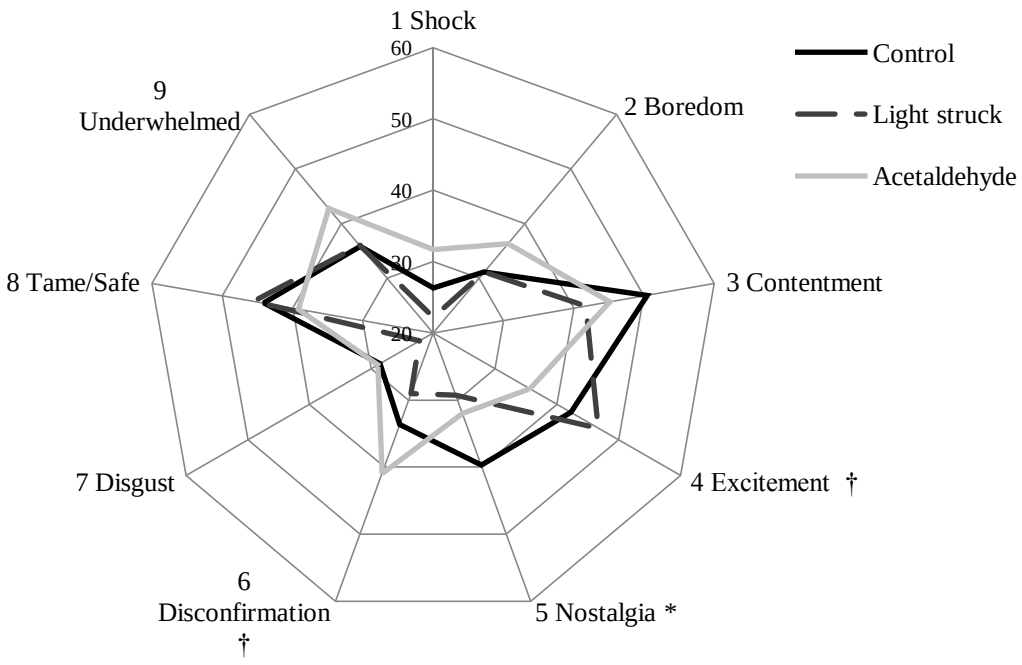


Figure 3.2b Spider plot showing mean score for all 9 emotion categories for Control, Light struck, and Acetaldehyde samples. As per post hoc Tukey's HSD tests ($p < 0.05$), * denotes a significant difference between the Control and Light struck samples, and † denotes a significant difference between the Light struck and Acetaldehyde samples (there were no significant differences between the Control and Acetaldehyde samples).

9 emotion categories discriminated between the Low CO_2 and High CO_2 samples, although nor did liking or familiarity. In fact, liking was found to be less discriminable between samples than the emotion categories. Familiarity, on the other hand, was found to discriminate well between most samples, but differences in familiarity did not necessarily equate to differences in emotional response.

3.4.3 Consumer group effects

This section further explores the effectiveness of the reduced product-specific consumer-led lexicon through its ability to differentiate between the emotional

responses of consumer groups – namely, between genders and between age groups – in response to the 14 samples and their associated sensory properties.

3.4.3.1 Gender

A significant main effect of gender was found for the emotion categories Disconfirmation, Shock, Nostalgia, Excitement, Boredom, and Underwhelmed ($p < 0.05$; Table 3.3). In each case, women gave lower mean ratings than men. This was also true of familiarity but not liking. Despite overall differences in ratings between genders for some emotion categories, there were no significant interactions ($p < 0.05$; Table 3.3) between gender and sample for any emotion category (or liking and familiarity), showing that males and females used the emotion categories to respond to the individual samples similarly despite an overall difference in scale usage for some emotion categories.

3.4.3.2 Age group

Main effects of age group ($p < 0.05$; Table 3.3) were found for emotion categories Disgust, Shock, Contentment, Excitement, Nostalgia, Boredom, and Underwhelmed (but not liking or familiarity). Emotion categories associated with the pleasantness/pleasure dimension (Disgust, Shock, Contentment, Excitement, Nostalgia) received higher mean ratings from 18-34s than 35+ year old consumers. Emotion categories associated with the arousal/engagement/activation dimension (Boredom and Underwhelmed) were rated higher by the older group than the younger group.

Just one emotion category - Excitement - showed a significant interaction ($p < 0.05$; Table 3.3) between age group and sample (Figure 3.3). Further analyses

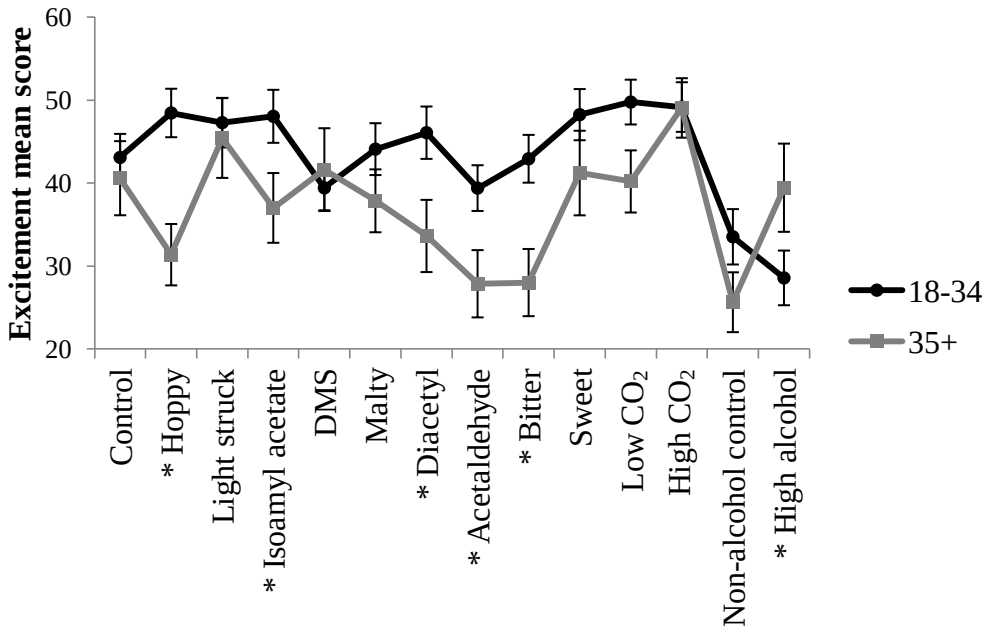


Figure 3.3 Mean ratings of Excitement (and SEM) for each sample by age group. As per Tukey's HSD post hoc tests, * denotes a significant difference in the rating of Excitement between age groups ($p < 0.05$).

showed that that the Diacetyl, Bitter, Isoamyl acetate, Acetaldehyde, and Hoppy samples were rated higher by the younger age group, contributing to the main effect of age group for Excitement. Contrary to this overall trend, the Non-alcohol control, High CO_2, Sweet, Malty, Light struck and DMS samples were not rated significantly differently between age groups. The High alcohol sample also bucked the overall trend in that significantly higher ratings of Excitement were assigned to this sample by 35+ year old consumers than 18-34 year old consumers. Liking also showed a significant interaction between age group and sample, but only for a subset of the samples that interacted between sample and age group in Excitement. As in Excitement, the Bitter and Hoppy samples received significantly lower ratings from the older age group and the High alcohol sample was scored significantly higher by the older consumers.

3.5 Discussion

The first objective of this study was to create an approach for the generation of a reduced product-specific consumer-led emotion lexicon. Firstly, it was found that the process was accelerated through 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 full lexicon of 43 terms was comparable in length to other published emotional lexicons (e.g. Ng et al. (2013); King and Meiselman (2010); Chrea et al. (2009)). Cluster analysis proved a useful tool for grouping terms into emotion categories of similar terms. As only 17 participants generated and rated the terms ahead of the cluster analysis, there is a low probability that this same clustering would be achieved by another group of participants. However, cluster analysis allowed subtle modifications at the hands of the researcher to reduce overlap and confusion between categories. With reference to each emotion category's internal validity (i.e. Cronbach's α) and by using linguistic checks, the final 9 emotion categories were clearly defined, whilst allowing for a breadth of emotions to be reported by consumers. Whilst another group of participants may have generated slightly different terms and their ratings been clustered slightly differently, the researchers feel confident that, in each case, very comparable emotion categories would be generated after following the same process. Of course, this can only be demonstrated by conducting further work with one or more naïve groups of participants.

The second objective of this study was to apply the reduced product-specific consumer-led emotion lexicon to the discrimination of the emotional responses elicited by sensorially-distinct samples. Firstly, it should be noted

that the use of the reduced product category-specific consumer-led lexicon permitted relatively quick emotion assessment for each individual consumer, reducing the potential for fatigue and boredom associated with the task.

With regard to the results themselves, 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); see section 1.1.2) 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. The PCA plot provided a useful guide for comparing sample differentiation by each emotion category because categories co-located in the emotional space elicited similar patterns of response to the sample set, as shown by post hoc tests. By comparing the relative abilities of similar emotion categories (and liking and familiarity) to discriminate between samples, it was possible to evaluate the effectiveness of the reduced product-specific consumer-led lexicon.

Four emotion categories were identified by the PCA as pleasant (Nostalgia, Contentment, Tame/Safe, Excitement) and three as unpleasant (Disconfirmation, Disgust, Shock). These emotion categories showed a general distinction between the two base beers, with the samples based upon the Non-alcohol control (i.e. non-alcohol commercial lager) receiving higher ratings in unpleasant emotion categories and lower ratings in pleasant emotion categories than the majority of samples based upon the Control (i.e. commercial lager).

However, very few differences were observed between the controls and the samples based upon them. Two of the three unpleasant emotion categories (Disgust and Shock) were able to discriminate between the Non-alcohol control and High alcohol samples but none of the four pleasant emotion categories. Just one emotion category revealed a significant difference between the Control and the samples which were based upon it, with lower ratings of Nostalgia given to the Light struck sample compared to the Control. Interestingly for this particular emotion category, this was in spite of no significant difference in familiarity between these samples. This could be related to the description of nostalgia as referring to a preference for objects that were more common when one was younger (Holbrook and Schindler, 1991). Therefore, whilst familiarity may have been similar between these two samples, the light struck aroma may have been less evocative of the consumers' youths. Although differences were not generally found for the manipulated samples relative to the Control sample, the individual manipulated samples were found to commonly differ to one another in a number of emotion categories, showing the discriminability of the reduced product-specific consumer-led emotion lexicon.

Boredom and Underwhelmed - the emotion categories associated with PC2 (arousal/engagement/activation) - were less discriminating than those associated with PC1 and showed significant differences between just a few samples. For both Boredom and Underwhelmed, lower ratings were assigned to the High alcohol sample than both the Bitter and Acetaldehyde samples. For Underwhelmed only, the Light struck sample was also scored significantly lower than the Bitter and Acetaldehyde. As an aside, it might, at first glance,

be surprising that the Bitter sample was among the least engaging/arousing/activating because there is innate rejection of bitter substances and it would be expected that there would be an alerting function when consuming a bitter substance. However, it must be remembered that beer consumers were participating in this study and they have likely adapted over time to accept the bitterness associated with beer. If the same Bitter sample were presented to a non-beer consumer then it would be reasonable to expect that this sample would be among the most engaging/arousing/activating because the period of adaptation and acceptance has not occurred.

Returning to the low discriminability of Boredom and Underwhelmed, consultation with consumers exposed a difficulty in rating emotion categories associated with PC2 (particularly Underwhelmed). Consumers reported that it was counterintuitive to give high ratings when feeling these unactivated emotions. This was evident in the range of mean ratings assigned to these two emotion categories; there was a difference of just 8.9 and 11.6 points between the highest and lowest rated samples for the Boredom and Underwhelmed emotion categories respectively. This is compared to a range of 18.1-36.1 points for the other emotion categories. This is particularly interesting given comparison with Spanish consumers in Chapter 5 whose language appears to better allow consumers to reflect their emotions on the arousal/engagement/activation dimension and, in fact, provide greater discriminability as compared to other emotion categories (for a full discussion, see section 5.6).

None of the 9 emotion categories discriminated between the Low CO_2 and High CO_2 samples, suggesting no difference in emotional response associated with sensory properties related to carbonation. This is in contrast to previous research where it was shown that more highly carbonated commercial beers were associated with more pleasant and engaging emotions (Chaya et al., 2015). This result was found in spite of a much narrower range of carbonation (2.5 - 2.8 volumes) than the present study (~1.6 - ~4 volumes). Perhaps carbonation was a more salient sensory property in driving emotional response for the set of beers included in the Chaya et al. (2015) study, giving rise to relative emotion discriminations between those products. The implication of this is that subtle changes to the sensory properties of a product may have a large impact on emotional response in some contexts (e.g. different sensory properties included within the product or in comparison to other products) but relatively large changes may have very little or no impact on emotional response in other contexts. Further research should probe this in more detail.

Overall, there were fewer effects of manipulating the control samples than expected. An observation that may be pertinent to this is that the Non-alcohol control and High alcohol samples were very different in how consumers responded emotionally to them as compared to the other samples. This was clearly seen on the PCA where these two samples were separated from the others on PC1. More detailed analysis using post hoc tests also showed that these two samples were typically grouped separately to the other samples. Whilst relative differences between the other samples can be seen in the PCA and directional differences in their mean scores, consumers' emotional responses to these samples were largely homogenous. It is a distinctly possible

that the presence of the Non-alcohol control and High alcohol samples lead to convergence of scores for other samples (e.g. the DMS sample may have had more associations with disgust than the Control, but after consumption of the Non-alcohol control or High alcohol samples, the DMS and Control samples would be evaluated by the consumer to relatively similar for this emotion). Therefore, the inclusion of the atypical Non-alcohol control and High alcohol samples is likely to have affected the discriminability of emotional response between the other samples which is a major limitation of the work presented. Nevertheless, it is promising for the method itself that some differences between samples based on the Control were observed.

In order to further assess the effectiveness of the reduced product-specific consumer-led emotion lexicon, its ability to discriminate between the responses of different consumer groups was evaluated. It was common across a number of emotion categories (associated with both pleasantness and engagement) that there were differences in overall ratings between consumer groups. Across emotion categories, it was seen that, where main effects of gender were found, women generally gave lower ratings than men. This was surprising as women are stereotyped as more emotional than men (Fabes and Martin (1991); Plant et al. (2000); Timmers et al. (2003)), with females also exceeding males in reported emotionality and emotion expressivity (Allen and Haccoun (1976); Gross and John (1995)). There could be an effect of familiarity in that women indicated that they were generally less familiar with the samples than men. If this is the case, it is interesting that familiarity had an effect on some reports of emotion intensity but not on liking. However, Sester et al. (2013) did not note any particular relationship between familiarity with

beer and affect. Added to the fact that differentiation between samples in familiarity had little effect on the differentiation in emotional response, this suggests that familiarity was not involved in the differences in ratings of emotion responses between genders. Nevertheless, further investigation of the relationships between familiarity and emotion are needed to understand the potential effect of familiarity on the differences observed between consumer groups.

An alternative explanation for the lower emotional ratings by women is gender roles. These have been discussed as playing an important role in emotions (King et al. (2010); Fischer (1993); Grossman and Wood (1993)) and gender role characteristics have indeed been found to moderate the relationship between gender and emotion expressivity (Kring and Gordon, 1998). As beer is viewed as a relatively masculine beverage (Landrine et al., 1988), it could mean that men have a gender role to be emotionally involved with beer where women do not, giving the differences in the intensity of reported emotion. It would be of particular interest to explore the relationship between experienced and reported emotion in response. In relation to the present research, it could be asked if men both feel and report higher emotions in response to beer than women or if men feel emotions similarly to women but report them higher. If the latter were found, it would be an indication of an effect of gender roles on emotional ratings of beer. Another avenue of research would be to investigate the relationship between the intensity of reported emotion and gender for neutral or typically feminine product categories. Lower emotional ratings by men to culturally feminine product categories may also be an indication of gender roles playing a part in reports of emotional response.

The literature reveals a trend for adults to experience more positive affect and less negative affect with increasing age (Mroczek and Kolarz, 1998). Whilst the 35+ age group, in general, assigned lower ratings than the 18-34 to the unpleasant emotion categories (Disgust and Shock), the younger age group tended to score samples higher for pleasant emotion categories (Excitement, Nostalgia, and Contentment), meaning the expected positive affect associated with age was not seen. However, for the emotion categories associated with arousal/engagement/activation (Boredom and Underwhelmed), higher scores were generally assigned to samples by the older age group. These differences were not related to familiarity as there was no difference found between age groups in this measure. It could be that these results are related to reduced emotional expressivity in older adults (Gross et al., 1997), with experienced pleasant or unpleasant emotions not reported as intensely by the older consumers. However, this does not explain the higher ratings of Underwhelmed and Bored. As was the case between genders, it would be informative to explore the relationship between experienced and reported emotion across age groups to learn about whether the consumer group differences result from the experience of emotion elicited by samples or the emotion experience is similar and it is merely the report that differs.

Putting the differences in overall ratings of emotion categories between consumer groups aside, most emotion categories showed no significant interactions between sample and gender or age group. This means that, whilst scale usage may have been different between consumer segments, the discrimination between the samples included in this study was generally similar irrespective of gender or age group when measured by the reduced

product-specific consumer-led emotion lexicon. The only emotion category to reveal a significant interaction between sample and consumer group was Excitement, which showed that sample and age group interacted. Some samples followed the overall trend identified by main effects of age group and were rated significantly lower by 35+ than 18-34 consumers (e.g. Bitter, Hoppy), some samples did not differ in ratings by age group (e.g. Control, Malty), and just the High alcohol sample was rated significantly higher in Excitement by the older consumer group.

Liking also showed a significant interaction between sample and age group and analyses revealed that Excitement captured the same differences in response to the samples between age groups as liking. However, liking did not show any differences between genders where some emotion categories revealed differences in scale usage between genders. This is added to the fact that liking was unable to discriminate between base beers and their associated manipulated samples at all (e.g. Shock and Disconfirmation were rated significantly higher in the High alcohol sample than the Non-alcohol control, whereas there were no significant differences in liking), showing that the sum of 9 emotion categories were able to match or go beyond the sample discriminability of liking. This confirms what has been reported in previous research (Ng et al., 2013). The complexity of emotional response as compared to liking is in keeping with the descriptions of preference and emotion given by Scherer (2005) in that preferences have a low behavioural impact (other than approach or avoidance), whereas emotions have a high behavioural impact, affecting more complex and varied behaviours.

Familiarity was very discriminative between samples but was found to have very little relationship with observed emotional responses beyond the fact that more familiar samples were generally associated with more pleasant emotions than less familiar samples. This was more perhaps related to the low familiarity with the Non-alcohol control and High alcohol samples; it is likely that such a strong relationship with pleasantness would not be observed for familiarity had these two samples been absent based on the fact that familiarity was not a good predictor of the groupings of sample for each emotion category.

3.6 Conclusions

This chapter's first main objective was to create an approach for the development of a product-specific reduced consumer-led emotion lexicon. Group interviews sped up the lexicon generation process, whilst maintaining the quantity of generated terms. Cluster analysis proved an effective approach for reducing the lengthy lexicon to a number of emotion categories of similar emotion terms. In addition to savings in time and resources, the use of a product category-specific reduced consumer-led lexicon permitted relatively quick emotion assessment for each individual consumer, reducing the potential for fatigue and boredom associated with the task. Taking this together, this chapter's first objective was successfully met.

The second main objective was to use the reduced product-specific consumer-led emotion lexicon to discriminate across a range of beer samples specifically designed to elicit specific sensory properties. Emotion categories co-located in emotional space discriminated between samples similarly but, importantly,

there were subtle differences. There were very few differences in emotional response observed between the control samples (Control, Non-alcohol control, Low CO_2) and the samples which were manipulated from them, owing perhaps to a convergence effect due to the atypicality of emotional response associated with the Non-alcohol control and High alcohol samples. However, the approach did discriminate between a number of the individual manipulated samples, demonstrating the discriminability of the method. Furthermore, the reduced product-specific consumer-led emotion lexicon was able to discriminate beyond liking.

The reduced product-specific consumer-led emotion lexicon was also able to show potentially interesting differences in overall ratings of emotion categories between genders and between age groups, especially given that no such differences were found for liking. Between genders, this may be related to familiarity or gender roles. Across age groups, differences in emotional expressivity were discussed. It was recommended that research exploring the relationship between experienced and reported emotion should be carried out to inform further about the differences in self-report between consumer groups. Despite revealing differences in the overall ratings of emotion categories, the reduced emotion lexicon was relatively ineffective at showing differences in emotional response to individual samples across different consumer segments. Nevertheless, one emotion category showed interactions between sample and age group and these went beyond those identified for liking.

Although the reduced product-specific consumer-led emotion lexicon was more effective than liking at revealing differences between samples as well as showing differences between consumer groups, differences in emotional response to samples and between the responses of different consumer groups were not as frequent as might have been expected. As previously discussed, it is likely that the Non-alcohol control and High alcohol samples had a large part to play on this. However, these observations may also owe somewhat to at least one of two factors:

1. There is limited effect of manipulating the selected sensory properties on consumer emotional response
2. The rating of the reduced consumer-led emotion lexicon emotion categories was not sensitive enough to reveal existing differences in emotional responses

A possibility relating to the second factor is that, by reducing the full lexicon of 43 terms to just 9 emotion categories through modified cluster analysis, there was a compromise in the level of detail acquired about differences in consumer emotional responses to different sensory properties. In order to explore this further, a direct comparison between consumer assessments of samples using both the full and reduced lexicons was required. This will be the focus of Chapter 4.

Chapter 4: Comparing the effectiveness of full and reduced emotion lexicons

4.1 Introduction

There is a history in psychology of successful reductions of affect measurement questionnaires and a current trend for rapid methods in sensory and consumer science, making the development of short forms inevitable. However, it is important that these shorter approaches do not overly compromise their effectiveness. This section reviews the relative efficacy of reduced emotion forms as compared to full emotion measurement approaches.

4.1.1 Reduced emotion forms in psychology

There is an extensive history of psychological affective questionnaires (see section 1.3.1.1) being successfully modified to offer savings in time where often multiple measurement instruments are employed for clinical application. The Multiple Affect Adjective Check List (MAACL; Zuckerman (1960)) is an example of an affect measurement tool with significant popularity. By the time the revised version was published some two decades later (MAACL-R; Zuckerman and Lubin (1985)), the authors had identified 716 published articles and doctoral dissertations that had made use of MAACL. With a view to saving time in completing the questionnaire, a shorter version comprising just 66 items was tested and shown to be equivalent to the original form in both reliability and validity (Lubin et al., 2001). Subsequently developed affect questionnaires have also been successfully reduced. For example, Profile of Mood States (POMS; McNair et al. (1971)) is a 65-item scale to measure psychological distress on six dimensions (anger–hostility, vigour–activity, tension–anxiety, depression-dejection, fatigue–inertia, and confusion–

bewilderment). A reduced 37-item version was developed (Shacham, 1983) and shown to be comparable to the original version (Curran et al., 1995). A further example is the Positive and Negative Affect Schedule (PANAS) which is a 20-item self-report measure of affect (Watson et al., 1988) but was expanded to include 60 items (PANAS-X, Watson and Clark (1991)). However, a short form of just 10 items was subsequently developed and validated (I-PANAS-SF; Thompson (2007)). In fact, this short form is so effective that it has been successfully applied to non-native English speakers (Karim et al., 2011).

4.1.2 Reduced emotion forms in sensory and consumer science

Given the history in psychology of shortening affective questionnaires, it is perhaps no surprise that a trend towards reducing forms has already developed. This is also in keeping with a more general trend in sensory and consumer science for rapid methods (for a review, see Valentin et al. (2012)). EsSense Profile (King and Meiselman, 2010) has proven a popular self-report questionnaire in application to sensory and consumer science (see section 1.3.1.1) and has undergone a reduction through the development of shorter 25-term version named EsSense25 (Nestrud et al., 2013). Researchers printed the 39 EsSense terms on individual cards and asked participants to sort the cards into 2 or more groups. Participants were then asked which word best summarised each group. Cluster analysis was performed on the responses to the 25 terms which were then validated by having a new cohort of participants sort the terms. It was found that the clusters were the same as when there were 39 cards to sort. Subsequently, this approach was used to reduce the lexicon further to just 10 representative terms (Cardello et al., 2014). However, it was

found that a probable demand characteristic was evident in that the more terms that were available to the consumer, the more terms were rated or checked. Shorter lexicons also showed a presumable halo dumping effect when consumers used a rating approach as higher scores were given when terms were included in a shorter form. This suggests that, where an elicited emotion is not included in the form, consumers 'dump' this emotion onto other scales in an attempt to reflect their response. Therefore, whilst the described approach offers a reliable way to reduce a lexicon to a shorter list of terms, it is liable to response biases.

In the reduction of GEOS to ScentMove (see section 3.1.2), groups of three representative terms for each of the six GEOS dimensions were rated by participants, with a high correlation between the original and modified questionnaires reported (Porcherot et al., 2010). Whilst the number of assessments was reduced, many of the original emotion terms were included due to the grouping. Therefore, this could potentially avoid halo dumping to some extent. The approach created in Chapter 3 of this thesis presents an even smaller potential for halo dumping as the full lexicon was preserved in the emotion categories.

4.1.3 Objectives

The main objective of the research described in this chapter was to compare the relative efficacy of the previously described product-specific reduced consumer-led emotional lexicon (Chapter 3) to the full emotional lexicon upon which it was based. This was assessed by comparing (a) the relative discriminability between samples of each approach, and (b) the ability of the

reduced form emotion categories to differentiate between the responses of different consumer groups as compared to the full form. Further objectives were (c) to explore the potential effect of form reduction on halo dumping, and (d) to assess the effectiveness of the modified cluster analysis for grouping terms that elicit similar patterns of response to one another as well as to the reduced form emotion category to which they belonged.

4.2 Materials and method

4.2.1 Procedure

The 109 consumers who assessed the 14 samples using the reduced form (see section 3.3.1) also assessed these same samples using the full form. In the interest of counterbalancing, 57 consumers completed the full form first and the reduced form second whilst 52 consumers completed the forms in the opposite order.

The full form consisted of all 43 individual emotion terms (see Table 3.2b), each associated with its own 150mm continuous line scale. In all other regards, the assessment procedures for the full form mirrored those of the reduced form (see section 3.3.2).

Two-way between subjects ANOVA was applied to each emotion term (and liking and familiarity) to consider the effects of form order on ratings (IBM SPSS Statistics 22). For the ‘bored’, ‘calm’, ‘overwhelmed’, ‘safe’, ‘shocked’, and ‘tame’ emotion terms (in addition to familiarity), consumers that completed the full form first gave significantly higher scores ($p < 0.05$) than those who completed it second. In contrast, ‘relieved’ was rated significantly lower ($p < 0.05$) by consumers who completed the full form second. Despite

these observed magnitude differences, there were no significant interactions between form order and sample ($p > 0.05$), showing that there was minimal effect on reported emotion using the full form no matter whether it was completed before or after the reduced form.

4.2.2 Data analysis

The mean ratings assigned to every sample for each individual full form emotion term and each reduced form emotion category subjected to MFA (XLSTAT, Version 2009.6.03), which can be seen as an extension of PCA in that allows the comparison of more than one dataset in the same space. Therefore, this permitted the identification of the relative locations of individual emotion terms and emotion categories in the previously identified circumplex emotional space (see section 3.4.1) and differences between samples in their projections in the space depending on the form used. Further analysis was performed on each of the full form emotion terms (and liking and familiarity) using mixed model ANOVA, with sample as a fixed factor and subject as a random factor (IBM SPSS Statistics 22). Post hoc Tukey's HSD tests were applied where relevant to determine which samples significantly differed from one another. Post hoc sample discriminations were then compared to those previously generated from the reduced form (see Table 3.4).

Additional analyses were carried out for the single term emotion categories Underwhelmed and Boredom to explore potential effects of halo dumping as observed for shorter forms in the literature (i.e. higher ratings assigned to the short form). Ratings of the terms 'underwhelmed' and 'bored' were compared

using ANOVA with form and sample as fixed factors. In this way, main effects of form and interactions between form and sample could be investigated. Halo dumping would be evident if the reduced form received significantly higher ratings of either 'underwhelmed' or 'bored' than the full form. Significant interactions would show a more complex relationship between form length and sample rating.

In order to investigate consumer group effects, further ANOVAs were carried out with fixed effects of sample, gender and age group for each emotion term (and liking and familiarity). Interactions between sample and gender and between sample and age group were explored in order to investigate the effects of individual sensory properties on consumer group ratings of emotion terms (SPSS Statistics 22, IBM, USA). The ability of the full form emotion terms to differentiate between consumer groups was then compared to the reduced form's ability (see Table 3.3).

4.3 Results

4.3.1 Emotion space

The mean ratings assigned to every sample for each individual full form emotion term and each reduced form emotion category were subjected to MFA. A high RV coefficient of 0.791 indicated that the two datasets were relatively closely aligned. As previously observed for the reduced form alone (section 3.4.1), the emotional space was consistent with circumplex models of emotion (see section 1.1.2) with the majority of the data variance was accounted for by the first two dimensions (88.2%; Figure 4.1a), with dimension 1 (73.89%) correlated with emotions associated with

pleasure/pleasantness and dimension 2 (14.31%) correlated with emotions associated with arousal/engagement/activation.

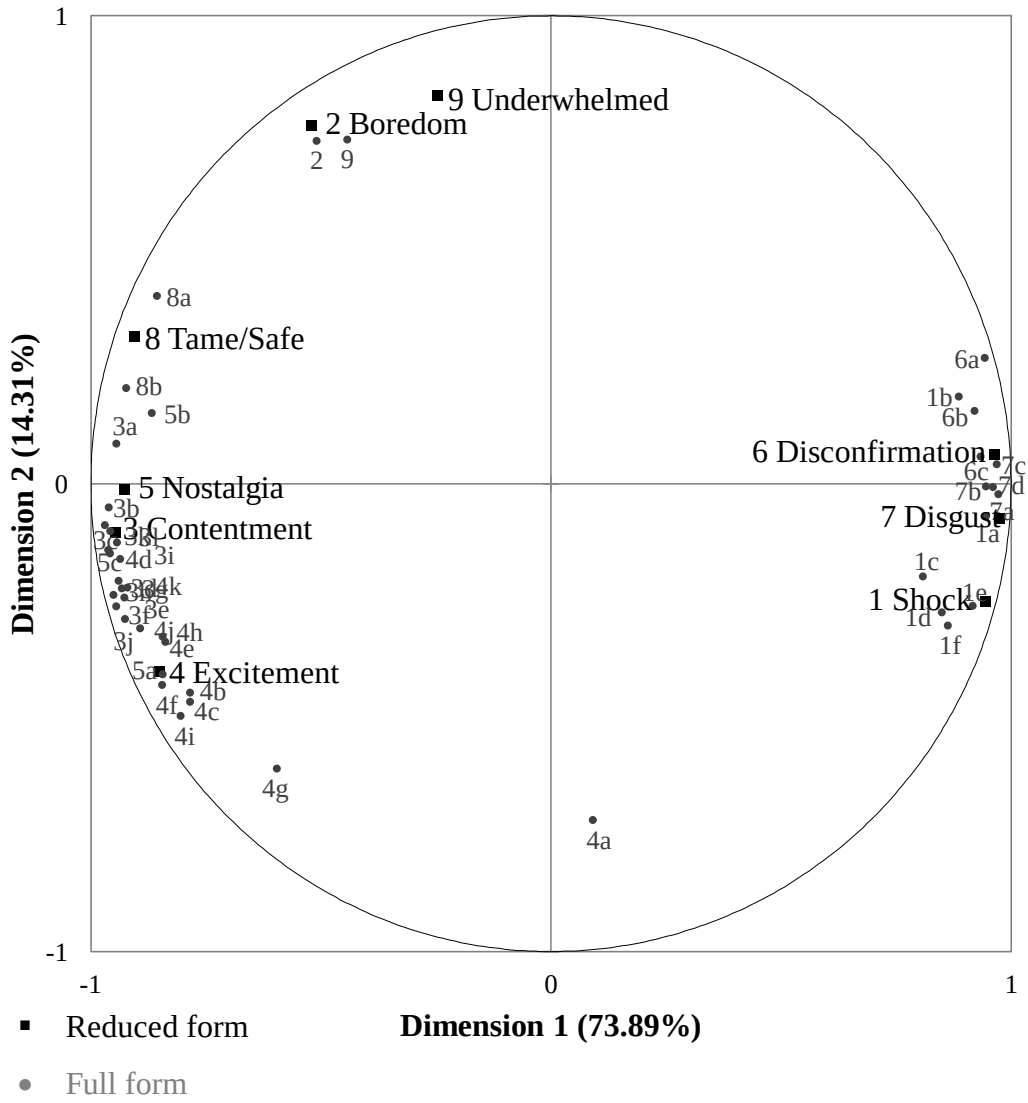


Figure 4.1a MFA plot indicating the positioning of individual emotion terms (full form, labelled by emotion term according to Table 3.2b) and emotion categories (reduced form, emotion categories labelled) in 2-dimensional emotion space.

Individual emotion terms were very closely located in this 2-dimensional emotional space to the emotion categories to which they belonged. The most prominent example of a discrepancy between an emotion term and its emotion

category was the term ‘curious’ (4a; Figure 4.1a) which belonged to the Excitement emotion category in the reduced form. Compared to the majority of the other terms that belong to this emotion category which were highly negatively correlated with the first dimension, ‘curious’ (4a; Figure 4.1a) did not load onto the first dimension but instead loaded highly negatively onto the second dimension. A similar effect was also found for the term ‘interested’ (4g; Figure 4.1a) - which also belonged to the Excitement emotion category of the reduced form - except that ‘interested’ loaded approximately equally onto the first two dimensions.

When mapping samples in this emotional space (Figure 4.1b), most were relatively closely co-located whether assessed using the reduced or full form. The relatively small differences observed in dimension 1 generally showed that the reduced form elicited more pleasant/pleasurable emotions than the full form (e.g. DMS, Low CO_2, Bitter). However, the largest differences between forms were found in dimension 2 (arousal/engagement/activation). The Non-alcohol control and High alcohol samples were perceived as less emotionally engaging when the full form was used, whereas the Acetaldehyde and Bitter samples were more engaging when assessed with the full form.

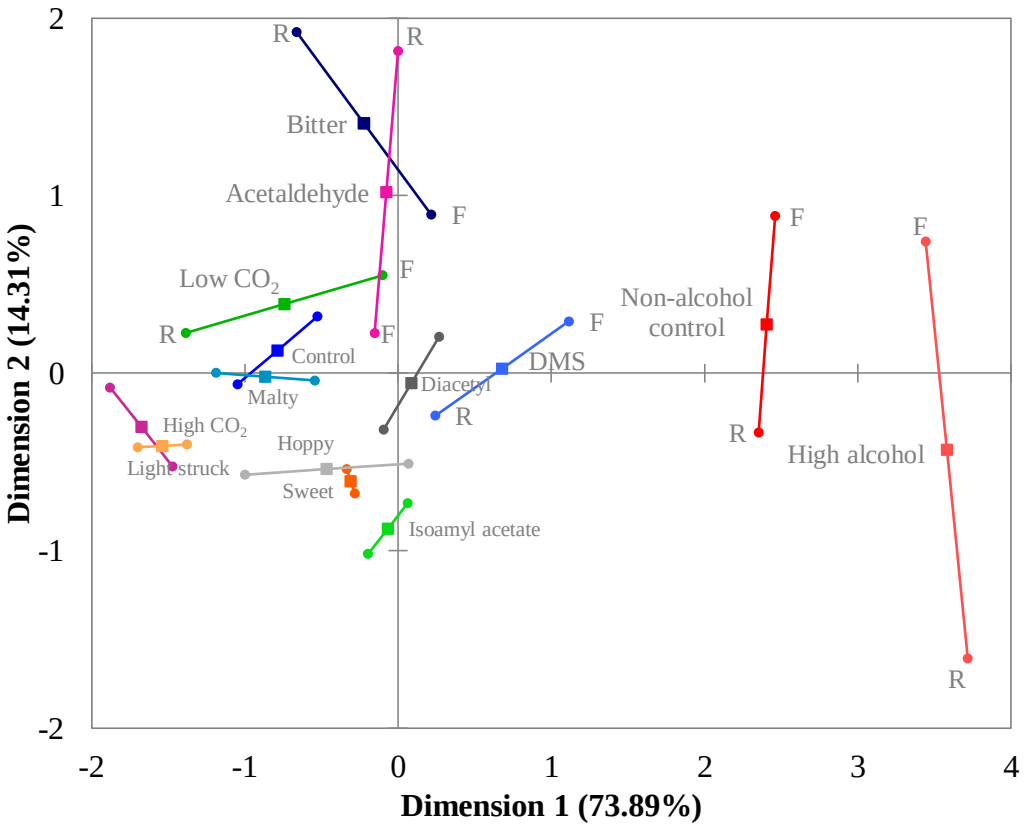


Figure 4.1b MFA plot showing relative positions of samples in the 2-dimensional emotion space according to reduced and full forms. F refers to the full form and R to the reduced form (selected samples of interest only).

4.3.2 Comparing the discrimination abilities of the full and reduced forms

As would be expected due to the product-specific nature of the generated lexicon, every individual emotion term (as well as liking and familiarity) was shown by ANOVA to give a significant effect of sample (Table 4.1). Therefore, post hoc comparisons were made to indicate which samples significantly differed in their ratings from one another (Tables 4.2a-j). The ways that individual emotion terms from the full form discriminated between samples was compared to the discriminability of the reduced form's emotion

Table 4.1 p-values for main effects of sample, gender, and age group, and interactions between sample*gender and sample*age group for each full form emotion term (and liking and familiarity).

Emotion category	Emotion term	Sample	Gender	Age	Sample* Gender	Sample* Age
1 Shock	(a) Alarmed	**<0.001**	**0.032**	**<0.001**	0.114	0.055
	(b) Cheated	**<0.001**	0.341	**0.002**	0.686	0.116
	(c) Confused	**<0.001**	**0.008**	**<0.001**	0.530	0.830
	(d) Overwhelmed	**<0.001**	**<0.001**	**0.011**	0.566	0.535
	(e) Shocked	**<0.001**	**<0.001**	**<0.001**	0.406	0.183
	(f) Strange/weird	**<0.001**	**<0.001**	**<0.001**	**0.020**	**0.025**
2 Boredom	Bored	**<0.001**	**<0.001**	**0.012**	0.301	0.506
3 Contentment	(a) Calm	**<0.001**	**<0.001**	**<0.001**	0.885	0.443
	(b) Comfortable	**<0.001**	**<0.001**	**0.018**	0.774	0.117
	(c) Comforted	**<0.001**	**<0.001**	**0.001**	0.734	0.141
	(d) Content	**<0.001**	**0.001**	**<0.001**	0.624	**0.008**
	(e) Enjoyment	**<0.001**	**<0.001**	0.175	0.622	0.373
	(f) Good	**<0.001**	**<0.001**	0.059	0.802	**0.008**
	(g) Happy	**<0.001**	**<0.001**	**<0.001**	0.387	**0.019**
	(h) Nice	**<0.001**	**<0.001**	**0.005**	0.736	**0.012**
	(i) Pleasant	**<0.001**	**<0.001**	**0.029**	0.856	0.077
	(j) Pleased	**<0.001**	**<0.001**	**0.009**	0.896	0.058
	(k) Relaxed	**<0.001**	**<0.001**	**0.003**	0.957	0.194
	(l) Satisfied	**<0.001**	**<0.001**	**0.025**	0.875	0.608
4 Excitement	(a) Curious	**<0.001**	**<0.001**	**<0.001**	0.059	0.058
	(b) Enthusiastic	**<0.001**	**<0.001**	**0.018**	0.644	**0.040**
	(c) Excited	**<0.001**	**<0.001**	**<0.001**	0.089	**0.009**
	(d) Fulfilled	**<0.001**	**<0.001**	**0.007**	0.553	**0.039**
	(e) Fun	**<0.001**	**<0.001**	**0.001**	0.859	**0.014**
	(f) Impressed	**<0.001**	**<0.001**	0.764	0.785	0.112
	(g) Interested	**<0.001**	**<0.001**	**0.011**	0.903	0.033
	(h) Optimistic	**<0.001**	**<0.001**	**<0.001**	0.926	0.070
	(i) Pleasantly surprised	**<0.001**	**<0.001**	0.200	0.939	0.076
	(j) Want	**<0.001**	**<0.001**	**0.001**	0.598	**0.001**
	(k) Warm	**<0.001**	**<0.001**	**<0.001**	0.831	0.457
5 Nostalgia	(a) Desirous	**<0.001**	**<0.001**	**0.004**	0.819	0.144
	(b) Nostalgic	**<0.001**	**<0.001**	**0.024**	0.603	0.103
	(c) Relieved	**<0.001**	**<0.001**	**<0.001**	0.779	0.590
6 Disconfirmation	(a) Disappointed	**<0.001**	0.725	**0.024**	0.401	**0.013**
	(b) Dissatisfied	**<0.001**	0.415	**0.009**	0.207	**0.012**
	(c) Unpleasantly surprised	**<0.001**	0.733	**<0.001**	0.909	**0.007**
7 Disgust	(a) Disgusted	**<0.001**	0.080	**<0.001**	0.341	0.133
	(b) Horrible	**<0.001**	0.221	**<0.001**	0.153	**0.007**
	(c) Repulsed/repelled	**<0.001**	0.389	**<0.001**	0.808	0.097
	(d) Unpleasant	**<0.001**	0.067	**<0.001**	0.100	0.102
8 Tame/Safe	(a) Tame	**<0.001**	**<0.001**	0.326	0.688	0.700
	(b) Safe	**<0.001**	**<0.001**	**0.008**	0.932	0.461
9 Underwhelmed	Underwhelmed	**0.015**	0.100	0.901	0.754	0.064
Liking		**<0.001**	**0.004**	0.450	0.381	**0.006**
Familiarity		**<0.001**	**<0.001**	0.432	0.956	0.079

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

Table 4.2a Mean scores for the 6 emotion terms belonging to the emotion category Shock across the 14 samples.

Sample	1 Shock	Full form					
		(a) Alarmed	(b) Cheated	(c) Confused	(d) Over-whelmed	(e) Shocked	(f) Strange/ weird
Control	26.3 ABC	19.0 AB	23.5 A	23.2 A	24.4 A	23.2 AB	26.2 AB
Hoppy	34.7 BC	20.0 AB	24.7 AB	28.9 ABC	24.2 A	27.1 ABC	33.6 BCD
Light struck	22.2 A	17.2 A	23.3 A	20.6 A	21.5 A	20.8 A	20.7 A
Isoamyl acetate	35.6 BC	28.0 BC	29.2 ABC	35.4 CD	30.3 AB	34.4 CD	45.6 EF
DMS	35.6 BC	28.0 BC	30.6 ABC	29.9 ABC	27.4 A	32.5 BCD	39.1 CDE
Malty	30.2 ABC	17.8 AB	22.9 A	23.3 AB	22.6 A	22.4 AB	24.5 AB
Diacetyl	36.9 CD	27.5 ABC	30.0 ABC	33.5 BCD	28.2 AB	31.0 ABCD	39.1 CDE
Acetaldehyde	31.7 ABC	25.2 AB	28.8 ABC	30.3 ABC	24.6 A	27.0 ABC	30.7 ABCD
Bitter	25.1 AB	22.7 AB	30.1 ABC	27.3 ABC	25.7 A	25.1 ABC	29.5 ABC
Sweet	34.5 BC	24.6 AB	27.5 ABC	34.0 CD	29.7 AB	35.4 CD	41.4 DE
Low CO_2	23.5 A	25.2 AB	31.3 ABC	29.2 ABC	24.4 A	27.1 ABC	29.8 ABC
High CO_2	22.0 A	20.9 AB	21.9 A	23.7 AB	23.7 A	22.3 AB	26.4 AB
Non-alcohol control	46.9 D	35.6 CD	34.9 BC	34.0 CD	29.0 AB	39.5 DE	46.8 EF
High alcohol	58.1 E	43.7 D	37.6 C	41.0 D	37.2 B	47.6 E	54.5 F

ABCDEF Letters within the same column indicate post hoc groupings by Tukey's HSD.
Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2b Mean scores for the 'bored' emotion term across the 14 samples.

Sample	2 Bor-edom	Full form Bored
Control	31.1 AB	30.6 AB
Hoppy	30.0 AB	28.7 AB
Light struck	31.1 AB	30.1 AB
Isoamyl acetate	28.2 AB	24.2 A
DMS	27.7 AB	26.0 AB
Malty	28.9 AB	31.8 AB
Diacetyl	30.3 AB	28.8 AB
Acetaldehyde	36.3 B	32.4 AB
Bitter	36.2 B	35.0 B
Sweet	27.4 AB	24.7 A
Low CO_2	32.0 AB	35.2 B
High CO_2	31.2 AB	28.3 AB
Non-alcohol control	29.9 AB	25.9 AB
High alcohol	23.0 A	24.2 A

AB Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2c Mean scores for the 12 emotion terms belonging to the emotion category Contentment across the 14 samples.

Sample	3 Conte-ntment	Full form											
		(a) Calm	(b) Com-fortable	(c) Com-forted	(d) Content	(e) Enjo-yment	(f) Good	(g) Happy	(h) Nice	(i) Ple-asant	(j) Pleased	(k) Relaxed	(l) Sati-sfied
Control	50.4 CD	46.5 CD	47.3 CD	43.5 CD	47.5 CDE	47.5 CD	48.4 CD	47.7 BCD	49.8 D	50.0 CD	47.5 CDE	47.7 D	45.5 CD
Hoppy	44.0 CD	46.3 CD	48.7 CD	41.7 CD	47.9 CDE	49.3 CD	49.5 CD	46.3 BCD	49.9 D	52.0 D	50.0 DE	45.3 CD	46.0 CD
Light struck	41.7 BC	45.8 CD	52.5 D	47.8 D	49.0 DE	51.1 D	53.2 D	51.1 D	51.8 D	50.5 D	49.6 DE	51.0 D	48.4 D
Isoamyl acetate	50.7 CD	40.9 BCD	43.3 BCD	39.6 BCD	42.9 BCDE	47.7 CD	49.1 CD	46.2 BCD	46.0 CD	47.3 CD	45.7 CDE	41.7 BCD	42.1 CD
DMS	44.4 CD	38.5 BC	39.0 ABC	34.5 ABC	37.8 ABC	39.9 ABC	41.4 ABC	39.6 ABC	38.9 ABC	39.1 BC	37.3 ABC	37.4 ABC	36.8 ABC
Malty	45.9 CD	49.5 D	48.9 CD	46.3 D	47.6 CDE	49.3 CD	51.2 CD	46.9 BCD	47.8 CD	49.4 CD	47.0 CDE	48.4 D	49.4 D
Diacetyl	44.1 CD	42.9 CD	45.7 CD	38.1 BCD	43.7 BCDE	44.7 BCD	44.6 BCD	45.3 BCD	45.2 CD	45.3 CD	43.7 CDE	45.0 CD	42.8 CD
Acetaldehyde	45.1 CD	43.5 CD	45.0 BCD	39.1 BCD	41.2 BCDE	46.2 CD	45.0 BCD	43.7 BCD	43.8 BCD	46.1 CD	42.1 BCDE	42.7 CD	42.4 CD
Bitter	47.1 CD	44.1 CD	44.6 BCD	38.5 BCD	39.3 ABCD	41.6 ABCD	41.8 ABC	39.3 AB	41.3 BCD	42.4 BCD	40.4 BCD	42.7 CD	39.1 BCD
Sweet	53.5 D	44.3 CD	46.0 CD	40.0 CD	43.0 BCDE	45.8 CD	46.8 BCD	45.5 BCD	45.6 CD	47.7 CD	47.7 CDE	46.6 CD	43.5 CD
Low CO_2	47.4 CD	45.4 CD	44.6 BCD	40.3 CD	41.7 BCDE	43.5 ABCD	47.6 CD	44.8 BCD	43.4 BCD	44.7 CD	42.6 BCDE	45.6 CD	42.4 CD
High CO_2	54.3 D	45.5 CD	49.3 D	44.7 D	49.7 E	52.6 D	51.9 CD	49.6 CD	49.9 D	49.7 CD	51.8 E	51.3 D	49.5 D
Non-alcohol control	32.4 AB	32.7 AB	35.4 AB	30.1 AB	33.5 AB	33.8 AB	36.3 AB	32.8 A	34.1 AB	33.0 AB	32.1 AB	32.2 AB	30.1 AB
High alcohol	29.8 A	28.8 A	30.3 A	26.8 A	29.2 A	32.9 A	32.4 A	29.6 A	29.0 A	27.8 A	29.2 A	29.5 A	26.8 A

ABCDE Letters within the same column indicate post hoc groupings by Tukey's HSD.
Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2d Mean scores for the 11 emotion terms belonging to the emotion category Excitement across the 14 samples.

		Full form										
Sample	4 Excitement	(a) Curious	(b) Interested	(c) Enthusiastic	(d) Excited	(e) Fulfilled	(f) Fun	(g) Impressed	(h) Optimistic	(i) Pleasantly surprised	(j) Want	(k) Warm
Control	42.3 BCDE	40.1 ABC	42.9 ABCD	42.3 C	39.9 AB	42.0 C	40.8 CD	40.5 BCD	41.2 BCD	43.2 BCD	41.4 C	44.0 C
Hoppy	43.0 CDE	48.6 CD	47.2 ABCD	42.4 C	39.9 AB	40.9 C	41.9 CD	44.7 CD	42.3 CD	49.7 D	43.5 C	42.0 C
Light struck	46.7 DE	39.1 ABC	46.3 ABCD	43.8 C	41.3 B	43.8 C	43.7 CD	43.5 CD	43.4 D	46.6 CD	43.3 C	44.7 C
Isoamyl acetate	44.5 CDE	55.5 D	53.7 D	44.3 C	42.5 B	37.5 BC	44.9 D	44.3 CD	42.7 CD	47.7 CD	43.4 C	40.0 C
DMS	40.1 ABCDE	42.8 ABC	40.9 ABC	37.7 ABC	36.0 AB	34.2 ABC	33.9 ABC	34.4 ABC	32.7 ABC	35.6 ABC	34.1 ABC	37.5 ABC
Malty	42.1 BCDE	39.7 ABC	44.8 ABCD	39.2 ABC	38.5 AB	39.7 BC	40.4 CD	38.4 BCD	39.5 BCD	42.1 BCD	43.1 C	43.3 C
Diacetyl	42.1 BCDE	46.1 BCD	48.7 BCD	40.8 BC	38.1 AB	39.6 BC	40.7 CD	40.5 BCD	41.2 BCD	43.6 BCD	39.9 C	41.1 C
Acetaldehyde	35.7 ABC	42.2 ABC	42.5 ABCD	39.2 ABC	36.8 AB	37.4 BC	40.2 BCD	36.4 ABCD	39.0 BCD	37.8 ABCD	37.6 ABC	39.0 BC
Bitter	38.1 ABCD	34.8 A	39.0 ABC	34.6 ABC	35.2 AB	35.5 BC	36.8 ABCD	34.5 ABC	35.2 ABC	34.1 AB	36.5 ABC	36.9 ABC
Sweet	46.0 CDE	46.3 BCD	48.2 ABCD	44.1 C	43.2 B	38.8 BC	44.0 CD	41.7 CD	40.9 BCD	44.0 BCD	39.5 BC	39.9 C
Low CO_2	46.7 DE	37.5 AB	41.1 ABC	35.6 ABC	35.4 AB	38.1 BC	38.2 ABCD	37.0 ABCD	36.5 ABCD	38.4 ABCD	36.5 ABC	38.0 BC
High CO_2	49.1 E	42.7 ABC	49.8 CD	42.3 C	40.6 AB	43.3 C	42.6 CD	45.9 D	41.9 CD	47.3 CD	44.3 C	45.5 C
Non-alcohol control	32.1 AB	39.8 ABC	36.9 A	30.3 AB	30.4 A	29.3 AB	29.2 A	29.5 AB	31.2 AB	29.2 A	28.8 AB	28.2 A
High alcohol	31.0 A	43.9 ABC	37.5 AB	29.5 A	30.5 A	24.9 A	29.9 AB	26.6 A	28.2 A	27.6 A	27.5 A	29.8 AB

ABCDE Letters within the same column indicate post hoc groupings by Tukey's HSD.
Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2e Mean scores for the 3 emotion terms belonging to the emotion category Nostalgia across the 14 samples.

Sample	5 Nostalgia	Full form		
		(a) Desirous	(b) Nostalgic	(c) Relieved
Control	39.7 D	36.7 BCD	35.8 BCD	39.1 BC
Hoppy	32.5 BCD	40.4 CD	33.5 BCD	42.0 C
Light struck	29.3 ABC	40.7 D	40.3 D	41.7 C
Isoamyl acetate	33.9 BCD	37.9 CD	29.4 ABC	36.3 BC
DMS	34.0 BCD	30.7 ABC	26.1 AB	31.3 AB
Malty	34.0 BCD	36.0 BCD	32.0 BCD	43.1 C
Diacetyl	32.1 BCD	36.9 CD	33.4 BCD	35.7 BC
Acetaldehyde	32.0 BCD	31.6 ABCD	28.7 ABC	37.3 BC
Bitter	34.6 CD	31.2 ABCD	37.5 CD	35.6 BC
Sweet	38.8 CD	38.1 CD	29.3 ABC	40.7 BC
Low CO_2	35.7 CD	36.0 BCD	35.4 BCD	36.9 BC
High CO_2	41.3 D	36.3 BCD	38.1 CD	41.7 C
Non-alcohol control	24.7 AB	26.8 AB	25.4 AB	25.7 A
High alcohol	20.3 A	24.6 A	20.9 A	22.9 A

ABCD Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2f Mean scores for the 3 emotion terms belonging to the emotion category Disconfirmation across the 14 samples.

Sample	6 Discon-firmation	Full form		
		(a) Disappointed	(b) Dissatisfied	(c) Unpleasantly surprised
Control	33.7 ABC	34.7 AB	34.0 ABC	26.4 A
Hoppy	34.3 ABC	33.3 A	31.8 ABC	27.8 A
Light struck	29.0 AB	29.0 A	30.5 AB	26.1 A
Isoamyl acetate	36.3 ABC	35.8 AB	34.8 ABC	29.8 AB
DMS	37.7 BC	40.8 ABC	43.4 CDE	29.9 AB
Malty	34.2 ABC	33.6 A	29.8 AB	27.3 A
Diacetyl	40.5 BCD	37.1 ABC	35.8 ABC	34.4 ABC
Acetaldehyde	40.9 CD	37.9 ABC	33.3 ABC	40.3 BCD
Bitter	38.3 BC	40.0 ABC	38.8 BCD	33.6 ABC
Sweet	37.8 BC	35.1 AB	33.0 ABC	32.6 AB
Low CO_2	29.1 AB	38.2 ABC	40.0 BCD	26.1 A
High CO_2	25.7 A	28.8 A	24.7 A	22.8 A
Non-alcohol control	51.4 DE	47.2 BC	49.4 DE	51.2 D
High alcohol	60.0 E	49.2 C	53.3 E	45.2 CD

ABCDE Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2g Mean scores for the 4 emotion terms belonging to the emotion category Disgust across the 14 samples.

Sample	7 Disgust	Full form			
		(a) Disgusted	(b) Horrible	(c) Repulsed/ repelled	(d) Unpleasant
Control	28.6 AB	22.6 AB	23.9 AB	22.6 AB	24.6 A
Hoppy	28.7 AB	23.7 ABC	25.0 AB	24.2 AB	28.6 AB
Light struck	22.2 AB	17.0 A	21.7 A	16.4 A	23.0 A
Isoamyl acetate	29.5 AB	27.3 ABC	27.1 AB	26.1 ABC	32.6 AB
DMS	31.4 B	34.1 CD	34.5 BC	36.7 CD	39.0 BC
Malty	25.9 AB	20.7 AB	19.7 A	20.8 AB	26.1 A
Diacetyl	32.7 BC	27.7 ABC	28.3 AB	28.2 BC	31.5 AB
Acetaldehyde	28.9 AB	24.4 ABC	26.2 AB	22.0 AB	29.1 AB
Bitter	28.7 AB	27.2 ABC	29.9 AB	28.4 BC	32.8 AB
Sweet	28.9 AB	28.6 BC	28.6 AB	29.3 BC	30.8 AB
Low CO_2	21.9 AB	25.8 ABC	27.2 AB	28.1 BC	32.9 AB
High CO_2	20.3 A	21.4 AB	20.7 A	21.2 AB	26.8 A
Non-alcohol control	43.5 C	42.3 DE	42.3 CD	43.7 DE	46.2 CD
High alcohol	54.8 D	47.7 E	47.4 D	48.2 E	54.1 D

ABCDE Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2h Mean scores for the 2 emotion terms belonging to the emotion category Tame/Safe across the 14 samples.

Sample	8 Tame/Safe	Full form	
		(a) Tame	(b) Safe
Control	44.1 CDE	41.7 BC	45.6 C
Hoppy	35.9 BC	37.5 BC	42.0 BC
Light struck	45.5 CDE	40.2 BC	44.6 C
Isoamyl acetate	42.8 CDE	33.0 ABC	38.5 BC
DMS	37.7 BCD	34.4 BC	36.5 ABC
Malty	47.6 DE	43.3 C	46.0 C
Diacetyl	37.5 BCD	33.1 ABC	39.1 BC
Acetaldehyde	39.2 CD	40.0 BC	42.4 BC
Bitter	41.5 CDE	40.7 BC	42.1 BC
Sweet	44.6 CDE	33.5 ABC	39.6 BC
Low CO_2	45.2 CDE	40.9 BC	42.5 BC
High CO_2	49.4 E	40.2 BC	41.7 BC
Non-alcohol control	28.7 AB	31.3 AB	32.9 AB
High alcohol	21.4 A	23.4 A	27.9 A

ABCDE Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2i Mean scores for the 'underwhelmed' emotion term across the 14 samples.

Sample	9 Under-whelmed	Full form Underwhelmed
Control	36.7 AB	36.7 AB
Hoppy	32.5 AB	34.4 AB
Light struck	31.2 A	32.9 AB
Isoamyl acetate	32.5 AB	28.6 A
DMS	36.1 AB	31.3 AB
Malty	35.4 AB	36.3 AB
Diacetyl	38.1 AB	33.8 AB
Acetaldehyde	41.6 AB	34.9 AB
Bitter	42.8 B	42.0 B
Sweet	34.9 AB	31.3 AB
Low CO_2	35.9 AB	39.8 AB
High CO_2	34.0 AB	35.6 AB
Non-alcohol control	32.2 AB	32.4 AB
High alcohol	30.8 A	30.0 AB

AB Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

Table 4.2j Mean scores for the liking and familiarity across the 14 samples when included in the reduced and full forms.

Sample	Liking		Familiarity	
	Reduced form	**Full form**	**Reduced form**	**Full form**
Control	49.2 CDE	52.2 CD	43.9 CDE	49.7 FG
Hoppy	48.6 CDE	54.2 D	35.7 BC	43.4 CDEF
Light struck	56.1 DE	56.3 D	55.3 F	55.5 G
Isoamyl acetate	48.5 CDE	52.2 CD	34.8 BC	30.4 AB
DMS	46.0 CD	42.4 BC	38.3 BCD	32.9 AB
Malty	51.7 CDE	54.7 D	46.5 DEF	48.2 FG
Diacetyl	44.1 BC	49.2 CD	35.4 BC	36.0 ABCD
Acetaldehyde	43.1 BC	47.4 CD	36.0 BC	38.1 BCDE
Bitter	45.9 CD	46.1 BC	44.3 CDE	45.7 DEFG
Sweet	50.6 CDE	50.4 CD	32.3 B	35.5 ABC
Low CO_2	53.2 CDE	47.0 CD	51.3 EF	47.8 EFG
High CO_2	58.5 E	53.9 D	54.9 F	50.6 FG
Non-alcohol control	34.0 AB	35.8 AB	34.6 BC	35.8 ABCD
High alcohol	30.0 A	31.1 A	22.0 A	26.5 A

ABCDEFG Letters within the same column indicate post hoc groupings by Tukey's HSD. Different letters represent significant differences ($p < 0.05$) in rating between samples

categories to which they belonged in order to reveal any differences between the two approaches.

Boredom and Underwhelmed (and liking and familiarity)

The emotion categories Boredom and Underwhelmed offered interesting comparisons between the full and reduced forms because both included a single eponymous emotion term, allowing an investigation of how form length affected consumer ratings. Upon inspection of the MFA plot (Figure 4.1a), it was seen that the emotion terms 'bored' and 'underwhelmed' were closely co-located in the upper-left quadrant (pleasant, low engagement) whether included as part of the full or reduced forms. Post hoc tests showed that the terms were not discriminating between most samples, regardless of form length (Table 4.2b and Table 4.2i).

When 'bored' was included as part of the full form, post hoc tests showed Bitter and Low CO_2 samples were rated significantly higher than the High alcohol, Isoamyl acetate, and Sweet samples. Samples were rated differently when 'bored' was included as part of the reduced form, with Bitter and Acetaldehyde samples rated significantly higher than just the High alcohol sample. The only consistent difference between the two form lengths was that the Bitter sample was found to be rated significantly higher than the High alcohol sample.

'Underwhelmed' also returned somewhat different results based on which form the term was included in. In both forms, the Bitter sample received the highest mean rating in Underwhelmed. However, the full form showed the Bitter sample was significantly more underwhelming than just the Isoamyl

acetate sample, whereas the reduced form showed that the Bitter sample was significantly more underwhelming than both the Non-alcohol control and Light struck samples. Taken together, the results from emotion categories Boredom and Underwhelmed showed that consumers were not particularly discriminating for this sample set using these terms and, where differences were found, there were inconsistencies between forms. These inconsistencies were related to the projection of samples onto the second dimension of the MFA (arousal/engagement/activation; Figure 4.1b), with Bitter and Acetaldehyde samples more opposed to the High alcohol sample when included in the reduced form than in the full form. As a result, there were more differences found between these samples in the post hoc tests of the arousal/engagement/activation-related emotions 'bored' and 'underwhelmed' when included in the reduced form.

It is informing at this point to compare liking and familiarity when rated after completion of the reduced and full forms (Table 4.2j) as the task was identical except for the length of emotion form that preceded it. There were several differences in groupings of samples according to post hoc tests when liking and familiarity were rated after the full or reduced forms. For example, the Non-alcohol control was significantly less liked than the Bitter sample when included in the reduced form but not the full form. Familiarity was scored significantly lower for the Isoamyl acetate sample as compared to the Control sample when included as part of the full form but not the reduced form. Like 'bored' and 'underwhelmed' therefore, there were inconsistencies in consumer evaluations of liking and familiarity dependent on form length.

Further ANOVAs were conducted, comparing ratings of 'bored' (Figure 4.2a) and 'underwhelmed' (Figure 4.2b) with form and sample as fixed factors. Analysis showed no main effects of form ($p < 0.05$; i.e. no significant differences in overall ratings of samples depending on which form was used) and there were no significant interactions between form and sample ($p < 0.05$; i.e. ratings of individual samples did not significantly differ between forms). This suggests that the reduced product-specific consumer-led emotion lexicon did not produce a halo dumping effect, at least for these two emotion terms.

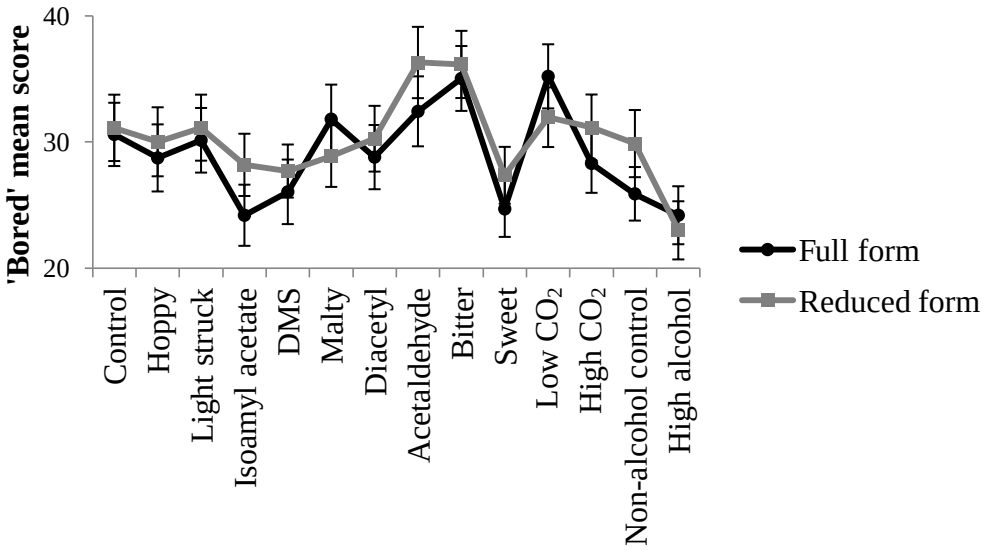


Figure 4.2a Mean ratings of 'bored' for each of the 14 samples when the term was included as part of the full and reduced emotion forms.

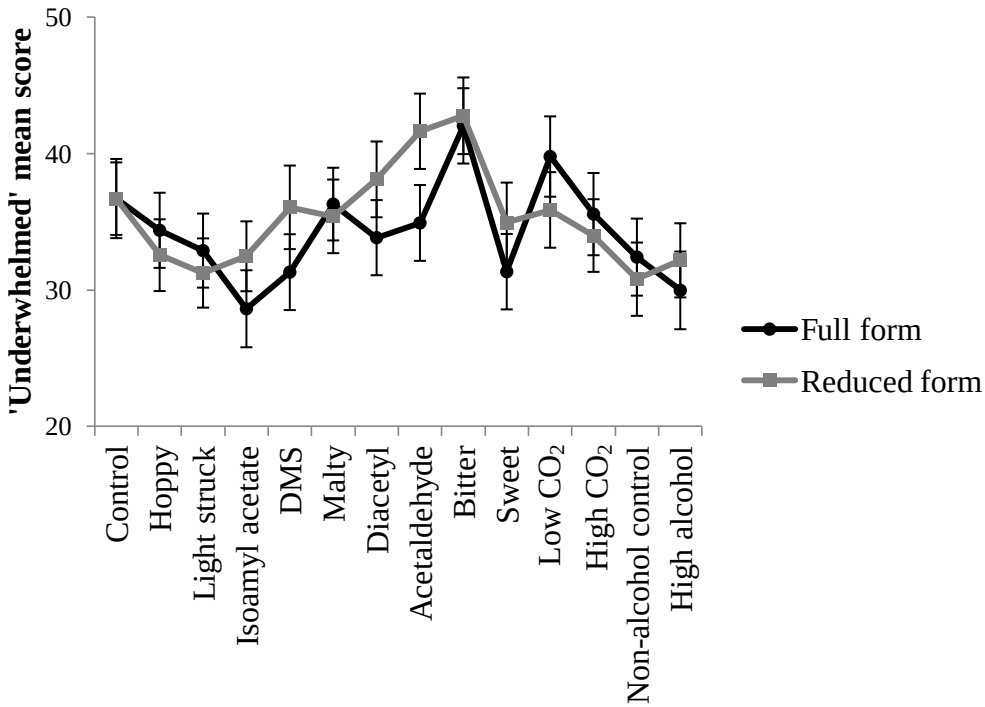


Figure 4.2b Mean ratings of 'underwhelmed' for each of the 14 samples when the term was included as part of the full and reduced emotion forms.

Tame/Safe

The emotion category Tame/Safe also offered an interesting comparison between the two forms as it included just a pair of emotion terms. Figure 4.1a shows that the two emotion terms flank the emotion category and all three points were extremely closely grouped in the 2-dimensional emotion space, although 'safe' was slightly more engaging than 'tame'. Post hoc tests showed that, individually, 'tame' and 'safe' grouped samples very similarly (Table 4.2h). For both, Malty was the highest mean rated sample and was rated significantly higher than Non-alcohol control and High alcohol samples. 'Safe' went further and also showed the Control and Light struck samples were rated significantly higher than the Non-alcohol control and High alcohol samples. The High alcohol sample was the lowest scored sample for both

emotion terms and was rated significantly lower than the majority of samples (Hoppy, Acetaldehyde, High CO_2, Light struck, Bitter, Low CO_2, Control).

The DMS sample was also rated significantly higher than the High alcohol sample for 'tame', whilst the Isoamyl acetate and Diacetyl samples were rated significantly higher than the High alcohol sample 'safe'. Considering these emotion terms separately, it is clear that the two are closely related but do give slightly different groupings of samples. Inevitably then, by combining these terms into an emotion category, information is lost about the differences in consumer emotional responses. However, it was found that the Tame/Safe emotion category was able to discriminate between a number of other samples that the individual terms could not. For example, Tame/Safe showed that the High CO_2 sample was rated significantly higher than the DMS, Acetaldehyde, Diacetyl, Hoppy, and Non-alcohol control samples but these was not found when consumers used individual terms on the full form. In addition, the reduced form showed the Bitter sample to be rated significantly higher than the Non-alcohol control sample which was not the case for 'tame' or 'safe'. The Low CO_2, High CO_2, Sweet, Bitter, Acetaldehyde, and Isoamyl acetate samples received significantly higher ratings of Tame/Safe than the Non-alcohol control but this was not the case for the full form terms. Therefore, by combining the terms 'tame' and 'safe' into a single emotion category, sample discriminability increased.

Nostalgia

The emotion category Nostalgia included just three emotions terms: 'desirous', 'nostalgic', and 'relieved'. Like the reduced form emotion category, none showed significant effects when comparing the base Control,

Non-alcohol control, or Low CO_2 samples to their associated manipulated samples. However, there were subtle differences in the way the terms discriminated between samples which were missed by the reduced form. For example, the individual terms all showed that the Light struck sample received higher ratings than the DMS sample (Table 4.2e) but did not when combined into the Nostalgia emotion category. However, the reduced form revealed some sample differences that were not shown for individual emotion terms (e.g. High CO_2 received significantly higher ratings than the Light struck sample). This demonstrates both losses and gain in discriminability associated with the use of the reduced form.

Shock

For the Shock emotion category, neither the reduced form nor the individual full form emotion terms discriminated between Low CO_2 and High CO_2 samples (Table 4.2a). However, a number of individual emotion terms were able to show a significant difference between the Control and its associated manipulated samples where the reduced form emotion category was not. 'Strange/weird' was rated significantly higher for Isoamyl acetate, Sweet, Diacetyl, and DMS samples. 'Confused' also received higher ratings than the Control for Isoamyl acetate, Sweet, and Diacetyl samples. Furthermore, 'shocked' was rated higher in Isoamyl acetate and Sweet samples than the Control. These three terms ('strange/weird' (1f), 'confused' (1c), and 'shocked' (1e) (Figure 4.1a) were also among the most arousing/engaging/activating that belonged to the Shock emotion category. However, none of the individual emotions terms discriminated between the Non-alcohol control and High alcohol samples, whereas the reduced form

showed the High alcohol sample was rated significantly higher in Shock than the Non-alcohol control. Therefore, Shock emotion category was more discriminating than its constituent terms in the full form between some samples, but less discriminating between others.

Disgust

Similarly to Shock, the Disgust emotion category was able to discriminate between the Non-alcohol control and High alcohol samples (the addition of alcohol was rated higher in Disgust) but none of the individual terms belonging to this category were able to make such a distinction (Table 4.2g). Nevertheless, the reduced form showed no difference between Control and DMS samples in Disgust, whereas 'disgusted', 'repulsed/repelled', and 'unpleasant' were rated significantly higher for the DMS sample than the Control in the full form.

Disconfirmation, Contentment, and Excitement

The Disconfirmation (Table 4.2f), Contentment (Table 4.2c), and Excitement (Table 4.2d) emotion categories, like Disgust, also showed no significant differences between the Control and DMS samples, where 'unpleasantly surprised' (belonging to the Disconfirmation category) was rated higher for DMS than the Control and 'nice' and 'relaxed' (belonging to the Contentment category) were rated significantly lower for DMS than the Control. In addition, 'dissatisfied' (belonging to the Disconfirmation category) received a higher rating for the Low CO_2 sample than the High CO_2 sample. Isoamyl acetate was rated higher than the Control in 'curious' but not in Excitement (the emotion category to which 'curious' belongs).

4.3.3 Consumer group effects

4.3.3.1 Gender

ANOVA showed significant main effects of gender for a number of the full form emotion terms, as well as liking and familiarity (Table 4.1). Pertinently, a number of individual emotion terms showed main effects of gender where the emotion category to which they belonged did not. For example, all the constituent terms of the emotion category Contentment showed main effects of gender, with higher scores obtained from males. This was also the case for the Tame/Safe emotion category. Interestingly, there were no differences between genders in ratings of liking after completion of the reduced emotion form, whereas there was a main effect of liking following assessment of samples using the full form, with males generally scoring samples higher in liking than females. It must be noted that differences between forms were not found in every case; the emotion category Disgust was not rated significantly differently between genders and neither were the category's four constituent individual emotion terms in the full form.

No reduced form emotion category (or liking or familiarity) showed significant interactions between sample and gender and just a single full form emotion term showed such interactions (Table 4.1). This term was 'strange/weird' (belonging to the Shock emotion category), which was rated higher by males for the Diacetyl, Bitter, Acetaldehyde, Hoppy, and DMS samples (Figure 4.3).

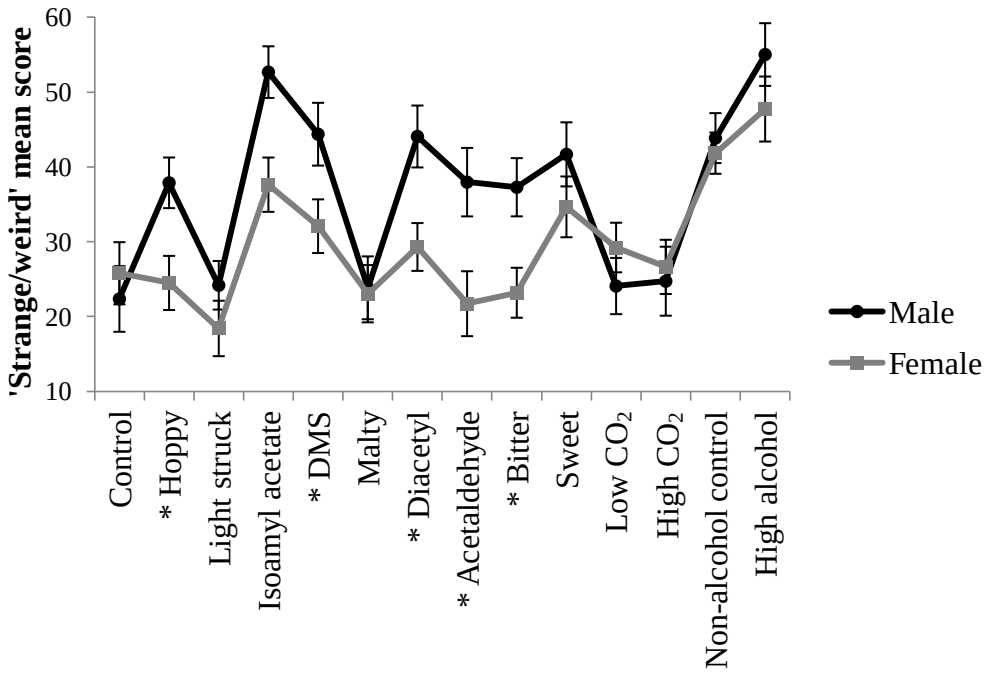


Figure 4.3 Mean ratings of 'strange/weird' (and SEM) for each sample by gender. As per Tukey's HSD post hoc tests, * denotes a significant difference in the rating of 'strange/weird' between genders.

4.3.3.2 Age group

A number of significant main effects of age group were found in the use of individual emotion terms in the full form (Table 4.1). As for gender, the full form revealed age differences in ratings where the reduced form did not. There were no main effects of age group for the Disconfirmation emotion category, whereas the three constituent emotion terms did show effects when included in the full form, with 18-34s scoring 'disappointed', 'dissatisified', and 'unpleasantly surprised' higher than 35+ consumers. Similarly, the emotion category Tame/Safe showed no differences in ratings between age groups, whereas 'safe' was rated significantly higher by 18-34 year old consumers than 35+ consumers. Main effects of age group were not found for liking or familiarity whether included in the reduced or full form.

Just a single emotion category – Excitement - (and liking) showed significant interactions between sample and age group: Excitement (Table 4.1). Accordingly, six of the individual terms associated with Excitement (enthusiastic, excited, fulfilled, fun, interested, want) also showed interaction (see the example of 'fulfilled' in Figure 4.4a). There were many similarities in the nature of these interactions between these individual full form emotion terms and the Excitement reduced form emotion category, which will be described here. All six terms identified that the Isoamyl acetate sample was scored higher by the 18-34s than the 35+ group. Furthermore, 'fulfilled', 'fun', and 'want' were scored higher by the 35+ group for the High alcohol sample.

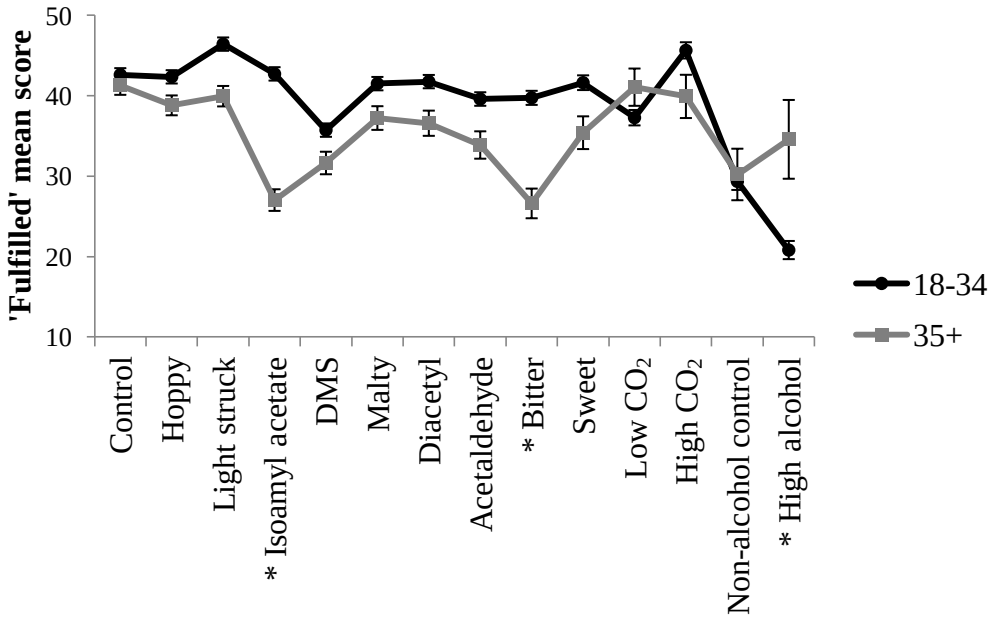


Figure 4.4a Mean ratings of 'fulfilled' (and SEM) for each sample by age group. As per Tukey's HSD post hoc tests, * denotes a significant difference in the rating of 'fulfilled' between age groups.

Like the Excitement emotion category, four of the terms (all excluding 'enthusiastic' and 'interested') showed the Bitter sample was rated higher by the younger age group. 'Fun' and 'want' were rated higher by the 18-34 age

group for the Acetaldehyde sample and this was also reflected by the Excitement emotion category. Just ‘excited’ on the full form showed that there was a difference in ratings of the Hoppy sample between age groups, with the sample rated lower in ‘excited’ by the older consumers. The Excitement emotion category also identified that the Diacetyl sample was rated higher by 18-34s where the full form terms did not. However, in response to the Sweet sample, ‘excited’, ‘fun’, and ‘interested’ received higher scores from 18-34s than the 35+ group and this was not reflected by the Excitement emotion category. Excitement was able to encapsulate all of the interactions revealed by liking when included in the reduced form and this was the same for the full form terms. It was seen that liking showed an interaction between the Bitter sample and age group when included at the end of the reduced form where no such interaction was found when liking was included at the end of the full form.

All of the emotion terms associated with the Disconfirmation emotion category (disappointed, dissatisfied, unpleasantly surprised) showed interactions where Disconfirmation did not (see the example of ‘disappointed’ in Figure 4.4b). In each case, the High alcohol sample was rated higher by the younger consumers. For ‘disappointed’ and ‘dissatisfied’, the Low CO_2 sample was also rated higher by the 18-34s. Also, ‘unpleasantly surprised’ was scored higher in response to the Non-alcohol control and Diacetyl samples by the younger age group. In contrast, the Isoamyl acetate sample was rated higher by the 35+ age group in ‘disappointed’.

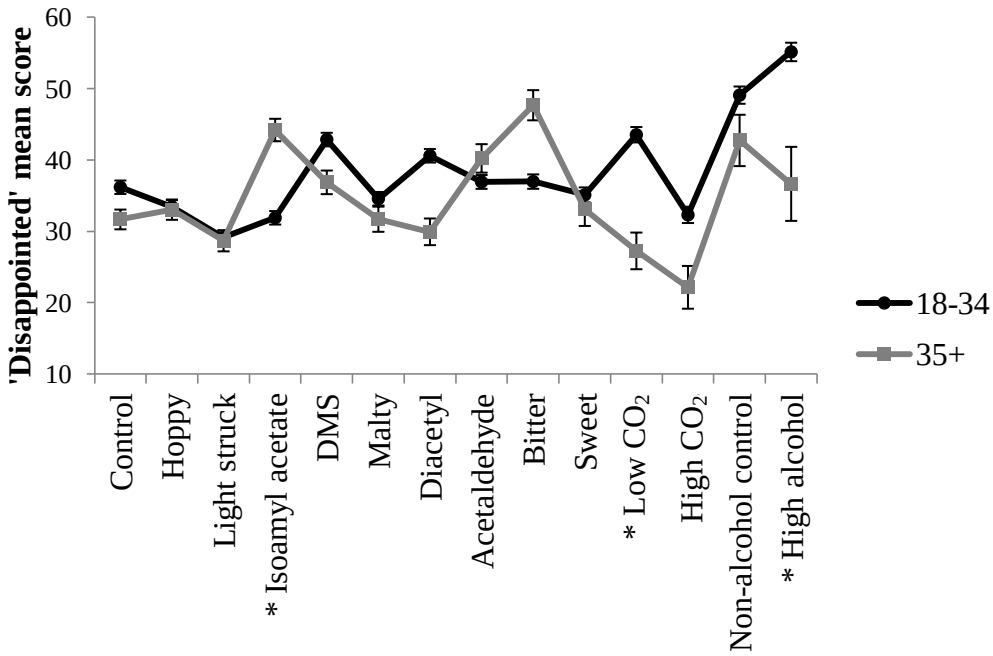


Figure 4.4b Mean ratings of 'disappointed' (and SEM) for each sample by age group. As per Tukey's HSD post hoc tests, * denotes a significant difference in the rating of 'disappointed' between age groups.

A number of the emotion terms associated with Contentment (content, good, happy, nice) revealed interactions between sample and age group (see the example of 'happy' in Figure 4.4c) where the emotion category included in the reduced form did not. In all cases, the Isoamyl acetate sample was rated higher by younger consumers and the High alcohol sample was scored more highly by the older consumers. There were also other examples of 18-34s rating these emotion terms higher: Acetaldehyde in the case of 'content', Sweet, Acetaldehyde, Bitter, and Hoppy for the term 'happy', and Bitter in 'nice'.

Horrible' (belonging to the Disgust emotion category) and 'strange/weird' (belonging to the Shock emotion category) also showed significant interactions between sample and age group where the related emotion categories did not. For both terms, the Non-alcohol control, Low CO_2, and

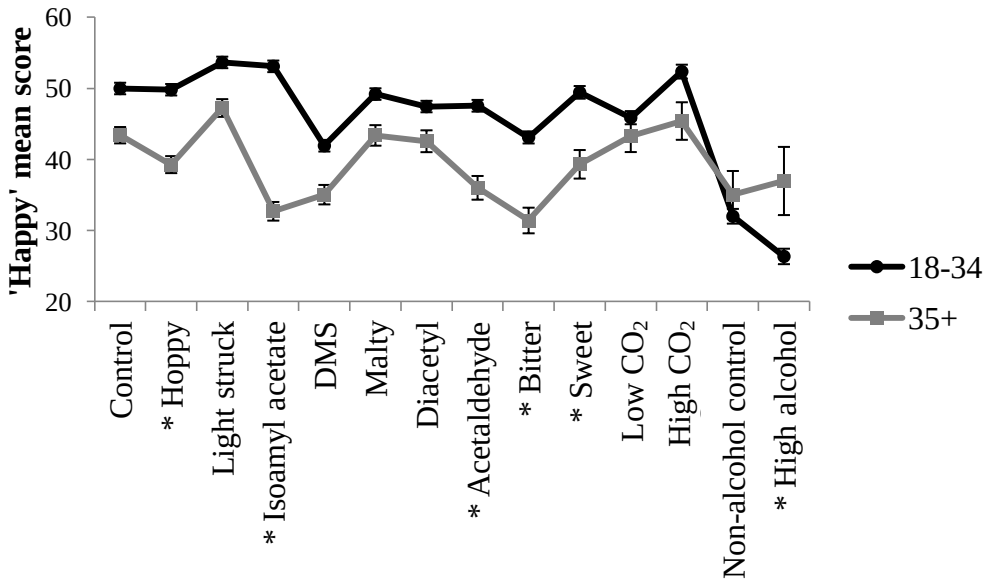


Figure 4.4c Mean ratings of 'happy' (and SEM) for each sample by age group. As per Tukey's HSD post hoc tests, * denotes a significant difference in the rating of 'happy' between age groups.

High alcohol samples received higher ratings from the 18-34 age group than the 35+ group. Younger consumers also scored 'strange/weird' higher for the Diacetyl, Sweet, and Hoppy samples.

4.4 Discussion

Multivariate analysis showed that the full form emotion terms and reduced form's emotion categories were closely aligned, as evidenced by a relatively high RV coefficient, and comparable in their positioning in circumplex emotion space. In particular, there was close positioning between emotion categories and their constituent emotion terms on the pleasure/pleasantness dimension, which accounted for the majority of the variance. Accordingly, there was little difference between samples on this dimension. Full form emotion terms tended to differ in this emotion space from the emotion category to which they belonged in the arousal/engagement/activation

dimension, which accounted for just a small amount of the variance in the data. This seemed to affect, for example, the terms belonging to the Shock emotion category as the most arousing/engaging/activating emotion terms were also the most discriminating. Overall though, the positioning of emotion terms on dimension 2 seemed to have little relation to post hoc groupings. A number of samples were positioned differently on this dimension dependent on the form used and post hoc tests showed this to be related to the low arousal/engagement/activation 'bored' and 'underwhelmed' emotions.

As might be expected from combining individual terms into emotion categories, some of the subtle abilities of individual emotion terms to discriminate between consumers' experiences of the samples were lost. In particular, the terms belonging to the Shock, Disgust, Contentment, Disconfirmation, and Excitement emotion categories were able to discriminate between the Control and samples manipulated from the Control where the reduced form was unable to achieve this. Furthermore, the reduced form was unable to discriminate between the Low CO_2 and High CO_2 samples, whereas one individual term ('dissatisfied' belonging to the emotion category Disconfirmation) did differentiate between differing levels of carbonation. Therefore, this discrimination was lost when consumers completed the reduced form. This demonstrates that the rich detail of information that can be gained from use of a full lexicon is compromised when using a reduced lexicon, meaning that important product information may be lost. However, it is worth noting that there were more comparisons made when using individual emotion terms as compared to emotion categories, meaning that, by chance

alone, some significant differences between samples would be expected to be found by chance alone.

However, this does not tell the full story because, in some cases, the reduced form was able to provide greater discriminability between samples than the individual emotion terms of the full form. Most prominently, the emotion categories Shock and Disgust were able to discriminate between the Non-alcohol control and High alcohol samples where none of the individual terms that belonged to these categories could. Furthermore, the emotion category Tame/Safe was more discriminating as a result of the combination of its two constituent terms.

The emotion category Nostalgia summed up the best and worst of both worlds. Whilst the subtleties that were revealed from individual emotion terms were lost in the employment of a single emotion category, discriminations between samples were shown by the reduced form that were not by the individual constituent terms. This is closely related to a common observation that the reduced and full forms did not always exactly match up in their discriminations. This was particularly clear for the Boredom and Underwhelmed emotion categories which each contained just a single term, allowing a direct comparison of form length on consumers' responses. Whilst neither 'bored' nor 'underwhelmed' were particularly discriminating when included as part of either form, the differences between samples that were found were inconsistent. This was related to the differences between forms in product projections in the emotional space generated by MFA. Pertinently, there were also inconsistencies in the grouping of samples according to both

liking and familiarity according to the length of emotion form that preceded their ratings. This was perhaps no surprise as it has been previously shown that the position of an overall acceptability question has an effect on the response (King et al., 2013). Furthermore, it was seen that males scored samples higher in liking when this was included at the end of the full form where this was not the case for the reduced form. For liking, the full form also did not reveal as many interactions between sample and age group as the reduced form. Taken together, these inconsistencies between form lengths may have implications for the employment of reduced emotion lexicons. At the very least, this suggests that data obtained from full and reduced emotion forms should not be mixed in their practical application.

No halo dumping effects were evident from form reduction for the 'bored' and 'underwhelmed' emotions. This is perhaps because, unlike Cardello et al. (2014) who found such effects, no emotion terms were excluded from the full form when producing the reduced form. By grouping similar terms together, consumers were still able to reflect their emotional response and, therefore, did not need to dump their responses as was the case when individual terms were removed by Cardello et al. (2014) to reduce their form. Nevertheless, it cannot be ruled out that there were halo dumping effects on other emotion categories that could not be directly compared between form lengths here. In addition, it has previously been discussed that 'bored' and 'underwhelmed' proved to be the most difficult emotions for consumers to rate (see section 3.5), meaning that the responses to these emotions were perhaps not representative of those for the other emotions.

When considering the relative abilities of the reduced and full forms to differentiate between the responses of consumer groups, it appeared that there was not much difference in terms of gender. Nevertheless, some differences in overall rating of individual emotion terms included in the full form were found between genders, where the associated emotion categories revealed no such differences. These followed the previously identified pattern, with males giving higher scores than females (see section 3.5). No reduced form emotion category was able to show interactions between sample and just one of the 43 full form emotion terms (strange/weird) showed interactions between sample and gender, meaning that there was not much difference in the reported response to individual sensory properties between genders regardless of form length.

Differences in ratings of individual terms between age groups were also identified by the full form, which were not by the reduced form. The most pertinent example was found for the three terms belonging to the Disconfirmation emotion category, ‘disappointed’, ‘dissatisfied’, and ‘unpleasantly surprised’, which were all scored higher by the younger consumers, whereas no difference was found for the reduced form emotion category. Furthermore, many interactions between sample and age group were found when employing the full form, whereas this was the case for just one reduced form emotion category. This shows that the full form was a more effective approach to differentiate between age groups.

The comparison of the full and reduced emotion lexicons allowed an opportunity to assess the effectiveness of the modified cluster analysis (see

section 3.2.3) for grouping terms that elicited similar patterns of response to one another as well as to the emotion category when grouped together in the reduced form. The results discussed above largely give a positive account of the cluster analysis approach as similar sample discriminations were given by emotion terms that belonged to the same emotion category. However, there are a few points to note. Firstly, MFA showed that 'curious' and to a lesser extent 'interested' appeared less closely grouped in the emotional space than other terms belonging to the category Excitement. The cluster analysis upon which the emotion categories were based grouped 'curious' and 'interested' into a separate group when including one further cluster. Based on just this evidence, it appears that one more cluster should have been included at the lexicon reduction stage to give a total of 10 emotion categories. Yet, post hoc tests showed that 'curious' and 'interested' were not so different in its discriminations to the other 10 terms included in the Excitement emotion category, suggesting that it may well have been correct to stop at 9 clusters and emotion categories.

This example highlights a difficulty for researchers in knowing where to 'draw the line' when using cluster analysis to group similar emotion terms, which could be levelled as a criticism of the lexicon reduction. On the other side of this coin, cluster analysis was chosen for lexicon reduction due to the practical advantage of having experimenter input to clearly define categories (see section 3.2.3). Chrea et al. (2009) made use of exploratory factor analysis in the first stage of creating the GEOS and this initially gave five emotion factors. However, further exploratory factor analysis on the responses of a larger group of consumers to a wider range of odorants gave a different factor

structure. It was with modifications of the two factor structures that the final GEOS six factor structure was generated. Therefore, researcher input into the grouping of emotion terms appears necessary regardless of the initial statistical approach. It is suggested that the employment of cluster analysis is advantageous in that modifications can be made earlier in the process than, for example, the factor analyses used by Chrea et al. (2009).

4.5 Conclusions

The results presented here compared the relative efficacies of the reduced product-specific consumer-led emotion lexicon to the full lexicon upon which it was based. It was demonstrated that the reduction of the full lexicon to a reduced lexicon gave relatively comparable results, reflecting the history of successful emotion form reductions in the literature (Lubin et al. (2001); Shacham (1983); Thompson (2007); Nestrud et al. (2013); Porcherot et al. (2010)). Nevertheless, there were relative merits of using reduced or full product-specific emotion lexicons. Whilst more detailed emotion information is potentially lost through the employment of the reduced form, consumers appeared to be able to use some emotion categories to more effectively discriminate between the samples.

There was not a great deal of difference in the ability of the two form lengths to differentiate between gender responses to the samples, although the full form provided slightly more detail. However, the full form was more effective at revealing differences in emotion response to individual samples between the two age groups included in this study.

Therefore, the use of a reduced emotion form may be preferable for product comparisons given its relative similarity to the full form for sample discriminability and the significant savings in both time and resources. However, if the aim of emotional research is to differentiate between consumer segments, particularly between age groups, then it may be preferable to use a full emotion lexicon. This, of course, only applies to the samples included within this study and may be different for beers with other sensory properties or other product categories.

An important point was that, where direct comparisons between forms were available (namely for 'bored', 'underwhelmed', liking, and familiarity), there were inconsistencies. This may have implications for practitioners in that only reduced or only full forms should be used for direct comparisons. Nevertheless, there were no observed halo dumping effects using the reduced lexicon, demonstrating the potential of the reduced product-specific consumer-led emotion lexicon approach to be more effective than simply removing terms as has been done in the literature (Nestrud et al. (2013); Porcherot et al. (2010)). However, a full comparison between forms was not possible, meaning the possibility of halo dumping cannot be fully excluded. Nevertheless, these results are preliminarily promising.

The research presented in this chapter also afforded an opportunity to evaluate the use of cluster analysis as a lexicon reducing technique. Overall, it was seen that emotion terms belonging to a given emotion category grouped samples very similarly to one another, as well as to the reduced form emotion category to which they belonged. This supports the use of cluster analysis for grouping

terms. The disadvantages of the requirement for researcher input into categorisation were discussed, although cluster analysis was still recommended in comparison to other statistical approaches for grouping emotion terms.

This chapter has shown that the reduced product-specific consumer-led emotion lexicon approach is able to discriminate relatively well between samples when compared to a full form. This opens up the potential for extended uses of the approach as a verbal self-report measure of emotion in sensory and consumer science. One such possibility is its application to cross-cultural verbal comparisons and this will be explored in the next chapter.

Chapter 5: Cross-cultural comparison of emotional response using reduced emotion lexicons

5.1 Introduction

In the modern food and beverage marketplace, it is common for products to compete at a multinational level. As such, there is a growing necessity for consumer research to be conducted in more than one country (Thomson and Crocker, 2013). With the continued and growing interest in emotion research in sensory and consumer science, it is important to understand the variation in emotional response across cultures for global products. This section will firstly consider general cross-cultural similarities and differences in emotional response before outlining and evaluating the proposed application of the reduced product-specific consumer-led emotion lexicon approach for comparing emotional response between two cultures.

5.1.1 Universality of emotions

It is widely acknowledged that there is an underlying universality to the expression of emotion. This was described as long ago as the time of Charles Darwin when he wrote in The Expression of the Emotions in Man and Animals that "...widely different races... express the same state of mind by the same movements" (Darwin, 1872). Almost a century later, Ekman and Friesen (1971) supported this observation in their seminal study in which it was demonstrated that members of an isolated culture were able to match American participants in their ability to accurately and reliably convey the emotion of a character by selecting a facial expression. Despite the general recognition of the universality of emotional expression, it is accepted that there is still a significant role for culture to play.

5.1.2 Cultural differences in emotional expression

Anecdotally, it is generally acknowledged that some cultures are more emotionally expressive than others. Indeed, Darwin commented of Victorian Europe that: "Englishmen rarely cry, except under the pressure of the acutest grief; whereas in some parts of the Continent the men shed tears much more readily and freely" (Darwin, 1872). An important aspect of emotional expression is language and cultural differences are clearly apparent when considering this. For example, van Zyl and Meiselman (2015) asked consumers from four English-speaking cultures (USA, UK, New Zealand, and Australia) and two Spanish-speaking cultures (Spain and Mexico) to describe their emotional responses to their favourite beverages. It was shown that the emotion words used by different cultures – even those sharing a language – were quite different, especially between the Spanish-speaking nations. This presents a significant challenge for the use of verbal self-report as a means for cross-cultural emotional comparisons.

The challenge increases when comparing between cultures with different languages. Where a common direct translation of an emotion word is readily available between languages, research has shown that there is potential variability in meaning between cultures. Hurtado de Mendoza (2007) (in Hurtado de Mendoza (2008)) highlighted such problems associated with one-to-one translations of individual emotion words. The researchers compared the emotional words 'shame' in the US and 'vergüenza' in Spain, which are given as frequent translations of one another. Participants from both countries were required to rate several related emotions on their "degree of typicality" of shame in the US and vergüenza in Spain. Significant differences in ratings of

typicality between the two countries were found for over 85% of emotions. For example, humiliation, guilt and regret were rated as highly typical features of shame by US participants whereas they were rated as much less typical of vergüenza by Spanish participants. In contrast, ridicule, shyness, and reluctance were among the most typical features of vergüenza but were not particularly highly rated in typicality for shame. Therefore, this study showed that equivalent translations of emotion terms between languages do not necessarily express equivalent emotional experiences.

Furthermore, there is the possibility of a lack of direct translations for some emotion terms between languages. This can be demonstrated by the English language's adoption of the German word Schadenfreude which describes the feeling of taking pleasure in another's failure or misfortune (literally translated as harm-joy). Similarly, the Mexican Rarámuri Indians have no word to accurately describe guilt, instead reporting shame in typically guilt-eliciting situations (Breugelmans et al., 2005). That is not to say that Rarámuri Indians are unable to feel guilt (and that English speakers are unable to feel Schadenfreude). On the contrary, it has been shown that Rarámuri Indians are able to differentiate between shame and guilt characteristics in a similar way to cultural populations that use distinct words for these two emotions (Breugelmans and Poortinga, 2006). Nevertheless, these studies show that accurate translations of emotion words are not always available between languages.

Much emotion research conducted in sensory and consumer science thus far has made use of verbal self-report and, as shown above, the use of language

presents the potential for significant challenges in cross-cultural research. Therefore, due consideration should be given when attempting to apply verbal self-report to the cross-cultural comparison of emotional responses.

5.1.3 Reduced lexicons as cross-cultural emotion comparison tools

The use of reduced emotion lexicons for cross-cultural comparison has already enjoyed success. As described in previous chapters, the GEOS was developed in French with Swiss participants to describe emotional experiences associated with odours using a smaller set of summary scales (Chrea et al., 2009). The research group then generated summary scales in the same way in both UK and Singapore (Ferdenzi et al., 2011). Common emotions were identified between the countries (e.g. disgust/irritation and happiness/well-being), showing that some responses are shared across cultures. Interestingly, a number of differences were also shown (e.g. spirituality was unique to Singapore), suggesting that there are differences across cultures in the reporting of emotion (at least to odours) even when using more general groups of emotion. The approach of these researchers was subsequently successfully applied to a number of other cultures (Ferdenzi et al., 2013), demonstrating the efficacy of using groups of emotion terms for cross-cultural comparisons. The researchers then went further in this paper to create UniGEOS, a universal scale comprised of some culturally shared and some culture-specific dimensions for use in cross-cultural research.

However, it has been suggested that language-based questionnaires should be administered in the native language of respondents due to the subtleties of expression in different languages (van Zyl and Meiselman, 2015). This chapter

proposes comparisons of verbal self-report using the approach outlined in Chapter 3 in more than one culture to allow comparison of emotional response. It is suggested that, by using groups of emotion terms for cross-cultural comparison, the difficulty of translating individual terms is avoided. Therefore, even if translation of terms proves difficult, there would be greater ease of comparison of more general emotion categories. In addition, respondents would be able to complete the questionnaire in their native language, capturing the subtleties of emotion associated with each language. In the research described in this chapter, Spanish consumers generated a reduced product-specific consumer-led emotion lexicon in response to a subset of the samples used in the UK. This Spanish reduced lexicon was used by a large number of beer consumers in Spain to rate their emotional responses to the samples. Differences in Spanish responses to the subset of samples revealed by the reduced Spanish lexicon were explored before cross-cultural comparisons between emotional response to the samples in the UK and Spain were made.

5.1.4 Objectives

The objectives of the present study were: 1) to generate a Spanish reduced consumer-led emotion lexicon and applying it to (a) discriminating across a range of beer samples designed to elicit specific sensory properties (b) revealing differences in emotional response across different consumer segments; and 2) to compare and contrast emotional response to the selected sensory properties of beer between UK and Spanish beer consumers by exploring the similarities/differences in (a) how emotion categories were associated with specific sensory properties of beer across the two cultures, and

(b) the extent to which the respective lexicons discriminated across consumer segments.

5.2 Materials and method

Following the method described in Chapter 3, Spanish-speaking colleagues conducted focus groups with Spanish consumers who generated emotion terms to describe their emotional responses to a subset of 10 beer samples that varied in selected sensory properties. These subjects then used this lexicon to rate their emotional responses to the samples. Linguistic checks and cluster analysis were performed on the data to group similar terms into distinct emotion categories. The emotion categories were then used by 113 Spanish consumers to rate their emotional response to the 10 beer samples. The key elements of the generation and application of the Spanish product-specific reduced consumer-led emotion lexicon are described below.

5.2.1 Samples

A subset of 10 of the 14 samples used in the UK was used in Spain due to resource limitations (Table 5.1). The excluded samples were the Acetaldehyde, Diacetyl, Malty, and High CO_2 samples, the latter of which was excluded due to a lack of equipment for carbonation in Spain. The others were excluded after consideration of the other sensory properties and the researchers' hypothesised drivers of differences in emotion response between the cultures studied. The included samples still represented key characteristic properties of beer, off-flavours and/or hypothesised drivers of emotional response as well as a range of modalities. It is important to note that, due to the exclusion of the High CO_2 sample, the Low CO_2 sample could only be compared to the Control. The trained beer sensory panel rated the Low CO_2

sample as significantly lower in bubbliness but not in tingliness (see section 2.5.2). This should be taken into account when considering the results of the following study.

Table 5.1 The ten beer samples included in the Spanish study and their treatments.

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/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 volumes
9	**Non-alcohol control**	Commercial non-alcohol lager
10	**High alcohol**	96% ethanol added to commercial non-alcohol lager (8% ABV)

5.2.2 Lexicon development

Lexicon development followed the method developed in the UK (see section 3.2), with three small focus groups of participants (n = 5-7 per group) generating an emotional lexicon in response to the subset of 10 samples. The participants then used this lexicon to rate the samples. Linguistic checks and cluster analysis were applied to their responses to group similar terms into emotion categories.

5.2.2.1 Subjects

To mirror the approach used in the UK (see section 3.2.1), 17 reasonably articulate Spanish consumers (aged 18-60 years), who consumed beer at least once per month (although all, in fact, consumed beer at least once per week), took part in this study after signing consent forms in line with local ethical procedures. As in the UK, most (70%) of the subjects recruited were female. Participants were divided into three groups of 5-7 subjects and attended a total of three 90min-2h sessions.

5.2.2.2 Procedure

Continuing to follow the UK approach (see section 3.2.2), participants first completed exercises to familiarise them with 'emotion' and were presented with warm-up samples of the Control and Non-alcohol control samples for contextualisation.

A total of 80 terms were generated in the elicitation phase. An initial reduction of terms was performed by combining synonymous terms using a thesaurus (Diccionario de la lengua española). This resulted in a condensed list of 53 terms. The focus group participants used the list of 53 terms to rate their emotional response to all 10 samples during the third and final session. At the end of this session, participants rated the relevance of each of the emotion terms for describing their emotional response to beer in general (for full details of how this was measured, see section 3.2.2).

5.2.2.3 Grouping of terms into emotion categories

As in the UK (see section 3.2.3), 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 (distressed, tired, sickly, infantile, afraid, suspicious, embittered, sad, and empty).

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 like ‘eager’ or ‘desirous’ or with negative connotations such as anxiety). Most consumers associated 'emocionado' with excitement and 'ansioso' with eagerness/desire. These responses were very useful when defining emotion categories.

The mean ratings of the remaining 43 terms (Table 5.2) were submitted to cluster analysis to group terms which produced similar patterns of data across the beer samples.

Eight clusters of terms were initially identified (Table 5.3a). However, it was considered by the Spanish researchers that some of the clusters were unclear and could potentially cause confusion. This observation was supported by a low Cronbach’s α associated with some clusters. To reduce confusion and increase internal consistency, a number of clusters were modified. For example, Cluster 1 (‘mild’, ‘bored’, and ‘indifferent’) had a very low Cronbach’s α of just 0.44. Based on the researchers’ discussions with participants, particularly at the elicitation phase, mild was deemed to be distinct in meaning from ‘bored’ and ‘indifferent’. Therefore, Cluster 1 was split into two categories: Mildness (including just ‘mild’) and Indifference

Table 5.2 Final lexicon of 43 Spanish terms translated into English (with original Spanish terms shown in italics).

Agreeable	*Conforme*	Excited	*Emocionado*	Objectionable	*Indeseable*
Annoyed	*Contrariado*	Festive	*Festivo*	Pleasant	*Placentero*
Appetised	*Apetecible*	Fresh	*Fresco*	Positive	*Positivo*
Authentic	*Auténtico*	Friendly	*Amistoso*	Relaxed	*Relajado/tranquilo*
Bad	*Mal*	Happy/	*Alegre/chispeante/*	Repulsed	*Reacio/rechazo/*
Bored	*Aburrido*	cheerful	*contento*		*repulsión*
Cheated	*Engañado*	Indifferent	*Indiferente*	Satisfied	*Satisfecho*
Classic	*Clásico*	Intense	*Intenso*	Shocked	*Sorprendido/*
Curious	*Curioso*	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*
Enjoyment	*Divertido*	Nostalgic	*Nostálgico*		

Table 5.3a Cluster analysis of the 43 Spanish terms (translated into English) grouped into 8 clusters with associated Cronbach's αs (adequate internal consistency > 0.8 (Streiner, 2003)).

Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Cronbach's α	**0.44**	**0.97**	**0.85**	-	**0.97**	**0.89**	**0.85**	**0.34**
	Bored	Nice	Happy/	Eager	Disgusted	Disillusioned/	Strong/	Excited
	Indifferent	Friendly	cheerful		Annoyed	disenchanted	powerful	Nostalgic
	Mild	Appetised	Lively		Disappointed	Unusual	Intense	
		Authentic	Curious		Unpleasant	Lacking in		
		Classic	Enjoyment		Unmotivated	appetite		
		Agreeable	Festive		Cheated	Dissatisfied		
		Normal			Uncomfortable			
		Fresh			Objectionable			
		Natural			Bad			
		Pleasant			Negative			
		Positive			Repulsed			
		Relaxed			Shocked			
		Satisfied						
		Traditional						

(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.). Finally, Cluster 8 (Cronbach’s $\alpha = 0.34$) was split into two categories: Excitement and Nostalgia because these terms were used to describe different emotions by the participants who had generated them.

The revised groupings resulted in a total of 12 emotion categories (Table 5.3b). 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 $\alpha > 0.8$; 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 affected by the number of items (Streiner, 2003).

5.2.3 Measuring consumer emotional response using the Spanish reduced emotion lexicon

In the second part of the study, Spanish consumers rated their emotional response to the 10 beer samples using the 12 emotion categories defined in the previous section.

5.2.3.1 Subjects

Native Spanish beer consumers were recruited to match the UK consumer sample for both gender and age (see section 3.3.1). One hundred and thirteen subjects who consumed beer at least once per month took part in this study (though 81% consumed beer at least once per week). Approximately half were

Table 5.3b Final grouping of the 43 Spanish terms (translated into English) into 12 clusters with associated Cronbach's αs (adequate internal consistency > 0.8 (Streiner, 2003)).

Emotion category	1 Mildness	2 Indifference	3 Pleasure	4 Classic	5 Fun	6 Desire	7 Disgust	8 Disillusionment	9 Disappointment	10 Intensity	11 Nostalgia	12 Excitement
Cronbach's α	-	**0.55**	**0.97**	**0.93**	**0.85**	-	**0.97**	**0.89**	**0.87**	**0.85**	-	-
	Mild	Bored	Nice	Authentic	Happy/	Eager	Disgusted	Disillusioned/	Disappointed	Strong/	Nostalgic	Excited
		Indifferent	Friendly	Classic	cheerful		Annoyed	disenchanted	Uncomfortable	powerful		
			Appetised	Natural	Lively		Unpleasant	Unusual	Dissatisfied	Intense		
			Agreeable	Normal	Curious		Unmotivated	Lacking in				
			Fresh	Traditional	Enjoyment		Cheated	appetite				
			Pleasant		Festive		Objectionable					
			Positive				Bad					
			Relaxed				Negative					
			Satisfied				Repulsed					
							Shocked					

male (52%) and most were aged 18-34 years (68%) with the remaining aged 35+ years.

5.2.3.1 Procedure

The procedure for the consumer study followed the UK (see section 3.3.2) except the subset of 10 samples was used (as opposed to the full 14 samples used in the UK) and the 12 Spanish emotion categories were used (as opposed to the 9 UK emotion categories).

5.2.4 Data analysis

Firstly, Spanish data was analysed with the aim of assessing the effectiveness of the reduced product-specific consumer-led emotion lexicon to discriminate between the subset of 10 samples as well as its ability to differentiate between the responses of consumer groups. The Spanish emotional response was then compared to the UK response to the 10 samples, with the aim of identifying key cross-cultural similarities and differences.

5.2.4.1 Spanish emotional response

Analysis of Spanish emotional response followed the analyses used in the UK (see section 3.3.3): first, the emotional space was mapped using PCA on the mean ratings of emotion categories across samples. Liking and familiarity were included as supplementary variables (i.e. the data generated from these measures was not used in generating the principal components but was subsequently projected onto them). In order to evaluate the ability of the Spanish reduced product-specific consumer-led emotion lexicon to discriminate across the range of beer samples, mixed model 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 ($p < 0.05$; SPSS Statistics 22, IBM, USA). Further ANOVAs were carried out with fixed effects of sample, gender and age group for each emotion category (and liking and familiarity). This allowed the observation of differences between consumer groups in overall ratings of emotion categories, as well as revealing interactions between sample and gender and between sample and age group. Where significant interactions were found, simple main effects analyses were conducted in order to ascertain which samples were rated significantly differently for a given emotion category (or liking or familiarity) between genders or between age groups (SPSS Statistics 22, IBM, USA).

5.2.4.2 Cross-cultural comparisons

PCA was re-calculated for the UK responses to the subset of 10 samples (as opposed to the full set of 14 samples presented in section 3.4.1) to enable a true comparison between the two countries (XLSTAT, Version 2009.6.03). Individually for each country, cluster analysis was performed on the mean scores of each emotion category across samples. Euclidean distances and Ward's criterion of aggregation were used to group samples (XLSTAT, Version 2009.6.03). Cluster analysis allowed the comparison of how the emotion categories grouped the 10 samples across cultures. These clusters were superimposed onto the associated PCA plots in order that the emotion categories driving these clusters could be identified. The interpretation of these results was supplemented by the calculation of Pearson's r coefficient in

order to ascertain correlations between emotion categories across cultures (XLSTAT, Version 2009.6.03).

5.3 Measuring Spanish consumer emotional response using the reduced emotion lexicon

The following sections demonstrate the ability of the generated 12 Spanish emotion categories to generate a discriminating emotion space between the subset of 10 samples. Furthermore, the ability of reduced lexicon to reveal differences in emotional response between genders and age groups in Spain is also investigated.

5.3.1 Emotion space

PCA enabled the visualisation of the emotional space for the samples tested. The first two principal components accounted for 95.34% of the data variance (Figure 5.1a). PC1 (72.65%) was highly positively correlated with the emotion categories Disgust, Disillusionment and Disappointment and negatively correlated with Pleasure, Classic, Fun, Desire, Nostalgia and Excitement. Liking and familiarity were supplementary variables in the PCA and were highly 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 (see section 1.1.2). That is to say, PC1 was associated with pleasure/pleasantness and PC2 was related to arousal/engagement/activation.

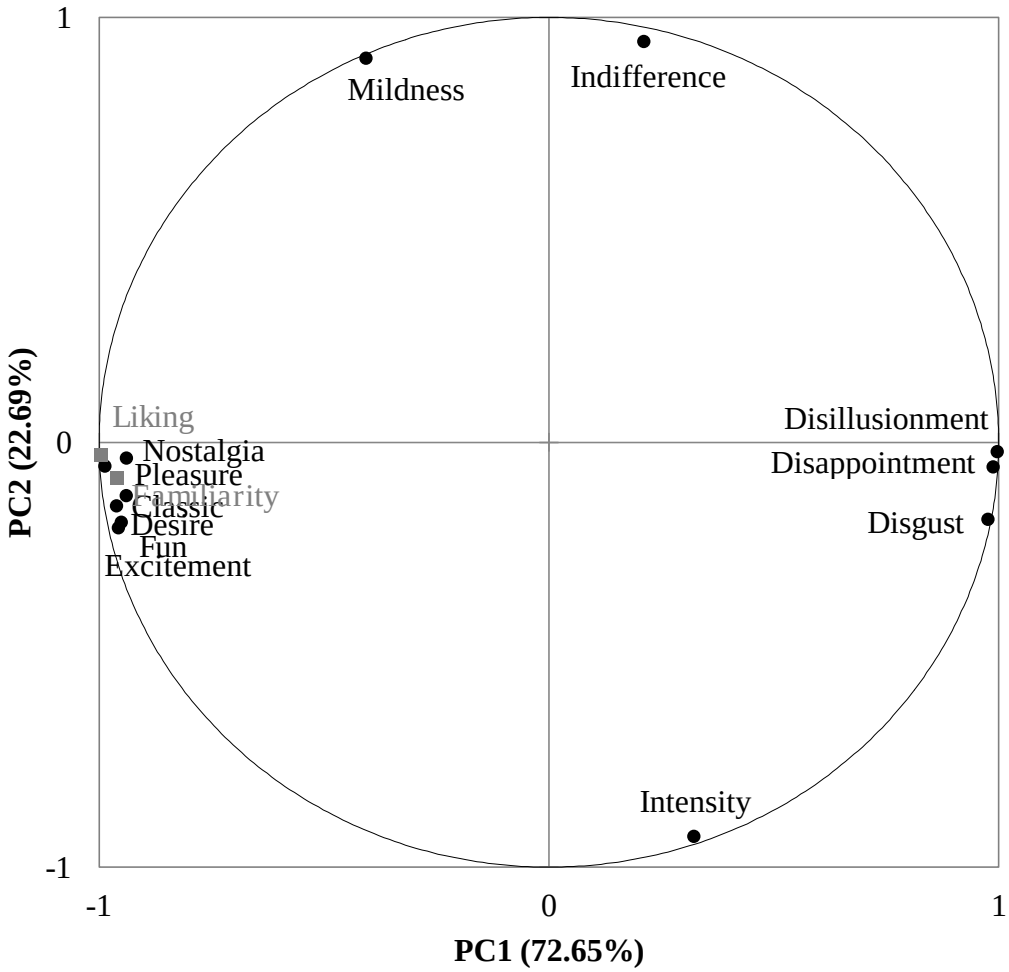


Figure 5.1a Loading of the 12 Spanish emotion categories on the first two principal components (liking and familiarity are included as supplementary variables).

When plotting the sample positions in the Spanish emotional space (Figure 5.1b), it was observed that the Non-alcohol control and High alcohol samples were projected highly positively onto PC1, with the Hoppy sample also loading somewhat on this dimension. Therefore, these samples were associated with unpleasant emotions. In contrast, the Control was highly negatively correlated with PC1, with the Low CO_2 and Light struck samples also projected somewhat in this direction. As such, these samples were associated with pleasant emotions. Both the Control and the High alcohol samples were projected negatively onto PC2, showing an association with

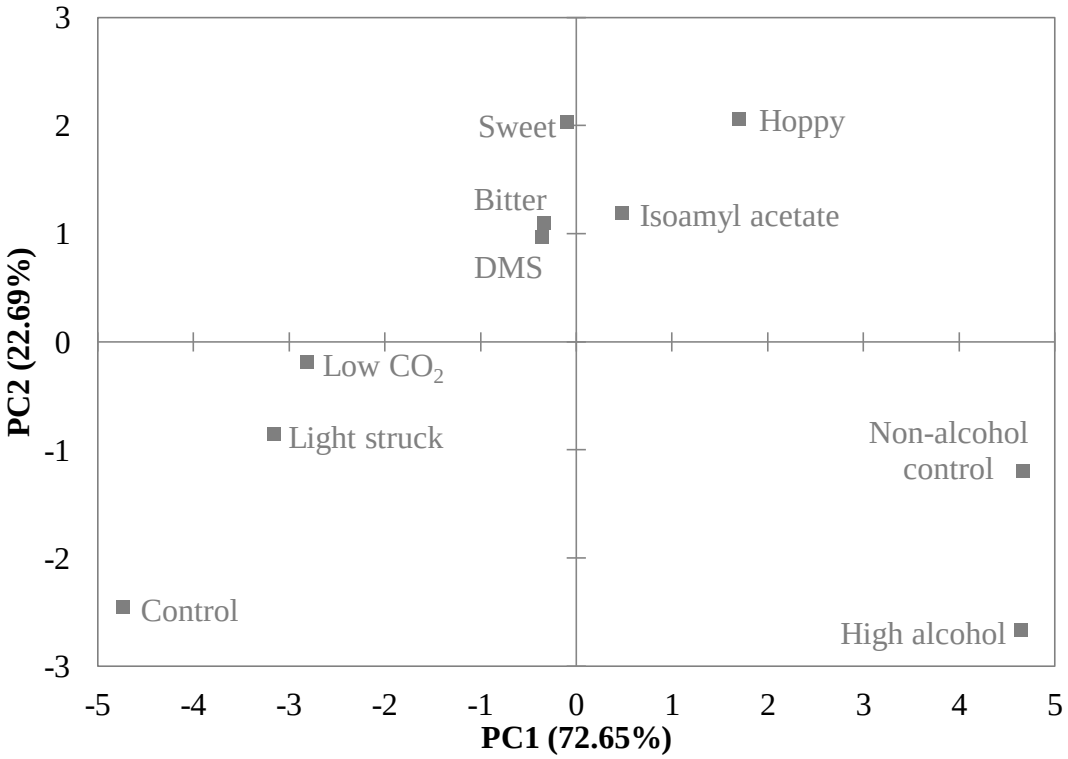


Figure 5.1b Projections of the 10 samples onto the first two principal components (Spanish data).

more arousing/engaging/activating emotions. The Hoppy and Sweet (and to a lesser extent Bitter, DMS, and Isoamyl acetate) samples were highly positively correlated with PC2 and hence less engaging emotions.

5.3.2 Discrimination ability of emotion categories between samples

ANOVA showed that all 12 emotion categories (and liking and familiarity) significantly discriminated between the beer samples (Table 5.4). Consequently, post hoc analyses were carried out for all emotion categories (and liking and familiarity) in order to identify which samples were rated significantly differently from one another (Table 5.5). Comparisons between the post hoc groupings for each emotion category highlighted patterns of sample groupings and, not surprisingly, this was related to how the emotion

Table 5.4 p-values for main effects of sample, gender, and age group, and interactions between sample*gender and sample*age group for each Spanish emotion category (and liking and familiarity).

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

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

Table 5.5 Mean scores for the 12 Spanish emotion categories (and liking and familiarity) across the 10 samples.

	Emotion category													
Sample	**1 Mildness**	**2 Indiff-erence**	**3 Pleasure**	**4 Classic**	**5 Fun**	**6 Desire**	**7 Disgust**	**8 Disillu-sionment**	**9 Disapp-ointment**	**10 Intensity**	**11 Nostalgia**	**12 Exci-tement**	**Liking**	**Famil-iarity**
Control	40.6 BC	30 A	59 E	52.7 E	54.5 D	52.2 D	26.7 A	28 A	27.9 A	54.8 D	38 C	51.9 D	60.5 E	55.4 F
Hoppy	53.4 DE	43.8 C	40.9 ABC	31.7 AB	35.9 AB	33.8 AB	40.6 B	43.4 CDE	43.8 CD	37.5 AB	32.1 ABC	34.6 AB	40.1 AB	29 B
Light struck	45.1 CD	34.9 ABC	52.7 DE	51.6 E	49 CD	43.3 CD	28.7 A	31.7 AB	30.6 AB	48.5 CD	38.2 C	46.1 CD	53.3 CDE	49.4 EF
Isoamyl acetate	50 CDE	44 C	44.9 CD	35.4 B	43.3 BC	38.8 ABC	40.4 B	41.6 BCD	40.3 BC	43.4 BC	30.9 ABC	40.3 BC	44.1 BC	33 BC
DMS	52.4 DE	42.1 BC	44.1 BCD	40.5 BCD	41.4 BC	40 BC	37.3 AB	39.1 BC	36.9 ABC	45.6 BC	33.6 ABC	42 BC	46.4 BCD	38.8 CD
Bitter	52.1 DE	40.5 BC	44.3 CD	45.4 CDE	39.6 AB	40.9 BC	36.7 AB	39.3 BC	39.2 BC	40 ABC	33.2 ABC	39.3 BC	46.4 BCD	43.9 DE
Sweet	57.4 E	41.5 BC	45.8 CD	37.2 BC	42.4 BC	39.5 BC	34.8 AB	40.5 BC	39.4 BC	33 A	30.6 ABC	39.4 BC	45.3 BCD	33.6 BC
Low CO_2	47.9 CDE	37.6 ABC	51.6 DE	49 DE	49.1 CD	47.4 CD	30.3 AB	32.1 AB	35.2 ABC	45.1 BC	36.9 BC	45.2 CD	54.4 DE	49.9 EF
Non-alcohol control	33.2 AB	36.7 ABC	31.6 A	32.5 AB	30.7 A	29.6 A	51.7 C	51.6 DE	53.2 D	57.5 D	28.9 AB	30.1 A	30.7 A	26.2 AB
High alcohol	29.8 A	32.8 AB	34.7 AB	26.3 A	35.8 AB	32.5 AB	55.6 C	53.3 E	52.5 D	66.8 E	26 A	34.7 AB	32.6 A	18.8 A

ABCDEF Letters within the same column indicate post hoc groupings by Tukey's HSD.
Different letters represent significant differences ($p < 0.05$) in rating between samples

categories and samples loaded onto the two dimensions identified by the PCA. The emotional space therefore offered a useful guide for comparing and contrasting the discrimination ability of individual emotion categories.

The three emotion categories identified as unpleasant by their positive loading on PC1 (Disappointment, Disgust, and Disillusionment; Figure 5.1a) were very similar in their sample groupings (Table 5.5). The Non-alcohol control and High alcohol samples were found to be rated significantly higher for these attributes than most other samples, but were not significantly different to one other, reflecting their relative positioning on PC1 of the PCA product plot (Figure 5.1b). The Hoppy sample loaded in the same direction on PC1 as the Non-alcohol control and High alcohol samples although not as highly and hence was not regarded as so unpleasant. Nevertheless, significant differences were found between the Hoppy sample and the samples identified as most pleasant by the PCA (Figure 5.1b), with the Control and Light struck samples rated significantly lower in unpleasant emotions than the Hoppy sample. In addition, the Hoppy sample was rated significantly higher in Disillusionment than the Low CO_2 sample (negatively correlated with PC1). For all three unpleasant emotion categories, the Isoamyl acetate sample (relatively neutral on PC1) was rated significantly higher than the Control (highly negatively correlated with PC1), whilst just Disappointment and Disillusionment revealed that the Control sample received significantly lower ratings than the Bitter and Sweet samples (both neutral on PC1). Disillusionment also revealed that DMS (relatively neutral on PC1) 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 significantly differ to the Control for all three unpleasant emotion categories.

A number of emotion categories loaded highly negatively on PC1 (Pleasure, Fun, Desire, Excitement, Classic, and Nostalgia; Figure 5.1a) and these pleasant emotion categories grouped samples similarly when considering post hoc tests (Table 5.5), as did liking and familiarity which were also negatively correlated with PC1 as supplementary variables. As with the unpleasant emotion categories, none of the pleasant emotion categories differentiated between Non-alcohol control and High alcohol samples and nor did liking or familiarity. The Control was rated significantly higher for Pleasure, Fun, Desire, Classic and Excitement (and liking and familiarity) than all other samples except Light struck and Low CO_2 (and also Bitter for Classic). This was evident in the PCA (Figure 5.1b) where only Light struck and Low CO_2 samples were located in the same quadrant as the Control. Other smaller differences between samples were also shown by Pleasure, Fun, Desire, Classic, and Excitement. For example, there were no significant differences in Fun between the High alcohol and Bitter samples, whereas there were in Pleasure, Desire, Classic, and Excitement (and liking and familiarity). Desire did not discriminate between High alcohol and Isoamyl acetate samples, whereas the other three emotion categories (and liking and familiarity) did. Nostalgia was markedly less discriminating than the other five pleasant emotion categories, with no significant effect of manipulating the Control sample's sensory properties. However, consumers did discriminate between samples in Nostalgia 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). Liking was unable to discriminate beyond the pleasant emotion categories, although familiarity did show significant differences that were not found by the pleasant emotion categories (e.g. the Bitter and Sweet samples were significantly different in familiarity). Thus, whilst familiarity was associated with the emotional pleasantness of the sample, a difference in familiarity did not necessarily mean there would be a difference in emotional response in the context of the Spanish reduced product-specific consumer-led emotion lexicon.

The low activation/engagement/arousal emotion categories associated with PC2 (Figure 5.1a), Mildness and Indifference, were rated significantly higher compared to the Control in Sweet, DMS, Hoppy, and Bitter samples (and also Isoamyl acetate in Indifference; Table 5.5). These samples can be seen to load positively (low activation/engagement/arousal) on PC2 and oppose the highly negatively loading (high activation/engagement/arousal) Control sample (Figure 5.1b). Intensity showed an inverse correlation with Mildness and Indifference as would be expected because it loads in the opposite direction on PC2 (high activation/engagement/arousal; Figure 5.1a). Therefore, Sweet, DMS, Hoppy, Bitter, and Isoamyl acetate samples were rated significantly lower than Control for Intensity. Discriminating more than the low activation/engagement/arousal emotions, Intensity also showed that the Low CO_2 received significantly lower ratings than the Control. In addition, Intensity was able to discriminate between the Non-alcohol control and High alcohol samples, with an increase in the sensory properties associated with alcohol content leading to a significant increase in ratings of Intensity. This is

reflected on PC2 (Figure 5.1b) where the High alcohol sample was projected more negatively than the Non-alcohol control.

Drawing together results from across emotion categories, the Spanish reduced product-specific consumer-led emotion lexicon revealed an individual emotional profile for almost all samples. The Control sample (Figure 5.2a) 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 5.2a). Only the Light struck sample was shown to have no significant emotional effects as compared to the Control (Figure 5.2a) although the samples were sensorially different. Many more differences were apparent between individual modified samples (see the example of the comparison between the Light struck and Hoppy samples in Figure 5.2a). 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 not rated significantly differently for 11 emotion categories but were significantly different in Intensity (Figure 5.2b).

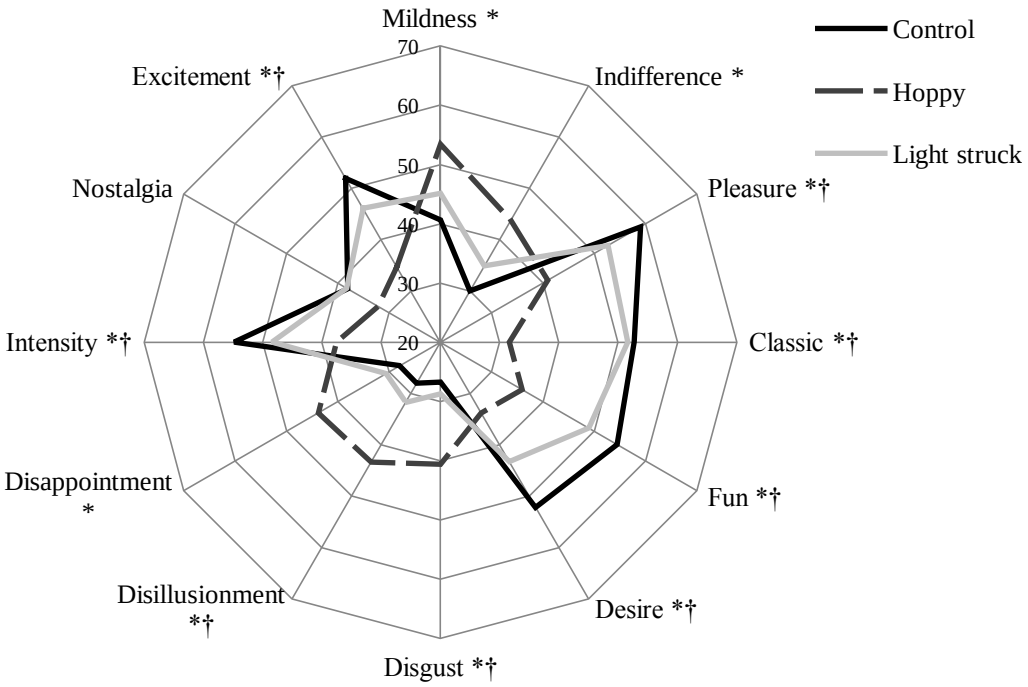


Figure 5.2a Spider plot showing mean scores of all 12 Spanish emotion categories for Control, Hoppy, and Light struck samples. As per post hoc Tukey's HSD tests ($p < 0.05$), * denotes a significant difference between the Control and Hoppy samples, and † denotes a significant difference between the Hoppy and Light struck samples (there were no significant differences between the Control and Light struck samples).

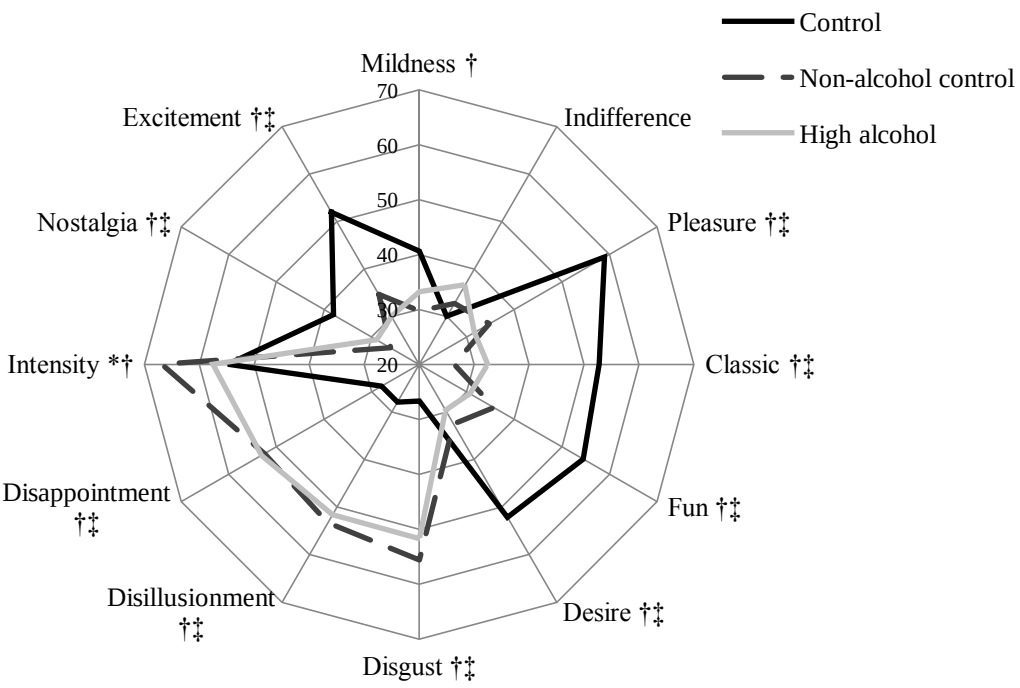


Figure 5.2b Spider plot showing mean scores of all 12 Spanish emotion categories for Non-alcohol control and High alcohol samples. As per post hoc Tukey's HSD tests ($p < 0.05$), * denotes a significant difference between Non-alcohol control and High alcohol samples, † denotes a significant difference ratings between High alcohol and Control samples, and ‡ denotes a significant difference between Non-alcohol control and Control samples.

5.3.3 Consumer group effects

The next section further explores the effectiveness of the reduced consumer-led emotion lexicon by showing how it was able to reveal differences in emotional response across Spanish consumer segments, namely between genders and between age groups.

5.3.3.1 Gender

A significant main effect of gender ($p < 0.05$) was found for the emotion categories Classic, Desire, Disappointment, Indifference, Intensity, Mildness, and Nostalgia, as well as familiarity (Table 5.4). In all except Intensity, Spanish women gave significantly lower ratings than Spanish men. Despite the numerous

differences in overall ratings between genders of emotion categories, such differences were not apparent in liking.

Furthermore, there were significant interactions ($p < 0.05$) between sample and gender for the emotion categories Disgust (Figure 5.3a) and Pleasure (Figure 5.3b). The key gender differences were driven by the responses to the Non-alcohol control and High alcohol samples. Although generally similar sample ratings were obtained for Pleasure, further analyses showed that the Non-alcohol control and High alcohol samples were rated significantly lower in Pleasure by women than men. Disgust was also scored similarly between genders for most samples, except for the High alcohol sample which females rated significantly higher in this emotion than males.

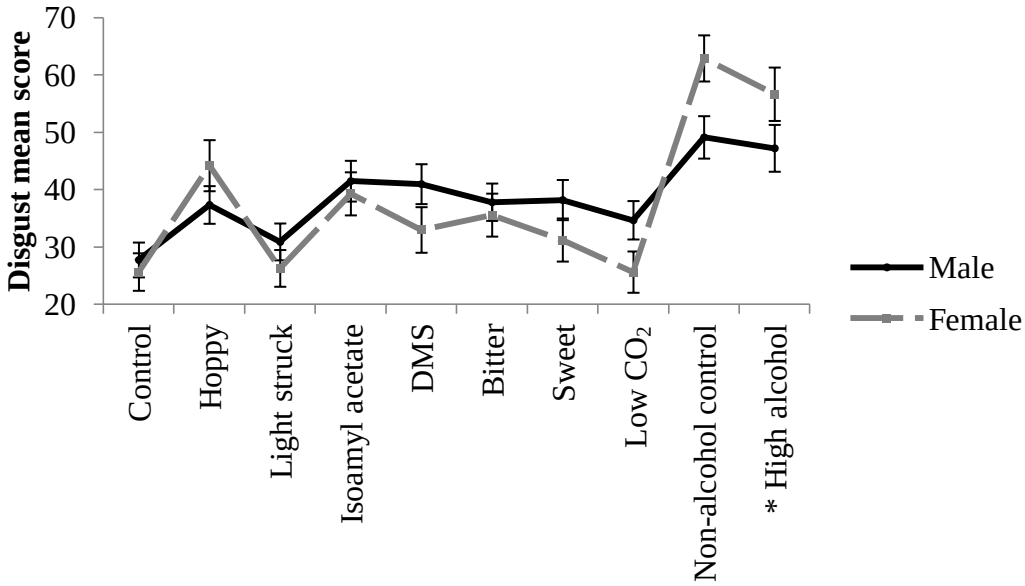


Figure 5.3a Mean ratings (and SEM) of Disgust for each sample by gender. As per post hoc Tukey's HSD tests, * denotes a significant difference ($p < 0.05$) in the rating of Disgust between males and females (Spanish data).

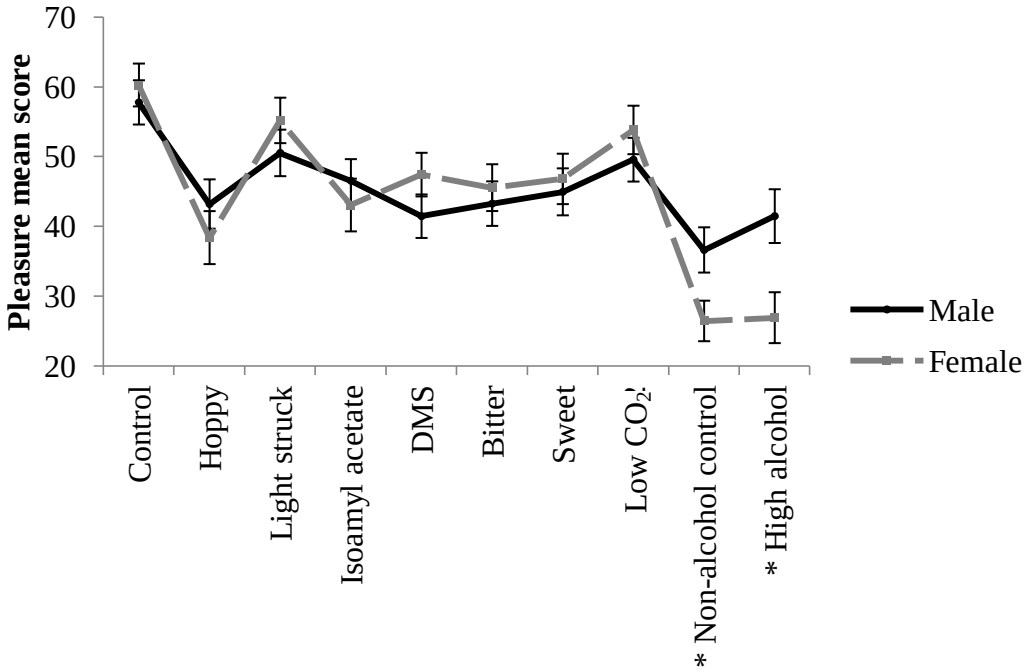


Figure 5.3b Mean ratings (and SEM) of Pleasure for each sample by gender. As per post hoc Tukey's HSD tests, * denotes a significant difference ($p < 0.05$) in the rating of Pleasure between males and females (Spanish data).

5.3.3.2 Age group

Significant main effects ($p < 0.05$) of age group were found for the emotion categories Desire, Disappointment, Disgust, Disillusionment, Excitement, Fun, Indifference, Intensity, Mildness, and Pleasure, as well as liking and familiarity (Table 5.4). On the whole, the 35+ age group scored the 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 amongst the younger age group.

Significant interactions ($p < 0.05$) were found between sample and age group for the emotion categories Classic (Figure 5.4a), Disappointment (Figure 5.4b) and Nostalgia (Figure 5.4c), as well as liking and familiarity. Further analyses showed that the Non-alcohol control and High alcohol samples received significantly higher ratings of Classic from the 35+ than the 18-34 age group. In contrast, the Bitter sample was rated significantly higher in Classic by 18-34 year old consumers. The 35+ age group assigned higher ratings of Disappointment to the Control, Bitter, Sweet, and Low CO_2 samples. These same four samples were also rated significantly lower by the older age group in liking and familiarity. For Nostalgia, the High alcohol sample which was rated significantly higher by the 35+ year old consumers and the Light struck sample which was rated significantly more nostalgic by the 18-34 year old consumers.

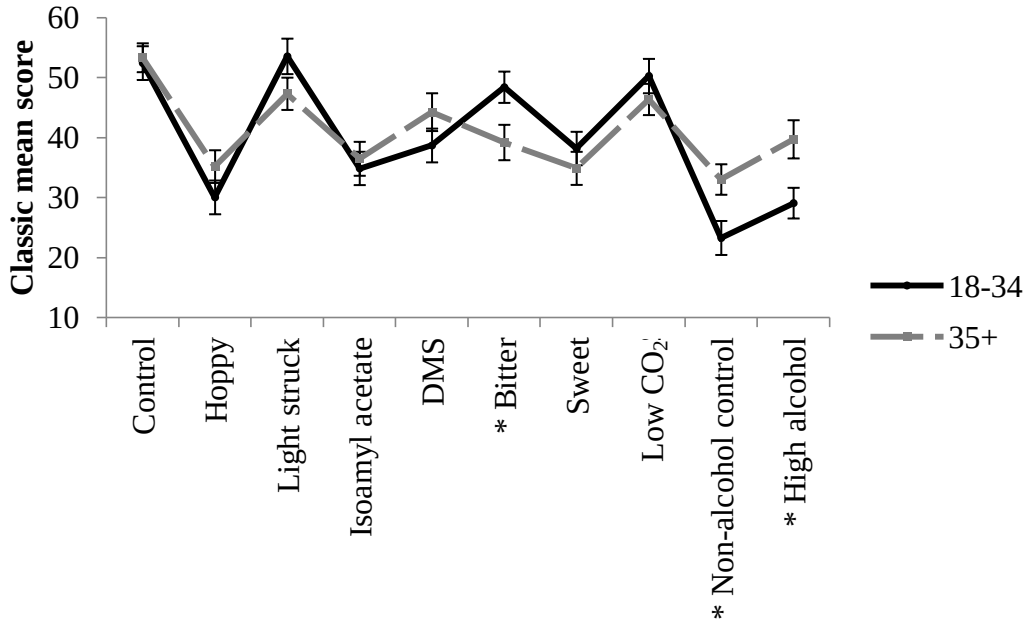


Figure 5.4a Mean ratings (and SEM) of Classic for each sample by age group. As per post hoc Tukey's HSD tests, * denotes a significant difference ($p < 0.05$) in the rating of Classic between age groups (Spanish data).

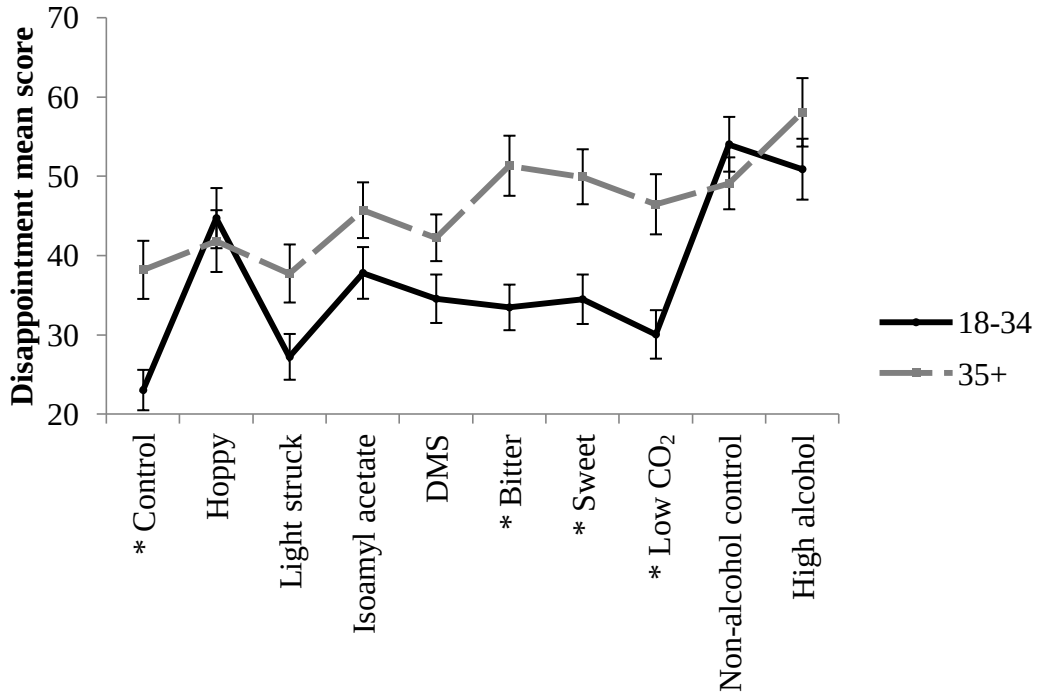


Figure 5.4b Mean ratings (and SEM) of Disappointment for each sample by age group. As per post hoc Tukey's HSD tests, * denotes a significant difference ($p < 0.05$) in the rating of Disappointment between age groups (Spanish data).

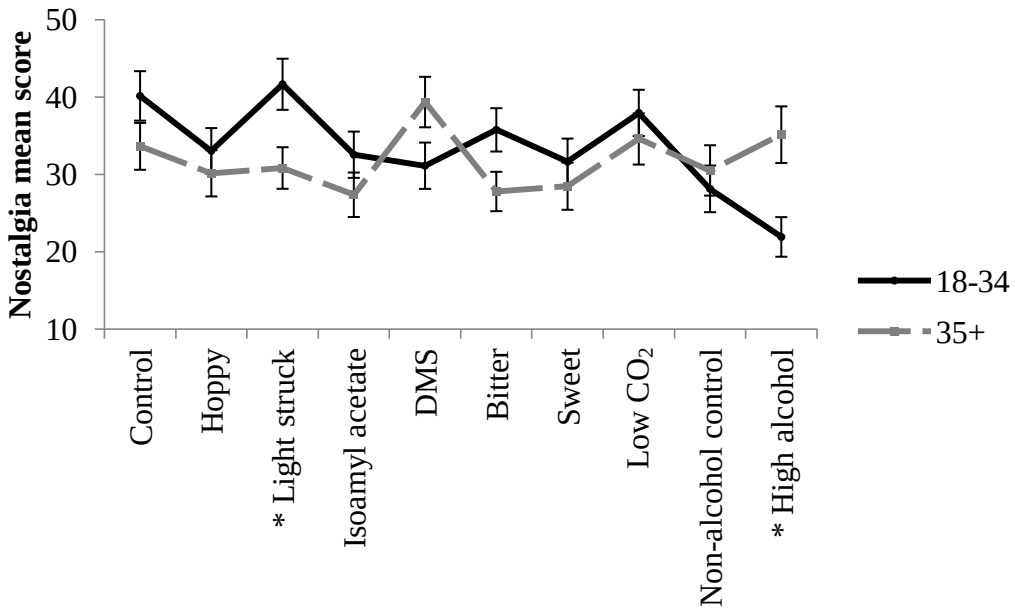


Figure 5.4c Mean ratings (and SEM) of Nostalgia for each sample by age group. As per post hoc Tukey's HSD tests, * denotes a significant difference ($p < 0.05$) in the rating of Nostalgia between age groups (Spanish data).

5.4 Discussion of the results obtained using the Spanish reduced product-specific consumer-led emotion lexicon

The use of focus groups to generate terms and cluster analysis to group similar terms into emotion categories was found to be successful with Spanish consumers. More modifications were required from the initial cluster analysis as compared to the reduced lexicon developed in the UK but the experimenter input was still minimal. Nevertheless, it is important to note that the low number of participants whose data was used to generate the cluster analysis offers at least a partial explanation as to why fewer clusters were generated than final emotion categories. It is suggested that, if less experimenter input is wanted, then certainly more participants should be included in the generation and rating of the initial lexicon.

The 2-dimensional structure of emotional space generated by the response to the 10 samples using the 12 emotion categories was consistent with the space generated by the UK reduced product-specific consumer-led emotion lexicon (section 3.4.1). This was described in terms of circumplex models of emotion (Russell (1980); Watson and Tellegen (1985); Larsen and Diener (1992); see section 1.1.2). Such models are also in line with previous sensory findings using 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 relationships between these emotions. Of the three, Disillusionment was the most discriminating. 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 Spanish women. This could be related to the finding that women are more sensitive to the alcohol burn associated with ethanol (Duffy et al., 2004). Women have also been reliably shown to score higher than men on disgust sensitivity scales (Haidt et al. (1994); Davey (1994); Curtis et al. (2004); Olatunji et al. (2005); Tybur et al. (2009)) so, although 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, and 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 and familiarity because the same four samples were found to be rated significantly lower in liking and familiarity 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. Nevertheless, the pleasant emotion categories together discriminated between samples better than liking. In addition, consumer group comparisons highlighted much larger differences between emotion categories. Nostalgia, although 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. Nostalgia has been described as referring to a preference for objects that were more common when one was younger (Holbrook and Schindler, 1991), and this may well be the case for higher alcohol beers for the older age group and light struck beers for the younger age group. For Classic, the Bitter sample

was rated higher by 18-34s, whereas the Non-alcohol control and High alcohol samples received lower ratings from the 18-34 year olds. Both Nostalgia and Classic were rated lower overall by females than males. For both emotion categories, a likely explanation seems to be the lesser familiarity with the sample set indicated by women.

A particularly noteworthy finding was that the difference in rating of the Light struck sample between age groups was the only case of a difference between the Light struck and Control samples throughout this Spanish study for all emotion categories, liking, and familiarity. This is surprising given the fact that light struck aroma is considered undesirable by the brewing industry (Stephenson and Bamforth, 2002). It seems then that the Spanish consumers were as familiar with the light struck sample as with the unmodified commercial lager and liked the two similarly with comparable emotions elicited by both. This could have implications in the brewing industry for future investment into light struck prevention.

A number of samples (Bitter, Sweet, Hoppy, Isoamyl acetate, and DMS) showed similar patterns to one another in that they were scored lower for pleasant emotions and higher for unpleasant emotion categories as compared to the Control sample. 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 could be viewed as detrimental and, consequently, would have affected consumer emotional response. 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, acetaldehyde) although at low concentrations can also be characteristic of some beers. For the emotion categories associated with pleasantness, just Classic was able to demonstrate any significant difference in rating between these samples (the Bitter was rated significantly higher than both the Isoamyl acetate and Hoppy samples), which is unsurprising given the lack of familiarity relative to the other samples indicated by Spanish consumers.

Intensity (high activation/engagement/arousal), Mildness, and Indifference (low activation/engagement/arousal) loaded very highly in opposite directions on PC2 and grouped samples comparably. 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 associated with pleasure/pleasantness based on the findings of Chaya et al. (2015). These authors used EsSense Profile to measure emotional response to beer and reported that increased carbonation or body associated with increased alcohol content elicited more pleasant emotions. However, it must be noted that the range of carbonation used was not very

large in the present study due to the use of a subset of the original samples. In fact, only bubbliness and not tingliness was found to significantly differ between the Control and Low CO_2 samples (see section 2.5.2). Also, the commercial samples used by Chaya et al. (2015) varied naturally in their alcohol contents, whereas the high alcohol sample in this study was a non-alcohol commercial lager with added ethanol. Due to the complexity of the brewing process (see section 1.5), there are many changes associated with an increase in alcohol content during fermentation, meaning that, by merely increasing the alcohol content of the sample, the resultant sample matrix is not representative of naturally brewed beers. Sweetness was perhaps more representative of naturally brewed beers because dextrose was chosen to increase the sweetness of the base beer as it was found in a preliminary study to offer the sweetness most typical of beer in this matrix (see section 2.2). Accordingly, sweetness was associated with less engaging emotions in the present study, agreeing with the findings of the previous authors.

As in the UK (see section 3.5), it must be acknowleged that the Non-alcohol control and High alcohol samples were very atypical in their associated emotional responses as compared to the other samples, meaning that they were very often scored very differently to the rest of the samples. This is likely to have had a convergence effect on the scores for other samples, meaning fewer significant differences between samples would have been observed than if the Non-alcohol control and High alcohol samples had been excluded. Nevertheless, differences were still picked up between the samples based on the Control by the use of the reduced form in Spain, underlining the effectiveness of this approach.Despite the discriminability between samples of

the three emotion categories associated with arousal/engagement/activation, there were no interaction effects between samples and gender or age group for any. However, it was found that 35+ year old consumers generally gave higher ratings for Mildness and Indifference than those aged 18-34, whereas Intensity was rated higher by the younger consumer group, showing a difference in the scoring of the arousal/engagement/activation emotion dimension between age groups.

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 et al. (2000); Timmers et al. (2003)) and exceed males in reported emotionality and emotion expressivity (Allen and Haccoun (1976); Gross and John (1995)). Therefore, higher ratings of emotion might be expected. There could be an effect of familiarity in that women indicated that they were generally less familiar with the samples than men. If this is the case, it is interesting that familiarity had an effect on some reports of emotion intensity but not on liking, although previous research did not find a relationship between familiarity with beer and affect (Sester et al., 2013).

An alternative explanation could be gender roles, which have been discussed as playing an important role in emotions (King et al. (2010); Fischer (1993); Grossman and Wood (1993); Kring and Gordon (1998)). As beer is viewed as a relatively masculine beverage (Landrine et al., 1988), Spanish males could be more emotionally involved with beer in playing out their gender. The effect

of gender roles in response to this masculine product could also give an explanation for the higher ratings of Intensity generally given by women as this emotion category included the terms ‘intense’, ‘strong’, and ‘powerful’ which could be associated with masculinity.

It is likely that there is a complex relationship between gender and ratings of emotions; this may be mediated by familiarity or gender roles. This is evidenced by the contradictory findings of Chaya (personal communication) who also studied the influence of gender on gender response to commercial beers. Using EsSense Profile in response to a set of commercial lagers, it was seen that women gave significantly higher scores to positive emotion terms and significantly lower scores to negative or unclear terms in response to beer. Suggestions for future research into the relationships between gender and emotion to food and beverage products have already been discussed in relation to the UK data (see section 3.5).

The literature also reveals a trend for adults to experience more positive affect and less negative affect with age (Mroczek and Kolarz, 1998), although this appears not to be the case in response to this sample set, with lower ratings assigned to pleasant emotion categories by the 35+ age group than the 18-34 age group. This also seemingly cannot be related to reduced emotional expressivity in older consumers (Gross et al., 1997), as higher ratings were given for unpleasant emotion categories by the older group. There could be an effect of liking and/or familiarity as both were rated significantly lower by 35+ consumers than 18-34 year old consumers. Thus, less positive emotions and more negative emotions were elicited in the older consumer group to this

set of relatively unfamiliar and disliked samples as compared to the younger consumer group.

These results have shown that, on the whole, the 12 emotion categories were able to discriminate across beer samples with varying sensory properties. This 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. Also, only Intensity was able to discriminate the Low CO_2 and Control samples and between the Non-alcohol control and High alcohol samples. No emotion category was able to discriminate between Light struck and Control samples. However, the consideration of consumer segments 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 a different behaviour between age groups for the Light struck sample but not for the Control. In fact, this was the only difference observed between the Light struck and Control samples for the whole Spanish study. Overall, this study has been able to show the efficacy of the reduced consumer-led lexicon through the demonstration of its ability to discriminate between beer samples with varying sensory properties, although the full value of the inclusion of 12 emotion categories was only fully evident when differences between consumer segments were considered. However, the approach was of limited efficacy until consumer segments were considered, at which point the full value of the inclusion of 12 emotion categories was shown.

5.5 Results of cross-cultural comparisons

The emotional responses to the subset of 10 samples was compared and contrasted between the UK and Spain by, firstly, exploring similarities/differences in the grouping of samples. Secondly, the relative abilities of each culture's reduced lexicon to reveal differences between the responses of different consumer segments were explored.

5.5.1 Comparing the grouping of samples using emotion categories across cultures

PCA was re-calculated using the UK data from just the subset of 10 samples used in Spain (Figure 5.5). This yielded a very comparable emotional space to the one generated when the data from all 14 samples were included (Figure 3.1a). In both instances, the majority of variance accounted for by the first two principal components (94.72% for the 10 samples). PC1 accounted for most of this variance (79.41% for the 10 samples) and this dimension was associated with high or low pleasantness/pleasure emotions, whilst PC2 (15.31% for the 10 samples) was related to high or low arousing/engaging/activating emotions.

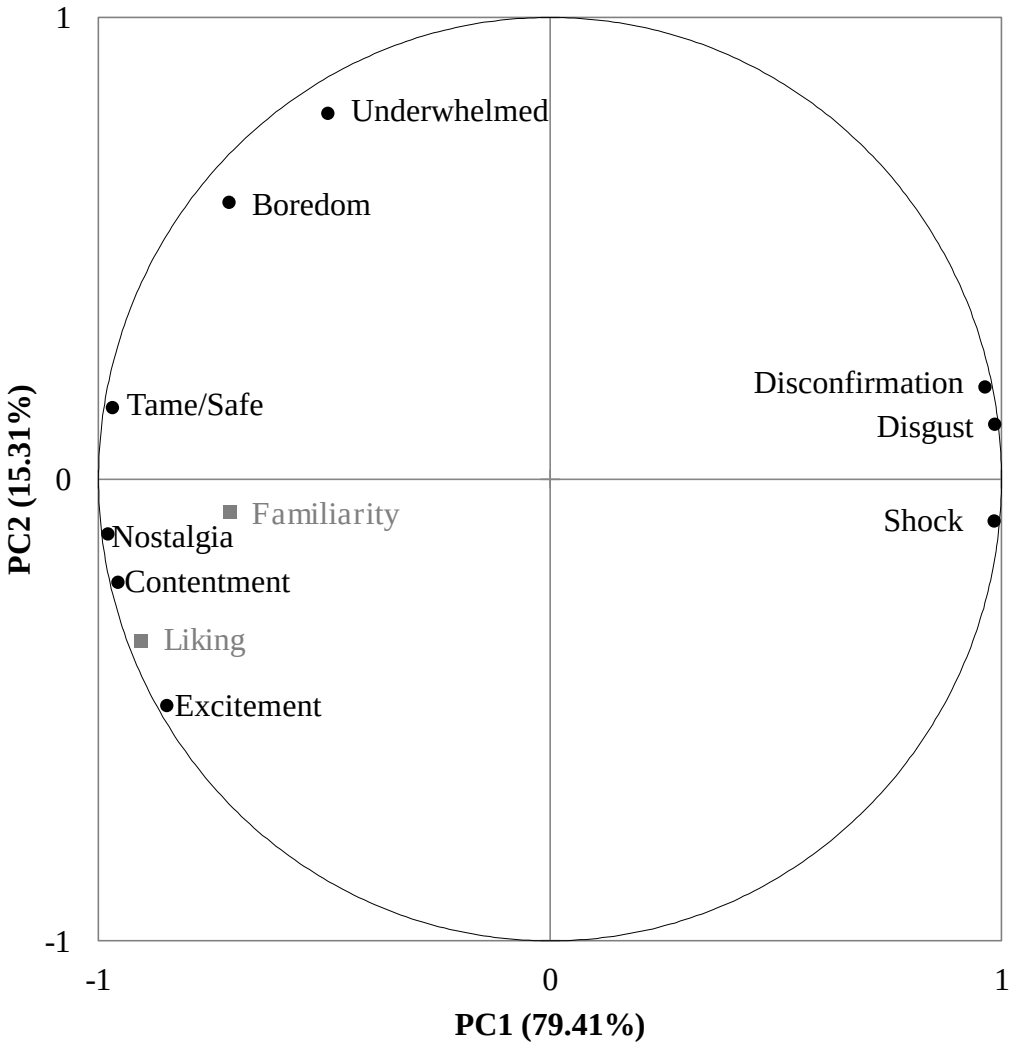


Figure 5.5 Loading of the 9 UK emotion categories on the first two principal components using data from just the subset of 10 samples (liking and familiarity are included as supplementary variables).

5.5.2 The association between emotion categories and sensory properties of beer across cultures

The cluster analyses of beer samples from each country (Figure 5.6) identified three major clusters. The first included the Non-alcohol control and High alcohol samples which were based on the non-alcohol commercial lager, whereas all other samples were based on the normal strength commercial lager (see Table 3.1). The next major cluster in both countries included the Control, Light struck, and Low CO_2 samples, with the Bitter sample also included

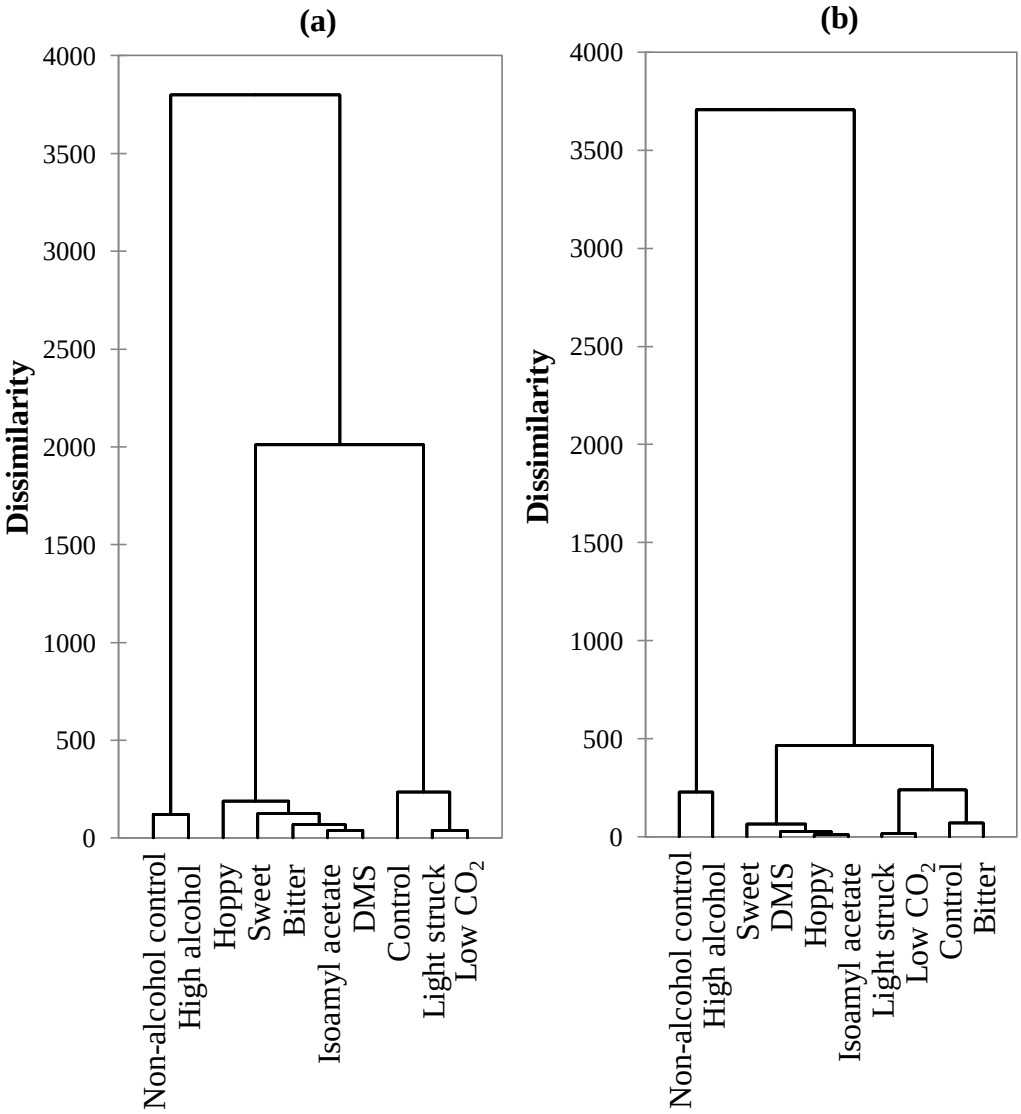


Figure 5.6 Dendograms of sample groupings generated by cluster analysis for (a) Spain and (b) the UK.

within this cluster in the UK. The remaining samples (Hoppy, Sweet, Isoamyl acetate, DMS) were then included in the third cluster (with Bitter also included in Spain). By superimposing these clusters onto the sample PCAs of each country (Figure 5.7a and Figure 5.7b), the emotion category drivers of these clusters were identified. In both countries, the Non-alcohol control and

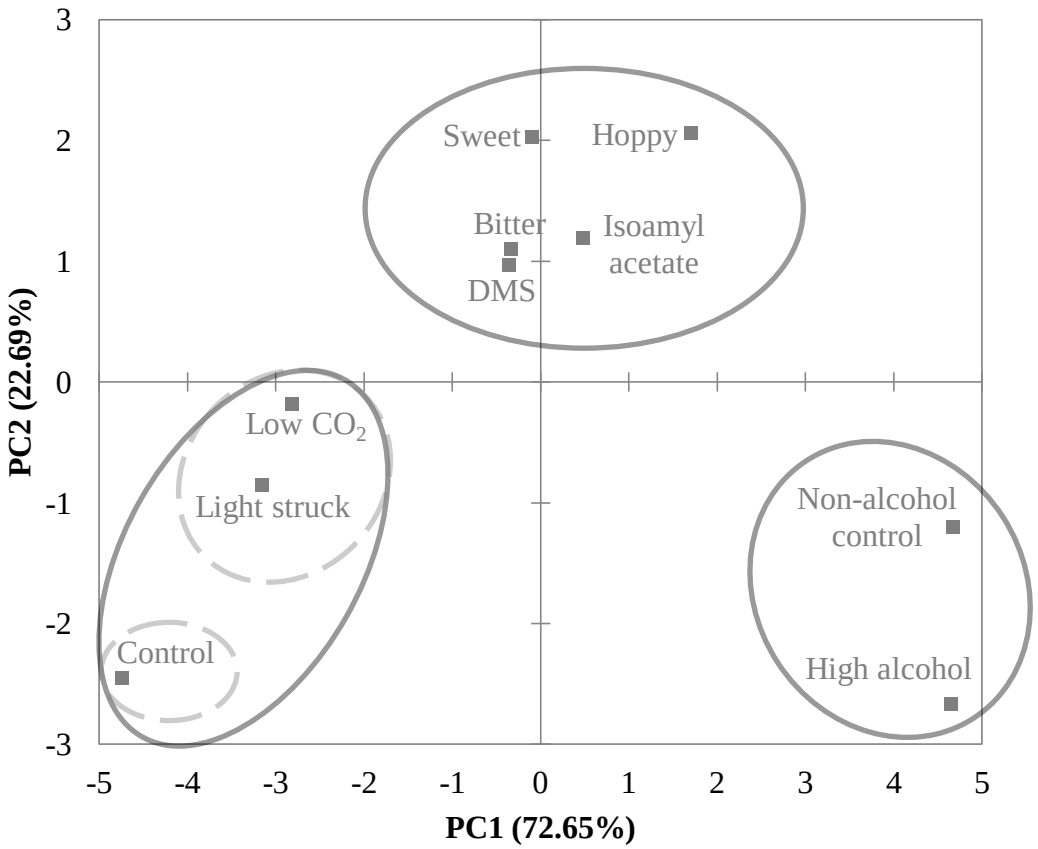


Figure 5.7a Spanish PCA sample plot with superimposed circled clusters. Solid lines indicate the first three clusters and dashed lines indicate the next division.

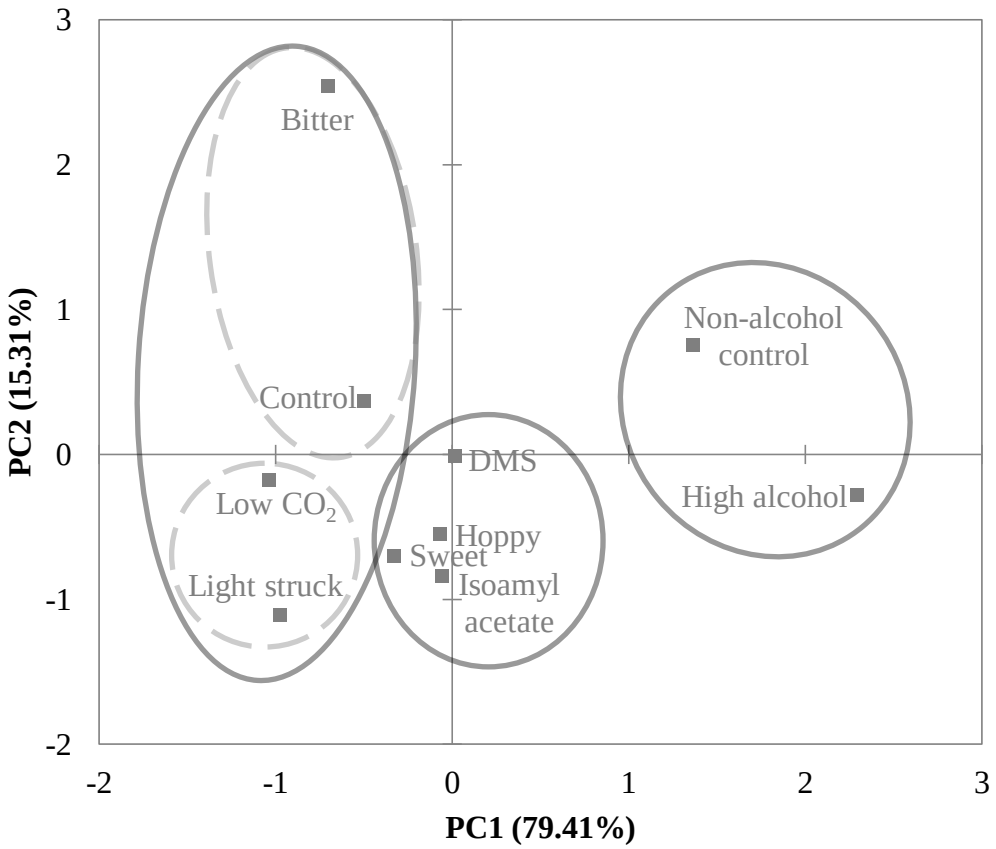


Figure 5.7b UK PCA sample plot with superimposed clusters. Solid lines indicate the first three clusters and dashed lines indicate the next division.

High alcohol cluster was separated from other samples on PC1, with the samples in this cluster being more positively correlated (unpleasant) with this dimension than others. Accordingly, higher scores were assigned to unpleasant emotions and lower scores assigned to pleasant emotions in both Spain (Table 5.5) and the UK (Table 3.4).

In contrast, the cluster including the Control, Light struck, and Low CO_2 samples (and also Bitter in the UK) was negatively correlated with PC1 (pleasantness; Figure 5.7a and Figure 5.7b) and generally received high ratings for pleasant emotions and low ratings for unpleasant emotion categories. When referring to the post hoc groupings of samples in Spain (see Table 5.5)

and the UK (see Table 3.4), it was seen that the Bitter sample was rated differently relative to the other samples in pleasure/pleasantness emotion categories across the two countries. In Spain, there were many significant differences in emotion categories related to pleasantness/pleasure between the Bitter samples and the Control, Low CO_2, and Light struck samples (to take just one example, the Bitter sample was rated significantly lower than the Control, Low CO_2, and Light struck samples for the emotion category Fun). This is in contrast to the UK, where there were no significant differences in the ratings of the Bitter sample compared to the Control, Low CO_2, and Light struck samples in the pleasantness/pleasure emotions.

It appears then that the emotion categories associated with pleasure/pleasantness in the two countries were used to respond to the 10 samples similarly, with the exception of the Bitter sample. Pearsons's r coefficients (Table 5.6) showed that there were high correlations between pleasant emotion categories across countries and also high correlations between unpleasant emotion categories between cultures (generally $r > 0.7$, although Excitement [UK] did not correlate as well with Classic [Spain] or Desire [Spain] and Tame/Safe [UK] did not correlate as well with Fun [Spain], but in all cases $r > 0.5$). Not surprisingly, this suggests a certain degree of relationship between emotion categories across the two cultures but also some differences, reflecting the qualitative differences between emotion categories.

Table 5.6 Pearson's correlation coefficients (r) between the emotion categories generated in Spain and the UK.

		UK								
		1 Shock	**2 Bor-edom**	**3 Conten-tment**	**4 Excit-ement**	**5 Nost-algia**	**6 Disconf-irmation**	**7 Disgust**	**8 Tame/ Safe**	**9 Under-whelmed**
Spain	**1 Mildness**	-0.600	0.347	**0.738**	**0.714**	**0.679**	**-0.695**	**-0.740**	**0.728**	0.432
	2 Indifference	-0.087	0.073	0.205	0.277	0.183	-0.244	-0.297	0.235	0.128
	3 Pleasure	**-0.829**	0.421	**0.825**	**0.766**	**0.818**	**-0.808**	**-0.767**	**0.751**	0.325
	4 Classic	**-0.893**	**0.673**	**0.765**	0.582	**0.830**	**-0.762**	**-0.760**	**0.804**	0.475
	5 Fun	**-0.728**	0.287	**0.745**	**0.718**	**0.738**	**-0.713**	**-0.664**	**0.645**	0.225
	6 Desire	**-0.779**	0.441	**0.726**	**0.641**	**0.757**	**-0.711**	**-0.679**	**0.710**	0.454
	7 Disgust	**0.937**	-0.563	**-0.934**	**-0.832**	**-0.939**	**0.919**	**0.907**	**-0.885**	-0.414
	8 Disillusionment	**0.908**	-0.546	**-0.863**	**-0.768**	**-0.898**	**0.877**	**0.848**	**-0.819**	-0.391
	9 Disappointment	**0.875**	-0.468	**-0.870**	**-0.773**	**-0.873**	**0.855**	**0.824**	**-0.810**	-0.374
	10 Intensity	0.581	-0.411	**-0.704**	**-0.679**	**-0.637**	**0.662**	**0.725**	**-0.715**	-0.396
	11 Nostalgia	**-0.907**	0.627	**0.796**	**0.673**	**0.883**	**-0.864**	**-0.831**	**0.763**	0.330
	12 Excitement	**-0.727**	0.313	**0.715**	**0.650**	**0.724**	**-0.692**	**-0.641**	**0.642**	0.306

Emboldened r values represent statistical significance ($p < 0.05$)

The third cluster including the Hoppy, Sweet, DMS, and Isoamyl acetate samples (and also Bitter in Spain) was relatively neutral on the first pleasantness/pleasure dimension for both countries. Instead, the cluster was characterised in Spain by its positive correlation with PC2 (Figure 5.7a) and associated higher ratings of Mildness and Indifference, and lower ratings of Intensity (Table 5.5). In the UK, this cluster was relatively neutral in terms of arousal/engagement/activation (Figure 5.7b).

Arousal/engagement/activation also proved important in differentiating between how the two countries grouped samples emotionally when including a fourth cluster. In the UK, the Light struck and Low CO_2 samples were scored relatively lower in Boredom and Underwhelmed (low arousing/engaging/activating emotions) than the Control and Bitter samples (Table 3.4). The opposite was found in Spain, with the Light struck and Low CO_2 samples associated with lower arousal/engagement/activation emotions than the Control sample. Referring to the means, significantly lower scores were assigned to the Low CO_2 sample in Intensity [Spain] than the Control (Table 5.5). These differences were perhaps unsurprising when referring to Pearson's r coefficients (Table 5.6) as the unengaging emotion categories in the UK (Boredom and Underwhelmed) were not well correlated with the unengaging emotion categories in Spain (Indifference and Mildness). In all cases, $r < 0.5$, although this value was typically much lower.

Intensity [Spain] was relatively correlated with a number of UK emotion categories, although there were no correlations of $r > 0.75$. The UK PCA (Figure 5.5) shows that there is no emotion category that loads particularly

highly on the arousal/engagement/activation dimension, where Intensity [Spain] loaded very highly in the equivalent dimension of the Spanish PCA (Figure 5.1a). Referring to the mean ratings of samples in Intensity, it was seen that this emotion category being used to score samples differently to any emotion category in the UK. To take one example, Excitement was the most arousing/engaging/activating of the UK emotion categories according to the PCA (Figure 5.5) and the Non-alcohol control and High alcohol samples received the lowest mean scores. However, these samples received the highest mean ratings in Intensity [Spain]. Conversely, the Sweet sample was rated relatively high in Excitement [UK] but was the lowest mean scored sample in Intensity [Spain].

There were three emotion categories that shared labels across cultures: Excitement, Nostalgia, and Disgust. Pearson's r between these shared name categories was high for Nostalgia (0.883; Table 5.6) and Disgust (0.907; Table 5.6), suggesting a similar response to the samples using these categories. However, Pearson's r was not particularly high for Excitement (0.65; Table 5.6). Referring to post hoc tests, it can be seen that there were particular differences in the responses to the Control and Hoppy samples. UK consumers did not particularly discriminate the Control and Hoppy samples from most other samples in Excitement [UK] (Table 3.4). Spanish consumers, on the other hand, rated the Control sample higher than most other samples in Excitement [Spain] and scored the Hoppy sample lower than a number of samples (Table 5.5). There may be a link to liking and familiarity here, with low ratings assigned to the Hoppy sample and high scores given to the Control

sample in these two measures by Spanish consumers, but approximately equivalent ratings given by UK consumers.

5.5.3 The relative abilities of each culture's reduced lexicon to discriminate across consumer segments

In both Spain and the UK, frequent differences in scale usage were found between genders and between age groups. In almost every instance, it was seen that, where there were significant differences between genders, women gave lower ratings than men (the only exception was Intensity [Spain]). In both cultures, the low arousal/engagement/activation emotion categories were rated highest by the consumers aged 35 and over. In contrast, the pleasant emotion categories were rated lowest by the older group. However, there was a difference between countries in that the 18-34 age group gave higher ratings to the unpleasant emotion categories in the UK, whereas the opposite was found in Spain.

Despite the frequent similarities in overall ratings of emotion categories between consumer groups across cultures, the abilities of the two reduced emotion lexicons to reveal interactions between samples and consumer groups was very different. The 9 UK emotion categories revealed no interactions between sample and gender and just a single UK emotion category – Excitement - showed a significant interaction between sample and age group, with higher ratings obtained from 18-34 year old consumers than 35+ year old consumers for the Bitter, Isoamyl acetate, and Hoppy samples. Conversely, the High alcohol sample received higher ratings of Excitement from the older consumer group (for full UK results, see section 3.4.3).

In Spain, frequent interactions between sample and consumer group were found. The High alcohol sample was given higher ratings of Disgust by females than males, and both the Non-alcohol control and High alcohol samples were rated significantly lower in Pleasure by women than men. These differences were found despite no interactions between sample and gender for either liking or familiarity. Additional interactions were found between sample and age group for the Classic, Disappointment, and Nostalgia emotion categories. Together, these three emotion categories drew out a total of 7 of the 10 samples as rated significantly differently between age groups. These interactions also were able to encapsulate the differences found by liking and familiarity (for full results, see section 5.3.3.2).

5.6 Discussion of cross-cultural comparisons

The product-specific consumer-led emotion lexicons generated in Spain and the UK were approximately comparable in the number of terms (43 in the UK, 43 in Spain). Whilst the quantity of terms was similar, the grouping of terms into emotion categories showed that there was a qualitative difference. In the UK, there were 9 distinct groups of emotion terms and in Spain there were 12. Differences in the number of emotion terms generated in the UK and Spain have previously been identified, hinting at a qualitative difference between these two cultures in their use of emotional language to describe their responses to beverages (van Zyl and Meiselman, 2015). Both cultures had three unpleasant emotion categories, which were well correlated. Spain had six pleasant emotion categories to the UK's four, with most correlating well across countries, although there were some exceptions. Accordingly, the emotion categories associated with the pleasure/pleasantness dimensions were

used by the consumers of each country to group the samples similarly into pleasant, neutral, and unpleasant emotion-eliciting groups of samples. The only difference in this regard was the response to the Bitter sample, which was grouped with pleasant emotion-eliciting samples in the UK but more neutral pleasure/pleasantness samples in Spain. However, on the whole, pleasant and unpleasant emotion categories were used comparably across cultures, showing a cross-cultural similarity between UK and Spanish consumers in response to this sample set.

These correlations could be largely dependent on the shared response between the two countries to the Non-alcohol control and High alcohol samples. In both countries, these samples were scored high for unpleasant emotion categories and low for pleasant emotion categories. The atypicality of these two samples in their emotional responses as compared to other samples (and their associations with unpleasant emotion categories in both countries) means that PC1 and the projected clusters may have looked quite different had the Non-alcohol control and High alcohol samples been excluded, especially had there not been the possible convergence of scale use for the other samples discussed individually for each country. Nevertheless, the comparison of the results obtained from reduced forms in both countries was still able to reveal differences.

The use of emotion categories associated with the arousal/engagement/activation dimension of emotion was very different between countries. The third cluster (Hoppy, Sweet, DMS, Isoamyl acetate, and also Bitter in Spain) was characterised by its relative neutrality on the first

two dimensions in the UK but correlation with low arousal/engagement/activation emotions in Spain. When including a fourth cluster, the Low CO_2 and Light struck samples were more engaging relative to the Control in the UK, whereas the opposite was found in Spain. The two Spanish low arousal/engagement/activation emotion categories Mildness and Indifference were not well correlated to the two UK low arousal/engagement/activation emotion categories Boredom and Underwhelmed. Post hoc tests also showed that the categories were indeed being used to respond very differently to samples across cultures. The single Spanish high arousal/engagement/activation emotion category Intensity was not particularly highly correlated with any single UK emotion category. This was not surprising given that the UK PCA showed no high loadings of emotion categories onto the arousal/engagement/activation dimension. It may be that, in the UK, such emotions were not relevant to beer, or at least these samples of beer. General cultural differences in emotionality, attitudes to and use of beer, context of consumption, to name but a few, are examples of potential drivers of the differences observed in this study. In particular, climate has an effect on beer consumption across cultures. For example, in the warm Spanish climate, beer plays a functional role of refreshment. As a result, there is a smaller range of beer styles generally available in Spain as compared to the UK (to illustrate this, 86% of beer consumed in Spain is lager (Euromonitor, 2014a), compared to 70% in the UK (Euromonitor, 2014b). Therefore, differences in the use of arousal/engagement/activation emotion categories between countries may have been very different between these cultures in this study as the UK consumers are more experienced with a wide

range of sensory properties like those included in this study. Further research including more countries could inform further about the role of culture on reports of emotion. For example, to investigate the effect of climate as a context for beer consumption, UK emotional response to beer could be compared to climactically similar countries like Ireland or the Netherlands, whilst Spain could be compared to its neighbour Portugal. It would be expected that more similar cultures would react more emotionally similarly to one another due to overlapping culture and, for example, shared climate which might impact on emotional response to beer. There has already been a suggestion that this is the case on response to odours on a more international scale, with European cultures generally more similar in their generated emotion dimensions than other cultures (Ferdenzi et al., 2013).

Another interpretation of the cross-cultural differences is linguistic. The Spanish emotion categories associated with arousal/engagement/activation (and their associated individual emotion terms) proved to be the most difficult to translate into English, relating to the point raised in the introduction to this chapter about the lack of direct translations sometimes found between languages. The present evidence suggests that more general emotional concepts, like individual emotion terms, can also prove difficult to directly translate. This possibility was anticipated due to the findings of Ferdenzi et al. (2013) who identified a number of differences in emotion dimensions between cultures in response to odours. For example, ‘spirituality’ was unique to the Singaporean sample and ‘melancholy’ was particular of the Chinese sample. Therefore, the research presented in this chapter tentatively suggests that the Spanish language has words that are able to reflect the

arousal/engagement/activation dimension of emotion where the English language does not. This has important implications for the comparison of emotional response between these cultures because Intensity in particular was a very effective emotion category for differentiating between the emotional responses of the 10 samples in Spain. Also, the Intensity emotion category was the only category to show a reversal of the general trend for lower ratings given to samples on the whole by women, which may offer important information about the emotional differences between genders in their emotional experiences of beer. Furthermore, Mildness and Indifference were also relatively effective at discriminating between samples in Spain, particularly when compared to the relatively undiscriminating UK low arousal/engagement/activation emotion categories, Boredom and Underwhelmed. As a result, a beer developed in Spain to increase feelings of Intensity could not be easily measured on this criterion in the UK and different emotional criteria would be needed to assess the responses of UK consumers.

A common finding across both cultures was that women generally rated emotion categories lower than men, suggesting a shared effect of gender across these two cultures of reports of emotion elicited by beer. Gender roles (King et al. (2010); Fischer (1993); Grossman and Wood (1993); (Kring and Gordon, 1998)) have been proposed as an explanation for this in the discussions of the results of both countries (see section 3.5 and section 5.4), with the masculine associations of beer drinking (Landrine et al., 1988), giving different gender experiences and/or reporting of emotions elicited by the samples. Further research is needed to advance the understanding of the relationship between gender and reported emotions to beer and other product

categories in either or both countries. It would be of particular interest to explore the emotional response to beer across genders in other cultures, to see if the varying gender roles of countries have differential effects on reported emotional experience. This could also be extended to other product categories which may be gender neutral in one culture but gender-specific in others.

Age groups were largely similar in their use of emotion categories relative to one another across cultures. In both countries, the older beer consumers gave lower scores for pleasant emotion categories than younger consumers. With reference to the UK data, this was attributed to reduced emotional expressivity in older adults (Gross et al., 1997) given that unpleasant emotion categories were also rated lower by older consumer in the UK. However, the reverse was found in Spain, with high scores given to unpleasant emotion categories by the 35+ age group. Added to this is the fact that low arousal/engagement/activation emotion categories generally received the highest ratings from the older consumer group in both countries. It appears than that there is a more complex relationship between age and emotional response to beer and that this is not completely consistent across cultures.

The big difference when it came to the abilities of each country's reduced lexicon to differentiate between consumer groups was seen when looking at interactions between sample and gender or age group. Several such interactions were observed in Spain, whereas just one was found in the UK. Although, as previously discussed, there were emotion categories in Spain for which there are no equivalents in the UK lexicon, these categories did not contribute to the differentiation between consumer groups in their emotional

responses to individual samples. This raises the question of why the Spanish reduced emotion lexicon was able to reveal differences in the responses of consumer groups to individual samples where the UK reduced emotion lexicon was largely unable. There are two possible reasons: first, consumer groups could simply be more heterogeneous in response to beer in Spain than in the UK. If the case, this could represent an important cross-cultural difference and would have implications for the practical use of emotional response in each country. It would be expected that a change in a sensory property of a product would have similar effects on emotional response between consumer groups in the UK. In contrast, a single sensory property could have very different emotional effects between genders or between age groups in Spain. At an extreme, this could mean the development of products to elicit specific emotional profiles in a particular gender or age group in Spain. At the very least, this finding suggests that the effects on emotional response in different consumer segments should be carefully considered before even subtly changing the sensory profile of a product in Spain.

The second reason for the difference in the abilities of each country's reduced lexicon to reveal interactions between sample and consumer group could be that the Spanish lexicon was more effective than the UK lexicon in its ability to differentiate between consumer groups in their emotional responses to individual samples. However, it was seen that the UK lexicon was comparable to the Spanish lexicon in its ability to discriminate between samples. It seems instead that the Spanish reduced emotion lexicon was more focussed on the subtleties of meaning between quite similar emotion concepts. For example, Disappointment and Disillusionment are very similar concepts and were found

to discriminate between samples extremely similarly. However, just Disappointment showed interactions between age group and sample, demonstrating the importance in Spain of having two separate categories instead of merging the two. In the UK, the emotion categories were arguably more distinct than in Spain.

5.7 Conclusions

This chapter has, firstly, shown the application of the reduced product-specific consumer-led emotion lexicon in Spain. The Spanish reduced lexicon of 12 emotion categories was found to discriminate effectively between the emotional responses to the subset of 10 samples included in this study. Although a number of these emotion categories differentiated very similarly between samples, it was seen that the inclusion of 12 emotion categories was important for identifying differences in emotional response to particular sensory properties between consumer groups. This underlines the effectiveness of this approach for generating and reducing a product-specific emotion lexicon in more than one country, culture, and language.

Cross-cultural comparisons showed that categories associated high or low pleasantness/pleasure were used to group samples relatively similarly across these two cultures. The noteworthy difference between cultures was the grouping of samples according to high or low arousal/engagement/activation emotion categories. This was discussed in relation to linguistic differences in emotion and/or the cultural relationship with beer as a product category between the two countries studied here.

Further differences were discussed in terms of the Spanish reduced emotion lexicon's greater ability than the UK reduced lexicon to effectively differentiate between the responses of different consumer groups to specific sensory properties of beer. This was discussed in relation to the subtleties of emotion captured by the higher number of emotion categories in the Spanish emotion form and possible increased heterogeneity between consumer groups in emotional response to beer in Spain.

Overall, the reduced product-specific consumer-led emotion lexicon was able to reveal similarities and expose key differences between these two cultures in their emotional responses to the sample set. This was whilst allowing the use of a relatively quick verbal self-report approach, carried out in the consumer's native language and avoiding the problems associated with direct translations of individual emotion terms. Expansion of this work to include more countries was suggested to inform further about the nature of the reported cross-cultural similarities and differences.

Chapter 6: General Discussion

The research presented in this thesis has defined and utilised an improved approach for verbal self-report of emotion, which allowed the investigation of the relationships between sensory properties of beer and consumer emotional response. This chapter draws together the main findings of this thesis and discusses their practical implications and suggests directions for future work.

6.1 Developing an approach for the measurement of emotional response

One of the primary objectives of this thesis was to create an approach for the development of a reduced consumer-led emotion lexicon, which was product-specific to beer samples that were characterised as being sensorially-distinct. The developed approach built on previous research by utilising group interviews to generate the lexicon, thereby reducing costs in time and resources as compared one-to-one interviews as well as promoting discussion for deeper probing of consumer emotion language. To the author's knowledge, this is the first specific emotional lexicon developed for this emotive product category. In both the UK and Spain, lexicon lengthy lexicon of individual emotion terms was reduced to a reduced number of emotion categories by having the focus group participants rate the selected beer samples using the individual emotion terms. This data was then subjected to cluster analysis and linguistic checks to give the final emotion categories (9 in the UK and 12 in Spain). In both countries, this was found to be a relatively time-efficient and simple approach for grouping similar terms and provided information about emotional response to different samples, as well as the differences in the responses of consumer groups.

An important consideration, especially with respect to other published methods, was the effect form reduction had on the approach's sensitivity to picking up differences in emotional response. In order to shed some light on this, the UK's reduced form of 9 emotion categories was compared against the full lexicon of 43 terms upon which it was based. The same 109 consumers completed the full version of the product-specific consumer-led emotion lexicon in response to the same 14 samples. Comparisons were made regarding the relative discrimination abilities of the reduced and full forms between samples as well as the abilities of each form to differentiate between the responses of consumer groups. It was seen that, for some emotions, subtle discriminations between samples were lost through the employment of the reduced form emotion categories as compared to the full form individual emotion terms. However, consumers used some reduced form emotion categories to discriminate more effectively between the samples than with individual emotion terms in the full form. The full form was found to be slightly better than the reduced form in its ability to discriminate between genders but the full form was much more able than the reduced form to identify interactions between sample and age group.

Overall, it was concluded that the reduced form was relatively comparable to the full form in its discriminability between samples, although it may be preferable to use a full emotion lexicon with consumers rating individual terms if the purpose of the research is to identify differences between consumer groups, particularly between age groups. Nevertheless, there are significant savings in time and resources, as well as decreased potential for consumer boredom and fatigue, with the additional benefit of greater ease of

product comparisons. This approach is therefore beneficial for sensory and consumer science practitioners in academia and industry for the application to food and beverages. As such the use of the reduced form was suggested to be appropriate for wider application.

6.1.1 Beyond self-report

Throughout this thesis, difficulties have been identified that may have arisen from the selection of self-report as a measure of emotion. For example, differences between consumers groups' reports of emotional response have been discussed with uncertainty about whether the differences were in experienced emotion and associated emotions or merely in scale usage. Furthermore, it was seen to be difficult to directly compare the arousal/engagement/activation dimension of emotion between the two cultures studied here.

Perhaps these issues could be addressed by making use of implicit or physiological methods (see section 1.3.2 and section 1.3.3), avoiding the need for self-report and accessing alternative emotional components outlined by Scherer (2005). In particular, physiological methods are associated with arousal/engagement/activation (Posner et al., 2005) and, as such, may provide valuable information for cross-cultural comparison (e.g. Tsai et al. (2002)) as this proved difficult to directly compare across cultures.

Implicit and physiological approaches could be important to the measurement of emotion in sensory and consumer science more generally due to their abilities to inform about unconscious components of emotional response. By its very nature, self-report is unable to access unconscious components of

emotion. This is significant, as it has been shown that unconscious emotional components can drive behaviour (Berridge and Winkielman, 2003) and, importantly, it has been suggested that our reactions to food and beverages may be largely unconscious (Thomson et al., 2010). Development of implicit and physiological methods would address this gap in the literature for understanding of consumers' unconscious emotional responses. Nevertheless, it is recommended that such approaches are conducted in tandem with self-report measures. This is because implicit and physiological methods are not very specific (e.g. is an increase in heart rate to be interpreted as excitement or disgust?) and, by referring to conscious components of emotion, a richer, more holistic interpretation of emotion, as recommended by Scherer (2005), may be acquired.

6.2 Emotional response to the sensory properties of beer

The research presented in this thesis represents as significant step forward in the understanding of the relationship between sensory properties and emotional response. This is, to the author's knowledge, the first research to link the manipulation of a product's individual sensory properties to emotional response. Examples have been shown in both the UK and Spain of beer samples with modified sensory properties eliciting different emotional responses from consumers. This confirms the suggestions of previous authors that sensory properties act as drivers for emotional response (Thomson et al. (2010); Ng et al. (2013); Sester et al. (2013)). In addition, it was seen that emotion exceeded liking in its sample discriminability, a result also previously observed in non-alcoholic beverages (Ng et al., 2013).

The method that has been described in this thesis was only able to go so far in informing about the nature of these differences. This is believed to be due to a major limitation of this research of including atypical sample in the sample set (namely the non-alcohol commercial lager and the sample based upon this with added ethanol). The emotional response to these samples was too different to the others, leading to convergence of ratings for the other samples for many emotion categories. This is an important learning for future research that intends to probe the effects of subtle differences in a product's sensory properties, for example, in a product development setting. One must be careful about which products are included otherwise the sensitivity of the measurement is compromised.

This is not to take away from the results presented here that have demonstrated for the first time that differences in individual sensory properties may have large effects on consumer emotional response that would not be picked up using traditional liking measures and this is a key finding of this research,

6.2.1 Beyond fixed intensities of sensory properties

The present study has confirmed the previously reported association between sensory properties and emotional response (Thomson et al. (2010); Ng et al. (2013); Sester et al. (2013)) through the experimental control of individual sensory properties and subsequently observed effects on reported emotion. This represents merely the tip of the iceberg for this type of research. For example, the impact of different levels of a given sensory property could be explored. This is particularly important for the example of beer, as some

sensory properties are considered off-flavours at high concentrations but can be characteristic of some beers at lower concentrations (e.g. DMS; Bamforth (2009)). It would be interesting to see if there is an emotional corollary to this. The suggestion from the present research would be that the distinction between a characteristic flavour and an off flavour would not be so clear cut in emotion because the off-flavours included in this thesis (e.g. DMS, acetaldehyde, light struck) were not generally found to be adverse to emotional response when compared to more characteristic flavours. However, this may be a suggestion that the compounds were not included in a high enough concentration to be considered as an off-flavour. Further research is suggested to probe this area.

6.2.2 Beyond single sensory properties

Another area for future research could be the exploration of the effects of interactions between several sensory properties. Of course, a change in a single sensory property as performed in the present study is rare within products. Even within a simple beer model system, modifying one parameter has been shown to have complex effects on the perception of other constant parameters (Clark et al., 2011). Therefore, it is important to understand how sensory properties drive emotional response in the presence or absence of other sensory properties.

This can be exemplified through the finding that relatively large differences in carbonation had no significant effect on emotional response in the UK. This was surprising given that previous research with commercial lagers found carbonation to be an important driver of emotional response, despite relatively

small differences in carbonation between products (Chaya et al., 2015). It was discussed that this may be due to the product contexts in which sensory properties were present. In other words, the other sensory properties in a product are posited to have an effect on how a particular sensory property drives emotional response, meaning that large variations may sometimes have little effect in some contexts but small variations in other contexts may be important drivers.

The lack of an appropriate context of other sensory properties appeared to have had a deleterious in the present study when increasing sensory properties associated with alcohol content. By only increasing the alcohol content and not increasing (or decreasing) other sensory properties associated with fermentation, it was seen that more unpleasant and fewer pleasant emotions were elicited in consumers.

These results suggest that it is not enough to understand the effects of individual sensory properties; their effects in the presence of other sensory properties must also be understood. This certainly needs to be researched further and is a logical next step in progressing the research in this thesis that has shown that changes in individual sensory properties can have significant effects on emotional response.

6.2.3 Beyond beer

The application of the reduced product-specific consumer-led emotion lexicon approach to product categories beyond beer is needed to ascertain if it is capable of differentiating between consumers responses to other products. It might be argued that other products are not so wide-ranging in their elicited

emotional experiences as beer, meaning perhaps a full lexicon might be more appropriate for finding subtle differences. On the other hand, it was shown in the present research that, for some emotions, the reduced form was more discriminating than the full form, meaning that perhaps a reduced lexicon may be more effective than a full lexicon for differentiating between some products. Only by using this approach with other product categories can this be revealed.

When applying the reduced product-specific consumer-led emotion lexicon to other products, it would be interesting to compare the reduced emotion lexicons generated in response to different sample sets to see how similar or different the lexicons might be. For example, it could be seen whether a reduced lexicon generated in response to a set of wines would be similar or different to the reduced lexicon generated in response to a set of beers. In turn, it could be observed if alcoholic beverages are more similar to each other in the reduced lexicons used to describe them than to non-alcohol beverages or to food products. If this was the case, it would be interesting to understand the differences in emotions used to describe some product categories over others. If, on the other hand, different products are relatively comparable in their generated emotion categories, perhaps the generation of a product-specific lexicon for each product category would be unnecessary, prompting the development of this approach instead into a pre-determined reduced emotion lexicon with application across food and beverage products, in order to save time and resources in initial development.

In addition, the emotional responses of different consumer groups to different product categories would be of interest. In the present research, certain patterns of response were identified for particular consumer groups. For example, in both Spain and the UK, females generally gave emotion categories lower ratings than males. King et al. (2010) suggested that gender differences in emotional response are dependent on the product category. As beer is a masculine product (Landrine et al., 1988), it would be particularly interesting to research more typically neutral and feminine products to see if the pattern observed in the present research persists.

6.3 Consumer groups

Although only a secondary objective of this research, this thesis also explored relatively untouched ground in terms of considering comparisons between consumer groups.

6.3.1 Gender

It was anticipated that there would be differences in emotional response between genders due to differences in emotionality, emotional expressivity, and gender roles (particularly related to the socially masculine nature of beer as a beverage). However, in the UK interactions between sample and gender were not found at all and in Spain were only found for the Non-alcohol control and High alcohol samples, which have already been discussed as being atypical of the sample set. In general then, it appeared emotional response was relatively similar to these samples between genders. The surprising finding was that, in both countries, females generally gave lower ratings than males for emotion categories. Based on the literature, it would have been expected

that had there been a difference between genders in their scores, then females would have been giving higher scores due to their higher expressivity. King et al. (2010) have previously reported that emotional response between genders is dependent on the product category. Perhaps then, for this particular product category, females were less emotionally 'engaged' and, although their relative responses to individual samples was comparable to that of males, they reflected this by assigning lower scores. It is certain that further research is required and that this thesis confirms that there may be interesting gender differences in emotion response to sensory-driven products, though this may not be in the way that was initially expected.

6.3.2 Age group

When comparing age groups in their emotional responses to these samples, pleasant emotion categories in both Spain and the UK received lower ratings from the older consumer group. Beyond this, interactions between sample and age group were frequently observed in Spain and many were found for one emotion category in the UK. It appears that age groups are more heterogeneous than genders when comparing and this is a significant finding in its implications for products. The suggestion is that targeting consumers from across all age groups in emotional response would prove a difficult task, at least for beer, whereas the results from comparisons of genders suggest that this would be less of a problem.

6.3.3 Beyond extrinsic grouping of consumers

The research has made some initial tentative steps in exploring similarities and differences in emotional response between consumers that was purely

demographic (i.e. extrinsic). Beyond the scope of this thesis was additional analysis exploring the differences between intrinsic groups of consumers. For example, it could have been investigated whether there were segments of consumers that responded emotionally similarly to one another in response to this sample set in much the same way that consumers can be segmented in their liking. In this way, much more detail would be obtained about how consumers differ in their responses to the samples as those that find a given sample disgusting, for example, would not be in effect 'cancelling out' those that did not find it disgusting. This has, to the author's knowledge, not yet been explored in the literature. The approach described in this thesis lends itself to such research because only a few emotion categories would need to be compared as opposed to many emotion terms in EsSense Profile, for example.

6.3.4 Cross-cultural comparisons

Cross-cultural comparisons between the UK and Spain were also explored in this thesis. Generally, the use of emotions associated with low or high pleasure/pleasantness in the two countries was very similar in that samples were grouped very similarly when using the associated emotion categories in each country. This revealed an important cross-cultural similarity. Differences between the two cultures were evident in the scoring of emotion categories associated with high or low arousal/engagement/activation. This was discussed in terms of cultural and linguistic differences between the two cultures. There were also differences in the extent to which the countries' respective lexicons discriminated across consumer segments, with the Spanish reduced lexicon being more able to differentiate between genders and between age groups than the UK reduced lexicon in response to these samples. This

was unrelated to the previously discussed differences in the use of emotions categories associated with the arousal/engagement/activation dimension of emotion. It was unclear whether this was due to differences in heterogeneity of consumer groups between countries or differences in the discrimination abilities of the UK and Spanish lexicons. Therefore, it was shown that the reduced product-specific consumer-led emotion lexicon approach was able to highlight key differences between Spanish and UK consumers in their emotional responses to the samples included in this study.

6.3.4.1 Beyond the UK and Spain

A potentially fruitful area for future research is the expansion of the work presented in Chapter 5 to include more cultures. It was discussed that some of the differences between the UK and Spanish emotional responses may stem from cultural and/or linguistic differences.

With relation to linguistic differences, it was suggested that the English language was less able than Spanish to encapsulate the arousal/engagement/activation dimension of emotion. A way to explore this idea further would be to compare cultures with a shared language (e.g. Spain vs. Mexico; UK vs. USA; Portugal vs. Brazil). Although, as described in the introduction to Chapter 5, use of shared languages to describe emotional experience is different between cultures (van Zyl and Meiselman, 2015), it is likely, for example, that Mexican Spanish includes emotion terms related to those that were present in Spanish but not in English (i.e. Mildness, Indifference, Intensity), owing to the shared history.

Beyond this, consumers in countries with overlapping or related cultures could be investigated to see if, for example, Mediterranean countries like Spain are more similar to one another than Northern European countries like the UK. It has already been shown on an international scale, with European cultures generally more similar in their emotion dimensions than they are to other cultures (Ferdenzi et al., 2013), owing to complex cultural factors.

6.3 Conclusions

To conclude, the new improved emotion measurement approach presented in this thesis has been described and demonstrated to be a time efficient, product-specific, consumer-led, verbal self-report measure. It was shown to be discriminating between sensorially-distinct beer samples, and even comparable in this regard to a full emotion lexicon. Furthermore, the reduced product-specific consumer-led emotion lexicon approach has been shown to be a valuable tool in the wider use of cross-cultural comparison of emotional response. Through the application of the new approach, it was also evidenced that emotional response was able to discriminate beyond liking in beer and that different consumer segments respond differently to beer with respect to their emotions. It is clear from the research described here that this approach provides researchers in both academia and industry with an enhanced consumer-led emotion methodology for use in many potential future applications across food and beverages categories in sensory and consumer science. This is particularly valuable at a time when research in this area continues to grow and effective approaches are needed to fulfil demand for information about consumer emotional response.

References

ALLEN, J. G. & HACCOUN, D. M. 1976. Sex differences in emotionality: a multidimensional approach. Human Relations, 29**,** 711-722.

BABCHUK, W. A., HAMES, R. B. & THOMPSON, R. A. 1985. Sex differences in the recognition of infant facial expressions of emotion - the primary caretaker hypothesis. Ethology and Sociobiology, 6**,** 89-101.

BAMFORTH, C. W. 2009. Beer: Tap into the Art and Science of Brewing, New York, Oxford University Press.

BARRETT, L. F. 2006. Solving the emotion paradox: Categorization and the experience of emotion. Personality and Social Psychology Review, 10**,** 20-46.

BARTOSHUK, L. M., DUFFY, V. B. & MILLER, I. J. 1994. PTC/PROP tasting - anatomy, psychophysics, and sex effects. Physiology & Behavior, 56**,** 1165-1171.

BERE, E., GLOMNES, E. S., VELDE, S. J. & KLEPP, K. I. 2008. Determinants of adolescents' soft drink consumption. Public Health Nutrition, 11**,** 49-56.

BERRIDGE, K. C. & WINKIELMAN, P. 2003. What is an unconscious emotion? (The case for unconscious "liking"). Cognition & Emotion, 17**,** 181-211.

BREUGELMANS, S. M. & POORTINGA, Y. H. 2006. Emotion without a word: Shame and guilt among Raramuri Indians and rural Javanese. Journal of Personality and Social Psychology, 91**,** 1111-1122.

BREUGELMANS, S. M., POORTINGA, Y. H., AMBADAR, Z., SETIADI, B., VACA, J. B., WIDIYANTO, P. & PHILIPPOT, P. 2005. Body sensations associated with emotions in Raramuri Indians, rural Javanese, and three student samples. Emotion, 5**,** 166-174.

BRIGGS, D. E., BOULTON, C. A., BROOKES, P. A. & STEVENS, R. 2004. Brewing: Science and Practice, Cambridge, Woodhead Publishing.

CARDELLO, A. V., BOGOGNO, M., CRAIG, C. & LESHER, L. L. 2014. Emotion questionnaires: The effect of the number of emotions on consumer responses. Poster presented at 6th European Conference on Sensory and Consumer Research. Copenhagen, Denmark.

CHAYA, C., PACOUD, J., NG, M., FENTON, A. & 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, 73**,** 49-60.

CHREA, C., GRANDJEAN, D., DELPLANQUE, S., CAYEUX, I., LE CALVE, B., AYMARD, L., VELAZCO, M. I., SANDER, D. & 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.

CLARK, R. A., HEWSON, L., BEALIN-KELLY, F. & HORT, J. 2011. The Interactions of CO2, Ethanol, Hop Acids and Sweetener on Flavour Perception in a Model Beer. Chemosensory Perception, 4**,** 42-54.

COOKE, L. J. & WARDLE, J. 2005. Age and gender differences in children's food preferences. British Journal of Nutrition, 93**,** 741-746.

CURRAN, S. L., ANDRYKOWSKI, M. A. & STUDTS, J. L. 1995. Short form of the Profile of Moof States (POMS-SF) - Psychometric information. Psychological Assessment, 7**,** 80-83.

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.

DAMASIO, A. 2006. Descartes' Error, London, Vintage.

DANNER, L., SIDORKINA, L., JOECHL, M. & DUERRSCHMID, K. 2014. Make a face! Implicit and explicit measurement of facial expressions elicited by orange juices using face reading technology. Food Quality and Preference, 32**,** 167-172.

DARWIN, C. 1872. The expression of emotions in man and animals, London, John Murray.

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.

DESMET, P. M. A., HEKKERT, P. & JACOBS, J. J. 2000. When a car makes you smile: Development and application of an instrument to measure product emotions. In: HOCH, S. J. & MEYER, R. J. (eds.) Advances in Consumer Research.

DESMET, P. M. A. & SCHIFFERSTEIN, H. N. J. 2008. Sources of positive and negative emotions in food experience Appetite, 50**,** 290-301.

DORADO, R., PICARD, A., PÉREZ-HUGALDE, C. & CHAYA, C. in preparation. First position effect on the emotional response of consumers: a case study in chocolate and beers.

DUFFY, V. B., PETERSON, J. M. & BARTOSHUK, L. M. 2004. Associations between taste genetics, oral sensation and alcohol intake. Physiology & Behavior, 82**,** 435-445.

EKMAN, P. 1994. Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique. Psychological Bulletin, 115**,** 268-287.

EKMAN, P. & FRIESEN, W. 1978. Facial action coding system: A technique for the measurement of facial movement, Palo Alto, California, Consulting Psychologists Press.

EKMAN, P. & FRIESEN, W. V. 1971. Constants across cultures in face and emotion. Journal of Personality and Social Psychology, 17**,** 124-129.

ETCOFF, N. L. & MAGEE, J. J. 1992. Categorical perception of facial expressions. Cognition, 44**,** 227-240.

EUROMONITOR 2014a. Beer in Spain. Category Briefing.

EUROMONITOR 2014b. Beer in the UK. Category Briefing.

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., ROBERTS, S. C., SCHIRMER, A., PORCHEROT, C., CAYEUX, I., SANDER, D. & 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., VELAZCO, M. I., SANDER, D., SCHERER, K. R. & 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, 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.

GIBSON, E. L. 2006. Emotional influences on food choice: Sensory, physiological and psychological pathways. Physiology & Behavior, 89**,** 53-61.

GROSS, J. J., CARSTENSEN, L. L., PASUPATHI, M., TSAI, J., SKORPEN, C. G. & HSU, A. Y. C. 1997. Emotion and aging: Experience, expression, and control. Psychology and Aging, 12**,** 590-599.

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.

GUINARD, J. X., UOTANI, B., MAZZUCCHELLI, R., TAGUCHI, A., MASUOKA, S. & FUJINO, S. 2000. Consumer testing of commercial lager beers in blind versus informed conditions: Relation with descriptive analysis and expert quality ratings. Journal of the Institute of Brewing, 106**,** 11-19.

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.

HOLLOWOOD, T. A. 2002. Taste-Aroma-Matrix Interactions Determine Flavour Perception. PhD, The University of Nottingham.

HU, S. Q., PLAYER, K. A., MCCHESNEY, K. A., DALISTAN, M. D., TYNER, C. A. & SCOZZAFAVA, J. E. 1999. Facial EMG as an indicator of palatability in humans. Physiology & Behavior, 68**,** 31-35.

HURTADO DE MENDOZA, A. 2007. Hacia una aproximación enciclopédica del concepto de vergüenza y el concepto de shame. Universidad Autónoma de Madrid, España.

HURTADO DE MENDOZA, A. 2008. The problem of translation in cross-cultural research of emotion concepts (commentary on Choi & Han). International Journal for Dialogical Science, 3**,** 241-248.

ISHIHARA 1972. Tests for color blindness, Tokyo, Kanehara Shuppan Co Ltd.

ISO8586-1:1993. Sensory analysis -- General guidance for the selection, training and monitoring of assessors -- Part 1: Selected assessors.

IZARD, C. E. 2010. The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. Emotion Review, 2**,** 363-370.

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.

KANEDA, H., HIDETOSHI, K. & WATARI, J. 2011. Novel Psychological and Neurophysiological Significance of Beer Aromas. Part 1: Measurement of Changes in Human Emotions During the Smelling of Hop and Ester Aromas Using a Measurement System for Brainwaves. Journal of the American Society of Brewing Chemists, 69**,** 67-74.

KANEDA, H., MEAESHIMA, K., GOTO, N., KOBAYAKAWA, T., AYABE-KANAMURA, S. & SACHIKO, S. 2000. Decline in Taste and Odor Discrimination Abilities with Age, and Relationship between Gustation and Olfaction. Chemical Senses, 25**,** 331-337.

KARIM, J., WEISZ, R. & REHMAN, S. U. 2011. International positive and negative affect schedule short-form (I-PANAS-SF): testing for factorial invariance across cultures. 3rd World Conference on Educational Sciences. Amsterdam: Elsevier Science.

KING, E. S., JOHNSON, T. E., BASTIAN, S. E. P., OSIDACZ, P. & LEIGH FRANCIS, I. 2012. Consumer liking of white wines: segmentation using self-reported wine liking and wine knowledge. International Journal of Wine Business Research, 24**,** 33-46.

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.

KOMIYAMA, O., WANG, K., SVENSSON, P., ARENDT-NIELSEN, L., KAWARA, M. & DE LAAT, A. 2009. Ethnic differences regarding sensory, pain, and reflex responses in the trigeminal region. Clinical Neurophysiology, 120**,** 384-389.

KREIBIG, S. D. 2010. Autonomic nervous system activity in emotion: A review. Biological Psychology, 84**,** 394-421.

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

KUESTEN, C., CHOPRA, P., BI, J. & MEISELMAN, H. L. 2014. A global study using PANAS (PA and NA) scales to measure consumer emotions associated with aromas of phytonutrient supplements. Food Quality and Preference, 33**,** 86-97.

LANDRINE, H., BARDWELL, S. & DEAN, T. 1988. Gender expectations for alcohol use - a study of the significance of the masculine role. Sex Roles, 19**,** 703-712.

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

LEON-CARRION, J., MCMANIS, M. H., CASTILLO, E. M. & PAPANICOLAOU, A. C. 2006. Time-locked brain activity associated with emotion: A pilot MEG study. Brain Injury, 20**,** 857-865.

LUBIN, B., VAN WHITLOCK, R., REDDY, D. & PETREN, S. 2001. A Comparison of the Short and Long Forms of the Multiple Affect Adjective Check List—Revised (MAACL-R). Journal of Clinical Psychology, 57**,** 411-416.

MCNAIR, D. M., LORR, M. & DROPPLEMAN, L. F. 1971. Manual for the Profile of Mood States, San Diego, CA, Educational and Industrial Training Services.

MEILGAARD, M. C., CIVILLE, G. V. & CARR, B. T. 2007. Sensory Evaluation Techniques, Boca Raton, Florida, CRC Press.

MINTEL 2013. Beer - UK - December 2013. London: Mintel.

MOSKOWITZ, H. R., BECKLEY, J. H. & RESURRECCION, A. V. A. 2012. Sensory and Consumer Research in Food Product Design and Development, Oxford, Blackwell Publishing Ltd.

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.

NESTRUD, M. A., MEISELMAN, H. L., KING, S. C., LESHER, L. L. & CARDELLO, A. V. 2013. A method for development of EsSense25 - a shorter version of the EsSense(R) Profile. Poster presented at 10th Pangborn Sensory Science Symposium. Rio de Janeiro, Brazil.

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.

PARTALA, T. & SURAKKA, V. 2003. Pupil size variation as an indication of affective processing. International Journal of Human-Computer Studies, 59**,** 185-198.

PHAN, L. K., WAGER, T., TAYLOR, S. F. & LIBERZON, I. 2002. Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. NeuroImage, 16**,** 331-348.

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.

POELS, K. & DEWITTE, S. 2006. How to capture the heart? Reviewing 20 years of emotion measurement in advertising. Journal of Advertising Research, 46**,** 18-37.

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.

POSNER, J., RUSSELL, J. A. & PETERSON, B. S. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17**,** 715-734.

ROBIN, O., ROUSMANS, S., DITTMAR, A. & VERNET-MAURY, E. 2003. Gender influence on emotional responses to primary tastes. Physiology & Behavior, 78**,** 385-393.

ROBINSON, M. D. & CLORE, G. L. 2002. Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychological Bulletin, 128**,** 934-960.

ROSENSTEIN, D. & OSTER, H. 1988. Differential facial responses to 4 basic tastes in newborns. Child Development, 59**,** 1555-1568.

ROSSITER, J. R. & BELLMAN, S. 2005. Marketing communications: Theory and applications, Frenchs Forest, NSW, Australia, Pearson Prentice-Hall.

ROUSMANS, S., ROBIN, O., DITTMAR, A. & VERNET-MAURY, E. 2000. Autonomic Nervous System Responses Associated with Primary Tastes. Chemical Senses, 25**,** 709-718.

RUSSELL, J. A. 1980. A circumplex model of affect. Journal of Personality and Social Psychology, 39**,** 1161-1178.

SCHERER, K. R. 2005. What are emotions? And how can they be measured? Social Science Information Sur Les Sciences Sociales, 44**,** 695-729.

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.

SHACHAM, S. 1983. A shortened version of the Profile of Mood States. Journal of Personality Assessment, 47**,** 305-306.

STEIN, L. J., NAGAI, H., NAKAGAWA, M. & BEAUCHAMP, G. K. 2003. Effects of repeated exposure and health-related information on hedonic evaluation and acceptance of a bitter beverage. Appetite, 40**,** 119-129.

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.

STONE, H. & SIDEL, J. L. 1993. Sensory Evaluation Practices, San Diego, Academic Press.

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

TEPPER, B. J., WHITE, E. A., KOELLIKER, Y., LANZARA, C., D'ADAMO, P. & GASPARINI, P. 2009. Genetic Variation in Taste Sensitivity to 6-n-Propylthiouracil and Its Relationship to Taste Perception and Food Selection. Annals of the New York Academy of Sciences, 1170**,** 126-139.

THOMPSON, E. R. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). Journal of Cross-Cultural Psychology, 38**,** 227-242.

THOMSON, D. M. H. & CROCKER, C. 2013. A data-driven classification of feelings. Food Quality and Preference, 27**,** 137-152.

THOMSON, D. M. H. & CROCKER, C. 2015. Application of conceptual profiling in brand, packaging and product development. Food Quality and Preference, 40**,** 343-353.

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.

TSAI, J. L., CHENTSOVA-DUTTON, Y., FREIRE-BEBEAU, L. & PRZYMUS, D. E. 2002. Emotional Expression and Physiology in European Americans and Hmong Americans. Emotion, 2**,** 380-397.

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.

VALENTIN, D., CHOLLET, S., LELIEVRE, M. & ABDI, H. 2012. Quick and dirty but still pretty good: a review of new descriptive methods in food science. International Journal of Food Science and Technology, 47**,** 1563-1578.

VAN ZYL, H. & MEISELMAN, H. L. 2015. The roles of culture and language in designing emotion lists: comparing the same language in different English and Spanish speaking contries. Food Quality and Preference.

WATSON, D. & CLARK, L. A. 1991. The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form.

WATSON, D., CLARK, L. A. & TELLEGEN, A. 1988. Development and validation of brief measures of positive and negative affect - the PANAS scales. Journal of Personality and Social Psychology, 54**,** 1063-1070.

WATSON, D. & TELLEGEN, A. 1985. Toward a consensual structure of mood. Psychological Bulletin, 98**,** 219-235.

WELGE-LÜSSEN, A., DORIG, P., WOLFENSBERGER, M., KRONE, F. & HUMMEL, T. 2011. A study about the frequency of taste disorders. Journal of Neurology, 258**,** 386-392.

ZUCKERMAN, M. 1960. The development of an affect adjective check list for the measurement of anxiety. Journal of Consulting Psychology, 24**,** 457-462.

ZUCKERMAN, M. & LUBIN, B. 1985. Manual for the Multiple Affect Adjective Check List, San Diego, CA, Educational and Industrial Testing Service.

Appendices

Appendix A: Pre-screening questionnaire

QUESTIONNAIRE (please return to: stxce1@exmail.nottingham.ac.uk)

Details

Name:

Address:

Email:

Telephone:

Date of Birth:

Occupation:

General availability

1. Please indicate (with 'X's) your general availability to attend tasting sessions at Sutton Bonington

	Monday	Tuesday	Wednesday	Thursday	Friday
Morning (10-1)					
Afternoon (2-5)					

2. Are there times of year where you may be away for more than 2 weeks? Delete as appropriate.

Yes

No

Food & drink

3. Provide a brief description (3-5 sentences) of your favourite alcoholic beverage.

4. What is your favourite food and why?

5. Write a brief description (3–5 sentences) about the last new food you tried.

Appendix A (continued)

6. What foods/drinks do you particularly dislike? Please explain why.

7. How often do you eat out?

8. Describe the smell of your favourite restaurant/public house.

9. How would you describe the difference between flavour and aroma?

10. How would you describe the difference in flavour between lager (e.g. Heineken) and stout (e.g. Guinness).

11a. Do you have any specific dietary requirements (e.g. Vegetarianism)? Delete as appropriate.

Yes

No

11b. If yes, please provide details below.

Health

12. Please indicate (with 'X's) if you have any of the following:

	Colour blindness
	Dentures
	Oral/gum disease
	Diabetes
	Hypertension
	Liver disease
	None of the above

Appendix A (continued)

13. Are you presently, or have you ever been, a smoker? Delete as appropriate.

Yes

No

14. Do you have any medical condition that may impair your sensory ability? Delete as appropriate.

Yes

No

15a. Do you have any food allergies? Delete as appropriate.

Yes

No

15b. If yes, please provide details below.

16. Are you currently on any long-term medication? Delete as appropriate.

Yes

No

16b. If yes, please provide details below.

17. Are you, or do you intend to become, pregnant? Delete as appropriate.

Yes

No

As successful applicants will be consuming alcohol, we have a duty of care concerning your health and so ask that you answer the following four questions honestly.

18. Have you ever felt that you should cut down on your drinking? Delete as appropriate.

Yes

No

Appendix A (continued)

19. Have people annoyed you by criticising your drinking? Delete as appropriate.

Yes

No

20. Have you ever felt bad or guilty about your drinking? Delete as appropriate.

Yes

No

21. Have you ever had a drink first thing in the morning (as an "eye opener") to steady your nerves or get rid of a hangover? Delete as appropriate.

Yes

No

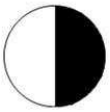
Further information

22. Please provide any other relevant information below

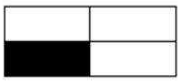
Appendix B: Scaling exercises (redrawn from Meilgaard et al. (2007))

Instructions: Move the lined provided on the right of each figure to mark on the line the proportion of the area of each shape that is shaded.

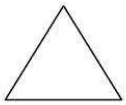
EXAMPLES



None — All

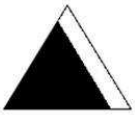


None — All



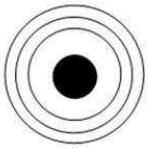
None — All

1



None — All

2



None — All

3

None — All

4

None — All

5

None — All

Appendix C: Panel sample assessment protocol

Aroma

General	Control	Hoppiness, diacetyl, DMS, isoamyl acetate, maltiness.

Light struck Skunky	Control	Aroma only. Use a couple of short, shallow sniffs.

Mouthful 1

Tingliness Prickliness	'Flat' control	Hold in mouth for 2-3 seconds. Assess by passing the liquid over the tip of the tongue.
Bitterness	Control	Hold in mouth for 5-6 seconds. Assess using back of the tongue/throat after swallow.

PALATE CLEANSE

Mouthful 2

Hoppiness Soapy, floral, citrus, hessian sack	Control	Wash around mouth for 2-3 seconds and breathe out on swallow.
Warming	Non-alcohol control	Becomes most intense about 4-5 seconds after swallowing in the chest.

PALATE CLEANSE

Mouthful 3

Bubbliness Sensation of bubbles, 'frothiness'	'Flat' control	Hold in mouth for 3-4 seconds. 2 "chews". Assess using whole mouth (particularly cheeks).
Diacetyl Buttery	Control	Wash around mouth for 2-3 seconds and breathe out on swallow.

Appendix C (continued)

Alcoholic flavour	Non-alcohol control	Wash around mouth and breathe out on swallow.

PALATE CLEANSE

Mouthful 4

Sweetness	Control	Wash around mouth, assessing using the tongue.
DMS Tinned sweetcorn, cabbage	Control	Top of scale apparent from aroma. Wash around mouth for 2-3 seconds and breathe out on swallow.
Isoamyl acetate Pear drops	Control	Wash around mouth for 2-3 seconds and breathe out on swallow.

PALATE CLEANSE
(water only suggested)

Mouthful 5

Acetaldehyde Green apples, emulsion paint, linseed oil	Control	Assess based on initial 1-2 second burst of flavour.
Body Fullness	Non-alcohol control	Coat mouth in liquid.
Maltiness Biscuity	Control	Wash around mouth for 5-6 seconds and breathe out on swallow.